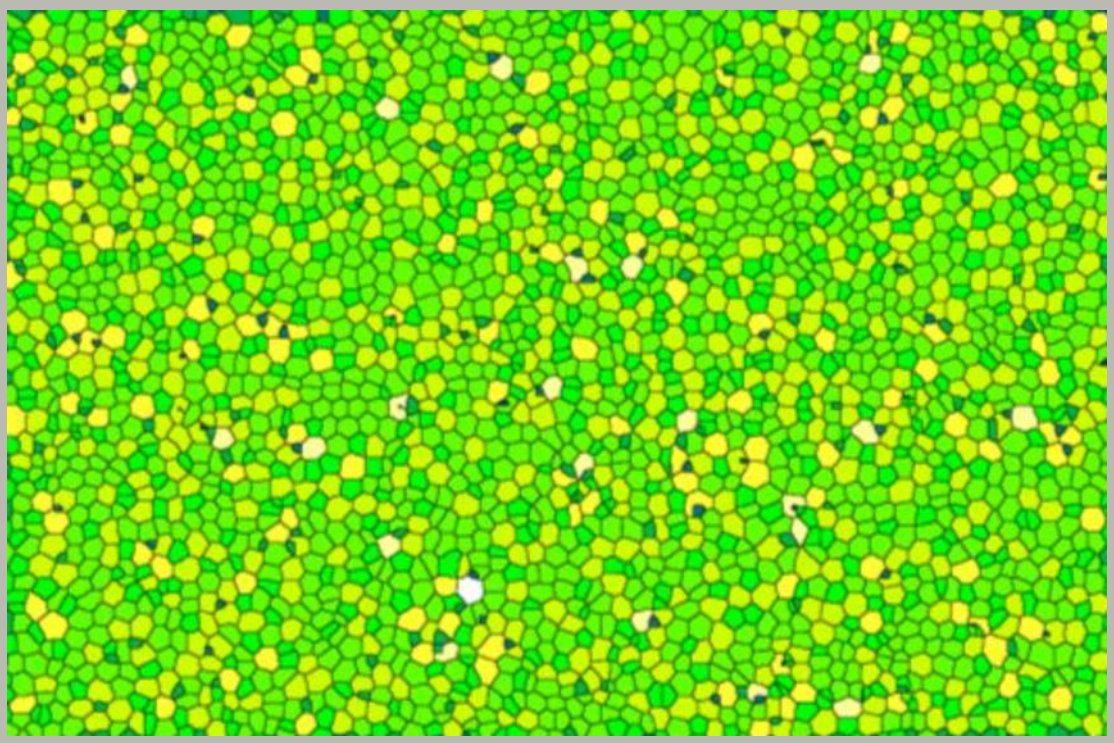
CHEMICAL ENGINEERING RESEARCH

Reports of the 4th year research projects in the Department of Chemical Engineering at Imperial College London

VOLUME 7



Edited by Erich A. Müller

February, 2025

CHEMICAL ENGINEERING RESEARCH

Reports of the 4th year research projects
in the Department of Chemical Engineering
at Imperial College London

Edited by Erich A. Müller

Volume 7

2025

9 781916 005068

© The Author(s) 2025.
Published by Imperial College London, London SW7 2AZ, UK
Contact details – e.muller@imperial.ac.uk

This book is a compilation of manuscripts created as part of a teaching assignment on the 4th year's Chemical Engineering CENG70001 course (Advanced Chemical Engineering Practice Research Project) . Copyright in each paper rests with its authors. The author of each paper appearing in this book is solely responsible for the content thereof; the inclusion of a paper in the book shall not constitute or be deemed to constitute any representation by Imperial College London that the data presented therein are correct or sufficient to support the conclusions reached or that the experiment design or methodology is adequate.



The book is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0). Under this licence, you may copy and redistribute the material in any medium or format on the condition that: you credit the author, do not use it for commercial purposes and do not distribute modified versions of the work. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law. https://creativecommons.org/licenses/by-nc-nd/3.0/.

The information contained in this publication is being distributed without warranty of any kind, either expressed or implied. The responsibility for the interpretation and use of the material lies with the reader. In no event shall Imperial College London be liable for damages arising from its use.

E-ISBN 978-1-9160050-6-8

First published in 2025

Preface

This volume of *Chemical Engineering Research* collects the unedited research project reports written by 4th year undergraduates (Class of 2025) of the M.Eng. course on Chemical Engineering in the Department of Chemical Engineering at Imperial College London. The research project spans one term (Autumn) during the last year of the career. It emphasises independence, the ability to plan and pursue original project work for an extended period, produce a high-quality report, and present the work to an audience using appropriate visual aids. Students are also expected to produce a literature survey and to place their work in the context of prior art. The papers presented showcase the diversity and depth of some of the research streams in the department but only touch on a small number of research groups and interests. For a complete description of the research at the department, the reader is referred to the departmental website[1].

The papers presented are in no particular order, and a manuscript number identifies them. Some papers refer to appendixes and/or supplementary information which are too lengthy to include. These files are available directly from the supervisors (see supervisor index at the end of the book). Some reports are missing and being embargoed, as they contain confidential information. A few reports correspond to industrial internships, called LINK projects, in collaboration with Shell.

The cover figure corresponds to a Voronoi neighbouring analysis of 0.01%AOT + 8%PS + DCM sample (taken from the work of Weng Kit Leong and Zewen Liu, manuscript 15).

London, February 2025

[1] https://www.imperial.ac.uk/chemical-engineering

Title Index

paper	Title	page
1	Electrodeposition of Manganese Dioxide from Lithium-ion Battery Leachate	1
2	Sulphur doped materials as electrocatalysts for the electrochemical reduction of O2	11
3	Optimised Multi-Vector Integration of a PTES System with Multi-Level Storage	(*)
4	Virus Mimicking LNP for Intracellular Delivery	(*)
5	Dye Extraction Decoded: The Role of Ionic Liquid Solvents and Process Tuning	19
6	Exergy analysis on sustainable aviation fuel production from air and water	29
7	Exploring Optimal Parameters for Graphene Liquid-Phase Exfoliation Using Taylor-Coutte Device	39
8	Wrinkle Morphology: An Investigation into Plasma Driven Cracking	48
9	A Hybrid Process for Adsorption-based CO_2 Capture	58
10	High-Temperature Electrolytic Synthesis of Carbon Nanotubes from Molten Li_2CO_3	68
11	Combined Effect of Water and Free Fatty Acids on Biodiesel Purity from Transesterification	78
12	Haemodynamic Analysis in Abdominal Aortic Aneurysms using Computational Fluid Dynamics	88
13	The Flexural Strength and Strength Development of Concrete with Carbon Addition	98

paper	Title	page
14	Pluronic F-127 Hydrogels for Delivery of Lipid Nanoparticles Containing mRNA	108
15	Surfactant Optimized Macroporous Polystyrene for Water Treatment Applications	118
16	Transfer Learning for Continuous Biopharmaceutical Manufacturing	128
17	Investigation into the Impact of Impurities on the Regeneration of Paracetamol Crystals	138
18	Comparison of Levasil CS50-28 and CS40-125 Colloidal Silicas to LUDOX TM-50 for the Purpose of Forming Shake-Gels	148
19	Hydrodynamic modelling to optimise hydrogen production in a photoelectrochemical reactor	158
20	Predicting the Competition of SO_2 with CO_2 in an MEA Capture Solvent using the SAFT-γ Mie Equation of State	168
21	Tuning Fluid Models to Predict Sustainable Gasoline Substitutes Properties	178
22	A Computational Framework for Modelling Combined Drug Resistance Mechanisms in Bacterial Cells	(*)
23	The Assessment of Readiness for Small-scale Carbon Capture & Storage Systems Applied on Diesel Fuel Power Generators in Remote Areas	188
24	PEG-based pH-responsive hydrogel for oral drug delivery	(*)
25	Halide Perovskite Solar Cells for Photoelectrochemical Water-Splitting Devices	(*)
26	Enhancing and Optimising TiO_2's Photothermal Properties as a Semiconductor for Clean Water Production	198
27	Applications of the Adversarial Robust Real-Time Optimization and Control Algorithm to Multi-Compressor Problems	208
28	Speed of Sound for Normal Hydrogen at Temperatures of 198 K to 298 K and Pressures of 2 to 50 MPa	216

paper	Title	page
29	Analysis of Policies to Facilitate Heat Pump Deployment for Building Decarbonisation	226
30	Selective Electrooxidation of Glycerol to Lactic Acid Using Platinum-Based Multicomponent Catalysts	236
31	Techno-Economic and Environmental Analysis of Liquid Direct Air Capture with Green and Grey Fuel Integration	246
32	Life Cycle Impact Assessment on the Green Hydrogen Supply Chain: Patagonia – Rotterdam	256
33	An experimentally validated numerical model of a hybrid spectral splitting concentrating photovoltaic-thermal collector	266
34	Design and Comparison of Dual-Use Terminals for Cold Energy Recovery during Liquid Hydrogen Regasification	276
35	Synthesis and testing of polymer membranes of intrinsic microporosity for use in water nanofiltration	(*)
36	Addressing compositional constraints, data imbalance, and machine learning interpretability in the diagnosis of inflammatory bowel disease and colorectal cancer using glycans	(*)
37	Characterisation of structural and kinetic properties of DETA-functionalised polymeric resin adsorbents for direct air capture of CO_2	296
38	Computational Modelling of Blood Clot Lysis: The Effect of Clot Heterogeneity on Lytic Efficacy	306
39	Optimisation of Feed Conditions of Catalysed Inverse Vulcanisation with Rapeseed Oil and Waste Cooking Oil	316
40	A Techno-Economic Analysis and Systematic Review of Hydrogen Storage and Transportation Technologies	325
41	Protein-based RNA integrity assay-relationship between mRNA capping ratio and reporter fluorescence	338
42	DEC Production Reactor Design at Commercial Scale	348
43	The effect of bentonite content on the equilibrium and kinetic CO_2 adsorption properties of zeolite 13X shaped adsorbents	357

paper	Title	page
44	Techno-Economic Assessment of Oxyfuel-Based Carbon Capture and Storage (CCS) in the Cement Industry	367
45	Reinforcement Learning for Constrained Continuous Stirred-Tank Reactors Systems	377
46	Assessing different methodologies for characterisation of carbon persistence in biochar	384
47	Energy System Modelling to Decarbonise Zambia's Transport Sector	394
48	Optimisation of Electrochemical Bleaching for Lignocellulosic Biomass	404
49	Superhydrophobic Cotton and PU Sponge Modified by Silica Nanoparticles Coating with Octadecyltrimethoxysilane	414
50	Screening the chemical space: An ML Approach to Predicting Synthesis Costs of Chemical Compounds	423
51	Incorporating Manufacturing Uncertainty and Sustainability into Supply Chain Modelling for Autologous CAR-T Cell Therapy	433
52	An imaged-based workflow to characterise multi-scale porosity in rocks	443
53	Quantifying Cellular Burden Using RNA Aptamer-based Fluorescent Biosensors	(*)
54	Investigating Liquid-Side Mass Transfer in Wavy-Laminar Falling Films	453
55	Evaluating future pathways for decarbonization in the UK's Road Vehicles sector through agent-based modelling	463
56	A Deep Deterministic Policy Gradient for Distributed Residential Energy Control	473
57	Numerical Approximation of Multivariate Functions using Kolmogorov-Arnold Representation Theorem	483
58	Expanding the applicability of the SAFT-γ Mie group-contribution approach to model thermodynamic properties of pure acetonitrile	493

paper	Title	page
59	Spontaneous Non-Planar Material Assembly via Frontal Photopolymerisation	501
60	Multi-fidelity Bayesian Optimisation for Solvent selection	(*)
61	Optimisation of hard carbon-based cathodes for sodium-ion batteries through varying binder content	511
62	A novel CoPx@C cocatalyst for enhanced selective visible light photocatalytic C-C coupling of methanol to ethylene glycol over CdS	520
63	Computational fluid dynamics modeling of Onyx liquid embolic system	530
64	Techno-Economic Analysis and Life Cycle Assessment of Sustainable Aviation Fuel: Comparing Direct Air Capture and Point-Source CO2 Capture Technologies	538
65	Process Flexibility Analysis of a Robotic Parallel Mini Bioreactor Platform for High-Throughput Bioprocess Development	548
66	Optimal Control of PV-Battery Systems in Commercial Buildings Using Tabular Q-Learning	558
67	ASPEN Integrated 1-Dimensional Non-Equilibrium Plasma Model for CO_2 Dissociation	568
68	A Preliminary Investigation of Large Language Models (LLMs) as Process Operator Assistants	577
69	Multi-Objective Real-Time Optimisation of Load Allocation for Gas-Oil Separation Facilities	587
70	Optimisation of Translational Efficiency in a Pichia Pastoris Cell Free System	597
71	Exploration of viscosity trends in ZIF-8–ionic liquid dispersions at varied mass loading for type III porous liquid development	607
72	Investigation of first-principles surrogate models for Protein crystallisation	617
73	Exploring the Growth and Fluorescent Protein Expression of Encapsulated Pichia with Varying Concentrations of Methanol	627

paper	Title	page
74	Effect of Agitation on Isothermal Paracetamol Crystal Regeneration	637

(*) These papers have been removed by request of the authors and/or supervisors

Author and Supervisor index at the end of the book.

Electrodeposition of Manganese Dioxide from Lithium-ion Battery Leachate

Yuan Xu, Zhenyu Xu
Department of Chemical Engineering, Imperial College London, U.K.

Abstract Nowadays, with lithium-ion batteries (LIBs) playing a crucial role in sustainable energy storage, the need for efficient recycling methods to address the scarcity of critical metals and mitigate environmental challenges is more demanding than ever. While traditional hydrometallurgy methods rely heavily on solvent extraction with highly toxic chemicals, techniques like electrodeposition offer a more sustainable alternative. LIBs contain valuable metals, including lithium (Li), nickel (Ni), manganese (Mn), and cobalt (Co), which are dissolved through acid leaching treatment. This research focuses on Mn as the valuable metal, with Mn-rich deposits prepared on both and within porous carbon paper substrate via electrodeposition in a three-electrode configuration. The electrochemical performance of the electrode materials was evaluated using software galvanostatic charge-discharge (GCD) and cyclic voltammetry (CV), while characterization techniques such as scanning electron microscopy with energy dispersive X-ray spectroscopy (SEM-EDS) were implemented to confirm successful deposition and provide insights into surface chemistry. The objective of this research was to test the capability for direct application of the electrodeposited material, thereby demonstrating the potential of electrodeposition as a viable method for future battery recycling. Under optimized conditions of 0.1 M NMC-111 solution at pH 5 and an applied potential of 1.0 V vs Ag/AgCl, combined with acid pre-treatment of the carbon paper to enhance the surface morphology, the resulting C/MnO_2 electrode achieved a high specific capacitance of approximately 300 F/g and demonstrated stable performance over 100 charge-discharge cycles. These findings underline the effectiveness of electrodeposition for producing high-performance electrode materials and highlight its potential as a sustainable and scalable approach to LIB recycling.
Key words: Electrodeposition, supercapacitor, MnO_2, lithium-ion batteries

Introduction:

Nowadays, lithium-ion batteries (LIBs) have become integral to modern energy solutions, owing to their high energy density [1], high coulombic efficiencies, great cyclability and ideal working thermal range. In particular, LIBs are the state-of-art for applications in electric vehicles (EVs) and renewable energy storage systems. According to International Energy Agency's (IEA) data [2], world LIB demand increased by about 65% in 2022 compared to 2021, especially in China, battery demand for vehicles grew over 70%, while electric car sales increased by 80%. It is anticipated that by 2030, renewables will contribute up to 80% of new power generation, with LIB playing a crucial role in sustainable energy storage.

Currently, there are five commercially available LIBs cathode chemistries, which is LCO ($LiCoO_2$), LMO ($LiMn_2O_4$), LFP ($LiFePO_4$), NCA ($LiNiCoAlO_2$) and NMC ($LiNiMnCoO_2$) listed in chronological order of their commercialisation [3]. Among the five types of LIBs, NMC batteries are the most widely used, owing to their high volumetric energy density and stable cyclability [4][5].

However, the widespread use of LIBs has brought challenges including the scarcity of critical raw materials such as lithium (Li), nickel (Ni), manganese (Mn) and cobalt (Co), as well as the environmental impact of mining and battery waste.

Previous research had indicated that LIB related waste will exceed 11 million tonnes from 2017 to 2030 [6]. Therefore, recycling spent LIBs, particularly NMC cathodes, offers a sustainable solution for addressing these issues while reducing the requirement of critical metals. The current existing recycling approaches are mainly pyrometallurgy and hydrometallurgy method. Both pyrometallurgy and hydrometallurgy approaches require a series of pre-treatments, including discharging and disassembling the battery, shredding it into the so-called black mass, which contains the active cathode and anode material [7]. In the pyrometallurgy route, the black mass is directly exposed to high temperatures above 1000 ℃, resulting in the formation of a lithium-rich slag and an alloy composed of elements such as Ni and Co. Pyrometallurgy has the advantages of burning off the hazardous compounds (electrolyte), being a mature process and scalable for industrial applications [8]. However, large input of energy is required, and large quality of toxic gases will be produced, resulting in extensive scrubbing afterward to prevent them from polluting the environment [9]. In traditional hydrometallurgy methods, the black mass undergoes a leaching process, where it is treated with acids to dissolve the valuable metals, such as Co, Mn and Ni. Those Metals are then be separated via one of the three processes: precipitation, solvent extraction or selective adsorption [10]. Among those three processes, solvent extraction is most widely used. This method offers advantages such as higher recovery purity of active materials and lower energy consumption [11]. However, the utilization of highly toxic organic solvents [12] and low selectivity of separating metals with similar chemical properties limit its

application. The electrodeposition method utilises electric current to oxidise/reduce dissolved metal cations, enabling the formation of a coherent metal/metal oxide coating on an electrode [13]. Ideally by controlling the operation potential, electrodeposition can extract any target chemicals from the leaching solutions, by allowing them to be deposited on a suitable substrate. Recently, Liu et al. demonstrated that with optimised conditions, electrodeposition could achieve MnO_2 of 99.52 % purity at the anode and a Ni-Co alloy of 99.80 % purity at the cathode, both of which are market ready [14]. This indicates that research needs to be put in to discover the full potential of electrodeposition.

Supercapacitors are electrochemical energy storage devices which store energy either via static adsorption or redox reaction. They exhibit unique features of high-power density, fast charging/discharging rates, long lifespan and safe operation [14]. Among various electrode materials, manganese dioxide (MnO_2) is a promising supercapacitor electrode material due to its ultra-high theoretical capacitance (1370 F/g) [16]. Notably, the potential use of MnO_2 obtained by electrodeposition as a high-performance supercapacitor material highlights the strengths of the process, offering a sustainable and scalable method for material recovery. Zhang et al. demonstrated that electrodeposited MnO_2 exhibits high specific capacitance (469 F/g at 1 A/g) and excellent cycling stability, with a retention rate of 83.9% after 2500 cycles [17]. According to Kang et al., a MnO_2/nanoporous gold/MnO_2 sandwich electrode can achieve high specific capacitance of 841 F/g at the scan rate of 5 mV/s [18]. This study focuses on the recovery of MnO_2 from a simulated leachate of waste LIBs, using a system designed to model NMC 111 composition, containing 0.1 M of $NiSO_4$, $MnSO_4$, and $CoSO_4$ along with 0.25 M Na_2SO_4, via electrodeposition. By optimizing electrodeposition conditions, including adjusting the solution pH and pre-treating the carbon paper substrate, the aim is to enhance the quality and electrochemical performance of MnO_2 for supercapacitor applications. This work contributes to sustainable recycling practices while showcasing the potential of MnO_2 as a high-performance energy storage material, paving the way for advancements in green energy solutions.

Background:

The previous work by Puru Goel (2023/2024 MSc student) and Dr. Xiaochu Wei (PDRA) focused on understanding the electrodeposition process and investigating the effects of pH and applied electrode potential on the purity of electrodeposited MnO_2 [19]. Their findings, supported by the Pourbaix diagrams of Ni, Co, and Mn (Figures 1a–1c), revealed the potential for undesired oxidation products apart from MnO_2 to form during the process. The relationship between SHE and Ag/AgCl can be expressed by the following equation:

$$E_{SHE}/V = E_{Ag/AgCl} + 0.205 \quad (1)$$

Within the range of pH from 3~7, the

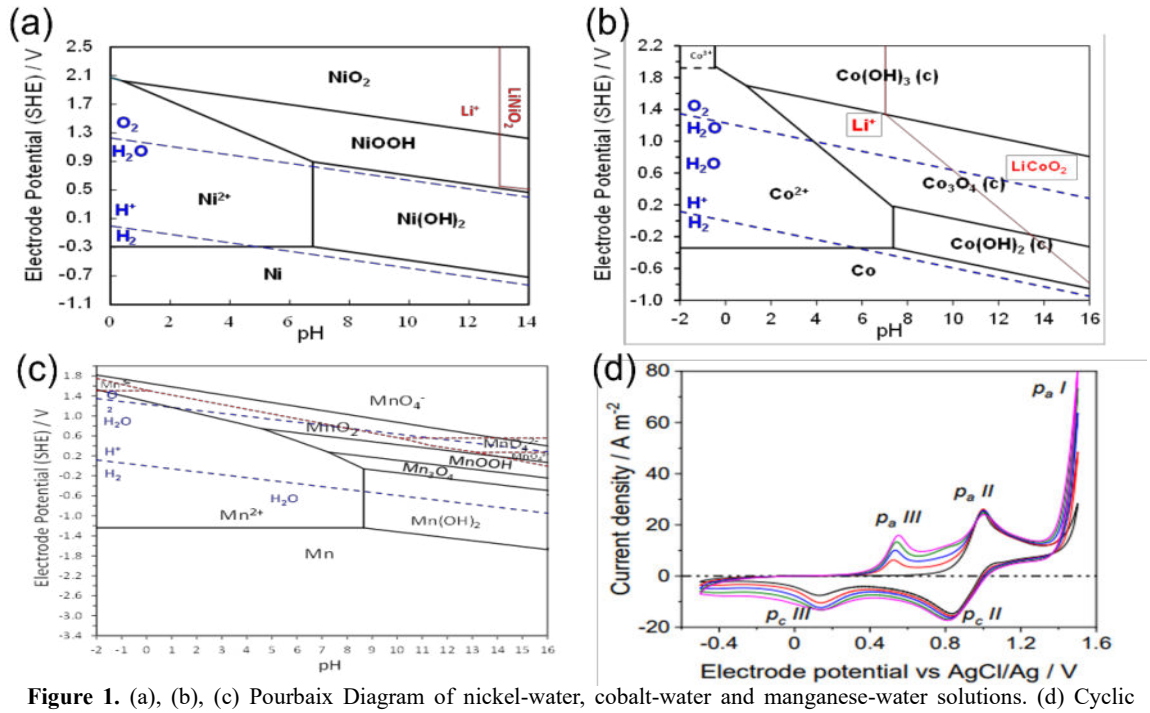


Figure 1. (a), (b), (c) Pourbaix Diagram of nickel-water, cobalt-water and manganese-water solutions. (d) Cyclic voltammetry Diagram for solution made by $NiSO_4$, $MnSO_4$ and $CoSO_4$, all fixed at 0.01 M, using 0.25 M Li_2SO_4 as electrolyte [19].

2

formation of undesirable oxidation products, such as nickel oxide hydroxide (NiOOH) and cobalt (II, III) oxide (Co_3O_4), were expected. These impurities are detrimental to the electrode material quality and negatively impact its electrochemical performance. Therefore, pH 3–6 was identified as the optimal range for the electrodeposition process.

The CV scan of the NMC-111 solution (Figure 1d) illustrated the electrochemical activity during electrodeposition. To produce high-purity MnO_2 and limit side reactions involving Ni and Co, the applied potential should remain below ± 1.3 V versus Ag/AgCl. These parameters provide a foundation for optimizing the electrodeposition process and improving the performance of MnO_2-carbon paper electrode materials.

Methodology

Materials

Sodium sulfate anhydrous (99%), nickel (II) sulfate hexahydrate (98%), cobalt (II) sulfate heptahydrate (98%), manganese (II) sulfate monohydrate (99%), sodium hydroxide (97%), nitric acid (68%) and 2×3 cm carbon paper.

Electrodeposition

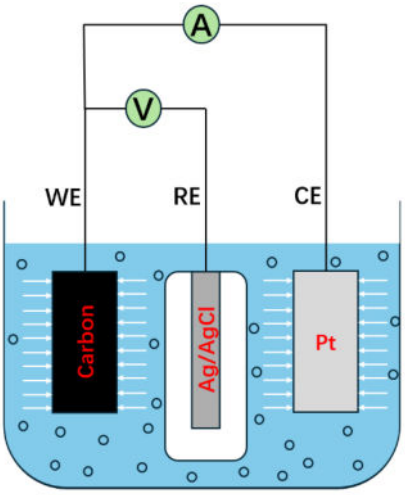


Figure 2. Electrodeposition and GCD test experimental lay-out.

The electrodeposition test was conducted in a 100 mL beaker, which contained 75 mL of NMC-111 solution with a three-electrode setup as shown in Figure 2. The working electrode is pre-heated carbon paper that has been heated to 300 °C for 24 hours and cut into 2D rectangular shape. The electronic connection to the carbon paper was done using an adhesive copper tape, and the connection part was sealed with waterproof tape to ensure the copper did not react with the solution, which would cause localised corrosion. The main purpose of the heat treatment is to remove impurities from the carbon paper and to change its surface structure by creating more pores to increase the surface area, which is critical in improving the performance of the electrodeposition process [20]. The exact deposition area cannot be determined directly due to the porous morphology of the carbon paper, so the geometric surface area of the carbon paper (3 cm^2) was used instead. Ag/AgCl was used as the reference electrode, and all applied potentials mentioned in the later sections are referenced against Ag/AgCl. Platinum was used as the counter electrode. The Pt electrode needs to be polished with abrasive silicon carbide before every experiment to remove all the impurities on its surface. It subsequently needs to be bathed in 1M sulfuric acid for 12 hours to remove any impurities from its surface. The stirring of the solution during electrodeposition was set to 200 RPM for maintaining the pH uniformity in the solution and improving mass transport.

The potentiostat was used to provide different potential for every test, which was controlled by using the software NOVA. The cut-off charge was set at 20 C. The mass of MnO_2 deposited was estimated by using Equation 2, where Q is the total charge supplied, F is the Faraday's constant which is 96485 C/mol and assuming all consumed charge were used for Mn^{2+}/MnO_2 transition:

$$m_{MnO_2} = \frac{Q}{2 \times F} \times M_{MnO_2} \quad (2)$$

Galvanostatic charge-discharge

The setup for the galvanostatic charge-discharge (GCD) tests was the same as the configuration used in the electrodeposition experiments, ensuring consistency in experimental methodology. The working electrode was carbon paper with coated electrodeposited MnO_2. To remove any impurities, the electrodeposited carbon paper was immersed into the deionized (DI) water for 12 hours, followed by 8 hours oven-dried process at 80℃ before the GCD testing. 1M sodium sulfate (Na_2SO_4) was selected as the electrolyte to facilitate ion transport during the GCD testing process.

The potentiostat was switched to supply current this time. The cut-off potential was set at 0.8V which is the typical value for electrochemical analysis [21].

The specific capacitance of the electrodeposited carbon paper was calculated by using Equation 3:

$$C_{sp} = i\frac{\Delta t}{\Delta V} \quad (3)$$

where i is the current density, which is capped at 1 A/g throughout the experiment, Δt is the discharging time and ΔV is the potential difference.

Pre-treatment of carbon paper

The acidic and basic treatments are capable of changing the surface morphology of carbon paper. Both acid and basic treatments will modify the surface of carbon paper, resulting in more pores and

3

increased surface area.

In the acid treatment process, the pre-heated carbon paper was immersed in 65% nitric acid for 2.5 hours to allow sufficient time for the acid to react with the carbon paper, leading to partial corrosion that increases surface porosity and creates more active sites for electrodeposition. Following acid treatment, the carbon paper was completely rinsed and soaked in DI water to remove residual acid. Finally, the carbon paper was dried in the oven for 8 hours, preparing for further experiments.

Similarly, the basic treatment followed the exact same procedures of the acid treatment, with the only difference being submerging the carbon paper in 2.5M sodium hydroxide solution instead of 65% nitric acid.

Results and Discussion

The result presented signifies the importance of optimised electrodeposition conditions for MnO_2 recovery. By systematically exploring the effects of pH, potential and substrate pre-treatment, we demonstrate the conditions under which optimal material properties and stable supercapacitor performance can be obtained.

Varying pH and potential:

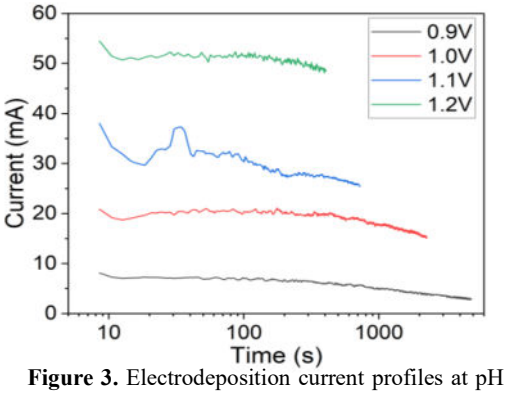


Figure 3. Electrodeposition current profiles at pH 5 under different applied potentials on carbon paper.

Figure 3 demonstrates the current trend during the electrodeposition process at different potentials. As expected, the current generally increases with potential. The 0.9V sample exhibited the lowest current throughout the process, consistently below 10 mA throughout the process. This observation is supported by the Pourbaix diagram of Mn (Figure 1c), which indicates that 0.9V vs Ag/AgCl is barely sufficient for Mn^{2+}/MnO_2 transition. In combination, these observations suggest slow reaction kinetics and an insufficient amount of MnO_2 formation. The currents of the 1.0V and 1.1V samples were relatively close, as these potentials were sufficient to enable Mn^{2+} oxidation to MnO_2 while minimizing undesired side reactions, as shown in Figure 1. In contrast, the 1.2V sample illustrates a significantly higher current, suggesting the occurrence of undesired side reactions, such as MnO_2 oxidation to MnO_4^-, as this potential is sufficient to activate the reaction. These side reactions possibly reduced the purity and performance of the deposited material. Consequently, 1.0V and 1.1V were identified as the optimal potentials, balancing efficient MnO_2 deposition with minimal side reactions, thereby ensuring stable and high-quality electrode materials for supercapacitor applications.

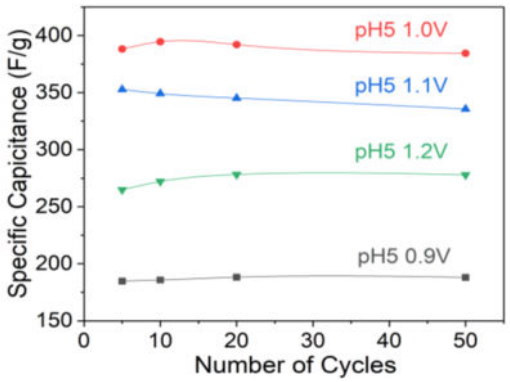


Figure 4. Specific capacitance of electro-deposited carbon paper under different applied potentials over 50 cycles of charge-discharge.

Figure 4 illustrates the specific capacitance of the electrodeposited samples at different potentials over 50 charge-discharge cycles. All samples exhibited supercapacitor properties with relatively stable capacitance, indicating good cyclability. The 0.9V sample shows the lowest specific capacitance, consistent with the earlier observations of lowest electrodeposition current, verified insufficient MnO_2 formation during electrodeposition. In contrast, the 1.0V and 1.1V samples achieve significantly higher specific capacitance, proved that these potentials are optimal for MnO_2 deposition with minimal side reactions. The specific capacitance of the 1.2V sample is about 275 F/g, while higher than 0.9V (180 F/g), is notably lower than those of 1.0V and 1.1V, likely due to side reactions at this higher potential which suppress MnO_2 formation.

Among four tested samples, the 1.0V sample exhibited the highest specific capacitance, reaching approximately 400 F/g, which is competitive compared to values reported in the literature [22]. This result highlights 1.0V as the optimised potential for achieving high-performance supercapacitor materials. Furthermore, the stable capacitance observed across all samples over 50 cycles underlined the promising cyclability of the electrodeposited MnO_2, solidifying its potential as a candidate for energy storage applications.

4

Table 1. SEM-EDS analysis of composition of electro-deposited metals on carbon paper at different potentials.

Potential	Mass % of metals on carbon paper			Mass % of metals on total metal deposited		
	Mn	Ni	Co	Mn	Ni	Co
1.0V	69.19	0.37	4.47	91.75	0.48	5.92
1.1V	64.33	4.76	9.14	81.02	5.99	11.51

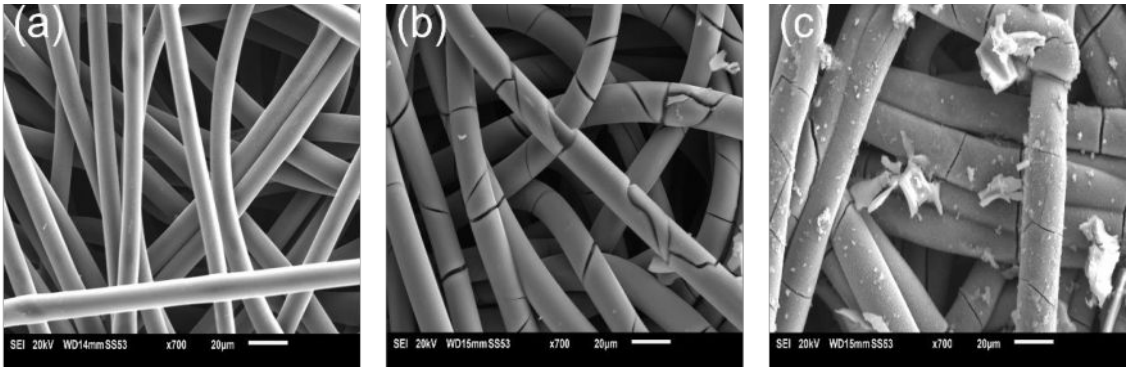


Figure 5. SEM images of carbon paper at x700 magnification: (a) Blank (no electrodeposition) (b) Electrodeposited at 1.0V (c) Electrodeposited at 1.1V.

Table 1 summarised the mass percentages of the key metal elements on the electrodeposited carbon paper. Mn is the dominant species under both 1.0V and 1.1V conditions, accounting for over 80 % of the total metal deposited in both samples. However, the 1.1V sample showed significantly higher percentages of Ni and Co at 5.99 % and 11.51 % respectively, compared to 0.48% and 5.82% in the 1.0V sample. This observation indicates that the higher applied potential promotes side reactions involving Ni and Co. Therefore, the lower specific capacitance of the 1.1V sample is due to its lower Mn composition, and higher levels of undesired products.

Figure 5 displays SEM images of the blank (no electrodeposition), 1.0V, and 1.1V electrodeposited samples. The blank sample exhibits a smooth and uniform carbon fibre surface, while the 1.0V and 1.1V samples show visible cracks and deposits on the surface, indicating electrodeposition had altered the surface morphology of the carbon paper. Although the 1.1V sample exhibited more pronounced deposits, previous EDS analysis reveals it contains higher proportions of impurities (Ni and Co) and lower Mn content compared to the 1.0V sample, meaning the deposits observed are more associated with these impurities rather than the desired MnO_2. In summary, the SEM-EDS data supports the conclusion that 1.0V is the optimal potential for achieving high-purity MnO_2 deposition and superior supercapacitor performance.

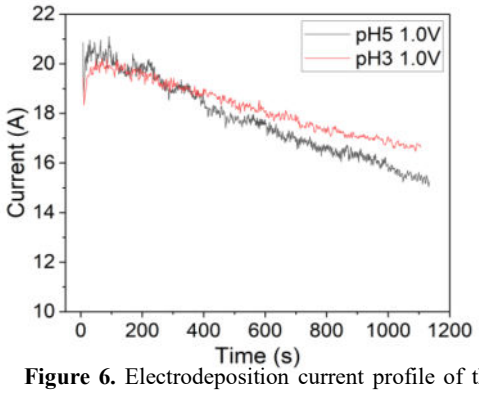


Figure 6. Electrodeposition current profile of the carbon paper under 1.0V applied potential at different pH.

After determining the optimal potential for electrodeposition, the effect of pH on the electrodeposition process and the resultant supercapacitor performance was evaluated. As shown in Figure 6, the electrodeposition currents for the pH 5 and pH 3 samples at 1.0V were similar, with negligible differences in the time required to reach the target charge. This indicated that monitoring the current profile alone during electrodeposition is insufficient to predict the performance of the resultant electrode materials as supercapacitors. Therefore, further electrochemical analysis, such as charge-discharge behaviour of the electrodeposited carbon paper, is necessary to comprehensively assess the influence of pH on the material's performance.

5

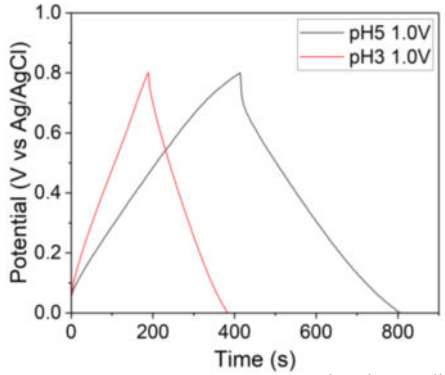


Figure 7. GCD curves compare the charge-discharge performance of the electro-deposited carbon paper under 1.0V at pH3 and pH5 (10^{th} Cycle).

As shown in Figure 7, the GCD curves at the 10th cycle were selected for both pH 5 1.0V and pH 3 1.0V samples. This selection is based on the specific capacitance data demonstrated in Figure 4, which indicated that approximately 10 cycles were required for the capacitive behaviour of the electrodeposited carbon paper to stabilize. The GCD curves of both samples display isosceles triangular shapes, indicating excellent reversibility, characteristics of capacitive behaviour [17]. The substantially longer charge-discharge duration observed for the pH 5 1.0V sample indicates a notably higher charge storage capacity compared to the pH 3 1.0V sample. This superior performance under slightly acidic conditions (pH 5) signifies the necessity of enhanced electrodeposition process, which likely promotes the formation of MnO_2 with improved electrochemical properties.

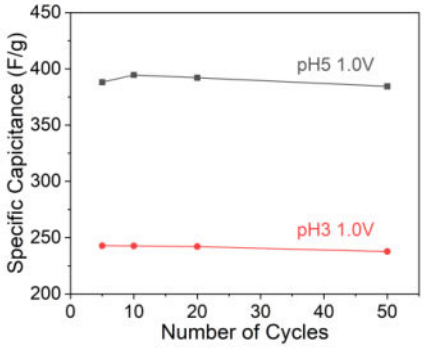


Figure 8. Specific capacitance of electro-deposited carbon paper under different pH over 50 cycles of charge-discharge.

To conclude, the supercapacitor performance of the electrodeposited samples under pH 3 1.0V and pH 5 1.0V conditions was compared, as shown in Figure 8. The results clearly demonstrated that the condition of pH 5, yielded a significantly higher specific capacitance compared to what obtained from the pH 3 environment. This indicates that pH 5 environment is more favourable for the formation of MnO_2 during the electrodeposition process, likely due to the suppression of undesired side reactions and the enhanced quality of the deposited material.

Although the pH 3 sample exhibited stable cyclability over 50 charge-discharge cycles, its 40 % lower specific capacitance compared to the pH 5 sample limits its applicability in further analysis. The findings reinforced the importance of optimizing pH conditions to achieve enhanced material properties and electrochemical performance, enabling the practical application of MnO_2- carbon paper electrodes in supercapacitors.

Acidic/Basic pre-treatment

From the tests above, the performance under the conditions of pH 5 and 1.0 V was determined to be optimal. Therefore, the subsequent experiments were conducted under these optimized conditions to investigate the effects of different pre-treatment methods. The objective is to find the pre-treatment that yields electrode materials with the best supercapacitor performance.

Additionally, the area of the carbon paper substrate and the cut-off charge were halved to 10 C, compared to the previous tests on pH and potential. This adjustment was made to minimize side reactions mentioned in background and to ensure that the working electrode and counter electrode areas are comparable, which is crucial for obtaining reliable and consistent electrochemical testing results. However, the halved charge was expected to drastically reduce the total amount of deposited metal, including MnO_2, which would likely result in a lower specific capacitance.

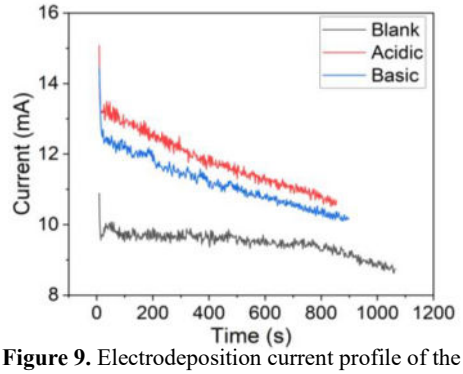


Figure 9. Electrodeposition current profile of the carbon paper under different pre-treatments, where Blank means no pre-treatment.

Figure 9 demonstrates the electrodeposition current profiles of the blank (no pre-treatment) carbon paper sample and acid/basic pre-treated carbon paper samples. From the figure, it's evident that both acidic- and basic- treated samples exhibited significantly higher current compared to the untreated sample throughout the deposition process. This suggests that the pre-treatment process enhances electrodeposition kinetics. A possible explanation for this phenomenon is that acidic and basic treatments modify the surface properties of the

6

Table 2. SEM-EDS analysis of composition of electro-deposited metals on carbon paper with different pre-treatment methods.

Pre-treatment	Mass % of metals on carbon paper			Mass % of metals on total metal deposited		
	Mn	Ni	Co	Mn	Ni	Co
Acidic	67.44	1.19	4.47	90.08	1.54	6.00
Basic	60.14	6.67	9.30	77.50	8.56	12.02

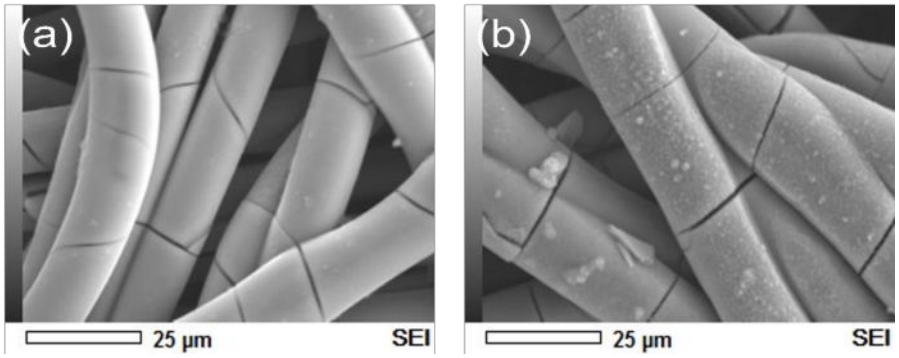


Figure 10. SEM images electrodeposited carbon paper under different pre-treatments: (a) Acidic treatment (b) Basic treatment.

carbon paper. The pre-treatment likely increases surface roughness, thereby introducing more active sites for MnO_2 nucleation and growth during the electrodeposition. These additional active sites facilitate improved charge transfer, leading to higher current measured. In contrast, blank carbon paper is likely to have fewer active sites and a smoother surface. This results in lower reaction kinetics and reduces current throughout the process. Although the deposited mass of MnO_2 is constrained by the charge limit at 10 C, quality and distribution of the MnO_2 layer formed could vary. Acidic and basic pre-treatment could promote the formation of a more uniform and adherent MnO_2 layer, which is crucial for the electrochemical performance of the material as a supercapacitor. These improvements may ultimately enhance the capacitance and stability of the electrode.

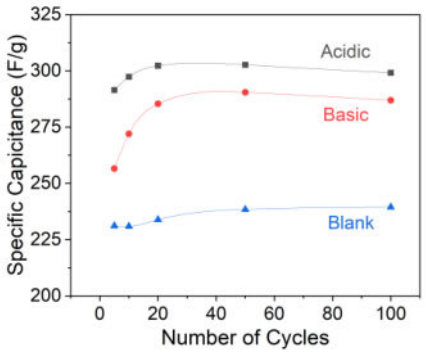


Figure 11. Specific capacitance of electro-deposited carbon paper with different pre-treatments over 100 cycles of charge-discharge.

As shown in Figure 11, all three electrodeposited carbon paper exhibited stable performance, with the acidic-treated sample achieving the highest specific capacitance (~300 F/g), followed by the basic-treated and blank samples. The enhanced performance of both acidic- and basic-treated samples than the blank one is attributed to increased porosity, resulting from the chemical interaction between the treatment solutions and the carbon paper substrate. This interaction creates more pores and voids in the carbon paper, providing additional active sites for MnO_2 deposition and enhancing interaction with the electrolyte during the GCD testing. This higher porosity also explains the approximately 20-cycle stabilization period observed for both the acidic- and basic-treated samples, as the abundance of active sites and pore space required more time for electrochemical behaviours to stabilise. Due to its stronger corrosive effect, acidic treatment generates greater porosity than basic treatment, which translates into improved capacitive performance. Additionally, acidic treatment introduces more oxygen-containing functional groups (e.g. carbonyl groups) onto the carbon paper surface, enhancing surface polarity and hydrophilicity. These properties in combination contribute to the improved electrochemical performance and energy storage capacity of the acid-treated carbon paper. [23]

7

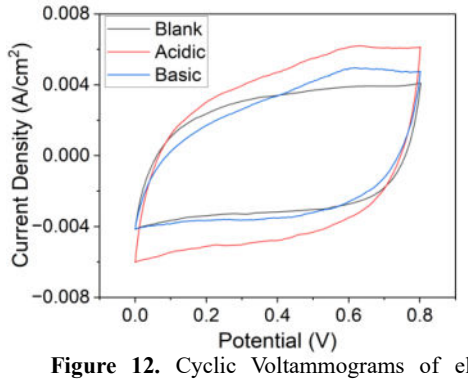


Figure 12. Cyclic Voltammograms of electro-deposited carbon papers with different pre-treatments at a scan rate of 10 mV/s.

Figure 12 compares the CV profiles of the three samples at a scan rate of 10 mV/s. It's evident that the three samples all displayed quasi-rectangular shape curves, which is a typical feature of capacitive behaviours, confirming the abilities of the electro-deposited carbon paper as supercapacitor materials [18]. The difference in current density can be attributed to surface modifications introduced by acidic and basic treatment, which is consistent with the hypothesis that pre-treatment improves material's electrochemical performance. Among the three samples, the acidic-treated sample has the highest current density, followed by the basic and the blank sample. The improved current density for acidic treatment indicates higher charge storage capacity, which correlates with GCD data where acidic treatment carbon paper had the most specific capacitance.

The SEM-EDS results shown in Table 2 confirm that the acidic pre-treatment of carbon paper is the optimal pre-treatment method for electrodeposition. Specifically, the acidic pre-treated carbon paper achieved the highest Mn deposition, with manganese making up 90.08% of the total deposited metal content. In contrast, the basic pre-treatment resulted in significantly lower Mn deposition of 77.50% and higher proportions of undesired metals: Ni and Co. Furthermore, the SEM images in Figure 10a-10b revealed no visible morphological differences apart from the more visible deposits on the basic-treated sample, which are consistent with earlier findings on the presence of higher impurity levels (e.g., Ni and Co), Similar to the comparison between 1.0V and 1.1V sample, this confirms the visible deposits are more related to impurities rather than desired MnO_2. This highlights the limitations of SEM in distinguishing structural or chemical differences.

These findings underline the importance of substrate pre-treatment in optimizing the electrodeposition process to maximize MnO_2 purity and enhance the electrochemical performance of the resulting materials.

Conclusion and Outlook

This study demonstrates that the optimal conditions for electrodeposition to achieve stable and high-performance supercapacitor electrodes involve pH 5 and a deposition potential of 1.0 V versus Ag/AgCl. Additionally, pre-treating the carbon paper substrate with acid significantly enhances the quality of the deposited material. Under these conditions, the resultant electrode material exhibited a high specific capacitance of approximately 300 F/g and maintained stable performance over 100 charge-discharge cycles. Characterization techniques such as SEM and EDS confirmed the uniformity and purity of the deposited MnO_2 layer, emphasizing the importance of deposition conditions in controlling material properties. The observed high capacitance and stability demonstrate the potential of MnO_2 recovered from NMC 111 solutions for use in supercapacitors.

However, it is important to acknowledge that SEM-EDS alone cannot fully explain the surface chemistry and structural details of the electrodeposited material. EDS, although effective for elemental composition determination, cannot provide detailed information about chemical bonding or functional groups. Future works should utilize advanced characterization techniques such as Brunauer-Emmett-Teller (BET) analysis to assess the specific surface area, X-ray photoelectron spectroscopy (XPS) to investigate surface chemistry, and electrochemical impedance test (EIS) to evaluate charge transfer properties. These techniques will provide deeper insights into the interface chemistry and electrochemical behavior of the electrodeposited materials, setting the stage for their further advancements in energy storage applications.

Acknowledgement

We would like to sincerely thank Dr. Xiaochu Wei for her invaluable support and guidance throughout this project. We are also deeply grateful to our supervisor, Dr. Anna Hankin, for her constructive feedback and encouragement, which has been crucial in the successful completion of this research.

Reference

[1] Kim, T.; Song, W.; Son, D.-Y.; Ono, L. K.; Qi, Y. Lithium-Ion Batteries: Outlook on Present, Future, and Hybridized Technologies. Journal of Materials Chemistry A 2019, 7 (7), 2942–2964. https://doi.org/10.1039/c8ta10513h.

[2] IEA. Global EV Outlook 2023 – Analysis. IEA. https://www.iea.org/reports/global-ev-outlook-2023 (accessed 2024-11-30).

[3] Zubi, G.; Dufo-López, R.; Carvalho, M.; Pasaoglu, G. The Lithium-Ion Battery: State of the Art and Future Perspectives. Renewable and Sustainable Energy Reviews 2018, 89, 292–308. https://doi.org/10.1016/j.rser.2018.03.002.

[4] Moiseev, I. A.; Savina, A. A.; Pavlova, A. D.; Abakumova, T. A.; Gorshkov, V. S.; Pazhetnov, E. M.; Abakumov, A. M. Single Crystal Ni-Rich NMC Cathode Materials for Lithium-Ion Batteries with Ultra-High Volumetric Energy Density. Energy Advances 2022, 1 (10), 677–681. https://doi.org/10.1039/d2ya00211f.

[5] Ran, A.; Chen, S.; Cheng, M.; Liang, Z.; Li, B.; Zhou, G.; Kang, F.; Zhang, X.; Wei, G. A Single-Crystal Nickel-Rich Material as a Highly Stable Cathode for Lithium-Ion Batteries. Journal of materials chemistry. A, Materials for energy and sustainability 2022, 10 (37), 19680–19689. https://doi.org/10.1039/d2ta01186g.

[6] Natarajan, S.; Aravindan, V. Recycling Strategies for Spent Li-Ion Battery Mixed Cathodes. ACS Energy Letters 2018, 3 (9), 2101–2103. https://doi.org/10.1021/acsenergylett.8b01233.

[7] Davis, K.; P. Demopoulos, G. Hydrometallurgical Recycling Technologies for NMC Li-Ion Battery Cathodes: Current Industrial Practice and New R&D Trends. RSC Sustainability 2023, 1 (8), 1932–1951. https://doi.org/10.1039/D3SU00142C.

[8] Yin, H.; Xing, P. Pyrometallurgical Routes for the Recycling of Spent Lithium-Ion Batteries. Springer eBooks 2019, 57–83. https://doi.org/10.1007/978-3-030-31834-5_3.

[9] Zhou, J.; Zhou, X.; Yu, W.; Shang, Z.; Xu, S. Towards Greener Recycling: Direct Repair of Cathode Materials in Spent Lithium-Ion Batteries. Electrochemical Energy Reviews 2024, 7 (1). https://doi.org/10.1007/s41918-023-00206-5.

[10] Bae, H.; Kim, Y. Technologies of Lithium Recycling from Waste Lithium-Ion Batteries: A Review. Materials Advances 2021, 2 (10). https://doi.org/10.1039/d1ma00216c.

[11] Wu, J.; Xiao, L.; Shen, L.; Ran, J.-J.; Zhong, H.; Zhu, Y.-R.; Chen, H. Recent Advancements in Hydrometallurgical Recycling Technologies of Spent Lithium-Ion Battery Cathode Materials. Rare metals/Rare Metals 2023, 43 (3), 879–899. https://doi.org/10.1007/s12598-023-02437-3.

[12] Sahu, S.; Devi, N. Hydrometallurgical Treatment of Spent Lithium Ion Batteries Using Environmentally Friendly Leachant and Extractant. Journal of Material Cycles and Waste Management 2023, 25 (6), 3303–3315. https://doi.org/10.1007/s10163-023-01754-0.

[13] Paunovic, M.; Schlesinger, M. Fundamentals of Electrochemical Deposition; John Wiley & Sons, 2006.

[14] Liu, G.; Chen, Z.; Luo, F.; Liu, T.; Xi, X.; Wang, Z.; Gao, Z.; Shao, P.; Wu, D.; Luo, X.; Yang, L. One-Step Nickel-Cobalt Alloy Electrodeposition from Spent Lithium-Ion Battery via Synergistic PH Adjustment and Mn2+ Supplementation. Separation and Purification Technology 2023, 314, 123581 – 123581. https://doi.org/10.1016/j.seppur.2023.123581.

[15] Han, Z.; Fang, R.; Chu, D.; Wang, D.-W.; Ostrikov, K. (Ken). Introduction to Supercapacitors. Nanoscale Advances 2023, 5 (16), 4015–4017. https://doi.org/10.1039/D3NA90074F.

[16] Zhu, J.; Zhang, D.; Zhu, Z.; Wu, Q.; Li, J. Review and Prospect of MnO2-Based Composite Materials for Supercapacitor Electrodes. Ionics 2021, 27 (9), 3699–3714. https://doi.org/10.1007/s11581-021-04139-1.

[17] Zhang, M.; Chen, Y.; Yang, D.; Li, J. High Performance MnO2 Supercapacitor Material Prepared by Modified Electrodeposition Method with Different Electrodeposition Voltages. Journal of Energy Storage 2020, 29, 101363–101363. https://doi.org/10.1016/j.est.2020.101363.

[18] Kang, J.; Chen, L.; Hou, Y.; Li, C.; Fujita, T.; Lang, X.; Hirata, A.; Chen, M. Electroplated Thick Manganese Oxide Films with Ultrahigh Capacitance. Advanced energy materials 2013, 3 (7), 857–863. https://doi.org/10.1002/aenm.201201046.

[19] Wei, X. Recovering Materials from End of Life Lithium-ion Batteries. PhD Thesis, Imperial College London, London, 2022.

[20] Li, T.; Tao, L.; Xu, L.; Meng, T.; Clifford, B. C.; Li, S.; Zhao, X.; Rao, J.; Lin, F.; Hu, L. Direct and Rapid High - Temperature Upcycling of Degraded Graphite. Advanced Functional Materials 2023, 33 (43). https://doi.org/10.1002/adfm.202302951.

[21] Yu, Z.; Duong, B.; Abbitt, D.; Thomas, J. Highly Ordered MnO2Nanopillars for Enhanced Supercapacitor Performance. Advanced Materials 2013, 25 (24), 3302–3306. https://doi.org/10.1002/adma.201300572.

[22] Liu, J.; Yang, L.; Song, Z.; Xu, C. Microstructures and Capacitance Performance of MnO2 Films Fabricated by Ultrasonic-Assisted Electrodeposition. Applied Surface

Science 2019, 478, 94 – 102. https://doi.org/10.1016/j.apsusc.2019.01.205.

[23] Liu, C.; Sun, C.; Gao, Y.; Lan, W.; Chen, S. Improving the Electrochemical Properties of Carbon Paper as Cathodes for Microfluidic Fuel Cells by the Electrochemical Activation in Different Solutions. ACS Omega 2021, 6 (29), 19153–19161. https://doi.org/10.1021/acsomega.1c02507.

10

Sulphur doped materials as electrocatalysts for the electrochemical reduction of O2

Jasmine Wang

Department of Chemical Engineering, Imperial College London, U.K.

Supervisors: Dr Jesús Barrio Hermida[a], Professor Magda Titirici[a]

Abstract

Sulfur-doped carbons oxygen reduction electrocatalysts with different sulphur contents have been synthesized by high temperature pyrolysis of a sulfur-containing cross-linked precursor in the presence of $MgCl_2.6H_2O$. The cross-linked reactant was synthesised by pyrolysis at high temperature of a cross-linked precursor consisting of hexaketocyclohexane octahydrate and elemental sulphur (Hexasulfur). And the carbon materials were characterized by X-ray diffraction (XRD) and N_2 physisorption. The electrocatalytic activity for the oxygen reduction reaction (ORR) was then assessed in a rotating ring disk electrode (RRDE) setup via cyclic voltammetry (CV) in alkaline electrolyte and compare to a reference material by Mg-meidiated pyrolysis of hexaketocyclohexane. Notably, the reference material, which is presumed to contain oxygen functional groups, exhibited the highest performance (with a kinetic mass activity of -1.27 Ag^{-1}). However, this synthesis approach leads the way to synthesize sulfur coordinated metallic single sites that mimic the active sites of enzymes such as nitrogenase [22].

1. Introduction

The electrocatalytic oxygen reduction reaction (OOR) is a significant process in sustainable electrochemical energy conversion technologies including fuel cells and metal air batteries, where O_2 is reduced to either H_2O_2 via the transfer of 2 protons or H_2O via the transfer of 4 protons [1,2,3] . However, the application of those techniques still significantly hampered by several factors including the high cost, unsatisfactory long-term durability related to the utilisation of highly active but costly Pt based electrocatalysis[4,5]. Under such conditions, great efforts have been made in recent decades to investigate the low-cost and abundant non-noble metal based electrocatalyst that could decrease the use of Pt-based electrocatalyst[6]. Earliest investigations concentrated on improving the ORR performance by reducing the particle size of Pt while lowering the cost. However, the inefficiency of the Pt nanoparticles for its tendency to detach from the carbon support and aggregate to large nanoparticles resulting in poor durability for a long term[7,8,9]. Therefore, the demand for highly efficient, inexpensive, stable non-metal based electrocatalyst is increased and has gain significant attention in material science research[10].

Carbon-based materials have been extensively studied due to their high specific surface area and the electronic conductivity, and the presence of pores and channels that create an efficient solid–liquid–gas phase region[11]. It offers tuneability through doping with heteroatoms, which induces charge redistribution and generates active sites, while enabling the coordination of metals to the lone electron pairs of the heteroatoms[12]. This structure enhances effective mass transfer of O_2 and electrolyte, thereby could further improving the performance of the oxygen reduction reaction (ORR) process[11]. Studies showed that the catalytic performance of the carbon-based material is affected by two primary factors, the intrinsic activity of a single active site and the density of active sites[11]. The method of introducing heteroatom into carbon materials was to improve the intrinsic activity since doping induced atom density could increase the rate of O_2 adsorption and better facilitating ORR. N-doping, has been widely investigated as an efficient approach to enhance the activity of carbon for ORR since N creates a positive charge on the neighbouring carbon atoms (C+) due to nitrogen's higher electronegativity (electronegativity of nitrogen: 3.04) compared to carbon's electronegativity (electronegativity of carbon: 2.55) enhancing the oxygen adsorption rate and then decrease the bond strength of the O-O[11,13,14]. Similarly, other heteroatoms with lower electronegativity comparing with nitrogen such as B (electronegativity of boron: 2.04), and P (electronegativity of phosphorus: 2.19) boosting the O_2 adsorption rate by creating the positive atoms of P^+ and B^+.[11,15,16] Sulfur, however, has a similar electronegativity to carbon (2.58 vs 2.55 respectively), however spin density can contribute to the catalytic efficiency therefore, sulphur-doped

carbon display a different catalytic mechanism to other heteroatom-doped carbons (i.e. N, P, B)[11, 17,18,19].

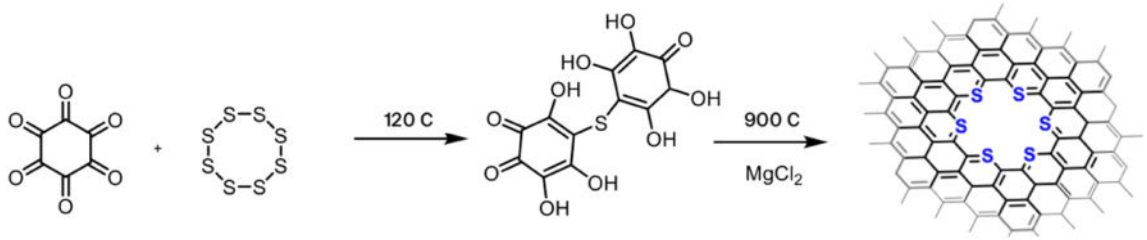
Additionally, Introducing heteroatoms enables their lone electron pairs to coordinate with metallic atoms (such as Fe), creating active sites that are effective for the oxygen reduction reaction.[20] This interaction is similar to the mechanisms found in natural enzymes, where metal centres coordinate with surrounding heteroatoms to facilitate catalytic processes.[21] These biologically inspired active sites enhance the material's ability to drive the oxygen reduction reaction efficiently, mimicking the functionality of enzymatic systems in nature[22]. As a result, carbon materials dopped with O (oxygen), N (nitrogen), S (sulphur), P (phosphorus), B (boron) have attracted widespread attention as ORR catalysts. In particular, Fe single sites coordinated to nitrogen doped-carbon materials showed the most promising alternative to platinum-group-metal catalysts[23].

Above those heteroatoms, nitrogen by far was most investigated and other heteroatoms are relatively rare reported. Recently, S-doped based material was investigated for its interests for battery applications given that sulphur is an electrochemically active element capable of reversible reactions with lithium (Li) or sodium (Na)[24] , Previous studies employing diverse synthetic approaches have explored sulfur-doped carbon-based materials for enhancing ORR activity. For instance, sulfur-doped carbon spheres have been synthesized through an in-situ doping strategy combined with a solvothermal method, using sucrose as the carbon source and benzyl disulfide as the sulfur source.[25] While variations in sulfur content were found to influence performance, the overall enhancement in ORR activity remains moderate, and the material does not perform particularly well. [25,26]

Although there is growing interest in heteroatom-doped carbon-based materials for catalytic applications, the potential of sulfur-doped carbon materials to enhance oxygen reduction reaction (ORR) performance remains largely underexplored. Besides, to date, the catalytic activity of sulfur-doped carbons remains relatively low, and the coordination of single active sites within fully sulfur-doped carbons has yet to be demonstrated [27]. Therefore, In this study, we present a sulfur doping method for synthesizing sulfur-doped carbon based catalyst materials for ORR reaction by using $MgCl_2$ to induce porosity, enabling efficient exposure of active sites and thereby achieving enhanced catalytic activity in facilitating the ORR process.

2. Background

Previous studies have extensively explored the development of sulfur-doped carbon-based electrocatalysts for the oxygen reduction reaction (ORR). These efforts utilized various sulfur sources, such as thiourea and L-glutathione, and demonstrated that the sulfur content in the final material is influenced by the sulfur precursor used. For example, Pajewska-Szmyt et al. showed that higher sulfur content in the precursor leads to increased sulfur incorporation into the doping material [28,29]. Similarly, approaches utilizing sucrose and benzyl disulfide as reagents have shown some promise but still fall short in catalytic efficiency, highlighting the need for further improvements[11].
Building on this foundation, we drew inspiration from the field of nitrogen-doped carbon materials, where synthesis methods have been developed and evaluated to maximize efficiency. Barrio et al[30], synthesised an eutectic cross-linked precursor employing hexaketocyclohexane, as carbon backbones and urea as nitrogen source, which was subsequently pyrolized at high temperature in the presence of lewis acid metal salts such as $ZnCl_2$ or $MgCl_2$. Additionally, Fechler et al[30] , screened different organic building blocks such as 1,2-benzoquinone or 2,5-dihydroxy,1,4-benzoquinone,



Scheme 1: Reaction scheme of making sulfur based carbon materials.

and observed that hexaketocyclohexane, owing to the formation of a deep eutectic liquid mixture, yielded the carbon material with the highest nitrogen content, reaching 28 wt%, demonstrating its potential as precursor for functional nitrogen-doped materials.[30] Inspired by the success of nitrogen-doped systems, we combined this approach with the utilization of elemental hexasulfur, which has been proven to allow the complete inclusion of sulfur moieties in polymers via nucleophilic, or electrophilic catalysis[29-31]. In our work, hexaketocyclohexane was selected as the carbon backbone, and hexasulfur was employed as the sulfur source (Scheme 1) and a cross-linked polymer was synthesized by solid-state reaction at mild temperatures (120 °). The cross-linked precursor was then pyrolyzed at 900°C in the presence of $MgCl_2 \cdot 6H_2O$, a well-known porogen that enhances porosity by creating high surface area and tubular channels in the resulting materials [31].

3. Methods

Material synthesis

Suphur-doped carbon materials were synthesized by pyrolysis at high temperature of a cross-linked precursor consisting of hexaketocyclohexane octahydrate (HSH, 97% Sigma Aldrich) and elemental sulphur (Hexasulfur). The mass of the HSH (963 mg, 3 mmol) were mixed with (96.18 mg, 3 mmol), (480 mg, 15 mmol), (48.1 mg, 1.5 mmol), (288.54 mg, 9 mmol) of elemental sulphur with a pestle and a mortar respectively until the homogeneous powder was obtained. The powder mixture was placed in a vial, and a homogeneous dark black cross-linked complex was obtained by heating the vial (HSH-S) at 120 °C in a silicone bath, the crosslinked precursors were labelled as HS_{XY}, where X and Y correspond to the molar ratio between HSH and S used.

The cross-linked complex then was ground to fine powder, with magnesium chloride hexahydrate (99% Sigma Aldrich) in a weight ratio of 1:8, and pyrolyzed in a ceramic crucible (filled 1/3 rd with material) at 900 °C for one hour under N_2 atmosphere (300 mL min^{-1}) with 5 °C min^{-1} heating rate. A reference material without sulphur was prepared by pyrolyzing HSH in the same conditions as a reference. The materials were grounded to fine powder and washed with 2M HCl (prepared by dilution of fuming 37% HCl, Merck) and was left overnight to further remove the MgCl2 and MgO species. The powder was filtrated and washed with distilled water, dried at 80 °C under vacuum and labelled as $HS_{XY}900$.

XRD patters were recorded with a powder X-ray diffractometer (PANanalytical X'PERT PRO). The BET surface area of $HS_{13}900$ and $HS_{15}900$ was determined from the adsorption isotherm in the relative pressure range of <0.03, with the optimal region for linear fitting identified using the Rouquerol method in 3 Flex software (version 5.02.2). The pore size distribution was calculated from the adsorption isotherm using the heterogeneous surface 2D-NLFDT method in SAIEUS software (version 3.06) [32]. XRD patterns were recorded with a powder X-ray diffractometer (PANanalytical X'PERT PRO).

Electrochemical measurements

Different ratio of the samples (4 mg) were prepared and dispersed in the mixture of isopropynal respectively (480 ul, ≥99.5% Honeywell, Fisher Scientific), Nafion (40 µL,5 wt%), and deionized water (480 µL), and was sonicated for 30 minutes at 30 °C, Electrochemical oxygen reduction tests were conducted using an AUTOLAB PGSTAT302N in a 1 M KOH (Suprapur, Merck) electrolyte saturated with either N_2 (≥99.99998% BIP Plus, Air Products) or O_2 (≥99.9998% Ultrapure Plus, Air Products). The experiments were performed in a compartmentalized glass cell with a three-electrode setup, employing a saturated Ag/AgCl (3 M KCl) electrode as the reference and a glassy carbon rod as the counter electrode. iR correction was applied after electrochemical measurements by determining the solution's Ohmic resistance (R = 23 Ω) from the intercept of the imaginary impedance axis in the Nyquist plot. This was done without fitting the data to an equivalent circuit, using electrochemical impedance spectroscopy in the frequency range of 10^{-5} to 10 Hz at open circuit potential. A 5 mm glassy carbon RRDE (Metrohm) was polished with a micropolish cloth and 0.05 µm alumina suspension (Buehler). The sample were drop-casted on to a

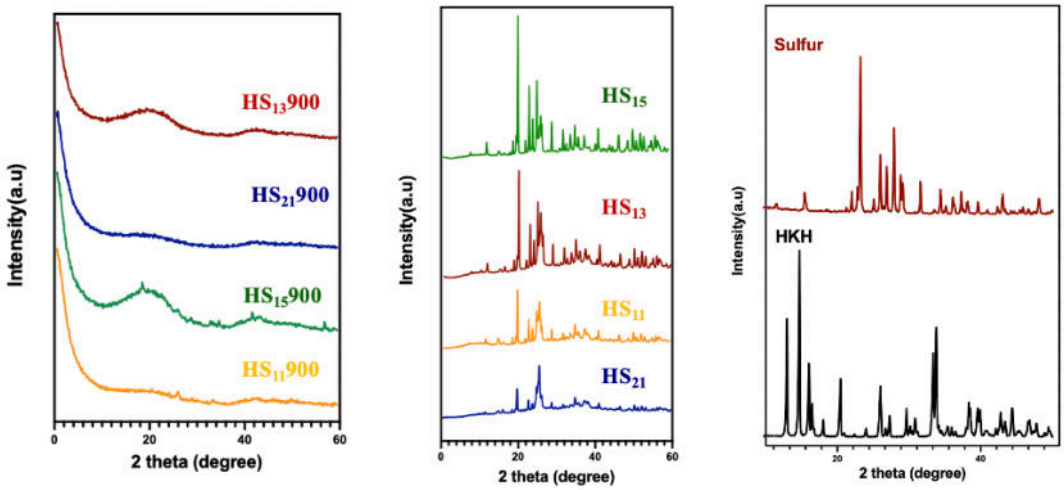


Figure 1: a) XRD spectra for the sample $HS_{13}900$, $HS_{21}900$, $HS_{15}900$, and HS11900. b) XRD spectra for sample HS_{13}, HS_{21}, HS_{21} and HS_{11} c)XRD spectra for the sample HSH (cyclohexanehexone) and Sulfur (hexasulfur).

freshly polished RRDE; The resulting catalyst sample was dried at 500 rpm for 5 min under a hairdryer. Cyclic voltammograms were recorded after purging the electrolyte for at least 10 minutes. Measurements were conducted at a scan rate of 50 mV s^{-1} at 0 rpm for N_2-saturated conditions) and at 10 mV s^{-1} at 1600 rpm for both O_2- and N_2-saturated conditions. Pseudo capacitance was corrected by subtracting the N_2-saturated measurements at 1600 rpm from the O_2-saturated results. For the best-performing material, this correction was observed in the cathodic scans of the cyclic voltammograms under O_2-saturated conditions at 1600 rpm [32].

Results and Discussion

In this work, we select hexasulfur as nucleophile, attacking and reducing C=O bond forming linkage between Sulfur and Carbon, hexaketocyclohexane (HSH) serve as carbon building block for S doped material. X-ray diffraction (XRD) analysis was performed on the HS precursors as well as on the products after pyrolysis (Figure 1), After pyrolysis at 900°C, the materials exhibited a highly morphous structure, attributed to the enhanced surface area in the presence of porogen ($MgCl_2$) [29].

Figure 1 a) displays similar XRD patterns for the sample of HS_{15} and HS_{13} , showing a striking resemblance to the XRD pattern of hexasulfur.
This similarity suggests that, at these specific precursor ratios, the sulfur remains largely unaltered during the reaction, maintaining its characteristic crystallographic structure. These results imply that the ratio of HKH to sulfur may influence the overall material morphology, but the sulfur component remains predominantly in its native form in both mixtures. Moving forward, solid state nuclear magnetic resonance could be used to find the nature of the carbon species and confirmed unambiguously whether a C-S bond has been established.

Furthermore, the XRD patterns obtained after pyrolysis (Figure 1 b) revealed the absence of any distinct peaks corresponding to nanoparticles. This suggests that the pyrolysis process did not result in the formation of crystalline nanoparticles in the material. Instead, the patterns exhibited broad, diffuse features typical of amorphous or disordered structures, further indicating that the material remains predominantly amorphous at the pyrolysis temperatures employed. The lack of sharp, well-defined diffraction peaks, which are typically associated with crystalline phases, supports the conclusion that the sulfur and carbon components do not undergo crystallization into nanoparticle forms under the conditions tested.

The mass activity at 0.8 V and 0.85 V for each sample were calculated (Table 1) and the kinetic current density (j_{kin}) at 0.80 and 0.85 VRHE was determined from the geometric disk current density (j_d) and geometric diffusion- limited current density (j_{lim}) at 0.3 VRHE using Koutecký– Levich (Equation (1))

$$j_{kin} = \frac{j_d * j_{lim}}{j_{lim} - j_d} \qquad (\textbf{equation 1})$$

Then the following the kinetic mass activity at 0.80 and 0.85 VRHE could be found:

$$m_{kin} = \frac{j_{kin}}{Loading\ cat} \qquad (equation\ 2)$$

where *m*kin was the kinetic mass activity at 0.80 or 0.85 VRHE and Loading cat was the catalyst loading on the RRDE (mg catalyst cm−2).

Table 1: The kinetic mass activity for each reagent ratio was calculated at current densities of 0.8 and 0.85 A·cm^{-2}.

Sample/Ratio	Mass activity at 0.8	Mass activity at 0.85
$HS_{11}900$	0.181	0.0115
$HS_{15}900$	0.0242	0.129
$HS_{21}900$	0.617	0.073
HSH	1.271	0.123

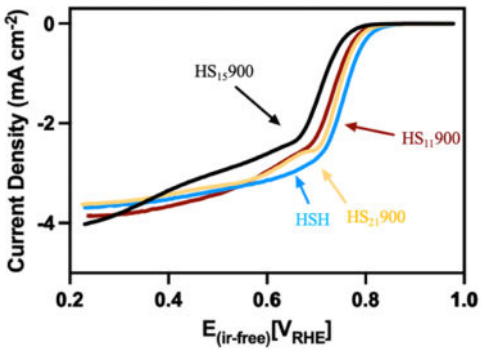


Figure 2: The RRDE results for sample of $HS_{11}900$ $HS_{21}900$, and the reference sample HSH900 without S atom doping, Capacitance-corrected cathodic scan of the cyclic voltammogram recorded in 0.1 M KOH (Suprapur) at a rotation rate of 1600 rpm and a scan rate of 10 mV s^{-1}. The data were obtained by subtracting the N_2-saturated cyclic voltammetry results (1600 rpm) from the O_2-saturated measurements. All experiments were conducted at room temperature. The trends and mass activity from Figure 2 and Table 1 indicates that the HKH sample demonstrates the highest kinetic efficiency, followed by $HS_{21}900$, and lastly $HS_{11}900$.

Interestingly, the reference sample having higher efficiency than other specific ratios suggests that under suboptimal ratios, the reaction efficiency is significantly hindered, potentially due to the lack of active site density or unfavourable reaction dynamics. To further investigate the factors affecting the observed phenomenon and the difference of the catalytic performance of different materials in different reagent ratio, BET studies was conducted for samples with ratios of $HS_{11}900$ and $HS_{15}900$ to investigate their porosity and surface chemistry (Figure 3).

In both materials ($HS_{11}900$ and $HS_{15}900$), a bimodal pore size distribution was observed. The smallest pore size is centred around 0.6–1 nm, suggesting the presence of appropriately sized micropores and large micropores also exists (1-5 nm). The BET results, shown in Table 2, indicates that the surface area for $HS_{11}900$ is higher than that of the $HS_{15}900$. This explains the lower kinetic efficiency of the $HS_{15}900$ compared to the $HS_{11}900$, which in turn has a higher material surface area.

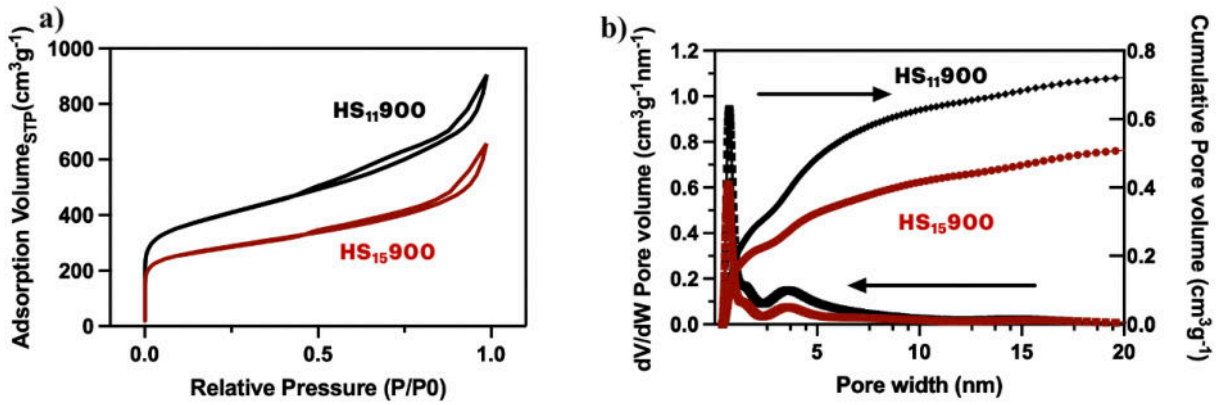


Figure 3. Nitrogen sorption isotherms with labelled BET specific surface areas (a) and the corresponding pore size distributions and cumulative pore volumes, calculated using the 2D-NLDFT heterogeneous surface carbon model in SAIEUS software (b) for the sample $HS_{11}900$ and $HS_{15}900$.

Table 1: The specific surface area of the materials prepared with different ratio of the reagent

Sample/Ratio	Specific surface area $(m^2\ g^{-1})$
$HS_{11}900$	1415 +/- 4
$HS_{15}900$	1017 +/- 2

Conclusion and outlook

In summary, in this study, we have prepared sulfur-doped carbon electrocatalysts with varying sulfur contents were synthesized via high-temperature pyrolysis of a sulfur-containing cross-linked precursor in the presence of $MgCl_2 \cdot 6H_2O$. Notably, the RRDE results revealed that the reference material without sulfur doping exhibited the highest mass activity and efficiency in facilitating the ORR reaction. This superior performance could potentially be attributed to the presence of oxygen atoms in the cyclohexanehexane-derived material. [33]

Future research could concentrate on:

1. Analysis of the ORR selectivity in sulfur doped carbons employing a rotating ring disk electrode setup.
2. Examining the impact of different pyrolysis temperatures on the material's electronic conductivity and performance.
3. The integration of iron to examine whether it can further increase catalytic activity by creating metallic single sites in the material.
4. The utilization of advanced characterization techniques such as TEM and Raman spectroscopy could be future conducted to provide valuable insights into the material's microstructure, while XPS analysis could help determine the chemical states of carbon and sulfur, offering a deeper understanding of the material's composition and catalytic properties.

Acknowledgements

I would like to acknowledgement and express my gratitude to my supervisors for their support, Dr. Jesús Barrio Hermida, Professor Magda Titirici, and also in particular thanks for JinJie Zhu (Jason) helping with the experimental setting and for the guidance all the time.

Reference

a. Imperial College Research Fellow Department of Chemical Engineering - Faculty of Engineering
b. Chair in Sustainable Energy Materials, Department of chemical engineering-Faculty of Enginering

1. Chanez Maouche, Zhou, Y., Wang, Y. and Yang, J. (2022). Recent advances of the key parameters of 3d block transition metal single and dual atoms catalysts: from their synthesis to their practical applications. *Materials Today Sustainability*, 21, pp.100288–100288. doi:https://doi.org/10.1016/j.mtsust.2022.100288.
2. Chen, G., An, Y., Liu, S., Sun, F., Qi, H., Wu, H., He, Y., Chen, M., Shi, R., Zhang, J., Kuc, A., Kaiser, U., Zhang, T., Heine, T., Wu, G. and Feng, X. (2022). Highly accessible and dense surface single metal FeN_4 active sites for promoting the oxygen reduction reaction. *Energy and Environmental Science*, 15(6), pp.2619–2628. doi:https://doi.org/10.1039/d2ee00542e.
3. Chang, Z.-W., Xu, J.-J. and Zhang, X.-B. (2017). Recent Progress in Electrocatalyst for Li-O_2Batteries. *Advanced Energy Materials*, 7(23), pp.1700875–1700875. doi:https://doi.org/10.1002/aenm.201700875.
4. Tian, X., Lu, X.F., Xia, B.Y. and Lou, X.W. (David) (2020). Advanced Electrocatalysts for the Oxygen Reduction Reaction in Energy Conversion Technologies. *Joule*, 4(1), pp.45–68. doi:https://doi.org/10.1016/j.joule.2019.12.014.
5. Zhang, C., Shen, X., Pan, Y. and Peng, Z. (2017). A review of Pt-based electrocatalysts for oxygen reduction reaction. *Frontiers in Energy*, 11(3), pp.268–285. doi:https://doi.org/10.1007/s11708-017-0466-6.
6. Lee, Y.-W., Cha, S., Park, K.-W., Sohn, J.I. and Kim, J.M. (2015). High Performance Electrocatalysts Based on Pt Nanoarchitecture for Fuel Cell Applications. *Journal of Nanomaterials*, 2015,pp.1-20doi:https://doi.org/10.1155/2015/273720.
7. Katsounaros, I., Cherevko, S., Zeradjanin, A.R. and Mayrhofer, K.J.J. (2013). Oxygen Electrochemistry as a Cornerstone for Sustainable Energy Conversion. *Angewandte Chemie International Edition*, 53(1), pp.102–121. doi:https://doi.org/10.1002/anie.201306588.
8. Yano, H., Kataoka, M., Yamashita, H., Uchida, H. and Watanabe, M. (2007). Oxygen Reduction Activity of Carbon-Supported Pt−M (M = V, Ni, Cr, Co, and Fe) Alloys Prepared by Nanocapsule Method. *Langmuir*, 23(11), pp.6438–6445. doi:https://doi.org/10.1021/la070078u.
9. Stamenkovic, V.R., Mun, B.S., Mayrhofer, K.J.J., Ross, P.N. and Markovic, N.M. (2006). Effect of Surface Composition on Electronic Structure, Stability, and Electrocatalytic Properties of Pt-Transition Metal Alloys: Pt-Skin versus Pt-Skeleton Surfaces. *Journal of the American Chemical Society*, 128(27), pp.8813–8819. doi:https://doi.org/10.1021/ja0600476.
10. Tan, H., Tang, J., Kim, J., Kaneti, Y.V., Kang, Y.-M., Sugahara, Y. and Yamauchi, Y. (2019). Rational design and construction of nanoporous iron- and nitrogen-doped carbon electrocatalysts for oxygen reduction reaction. *Journal of Materials Chemistry A*, [online] 7(4), pp.1380–1393. doi:https://doi.org/10.1039/C8TA08870E.
11. Ana Belen Jorge, Jervis, R., Arun Prakash Periasamy, Qiao, M., Feng, J., Linh Ngoc Tran and Maria-Magdalena Titirici (2019). 3D Carbon Materials for Efficient Oxygen and Hydrogen Electrocatalysis. *Advanced energy materials*, 10(11).
12. Ana Belen Jorge, Jervis, R., Arun Prakash Periasamy, Qiao, M., Feng, J., Linh Ngoc Tran and Maria-Magdalena Titirici (2019). 3D Carbon Materials for Efficient Oxygen and Hydrogen Electrocatalysis. *Advanced energy materials*, 10(11). doi:https://doi.org/10.1002/aenm.201902494.
13. Deng, Y., Luo, J., Chi, B., Tang, H., Li, J., Qiao, X., Shen, Y., Yang, Y., Jia, C., Rao, P., Liao, S. and Tian, X. (2021). Advanced Atomically Dispersed Metal–Nitrogen–Carbon Catalysts Toward Cathodic Oxygen Reduction in PEM Fuel Cells. *Advanced Energy Materials*, 11(37), p.2101222.
14. X. Hu, Y. Wu, H. Li, Z. Zhang, Adsorption and activation of O2 on nitrogen-doped carbon nanotubes, J. Phys. Chem. C 114 (2010) 9603–9607.
15. Z.-H. Sheng, H.-L. Gao, W.-J. Bao, Synthesis of boron doped graphene for oxygen reduction reaction in fuel cells, J. Mater. Chem. 22 (2012) 390–395.
16. J. Ozaki, N. Kimura, T. Anahara, Preparation and oxygen reduction activity of BN-doped carbons, Carbon 45 (2007) 1847–1853.
17. S. Yang, L. Zhi, K. Tang, X. Feng, J. Maier, K. Müllen, Efficient synthesis of heteroatom (N or S)-doped graphene based on ultrathin graphene oxide–

porous silica sheets for oxygen reduction reactions, Adv. Funct. Mater. 22 (2012) 3634–3640
18. Z. Yang, Z. Yao, G. Li, G. Fang, H. Nie, Z. Liu, X.M. Zhou, X. Chen, S.M. Huang, Sulfur-doped graphene as an efficient metal-free cathode catalyst for oxygen reduction, ACS Nano 6 (2011) 205–211.
19. I.-Y. Jeon, S. Zhang, L.P. Zhang, H.-J. Choi, J.-M. Seo, Z.H. Xia, L.M. Dai, J.-B. Baek, Edge-selectively sulfurized graphene nanoplatelets as efficient metal-free electrocatalysts for oxygen reduction reaction: The electron spin effect, Adv. Mater. 25 (2013) 6138–6145.
20. Asset, T. and Atanassov, P. (2020). Iron-Nitrogen-Carbon Catalysts for Proton Exchange Membrane Fuel Cells. *Joule*, 4(1), pp.33–44. doi:https://doi.org/10.1016/j.joule.2019.12.002.
21. Barrio, J., Pedersen, A., Favero, S., Luo, H., Wang, M., Sarma, S.C., Feng, J., Tran, L., Kellner, S., Li, A.Y., Belén, A. and Maria-Magdalena Titirici (2022). Bioinspired and Bioderived Aqueous Electrocatalysis. *Chemical Reviews*, 123(5), pp.2311–2348. doi:https://doi.org/10.1021/acs.chemrev.2c00429.
22: Kjaergaard, C.H., Rossmeisl, J. and Nørskov, J.K. (2010). Enzymatic versus inorganic oxygen reduction catalysts: comparison of the energy levels in a free-energy scheme. *Inorganic Chemistry*, [online] 49(8), pp.3567–3572. doi:https://doi.org/10.1021/ic900798q.
23. Pedersen, A., Bagger, A., Barrio, J., Maillard, F., Stephens, L. and Maria-Magdalena Titirici (2023). Atomic metal coordinated to nitrogen-doped carbon electrocatalysts for proton exchange membrane fuel cells: a perspective on progress, pitfalls and prospectives. *Journal of Materials Chemistry A*, [online] 11(43), pp.23211–23222. doi:https://doi.org/10.1039/d3ta04711c.
24. Tzadikov, J., Levy, N.R., Abisdris, L., Cohen, R., Weitman, M., Kaminker, I., Goldbourt, A., Ein-Eli, Y. and Shalom, M. (2020). Bottom-Up Synthesis of Advanced Carbonaceous Anode Materials Containing Sulfur for Na-Ion Batteries. *Advanced Functional Materials*, 30(19), p.2000592. doi:https://doi.org/10.1002/adfm.202000592.
25. Wang, H., Bo, X., Zhang, Y. and Guo, L. (2013). Sulfur-doped ordered mesoporous carbon with high electrocatalytic activity for oxygen reduction. *Electrochimica Acta*, 108, pp.404–411. doi:https://doi.org/10.1016/j.electacta.2013.06.133.
26. Sun, Y., Wu, J., Tian, J., Jin, C. and Yang, R. (2015). Sulfur-doped carbon spheres as efficient metal-free electrocatalysts for oxygen reduction reaction. *Electrochimica Acta*, [online] 178, pp.806–812.
27: Omann, L., Königs, D.F., Hendrik and Oestreich, M. (2017). Cooperative Catalysis at Metal–Sulfur Bonds. 50(5), pp.1258–1269. doi:https://doi.org/10.1021/acs.accounts.7b00089.
28. Ma, G., Ning, G. and Wei, Q. (2022). S-doped carbon materials: Synthesis, properties and applications. *Carbon*, 195, pp.328–340. doi:https://doi.org/10.1016/j.carbon.2022.03.043.
29. M. Pajewska-Szmyt, B. Buszewski, R. Gadzaa-Kopciuch, Sulphur and nitrogen doped carbon dots synthesis by microwave assisted method as quantitative analytical nano-tool for mercury ion sensing, Mater. Chem. Phys. 242 (2020), 122484.
30. Barrio, J. (2022). Metal coordination in C2N-like materials towards dual atom catalysts for oxygen reduction. *Zenodo (CERN European Organization for Nuclear Research)*. doi:https://doi.org/10.5281/zenodo.8111494.
31. M. Pajewska-Szmyt, B. Buszewski, R. Gadzaa-Kopciuch, Sulphur and nitrogen doped carbon dots synthesis by microwave assisted method as quantitative analytical nano-tool for mercury ion sensing, Mater. Chem. Phys. 242 (2020), 122484.
32. Barrio, J., Pedersen, A., Saurav Ch. Sarma, Bagger, A., Gong, M., Favero, S., Zhao Chang-xin, García-Serres, R., Li, A., Zhang, Q., Frédéric Jaouen, Maillard, F., Kucernak, A., Stephens, L. and Maria-Magdalena Titirici (2023). FeNC Oxygen Reduction Electrocatalyst with High Utilization Penta-Coordinated Sites. *Advanced Materials*, 35(14). doi:https://doi.org/10.1002/adma.202211022.
33:Smith, R.D.L., Pasquini, C., Loos, S., Chernev, P., Klingan, K., Kubella, P., Mohammadi, M.R., Gonzalez-Flores, D. and Dau, H. (2017). Spectroscopic identification of active sites for the oxygen evolution reaction on iron-cobalt oxides. *Nature Communications*, 8(1). doi:https://doi.org/10.1038/s41467-017-01949-8.

Dye Extraction Decoded: The Role of Ionic Liquid Solvents and Process Tuning

Braxton Goh and Xuanyong Tay
Department of Chemical Engineering, Imperial College London, U.K.

Abstract

Swift developments in the global textile industry have led to a wasteful culture of generating abundant textile waste with severe environment consequences. DyeRecycle bridges the sustainability gap and facilitates closed-loop processes for both fabrics and dyes with the use of ionic liquids (ILs). This study in collaboration with DyeRecycle focuses on using protic triethylammonium methanesulfonate (TEA-$MeSO_3$) to extract disperse dyes from polyester fabrics and examines the chemistry behind the dye-IL interactions at varying acid base ratios (ABRs). Extraction performance and regeneration cycle efficiency were also investigated via multistage operations and by varying IL water content. Due to balanced extents of cationic and anionic interactions, findings show that ABR 1.00 performs the best in dye extraction. However, non-equimolar ABRs are ideal for fabric decolourisation because of unintentional dye repartitioning during the washing of samples. Severe dye degradation can be caused by high proton activity due to these factors: (i) High ABR; (ii) Low water content. High temperature and long extraction time have also been observed to exacerbate dye degradation. Modelling of the continuous process highlights limitations in the regeneration and recycling of used IL with the use of activated carbon due to mass transfer limitations as well as the degradation due to prolonged exposure to heat. Optimal continuous extraction requires tailoring IL composition, as well as a moderate water content in a counter current mechanism to maximise mass transfer efficiency. There should also be an optimal operating temperature range and well-defined residence time that balances appreciable dye extraction with minimal degradation.

1. Introduction

The rapid rise of the fast fashion industry has undoubtedly reshaped the way that clothing is produced and purchased. Abundant raw materials and strong demand have made the latest fashion trends increasingly accessible to the average consumer. However, it is precisely because of this accessibility that an unsustainable throwaway culture of cheap textiles is being promoted in our world today. Global textile waste is currently estimated to be 92 million tonnes with a future projection of 134 million tonnes by 2030 (Beall, 2020).

Textile waste can be broadly classified into two categories: (i) Deadstock producer waste; (ii) Consumer discarded textiles. Deadstock fabrics are commonly incinerated as it is often the most cost-effective choice for some brand owners to maintain their brand image or to recover some energy from the production process (Baraniuk, 2023). Post-consumer waste is no better from an environmental standpoint with only 1% of the total amount being recycled into new clothes (Beall, 2020). Evidently, there is a noticeable sustainability gap between the current state of fashion and the ideal circular textile economy.

However, within this sustainability divide lies a USD 500 billion economic opportunity for innovative companies to seize (Ellen MacArthur Foundation, 2017). This is where DyeRecycle has strategically positioned itself to be a pioneer in developing a sustainable textile dyeing solution. In this novel approach, low-cost protic ionic liquids (IL) are utilised to extract disperse dyes from polyester-based fibres. Subsequently, decoloured fabric can be recycled as feedstock for new apparel production while the dye extract can be reused for dyeing fresh polyester or nylon fabrics. Once commercialised, DyeRecycle's closed loop process aims to reintegrate both the fabric and dye back into the supply chain to shrink the industry's sustainability gap (DyeRecycle, 2022). Key advantages of the DyeRecycle process include the elimination of fresh dye inputs, dispersion and swelling agents required in water-based polyester dyeing.

The objective of this investigation is to ascertain how dye extraction performance is influenced by the variation in process conditions. Findings made will be used to inform DyeRecycle's technological roadmap for developing its integrated decolourisation-dyeing chemical process.

2. Background

2.1 Polyester Fabrics and Disperse Dyes

Due to its cost-effectiveness during the production phase, along with its properties of high durability and resistance to wrinkles and creases (Aizenshtein, 2006), polyesters have been dominating the textile manufacturing industry for decades. This is evident in 2023 when it made up 57% of the 70.93 million tonnes of global fibre production (Textile Exchange, 2024).

A colouring agent that was formulated specifically to pair with the hydrophobic properties of polyesters is disperse dye. Disperse dyes are very small, non-polar molecules that are insoluble in water and undergo sublimation at high temperatures (Aspland, 1992).

A typical dyeing process on polyester fibres occurs under a high temperature (Patterson & Sheldon, 1959), where the water-insoluble dyes first undergo sublimation before dispersing as very fine particles into a liquid water carrier phase. They will then diffuse into the dense and tightly packed structure of synthetic fibres such as those of polyesters. Once dyed, the hydrophobicity of both components keeps the dye trapped within the fibre structure via Van der Waals forces and other weak intermolecular interactions (Ketema & Worku, 2020). This will lead to products that are highly durable and wear resistant.

On the molecular scale, more than 75% of disperse dyes are of monoazoic nature or are anthraquinonoid derivatives as seen in *Figure 1* (Aspland, 1992). They are described by heavily conjugated aromatic compounds that allow the delocalisation of electrons across the entire structure (Allen, 1971). This delocalisation and nature of other substituent groups then facilitates the absorption of light at a specific wavelength within the visible spectrum.

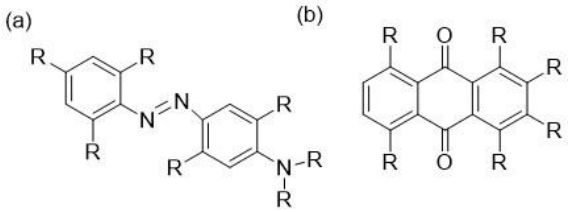


Figure 1: General chemical skeletal structures of (a) Monoazoic dyes and (b) Anthraquinone dyes. Made with ChemDraw 23.0.1.

2.2 Ionic Liquids

ILs can be classified as room-temperature molten salt compounds with a melting point below 100°C (Greaves & Drummond, 2007). They are also less flammable with a lower volatility and possess wide thermal stability windows with strong solvating potentials when compared to conventional organic solvents.

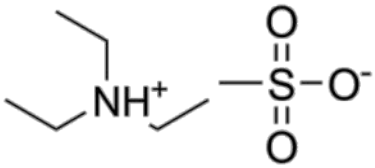


Figure 2: Chemical structure of triethylammonium methanesulfonate. Made with ChemDraw 23.0.1.

This study makes use of the protic IL: triethylammonium methanesulfonate (TEA-$MeSO_3$), as shown in *Figure 2*. Such ILs are desirable due to their low synthesis costs and easy preparation methods involving a neutralisation reaction. However, the proton transfer reaction does not always continue until completion. This results in a hybrid system of neutral acid and base molecules, their charged conjugate pairs, as well as an extensive network of hydrogen bonds and dissociated protons (Greaves & Drummond, 2007).

This combination of diverse intermolecular interactions such as ionic electrostatic attraction, hydrogen bonding and Van der Waals forces gives rise to distinctive dissolution properties of protic ILs brought about by nanoscale structuring (Schröder et al., 2000). For instance, organic species will preferentially localise in non-polar domains, whereas polar molecules will likely have stronger affinity towards charged regions.

2.3 Ionic Liquid – Dye Interactions

In this study, the complex bulk makeup of protic IL species will be utilised for dissolving and extracting disperse dyes out from polyester fabrics.

Molecular dynamics simulations conducted by Goloviznina et al., (2022) has postulated that the two most critical intermolecular interactions that determine disperse dye solubility in ILs are cation-π stacking as shown in *Figure 3* and ion dipole interactions.

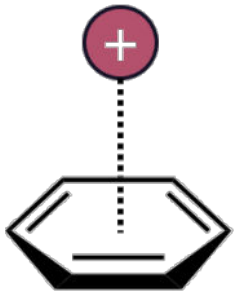


Figure 3: Cation-π stacking illustration. Made with ChemDraw 23.0.1.

Cation-π stacking is considered as the most significant amongst all the intermolecular interactions between IL and dye molecules. Moreover, there was competition between the anions and water for the amino group on the dye molecule, thus indicating that the presence of water induced hydrogen bonding disrupts the ion-dipole interactions between the anion and polar functional groups that would drive dye solubility (Goloviznina et al., 2022).

These chemical interactions likely also underpin the dynamics between the polyester-contained disperse dye and the protic TEA-$MeSO_3$ examined in this study. By integrating these chemical concepts with extraction process conditions, this study also seeks to provide insight into the behaviour of disperse dye-IL systems.

2.4 Activated Carbon Adsorption

In the context of the DyeRecycle continuous process, the final step involving ionic liquid recovery will incorporate the use of a commercial adsorbent.

Activated carbon is one of the most popular choices in industry for cleansing organic and inorganic impurities from aqueous solutions. This can mainly be attributed to its substantial adsorption capacity owing to its large internal surface area as well as a highly microporous structure (Lemus et al., 2013).

Commercially available activated carbon has been demonstrated to exhibit a low adsorption capacity for hydrophilic ionic liquids (Lemus et al., 2013), which TEA-$MeSO_3$ can be classified as. This is in conjunction with other studies that have used activated carbon to successfully purify wastewater streams containing disperse dyes (Mittal et al., 2022). This preferential adsorption towards disperse dyes therefore harmonises well with the final recovery stage as it streamlines the separation process to recycle and purify the IL for subsequent extractions.

Disperse dye adsorption on activated carbon has also been determined to be mainly a physisorption process driven by Van der Waals forces (Sahu & Singh, 2019). This binding mechanism has implications for the retrieval of dye particles as temperature increases would favour desorption and regeneration of the activated carbon adsorbent for more uses.

3. Methodology

The laboratory scale investigation setup of the IL mediated dye extraction involves a synthesis of TEA-$MeSO_3$. The chemical reagents used for the synthesis are 99% triethylamine, 99% methanesulfonic acid, as well as deionised (DI) water. The detailed procedures of IL synthesis are described in Subsection 3.2.1.

Dye extraction is conducted in batches in a pressure tube. 1.5g of fabric and 10g of IL are mixed thoroughly such that the fabric is entirely wetted by the IL. The proportion of the mix is fixed at 15% throughout the set-up of the investigation. This tube is then placed in a pre-heated oven at a specified temperature for 30 minutes and for one hour in separate setups. After which, the sample is to be cooled down before further analysis. The temperature at which extraction takes place is further elaborated in Subsection 3.2.2.

3.1 Performance Metrics

To quantify the extent of effectiveness of the dye extraction process from a commercial lilac coloured fabric, two main metrics are examined for each experimental setup: (i) the extent of dye extraction that is retained in the IL; (ii) the extent of fabric decolourisation.

3.1.1 Dye Concentration in IL

The concentration of coloured compounds in liquid samples can be determined with spectrophotometric analysis. This experimental procedure is described by the Beer-Lambert's Law (Swinehart, 1962) which correlates the absorbance of light to a sample's concentration.

$$A = \varepsilon bc$$

Equation 1: Beer-Lambert's Law

A is the absorbance; ε is the molar absorptivity; b is the path length of light; c is the concentration.

Utilising the relationship outlined in *Equation 1*, a calibration curve is constructed to determine the values for the molar absorptivity and characteristic wavelength for the disperse dye of interest. This is done by correlating known concentrations of the disperse dye in the IL solution to their maximum absorbances using a UV-Vis Shimadzu spectrophotometer. The proportionality constant: εb, corresponds to the gradient of the linear relationship between the two quantities.

For calibration purposes, disperse dye powder is obtained by means of IL extraction. After the process of dye extraction at 150°C for one hour, the coloured IL is diluted and centrifuged to separate the dyes out from the liquid phase. Powdered dye was obtained after overnight drying.

The powdered dye was dissolved and diluted down using the IL into multiple samples: – 50, 100, 200, 300, 400, and 500 ppm by mass. The absorbance values for each of these calibration samples are then taken from the common characteristic peak wavelength of the dye. Lastly, by plotting the absorbance data against dye concentration, a linear polynomial (that of Beer-Lambert's Law) is fitted across the calibration data to determine the proportionality constant – εb.

For experimental samples, the absorbance values for the liquid IL samples will be measured via the UV-Vis Shimadzu spectrophotometer across the visible light wavelength range. The corresponding absorbance value of the characteristic wavelength could be back correlated to the dye concentration using *Equation 1*. However, due to the limitations of the equipment accuracy at high dye concentrations, samples could be diluted down when necessary to ensure that all absorbance measurements are within the range of calibration.

3.1.2 Colour Intensity of Fabric (K/S Value)

The extent of fabric decolourisation can be defined as the change in its colour intensity. In the textile industry, a commonly used metric is the K/S value. K/S is a dimensionless number which corresponds to the colour strength of a solid sample. The determination of the K/S value could be obtained using the Kubelka-Munk function (Bajpai, 2018), shown in *Equation 2*, which expresses the quantity as a function of the reflectance value, which can be measured via UV-Vis spectroscopy.

$$\frac{K}{S} = \frac{(1 - R_{min})^2}{2R_{min}}$$

Equation 2: Kubelka-Munk Function

K is the absorption coefficient; S is the back-scattering coefficient; R_{min} is the lowest reflectance value of the sample at the characteristic wavelength.

In experiments, all fabric samples are retrieved from the pressure tubes after separation from the coloured IL liquid. The fabrics are then washed with DI water at a 1:5 mass ratio for a total of six times to remove any traces of IL that are retained on the fabric. The solid fabric samples are then dried overnight to evaporate any remaining liquid before conducting spectroscopy measurements.

To measure the reflectance value of the solid fabric samples, the UV-Vis spectrophotometer is configured for solid sample reflectance measurements, where the samples will be benchmarked against powdered barium sulfate as a baseline. The reflectance values for the samples are measured across the visible light spectrum, where the lowest reflectance value at the characteristic wavelength will be taken to calculate the corresponding K/S value of the samples using *Equation 2*.

3.2 Variables

In a process of dye extraction using IL, there exists a range of variables to be studied for the optimisation of the performance of dye uptake and recovery. This variability arises from the molecular dynamics between the disperse dye and IL molecules.

This study focuses on analysing the trends and effects of changing parameters on IL extraction performance. The aim is to provide insights into the behaviour of the system under varying conditions, contributing to a better understanding of the underlying mechanisms behind the disperse dye and IL molecules.

3.2.1 Acid Base Ratio (ABR) of the IL

A common parameter monitored during the IL synthesis process is the Acid Base Ratio (ABR) (Nakasu et al., 2020). The ABR is determined by the relative amount of acid and amine of the IL which then leads to a variability of the composition of charged ions and their corresponding intermolecular interactions.

TEA-$MeSO_3$ stock solutions are prepared at a range of different ABRs to investigate the effects of changing IL compositions against extraction performance. All synthesised ILs are also mixed with DI water such that the water contents are fixed at 20% by mass. The amount of each reagent to be used for the synthesis for these stock solutions are specified in *Table S1*.

For the synthesis, the amine and DI water are first mixed, in accordance with the amount that is detailed in *Table S1* in a round-bottom flask. After this, a set amount of acid is added dropwise into the flask using a funnel into a constantly stirred flask. As the neutralisation between TEA and $MeSO_3H$ is highly exothermic (Burrell et al., 2010), the neutralisation reaction is carried out at a very slow rate and conducted overnight in an ice bath. This is to ensure a controlled environment to prevent thermal runaway.

Upon completion of IL synthesis, each stock solution's water content and ABR are checked with a Volumetric Karl Fischer Titrator and NMR analysis (with D_2O as solvent) respectively.

Due to the inherent error present with NMR analysis (McLeod & Comisarow, 1989), the stock solutions are doubled checked with their apparent acidity via pH measurements when diluted down with DI water to a 1% by mass solution.

3.2.2 Temperature of Dye Extraction

The extraction process will have to be carried out under a high temperature for the dissolution of dye to take place. To study the extraction kinetics and maximise the recovery of dye from the fabric, this investigation aims to study the trade-off between these competing metrics at different temperatures.

This is done by conducting the extraction experiments at three different temperatures – low (130°C); moderate (150°C); and high (170°C). The varying operating temperatures serve as an additional dimension to the varying ABR of the IL, allowing in depth analysis on the impact due to the change in operating temperature on the extraction performance at different IL ABRs. This is to understand the dynamics of varying temperatures for ILs of differing ABRs and how this may affect the interactions with dye molecules.

3.2.3 Time

The industrial dye extraction will likely be a continuous process. To simulate the dynamics of this in the lab, multistage batch extractions will be conducted. This is based on the concept of discretisation of time to represent the continuous process using batch equipment (Hu, 2021). This involves the fixed fabric load repeatedly undergoing extraction, with the addition of fresh IL for each discrete time step.

For the purpose of this investigation, each time step will be of 30 minutes, at an oven temperature of 150°C. 1.5g of fabric is used throughout the investigation, with the 10g of IL used in the initial time step. After every extraction stage, the IL is squeezed out as much as possible from the pressure tubes and weighed to account for the residual IL that is trapped within the fabric. This mass will be the amount of fresh IL to be topped up into the pressure tube to ensure the fixed 15% fabric loading for each extraction stage.

Since the same fabric will be used throughout the investigation, the only performance metric that is monitored across the range of time steps will be the dye concentration in the IL at each extraction stage.

3.2.4 Regeneration Cycle

With regards to the IL regeneration and recycling process, the efficiency of which the IL has been 'cleaned' at each regeneration cycle by activated carbon is to be studied. To emulate the cycle of IL use, a set of investigations with operating variables akin to that of multistage extraction has been duplicated and modified to account for the use of activated carbon adsorption. This is done by using activated carbon to regenerate used IL at the end of each extraction stage. It will then be reused for a subsequent extraction.

For this investigation set-up, 60g of IL (ABR 1.00) is to be used to extract 9g of fabric in a pressure tube, adhering to the fixed fabric loading of 15%. The increase in the amount used is to ensure that the samples collected from each regeneration stage is sufficient as a substantial amount of IL will be retained within the fabric and activated carbon.

After each extraction stage, the IL is to be squeezed out, sampled, and weighed. It will be then mixed with activated carbon, amounting to 10% of the IL mass. This mixture is then centrifuged to separate the cleaned IL and activated carbon granules. The cleaned IL is to be analysed with UV-Vis spectroscopy and added back to the pressure tube for the subsequent stage of extraction. A small amount of fresh IL will be added to replenish any that has been lost. This maintains 60g of IL in the pressure tube at all stages.

3.2.5 Water Content of the IL

Another parameter to be investigated is the water content of the IL solution, which corresponds to the concentration of ionic species in the solution that is available to solvate the dye molecules. The extraction

efficiency of IL solutions at varying water contents is to be investigated to examine their effects on the intermolecular interactions in the system.

Three IL stock solutions of 100g each are synthesised with the procedures akin to that of Subsection 3.2.1. These solutions will all be of ABR 1.00; but with varying water contents (10%, 20% and 30% by mass) that constitute the make-up of the stock solutions. The exact amounts to be used for each reagent are as shown in *Table S2*.

Using the stock solutions for varying water content, experimental set-ups for dye extraction are to be repeated, at varying temperatures of 130°C, 150°C and 170°C. For all these cases, procedures are to follow that of Subsection 3.1.1, where the dye concentration in the IL is quantified using UV-Vis spectroscopy.

4. Results and Discussions

4.1 Dye Concentration Calibration

Following the procedures outlined in Subsection 3.1.1, the data obtained for a range of pre-determined dye concentrations is measured against their respective absorbance values, shown in *Figure S3*. A best fit line, with a zero intercept is constructed to determine the gradient that will correspond to the constant of proportionality in accordance with the Beer-Lambert's Law. Results show that the gradient, hence the constant of proportionality (εb) is **0.00080462ppm^{-1}**.

4.2 ABR Influence

4.2.1 Dye Degradation Trends

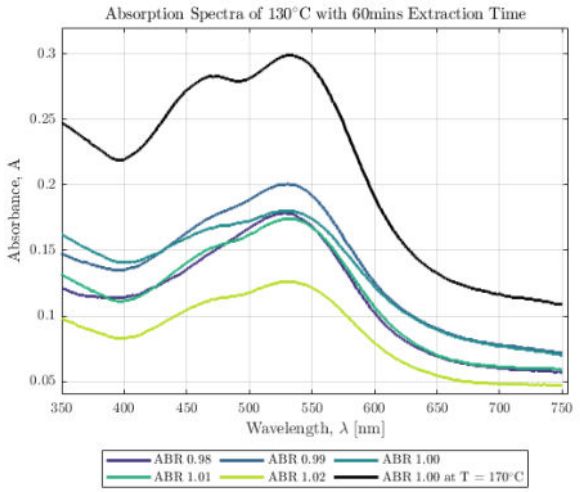


Figure 4: Dye extract UV-Vis absorbance spectra at 130°C

Figure 4 shows the dye extract absorbance spectra across a range of ABRs at 130°C. However, it can be seen that each profile gradually becomes distorted by a secondary inflection point around 475nm as the ABR increases. Similar UV-Vis disperse dye profiles in literature with altering absorption spectra can be characterised by dye degradation (Abdel-Monem, 2016).

In some samples such as ABR 1.00 at 170°C, there was even the emergence of a secondary local maximum. Such maxima in the absorption spectra indicate that the degradation products from the dyes are capable of absorbing UV-Vis light and are likely affecting the actual absorption of undamaged dye in the sample. As such, the formation of these products cannot be ruled out from other samples that exhibit an inflection point.

Increasing the ABR also gives a more pronounced inflection point which could mean that degradation becomes more significant at higher ABRs. Moreover, our measured pH values confirm that higher ABR ILs are indeed more acidic and contain more free protons that can then give rise to degraded dye by-products.

4.2.2 Dye Extraction Trends

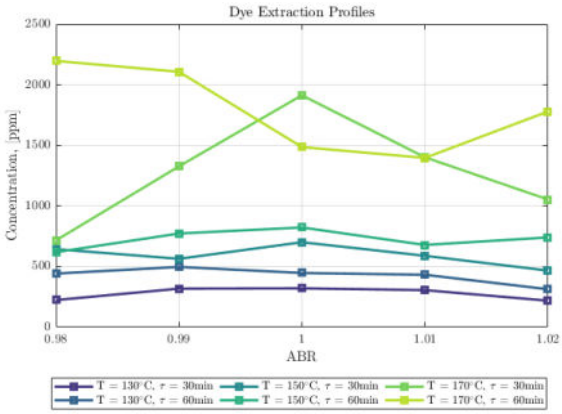


Figure 5: Dye extract concentrations with varying ABRs and temperatures.

Figure 5 outlines the effect of ABR on the concentration of the obtained dye extract. Looking within the temperature range of 130°C - 150°C, it is apparent that ABR 1.00 is the ideal choice for maximal dye extraction. This reflects the interplay between the triethylammonium cation and the sulfonate anion when it comes to dye solubility. The cation may be necessary for cation-π stacking interactions with the dye's aromatic groups while the anion will likely engage in ion-dipole interactions with any polar groups present in the structure. An equimolar amount of TEA and $MeSO_3$ (ABR 1.00) ensures that there will be a balance between these two intermolecular interactions and thus results in the most substantial amount of dye extraction from the polyester fabric.

In non-equimolar IL solutions, there is likely a lower degree of ionicity due to there being an excess of either acid or amine. This means that there will be less extensive ionic interactions that promote dye solubility. Dissolution of the dye from the polyester fabric therefore will likely be a challenge and this may be the reason that results in less concentrated dye extracts when the ABR is not equal to 1.00.

4.2.3 Fabric Decolourisation Trends

Another metric to analyse the influence of ABR on dye extraction is the extent of fabric decolourisation. This

can be defined as the reduction in a fabric's K/S value when compared to a control sample.

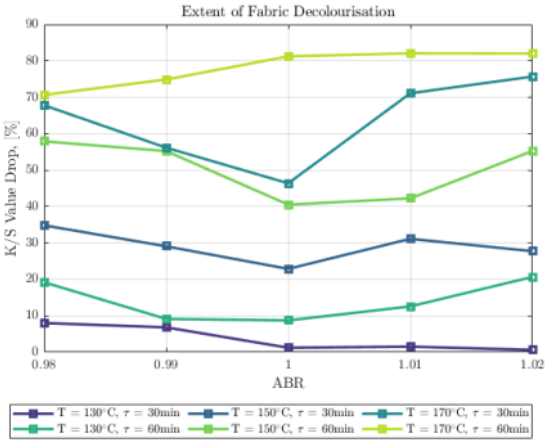


Figure 6: Extent of fabric decolourisation with reference to dry fabric control.

Figure 6 above indicates that the colour intensity of the fabric decreased the least when ABR 1.00 was used for extraction. At first glance, this reflectance drop trend seems to contradict the explanations that justified ABR 1.00 as the ideal candidate for dye extraction. Intuitively, a more concentrated dye extract should also be paired with a less intensely coloured fabric, but this is not the case.

A viable explanation for this phenomenon could be the partitioning of the dye back onto the fabric during the washing stage. The reflectance data seems to suggest that ABR 1.00 TEA-$MeSO_3$ interacts the most strongly with water and thus allows the Van der Waals forces between dye and fabric to override the dye-IL interactions in the presence of excess water. The ion dipole interactions between water and IL are not as significant at other ABRs due to an excess of free acids or bases. This entails that the dye will not be able to partition as effectively onto the fabric and will be flushed out along with the IL into the waste solution.

Furthermore, there was visual evidence that confirms this explanation. Dye was being flushed out along with the IL during the washing stage as shown in *Figure S4* in the supplementary section. This indicates that any flushed dye must have been dissolved in IL because its hydrophobicity would prevent it from interacting with water. As such, it can be interpreted that the outcome of the competition between the Van der Waals forces of the fabric and the hydrophilicity of IL determines the degree of dye partitioning back onto the fabric.

A study conducted by Minnick et al. (2016) that focused on cellulose solubility in ionic liquids when protic antisolvents were added helps support this explanation. Cellulose dissolution in ILs is driven by ionic interactions (Wang et al., 2012), which draws a parallel to the IL-dye systems explored in this study. Minnick et al. (2016) found that while protic antisolvents disrupt the IL-cellulose interactions and cause its precipitation, less polar antisolvents like alcohols cause a smaller decrease in cellulose solubility. This implies that more unfavourable IL-antisolvent interactions correspond to less recovery of the solute as is the case shown by the reflectance data.

In an operational context, DyeRecycle will have to consider the relationship and trade-off between dye extract concentration and fabric decolourisation. ABR 1.00 would theoretically yield the most concentrated dye extract but at the expense of fabric decolourisation.

A possible workaround would be to conduct these functions separately where one process unit would be optimised for extraction and another for fabric decolourisation. Operating with the highest ABR at 1.02 will enable the greatest extent of fabric decolourisation if dye degradation is not a concern. However, if dye degradation is an issue, then the choice would be to pick the lowest ABR at 0.98. Nevertheless, washing the fabric with DI water to remove any amine will be difficult as tertiary amines like TEA have low miscibility with water (Jessop et al., 2011).

4.3 Temperature Influence

4.3.1 Minimal Extraction Temperature

There is a lower temperature limit that must be surpassed for any appreciable dye extraction to occur.

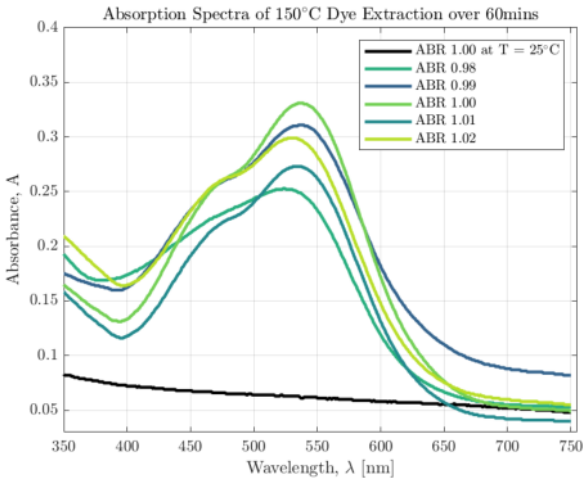


Figure 7: Comparison between successful and failed extraction runs

While dye extraction was successful for all ABRs at 150°C as seen in *Figure 7*, no visible absorption spectrum could be obtained when extraction was conducted at 25°C. The process of extraction would inherently encompass the dissolution of dyes out of the polyester fabric and into the bulk IL. This implies that IL-dye interactions are already intrinsically stronger than the Van der Waals forces that bind the dyes to the fabric. Temperature therefore can be deduced as exerting an influence on the physical nature of the polyester which then leads to the chemical interactions that mediate the transfer of dye from fabric to IL.

This is due to the free volume theory that underpins the diffusion characteristics of disperse dyes within polyester. Temperatures above its glass transition point mediate a morphological change in its overall structure. Individual chain segments will undergo reconfiguration and induce the formation of empty space where dye

molecules are able to diffuse into during commercial dyeing (Burkinshaw, 2023).

The IL therefore will not be able to gain access to these dye molecules until the temperature exceeds the glass transition point where the polyester's morphological change can occur. Below this temperature, no dye dissolution can take place and thus no extraction can happen as can be seen from the 25°C control.

4.3.2 General Temperature Trend

The trend of increasing temperature on the concentration of dye extract is obvious from *Figure 5* in Subsection 4.2.2. From 130°C to 150°C, there is more significant dye extraction across all ABRs tested. The extraction time variable is trivial in this case as more time logically caters for more dye molecules that can be transferred from fabric to solution.

This temperature trend could be explained by the free volume of polyester increasing when the temperature rises (White & Lipson, 2016). Larger porosity within the polyester's macromolecules makes the entrapped disperse dye molecules more accessible for the IL. This should then facilitate the dissolution and diffusion of dye molecules out into the IL solution.

The data point of particular interest in *Figure 5* is the ABR 1.00 sample (T = 170°C, τ = 60mins). This data point surprisingly falls below the sample that was obtained with an extraction time τ = 30mins at the same temperature. The reason for this outlier can be seen in *Figure 8* below which shows a marked increase in degradation in the 60min sample. This suggests that degradation is so significant that the decomposition products are disrupting the absorbance spectrum of undamaged dye in the sample.

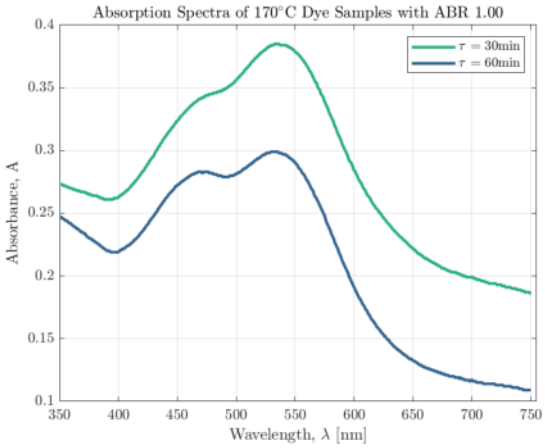


Figure 8: Absorption spectra of 170°C samples show that degradation becomes more significant with higher heat stress.

In the context of DyeRecycle's process, the IL's temperature should be maintained minimally above the glass transition point of polyester. This temperature should also ideally be as high as possible to ensure ample extraction of disperse dyes from the polyester but safely below an established degradation threshold to preserve the commercial feasibility of the dye extract.

4.4 Multistage Extraction – Elapsed Time

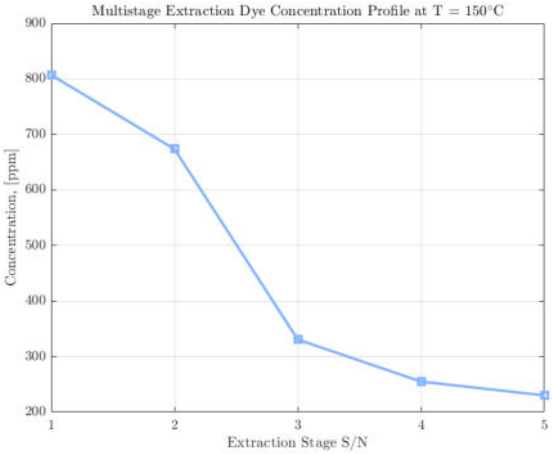


Figure 9: Dye concentration profile per extraction stage.

Figure 9 shows the change in performance of the dye extraction process with discretised stages by representing a span of continuous operating time. It is clear that the performance of extraction drops over time, demonstrating that there is a limitation to the extraction extent of dye from the fabric using a continuous supply of fresh IL. This is due to the depletion of accessible dye molecules within the fabric pores over an extended period of time which echoes the findings of Radei et al. (2018). In the process of continuous extraction, any accessible dye molecules to be extracted by the IL will gradually drop, leaving behind dye molecules that are more tightly bound within the pores of the fabric. This imposes an increasing mass transfer limitation on the fabric side.

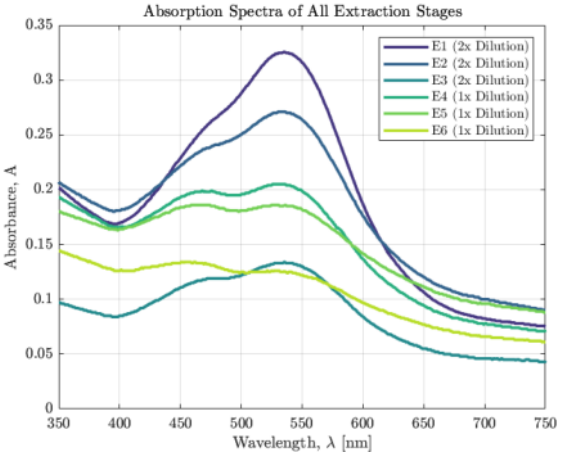


Figure 10: Absorption spectra of multiple extraction stages.

From *Figure 10*, it is observed that the extent of dye degradation eventually becomes more significant at later stages. This can be explained by the occurrence of thermal decomposition of the functional groups of dye molecules as explored by Kim & Lee (2023). However, this observation adds to the insights of dye molecule behaviour by showing the effects of prolonged heating. That is, even at a temperature that is below that of the threshold as discussed in Subsection 4.3.2, sustained heating can accumulate and provide sufficient energy to overcome the bond energy of the functional groups

in dye molecules. This causes changes to the chemical structures and leads to degraded dyes retained in the IL.

Building on from these findings, it is logical to consider a counter-current operational design in a continuous set-up to optimise dye extraction efficiency. In such a setup, the IL is to be introduced via continuous flow in a contactor, while the fabric is to be delivered from the opposite direction – possibly via a conveyor mechanism. This configuration ensures that the IL comes into contact with fabric with an optimal residence time that would maintain a substantial driving force for efficient dye extraction, whilst limiting thermal degradation due to prolonged exposure to heat.

4.5 Performance Loss across Regeneration Cycles via Activated Carbon Adsorption

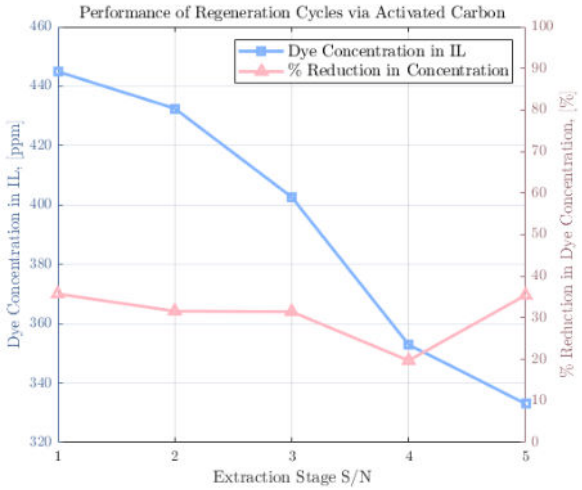


Figure 11: Dye concentration profile and % reduction in concentration across multiple regeneration cycles

The left axis in *Figure 11* shows that the dye concentration in the recycled IL decreases across multiple regeneration cycles. This indicates that the performance of dye extraction drops with extended use of the bulk IL. However, it is noted that the overall extraction performance with the use of activated carbon for IL regeneration is lower than that of constant addition of fresh IL, as in Subsection 4.4. This observation is a sign of the incomplete adsorption of dyes in between stages, where there seems to be residual dye molecules retained in the IL after each regeneration stage. This imposes a mass transfer limitation that reduces the dye extraction capabilities of recycled IL in subsequent stages.

Examining on the right axis in *Figure 11*, it is observed that the activated carbon steadily reduces the dye concentration in the IL by a third before showing more sporadic performance as the extraction proceeds. There will be a change in the chemical structure of the dyes upon degradation which may interfere with the effective adsorption of undamaged dye. This then explains the performance limitations of regeneration cycles as a trade-off to savings of IL feedstock costs via recycled use of IL.

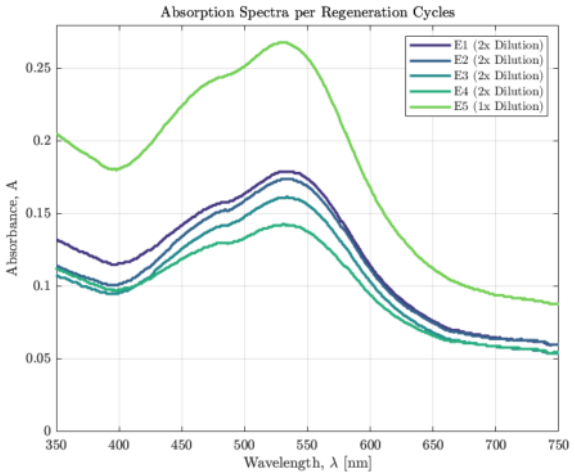


Figure 12: Absorption spectra of extraction stages with inclusion of regeneration cycles.

The effects of prolonged exposure to heat, similar to Subsection 4.4 can be seen again from *Figure 12*, when incorporating activated carbon into the cycle. There is the occurrence of thermal degradation of dyes at further stages of the process, showing the limitations to the repeated use of ILs even with activated carbon regeneration.

4.6 Effects of Varying IL Water Content

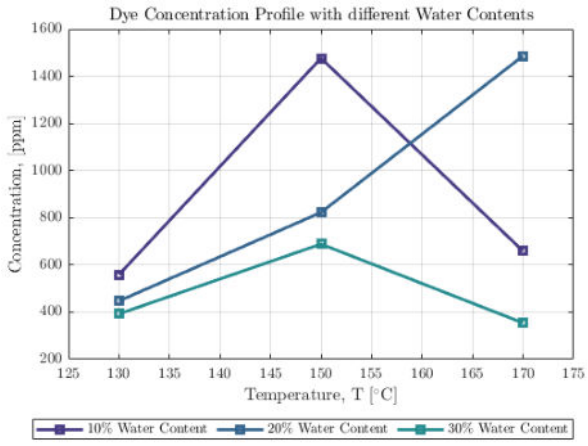


Figure 13: Dye concentration profile in of varying water content

The varying water content in the ILs can significantly affect the performance of dye extraction, as can be seen from *Figure 13*. Generally, the proportion of IL is lower in the solution at higher water contents. This then results in a lower extent of dye extraction. On a molecular scale, this could be rationalised by disruptive ion-dipole interactions between water and IL molecules that interfere with IL-dye interactions as discussed in Subsection 2.3.

However, as shown in *Figure 14*, all samples obtained from a low water content exhibit a significant extent of degradation. This demonstrates the effects of free protons within the IL solution at low water content that manipulate the chemical structures of dye molecules, leading to degradation (Pairetti et al., 2024). This means that sufficient water content in the solution is important to enhance the ionic conductivity of IL by

the means of proton scavenging (Zhou et al., 2018). This reduces the proton activity within the IL environment and limits the severity of degradation.

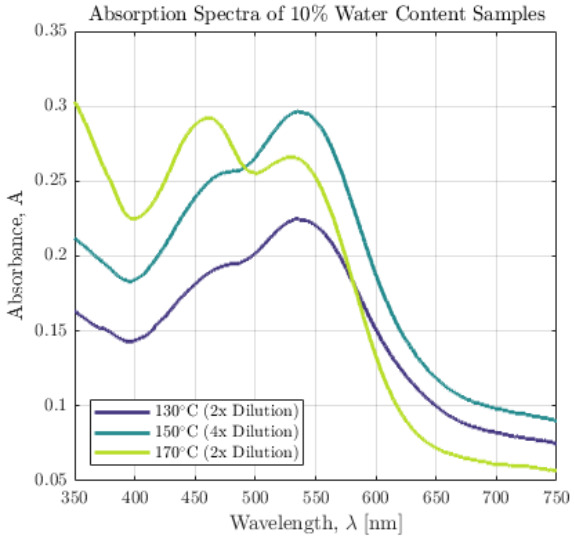


Figure 14: Absorption spectra of 10% water content samples across various temperatures

Therefore, in the synthesis of IL, it is crucial for the water content to be of a moderate level, because an optimal water content (that shown in 20%) establishes minimal proton activity within the IL environment for dye retention, while not compromising on its dye solvation capabilities due to over-dilution.

5. Conclusion

This investigation has used the metrics of dye extract concentration and extent of fabric decolourisation to analyse the performance of TEA-$MeSO_3$ mediated dye extraction from a commercial lilac coloured fabric.

The acid base ratio of the IL dictates its most appropriate function for DyeRecycle's process. ABR 1.00 was ascertained to give the highest dye extract concentration due to more considerable ionic interactions. However, non-equimolar ABRs seem to yield more noteworthy fabric decolourisation due to less dye partitioning in the washing stage. Higher ABRs lead to more decolourisation at the expense of dye degradation, while lower ABRs compromise fabric reusability for dye preservation.

Operating temperatures must be minimally above the glass transition point of polyester for any noticeable extraction to take place. A higher temperature would grant more dye extraction, but DyeRecycle should not exceed a certain thermal constraint if they want the recovered disperse dye to remain commercially viable.

The residence time of DyeRecycle's process must also be carefully calibrated. Even at moderate operating temperatures, sustained fabric heating inevitably leads to dye degradation as seen from the multistage extraction experiments. This implies that DyeRecycle should ideally operate its extraction system in a counter-current mode with a defined residence time and to ensure maximal dye mass transfer throughout the process.

There is a limit to using activated carbon to regenerate IL for successive extractions. Subjecting the dye-IL solution to prolonged heating leads to more significant degraded dye by-products. These by-products then seem to impose a dye mass transfer limitation to activated carbon adsorption. This limit becomes more apparent with more regeneration cycles. This means that the dye bath in DyeRecycle's continuous process must also have a defined residence time to preserve the IL regeneration capabilities of activated carbon which would offset new IL feedstock costs.

The water content of the IL also affects dye stability in the IL solvent. Low water contents correspond to more free protons that can react with certain dye functional groups and lead to degradation. High water contents however lead to lower IL dissolution properties due to competing ion-dipole interactions between the IL and water. To ensure minimal dye degradation but to achieve maximal solvating capacity, DyeRecycle should operate with a moderate water content of 20%.

6. Limitations and Recommendations

The findings of this study show common trends that lead to the characterisation of the influence of various variables on dye extraction. However, it must be caveated that more studies involving a wider range of dyes must be conducted to ensure repeatability and to build on the confidence of the deductions made in this investigation.

The dye re-partitioning discussion from Subsection 4.5 outlines the limitations of the K/S metric to quantify fabric decolourisation performance. There exist safety and practical restrictions of measuring the colour intensities of solid fabric samples at high temperatures in the lab. This affects the accuracy of the decolourisation results.

To circumvent the issue, further studies should utilise a continuous configuration in dye extraction to facilitate in-situ washing. The IL feed and the subsequent washing stage can be conducted at temperature to negate reabsorption of the dye-IL solution back to the fabric. This is also useful in modelling the process for the purpose of scaling up in a commercial operation.

Additionally, the certainty of the findings in regeneration cycles via activated carbon needs further investigation to quantify the difference between the adsorption extent of degraded dyes and that of nominal dyes. Another point of investigation can stem from the possible change in IL ABR during the regeneration stage due to disproportionate adsorption of ionic species that would lead to a change in ABR (Farooq et al., 2012). This might alter dye extraction performance in subsequent stages.

7. Acknowledgements

The authors would like to thank Dr. Antonio Ovejero-Pérez for his insightful discussions and technical advice, as well as the Hallett Group and the team at DyeRecycle for their guidance and support, including access to laboratory facilities essential for this research.

8. References

Abdel-Monem, Y.K. (2016) 'Efficient nanophotocatalyt of hydrothermally synthesized anatase tio2 nanoparticles from its analogue metal coordinated precursor', Journal of Materials Science: Materials in Electronics, 27(6), pp. 5723–5728. doi:10.1007/s10854-016-4484-7.

Aizenshtein, E.M. (2006). World production and consumption of polyester fibres and thread. Fibre Chemistry, 38(3), pp.264–271. doi:10.1007/s10692-006-0081-5.

Allen, R.L. (1971) 'The chemistry of Azo Dyes', Colour Chemistry, pp. 21–36. doi:10.1007/978-1-4615-6663-2_3.

Aspland, J.R. (1992), Textile Chemist & Colorist, 24(12)

Bajpai, P. (2018) Biermann's Handbook of Pulp and paper. volume 2, paper and Board making. Amsterdam, Netherlands: Elsevier.

Baraniuk, C. (2023). Will fashion firms stop burning clothes? [online] BBC Earth. Available at: https://www.bbcearth.com/news/will-fashion-firms-stop-burning-clothes.

Beall, A. (2020). Why Clothes Are so Hard to Recycle. [online] BBC. Available at: https://www.bbc.com/future/article/20200710-why-clothes-are-so-hard-to-recycle.

Burkinshaw, S.M. (2023) 'The roles of elevated temperature and carriers in the dyeing of polyester fibres using disperse dyes: Part 1 fundamental aspects', Coloration Technology, 140(2), pp. 149–207. doi:10.1111/cote.12715.

Burrell, G.L. et al. (2010) 'Preparation of protic ionic liquids with minimal water content and 15N NMR study of Proton Transfer', Physical Chemistry Chemical Physics, 12(7), p. 1571. doi:10.1039/b921432a.

DyeRecycle (2022). DyeRecycle. [online] DyeRecycle. Available at: https://www.dyerecycle.com/technology.

Ellen MacArthur Foundation (2017). A New Textiles Economy: Redesigning Fashion's Future. [online] Ellen MacArthur Foundation. Available at: https://www.ellenmacarthurfoundation.org/a-new-textiles-economy.

Farooq, A. et al. (2012) 'Adsorption of ionic liquids onto activated carbons: Effect of ph and temperature', Microporous and Mesoporous Materials, 158, pp. 55–63. doi:10.1016/j.micromeso.2012.03.008.

Goloviznina, K. et al. (2022) 'Enhancement of the solubility of organic dyes in aqueous ionic solvents doped with surfactants', Journal of Molecular Liquids, 357, p. 118958. doi:10.1016/j.molliq.2022.118958.

Greaves, T.L. and Drummond, C.J. (2007) 'Protic ionic liquids: properties and applications', Chemical Reviews, 108(1), pp. 206–237. doi:10.1021/cr068040u.

Hu, C. (2021) 'Reactor design and selection for effective continuous manufacturing of pharmaceuticals', Journal of Flow Chemistry, 11(3), pp. 243–263. doi:10.1007/s41981-021-00164-3.

Hussain, N. (2024). Imperial spinout wins big at IChemE Awards. [online] Imperial News. Available at: https://www.imperial.ac.uk/news/258939/imperial-spinout-wins-icheme-awards/ [Accessed 4 Dec. 2024].

Jessop, P.G. et al. (2011) 'Tertiary amine solvents having switchable hydrophilicity', Green Chemistry, 13(3), p. 619. doi:10.1039/c0gc00806k.

Ketema, A. and Worku, A. (2020) 'Review on intermolecular forces between dyes used for polyester dyeing and polyester fiber', Journal of Chemistry, 2020, pp. 1–7. doi:10.1155/2020/6628404.

Kim, Y.D. and Lee, J.J. (2023) 'An investigation on thermal decomposition behavior of water-soluble azo dyes', Fibers and Polymers, 24(8), pp. 2799–2806. doi:10.1007/s12221-023-00275-9.

Lemus, J. et al. (2013) 'Composition and structural effects on the adsorption of ionic liquids onto activated carbon', Environmental Science: Processes & Impacts, 15(9), p. 1752. doi:10.1039/c3em00230f.

Mahadevi, A.S. and Sastry, G.N. (2012) 'Cation–π interaction: Its role and relevance in chemistry, biology, and material science', Chemical Reviews, 113(3), pp. 2100–2138. doi:10.1021/cr300222d.

McLeod, K. and Comisarow, M.B. (1989) 'Systematic errors in the discrete integration of FT NMR spectra', Journal of Magnetic Resonance (1969), 84(3), pp. 490–500. doi:10.1016/0022-2364(89)90115-7.

Minnick, D.L. et al. (2016) 'Cellulose solubility in ionic liquid mixtures: Temperature, cosolvent, and antisolvent effects', The Journal of Physical Chemistry B, 120(32), pp. 7906–7919. doi:10.1021/acs.jpcb.6b04309.

Mittal, N., Kaur, M. and Singh, V. (2022) 'Adsorption studies on hydrophobic disperse dye using cellulose derived mesoporous activated carbon', Materials Today: Proceedings, 62, pp. 7595–7599. doi:10.1016/j.matpr.2022.04.765.

Nakasu, P.Y. et al. (2020) 'Interplay of acid–base ratio and recycling on the pretreatment performance of the protic Ionic Liquid monoethanolammonium acetate', ACS Sustainable Chemistry & Engineering, 8(21), pp. 7952–7961. doi:10.1021/acssuschemeng.0c01311.

Pairetti, Y. et al. (2024) 'Protic Poly(Diallylmethylammonium) poly(ionic liquid) proton exchange membranes with low fluorine content', Polymer Chemistry, 15(46), pp. 4763–4774. doi:10.1039/d4py01099j.

Patterson, D. and Sheldon, R.P. (1959) 'The dyeing of polyester fibres with disperse dyes. mechanism and kinetics of the process for purified dyes', Transactions of the Faraday Society, 55, p. 1254. doi:10.1039/tf9595501254.

Radei, S. et al. (2018) 'Kinetics of low temperature polyester dyeing with high molecular weight disperse dyes by solvent microemulsion and agrosourced auxiliaries', Polymers, 10(2), p. 200. doi:10.3390/polym10020200.

Sahu, O. and Singh, N. (2019) 'Significance of bioadsorption process on textile industry wastewater', The Impact and Prospects of Green Chemistry for Textile Technology, pp. 367–416. doi:10.1016/b978-0-08-102491-1.00013-7.

Schröder, U. et al. (2000) 'Water-induced accelerated ion diffusion: Voltammetric Studies in 1-methyl-3-[2,6-(s)-dimethylocten-2-yl]imidazolium tetrafluoroborate, 1-butyl-3-methylimidazolium tetrafluoroborate and hexafluorophosphate ionic liquids', New Journal of Chemistry, 24(12), pp. 1009–1015. doi:10.1039/b007172m.

Swinehart, D.F. (1962) 'The beer-lambert law', Journal of Chemical Education, 39(7), p. 333. doi:10.1021/ed039p333.

Textile Exchange (2024). Materials Market Report 2024. [online] Textile Exchange. Available at: https://textileexchange.org/app/uploads/2024/09/Materials-Market-Report-2024.pdf [Accessed 4 Dec. 2024].

Exergy analysis on sustainable aviation fuel production from air and water

Myungchan Kim
Department of Chemical Engineering, Imperial College London, U.K.

Abstract
This study examines the production of Sustainable Aviation Fuel (SAF) from air and water, focusing on the integration of prebuilt process models from prior work, which included Direct Air Capture (DAC), electrolysis, and Reverse Water-Gas Shift (RWGS), with Fischer–Tropsch synthesis (FT), hydrocracking, and separation units. This integration completes the SAF production pathway, enabling a comprehensive analysis of its efficiency. Exergy analysis was applied throughout to identify irreversibilities, with the FT reactor and PEM electrolyser identified as the primary sources of exergy destruction due to highly exothermic reactions and energy-intensive operations. The analysis demonstrated how introducing recycle streams in the FT section reduced the exergy destruction in upstream units by lowering fresh syngas requirements. To provide context, SAF production was compared to high-pressure hydrogen (700 bar) as an alternative fuel. Hydrogen production involved significantly fewer process units, leading to much lower exergy destruction (154 kW/MW of usable fuel compared to 2024 kW/MW for SAF) and reduced complexity. While SAF remains essential for aviation due to its compatibility with existing infrastructure, the results highlight the thermodynamic penalties of its production. This study underscores the value of exergy analysis in pinpointing inefficiencies and guiding process optimisation while proposing directions for future research to refine efficiency metrics and bridge enthalpy and exergy-based methodologies.

1.Introduction

In the pursuit of net-zero emissions, finding sustainable alternatives to conventional fossil fuels is critical, especially in sectors like aviation where decarbonization is challenging. Sustainable Aviation Fuel (SAF) presents a promising option as an alternative fuel that can significantly reduce lifecycle emission compared to conventional jet fuel [1]. Among various SAF production pathways, synthesizing fuel directly from air and water offers a potential sustainable approach, using CO2 captured by Direct Air Capture (DAC) units and green hydrogen produced via electrolysis. This process involves a sequence of steps: a Reverse Water-Gas Shift (RWGS) unit, followed by a Fischer-Tropsch (FT) reactor, to produce a carbon-neutral fuel source [2].

This work builds on prior summer research done at Shell Amsterdam, which focused on the SAF production process up to syngas generation, consisting of CO, H2, CO2 and H2O. The summer project included modelling Direct Air Capture (DAC), electrolysis, and RWGS units in Aspen Plus V11.1, with an emphasis on the application of exergy analysis and applying optimisations within this analysis framework. The main methodologies and findings from this summer work will be presented to contextualise the current work. The current phase focuses on modelling the FT reactor, integrating its results with the earlier syngas production model, and evaluating the overall SAF production pathway. Furthermore, given the complexity of exergy analysis, this work aims to evaluate the added value of applying exergy analysis.

The primary objective of this work is to model the FT reactor in detail and assess its integration into the SAF production pathway. By combining the FT reactor results with the earlier syngas production model, this study aims to evaluate the overall process efficiency and identify opportunities for optimisation. A key focus is the application of exergy analysis, which provides insights into the location, type, and magnitude of inefficiencies across the pathway. Additionally, this work critically evaluates the added value of exergy analysis by comparing its results with simpler, enthalpy-based assessments, considering the additional complexity required for its application.

2.1. Overview of the SAF Production Pathway

The production of Sustainable Aviation Fuel (SAF) from renewable resources, such as air and water, presents an innovative and sustainable approach to decarbonizing the aviation sector. This pathway leverages carbon dioxide captured from the atmosphere via DAC and green hydrogen produced through water electrolysis. These reactants are combined in a Reverse Water-Gas Shift (RWGS) unit to produce syngas, a mixture of H2, CO, CO2, and H2O. The syngas is then converted into hydrocarbons in a FT reactor, yielding SAF and water as products. In the summer work, SAF was assumed to be comprised of purely undecane (C11) alkanes.

This process involves several energy-intensive steps, with the efficiency of each unit operation having a significant impact on the overall sustainability and economic feasibility of SAF production. Figure 2.1 shows a simplified diagram of the overall process, denoting whether the whole process requires heating and/or electrical inputs.

2.2. Exergy Analysis

2.2.1 Exergy

Exergy is defined as the maximum useful work obtainable from a system as it moves towards thermodynamic equilibrium with a specified surrounding environment. As emphasized by Jan Szargut, the surrounding environment is typically defined as the standard environment, which reflects the most commonly occurring substances and conditions found in nature, such as atmospheric air at 25°C and 1 bar [3,4].

Unlike enthalpy, which only measures the total energy content of a system, exergy considers both the quantity and quality of energy, making it a measure of its potential to perform useful work in terms of both the first and second laws of thermodynamics. While an enthalpy balance treats two systems with the same total energy equivalently, exergy reveals differences in work potential. For example, 5 MW of energy in the form of lukewarm water at 40°C and 1 bar is not equivalent to 5 MW of high-temperature, high-pressure steam; exergy analysis captures this distinction by accounting for energy quality.

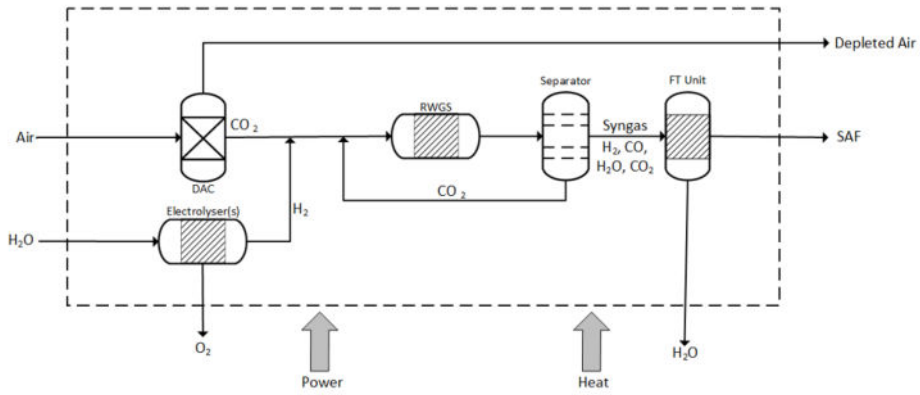


Figure 2.1: Simplified flow diagram for the overall SAF production process, self-made

A system's exergy becomes zero when it reaches equilibrium with its standard environment, as no work potential remains. Therefore, exergy quantifies how far each component or stream is from equilibrium, enabling a detailed assessment of irreversibility and inefficiencies throughout the process [3,5]. Unlike conventional energy balances, exergy analysis reveals the location, type, and magnitude of inefficiencies by capturing both thermo-mechanical and chemical contributions to energy losses.

Exergy is divided into 2 main components: physical exergy and chemical exergy, which together quantify the total work potential of a system. The total exergy of a system is simply the sum of these 2 parts:

$$Ex = Ex_{ph} + Ex_{ch} \quad (1)$$

In this study, kinetic and potential exergy contributions are neglected, as they are typically insignificant compared to the physical and chemical exergy terms in chemical process systems. This assumption is standard practice in exergy analysis for stationary systems, where velocity and elevation changes are minimal.

2.2.2. Physical Exergy

Physical exergy refers to the work potential arising from thermo-mechanical differences between the system and its reference environment. It accounts for deviations in temperature and pressure, capturing the energy available for useful work due to these differences. Physical exergy is determined using:

$$Ex_{ph} = (H - H_0) - T_0(S - S_0) \quad (2)$$

H_0 and S_0 are the values of enthalpy and entropy at the reference environment defined by T_0 and P_0. In this study, a reference environment of 25°C and 1 bar are used.

2.2.3. Chemical Exergy

Chemical exergy represents the work potential from chemical potential differences between a system and its reference environment, quantifying the energy available from chemical reactions and compositional changes. Following the methodology of Jan Szargut, the chemical exergy of a substance is calculated relative to the most stable and commonly occurring forms of elements in the standard environment [4]. For each species, a standard reaction, typically its most relevant chemical transformation (in this study: combustion), is defined. The Gibbs free energy change of the reaction is then calculated using values from NIST [7]. Usually, mixing effects are accounted for using the term $RT\sum x_i \, lnx_i$, where x_i represents the mole fraction of species i in the surroundings [3]. The general equation for chemical exergy is:

$$Ex_{ch,i} = -\Delta G_r + \sum_j RTln(x_j) \quad (3)$$

Here, $Ex_{ch,i}$ refers to some species i and j refers to all species present in the standard chemical reaction.

In this study, the mixing term was neglected because its contributions were numerically negligible compared to the Gibbs free energy change. This decision aligns with the nature of the process, which focuses on producing fuels for combustion. CO_2 and H_2O were assumed to have zero chemical exergy, as their equilibrium states already match the environment.

Table 2.1 lists the standard reactions assumed for all species participating in the process, along with their calculated chemical exergies. Using data from NIST and mole fractions of product streams, chemical exergies of FT reactor products were calculated up to C30 alkanes.

Table 2.1: All assumed standard reactions for all species present in the process

Species	Reaction	Chemical Exergy (kJ/mol)
CO	$CO + \frac{1}{2}O_2 \rightarrow CO_2$	274.5
H2	$H_2 + \frac{1}{2}O_2 \rightarrow H_2O(g)$	235.1
CO2, H2O	N/A: Species already present in the environment.	0
Hydro-carbons	$C_nH_{2n+2} + \frac{3n+1}{2}O_2 \rightarrow nCO_2 + (n+1)H_2O$	-

For this work, the standard environment is assumed to be atmospheric air with the following composition: 423 ppm CO2 [8]; 3% H2O; 20% O2; and 77% N2, at a temperature of 25°C and a pressure of 1 bar. This reference composition is used to calculate the chemical potential of components relative to their standard states.

2.2.4 Exergy balance

After physical and chemical exergies have been defined, the next step is to apply these concepts to quantify exergy destroyed and inefficiencies within the system. This is done using an exergy balance, which accounts for the total exergy entering and leaving a control volume, as well as the exergy destroyed due to irreversibility and hence entropy generation. For a steady-state system, the general form of the exergy balance is expressed as:

$$0 = Ex_Q + Ex_{elec} + \sum \dot{Ex}_{in} - \sum \dot{Ex}_{out} - Ex_d \quad (4)$$

Where Ex_{elec} represents net exergy input from electricity, where electricity is assumed to be pure exergy, Ex_d is the exergy destroyed due to irreversibilities in the system, and Ex_Q represents the net exergy input from heat calculated by:

$$Ex_Q = \left(1 - \frac{T_0}{T_b}\right) Q \quad (5)$$

T_0 refers to the reference temperature while T_b is the boundary temperature between the system and the surroundings.

Exergy destroyed, Ex_d, is the key metric in this analysis, as it directly quantifies the losses from irreversibility within the system due to entropy generation. Unlike energy balances (such as enthalpy balances), which are always conserved and cannot reflect these inefficiencies, exergy balances explicitly account for the degradation of energy quality during the process. I.e., energy balances lack a term equivalent to Ex_d, making it insufficient for diagnosing where and how losses occur. By calculating Ex_d for each unit, this study identifies the major sources of inefficiencies, providing valuable insights for process optimisation.

Additionally, both the current and summer work emphasise exergy destroyed rather than exergy loss from unutilised exergy in outlet streams. Exergy loss heavily depends on the specific flowsheet configuration, whereas exergy destroyed focuses on irreversibility inherent to the system. This distinction is important, as the study aims to address inefficiencies caused by entropy generation and irreversibility, independent of how unutilised exergy might be reclaimed or integrated in downstream processes.

2.3. Exergy Analysis in Aspen Plus V11.1

Aspen Plus, the primary process simulation tool used in this study, provides exergy values based on the concept of availability, defined as $A = H - T_0 S$, where H is the enthalpy, S is the entropy, and T_0 is the reference environment temperature [6]. While availability is often used interchangeably with exergy, it is important to note that in Aspen Plus, the reported values represent only physical exergy. These values reflect the difference in physical availability between the system and its reference state, effectively: $Ex_{ph} = A - A_0$, and do not account for chemical exergy.

This distinction is crucial for accurate analysis, as Aspen Plus reports does not account for chemical exergy. Additionally, Aspen allows users to specify the reference state's temperature (T_0) and pressure (P_0) independently, meaning the reported exergy values depend on the chosen environmental conditions.

During this study, manual verification of Aspen's reported values revealed that they accurately capture thermo-mechanical contributions but require supplementary calculations on Excel to account for chemical exergy. This is especially important for systems like SAF production, where chemical transformations significantly influence the overall exergy balance.

3. CO2 and Water to Syngas: Summer Work

To provide context for this study, a summary of the summer project, which focused on modelling the process up to syngas production, is included. This phase involved building a process simulation of all units except the FT reactor using Aspen Plus V11.1 and applying exergy analysis to identify inefficiencies and optimise the system. The findings from this work serve as a foundation for the current report, where the FT reactor is modelled and integrated into the SAF production process.

A process simulation was developed in Aspen Plus V11.1, integrating DAC, electrolysis, and RWGS units, with ambient air at 25°C and 1 bar serving as the reference environment. Initial exergy assessments revealed substantial irreversibilities in the heating and cooling units, where large temperature gradients contributed to significant exergy destruction. In response, an influent–effluent heat exchanger (HEATEX) was introduced to recuperate thermal energy internally, thereby reducing these inefficiencies. Figure 3.1 presents the improved process configuration, now featuring the HEATEX unit as a key optimisation step toward minimizing overall exergy destruction.

During the study, the following conditions and assumptions were used:

1)Reference state is ambient air at 25°C, 1 bara with composition as stated in section 2.2.3.

2)DAC: Solid-adsorbent-based system delivering pure CO2 at 80°C,20 bar [9,10].

3)Electrolyser: PEM used with a pure H2 product stream at 20 bar [11,12].

4)RWGS: operates at 20 bar and 520°C and a high selectivity to CO [13,14].

5)**The condenser** cools the stream to 40°C while 99% of water is knocked out at the SEP unit.

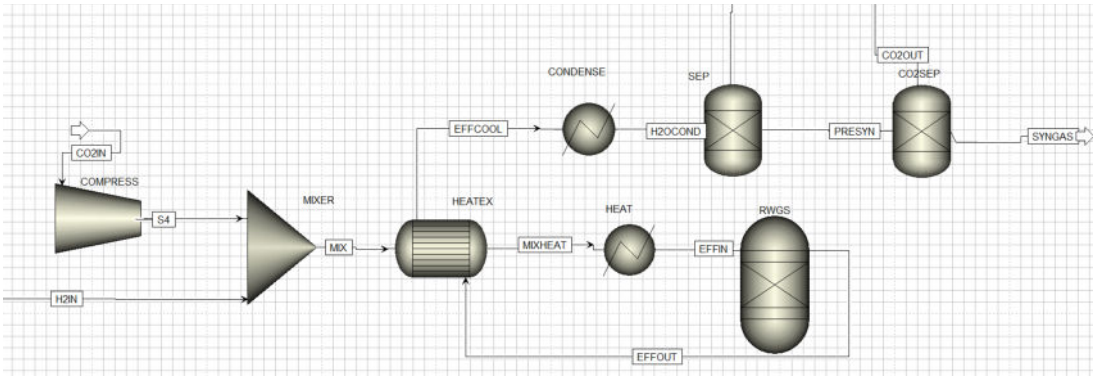


Figure 3.1: Aspen Flowsheet used for the whole process, excluding the FT reactor

3.1 Exergy balance results on both flowsheets

Exergy balances were applied to all units in both the baseline flowsheet and the improved configuration featuring the HEATEX unit. Figure 3.2 presents a side-by-side comparison of exergy destruction between the original (non-integrated) arrangement and the integrated flowsheet. The results highlight a considerable reduction in exergy destruction following the introduction of heat integration, confirming that integrating the HEATEX can substantially improve thermodynamic efficiency.

A sensitivity analysis was then conducted to examine the effect of varying the cold stream outlet temperature (i.e., adjusting the temperature approach) on exergy destruction. The findings indicate that as the temperature approach decreases toward 0 K, exergy destruction is further reduced, aligning with thermodynamic expectations based on enthalpy balances. However, even under near-ideal conditions, certain units—such as the condenser—continue to exhibit irreversibilities, emphasizing that not all inefficiencies can be eliminated solely through heat integration.

These results show the effectiveness of exergy analysis in identifying inefficiencies that conventional enthalpy balances cannot fully capture. With exergy analysis the process design can be optimised more thoroughly, providing a robust foundation for the subsequent integration of the FT reactor and downstream processes.

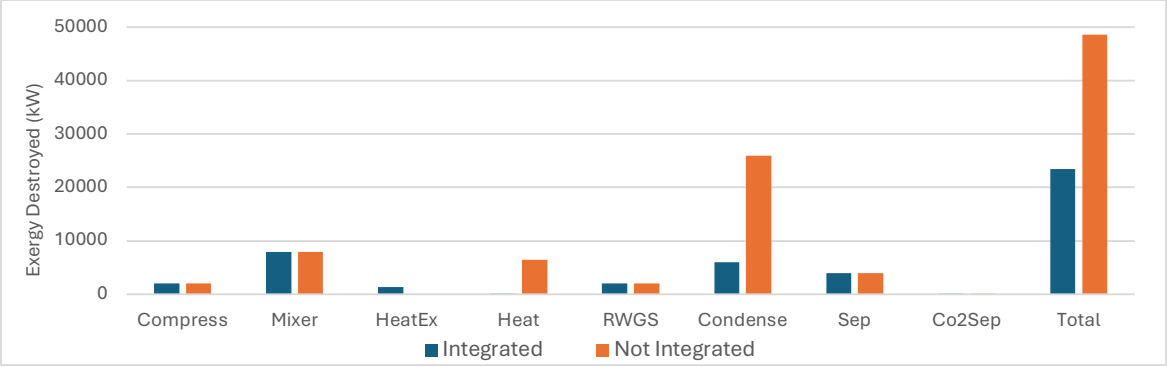


Figure 3.2: Comparison bar charts for final exergy analysis on heat integrated and non-integrated flowsheet

4. FT reactor: Current work

With the syngas generated from the summer work, the next step involves converting Syngas into SAF. This is achieved through FT synthesis. However, the direct product distribution from the FT reactor typically spans a wide range of carbon numbers and may not align with the desired SAF range. For computational tractability and availability of data, this work considers n-alkanes up to C30. Furthermore, the chosen operating condition yielded negligible amounts of heavier fractions.

To address this, the FT reactor is integrated with a hydrocracking unit, which selectively breaks down heavier hydrocarbons into lighter fractions. By carefully adjusting reaction conditions, the product distribution can be shifted toward the C8–C16 range, with a particular focus on maximizing the yield of the targeted SAF fraction (assumed to be C11 alkanes). After hydrocracking, separation units are introduced to purify the product streams, as well as to recover and recycle unreacted gases. Together, these steps refine the raw FT output into a fuel that meets SAF specifications while reducing process inefficiencies.

This chapter details the modelling of the FT reactor, the incorporation and optimisation of the hydrocracking stage, and the design of the subsequent separation train. Exergy analysis will again serve as a key diagnostic tool, enabling the identification of irreversibilities within these new units and guiding further improvements to the overall SAF production pathway. Also, the value of applying exergy analysis will be evaluated by comparing with enthalpy balances. Finally, using the results from exergy analysis, an alternative pathway—producing high-pressure hydrogen instead of SAF—will be explored to highlight the trade-offs between process complexity, efficiency, and fuel compatibility.

4.1 FT reactor: Chain Growth Model

To accurately model the FT process, it is essential to first calculate the expected product distribution to use as an input to Aspen, which is achieved using the Anderson-Schulz-Flory (ASF) model. The ASF model uses a chain growth probability parameter (α) to predict the distribution of hydrocarbons produced in the reactor. The mole fraction distribution of hydrocarbons is given by [15]:

$$x_n = n\alpha^{n-1}(1-\alpha)^2 \quad (4)$$

The chain growth probability α was calculated using the CO to H2 ratio [16]:

$$\left(0.2332 \cdot \left(\frac{y_{CO}}{y_{CO}+y_{H2}}\right)+0.633\right) \cdot (1-0.0039 \cdot ((T(°C)+273)-533)) \quad (5)$$

Where y refers to the mole fraction of the species and the CO to H2 ratio can be used to calculate $\left(\frac{y_{CO}}{y_{CO}+y_{H2}}\right)$. In the case where H2/CO ratio is 2.8, 200°C and 20 bara of the stream: $\alpha = 0.857$. Applying equation (4) to get the product distribution of the FT reactor, the figure 4.1 "FT Only" curve is obtained. Alternative methods for calculating chain growth, such as those suggested by Prins et al. [15] and Cheng et al. [17], rely on experimental kinetic rate constants that are not readily available. In contrast, the approach used in this study focuses on the H2/CO ratio, which is both controllable and easily obtainable, making it a practical and accessible method for modelling chain growth behaviour.

4.2 Hydrocracker unit

A hydrocracking unit was included in the analysis, utilising reaction kinetics obtained from experimental data [18] and shown in equations 6-8 for n-alkane cracking. In this study, symmetric cracking and primarily n-alkane cracking were considered. For symmetric splitting, if n is odd, the molecule is divided into two fragments represented by the floor and ceiling of n/2, corresponding to the two integers closest to the midpoint. For example, C15 alkane is assumed to split into equal parts C8 and C7 alkanes.

The use of the hydrocracking unit was necessary to shift the product distribution towards the C11 range, which is a target area for SAF production, rather than producing predominantly heavier hydrocarbons. Equations 9,10 show the general reaction mechanism occurring in the FT reactor followed by the Hydrocracking unit.

$$k_c = k_c^0(n) \cdot exp\left(\frac{-Ea_c(n)}{RT}\right) \quad (6)$$

$$k_c^0(n) = 2.80 \cdot 10^{29} + \frac{2.99 \cdot 10^{22} - 2.80 \cdot 10^{29}}{1 + exp\left(\frac{n - 25.25}{0.97}\right)} \quad (7)$$

$$Ea_c(n) = 186730.6 - 339264.9 \cdot \exp(-0.1988n) \quad (8)$$

$$nCO + (2n+1)\,H_2 \rightarrow C_nH_{2n+2} + n\,H_2O \quad (9)$$

$$C_nH_{2n+2} + H_2 \rightarrow 2C_{n/2}H_{n+2} \quad (10)$$

The product distribution of the product stream after hydrocracking is represented by the "After Hydrocracker" curve in Figure 4.1. The residence time of the hydrocracker was optimised so that the product stream had the highest mole fraction of C11 alkanes.

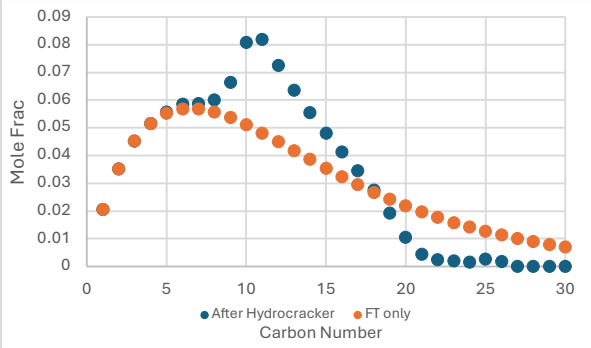


Figure 4.1: Product distribution of stream after FT and Hydrocracker

Figure 4.2 shows the temperature-dependent relative reactivity of various carbon numbers during hydrocracking, demonstrating how operating temperature affects selectivity and product distribution. A generally linear increase in reactivity with rising temperature is seen, while the entire curve shifts upward for hydrocarbons with a higher number of carbon atoms. This trend illustrates how both temperature and carbon chain length influence the cracking behaviour and product selectivity. Even when focusing solely on straight-chain alkanes, the model exhibits a strong resemblance to literature [18], confirming that temperature can serve as a reliable variable for optimising hydrocracking operation and achieving higher yields of hydrocarbons in the desired range.

4.3 Separation

After the FT reactor, the product stream is separated into gaseous and liquid phases. The gaseous fraction is assumed to contain C1-C4 alkanes and unreacted syngas (H2 and CO). To maximise resource efficiency, idealized separation units (COSEP and H2SEP in figure 4.3) are used to fully isolate hydrogen (H2) and carbon monoxide. Once separated, the CO the majority of the H2 streams are recycled back into the upstream inlet flows (streams COREC and H2REC), thereby reducing the demand for fresh syngas, hence reducing exergy destroyed from previous process units as mentioned in section 3. A portion of the purified H2 is directed to the hydrocracking unit (stream H2CRCK) to drive the reaction. The lighter hydrocarbons are then left as a product stream.

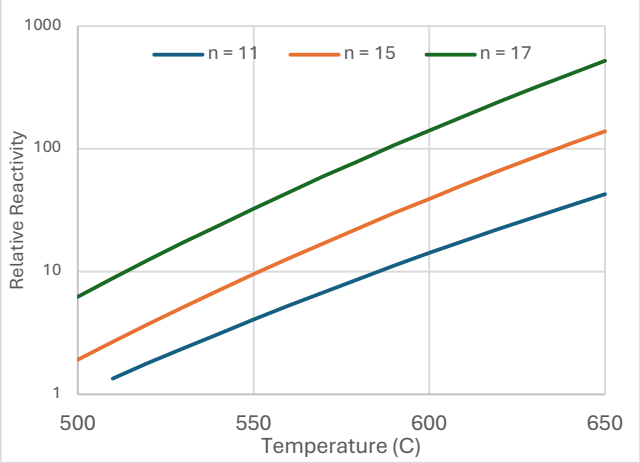


Figure 4.2: Relative reactivity based on temperature and chain hydrocarbon chain length

Meanwhile, the liquid fraction from the initial FT product separator (FTSEP) is heated and passed through the hydrocracking unit, where carefully controlled conditions facilitate the selective cleavage of heavier molecules. Subsequently, hydrocracking, an additional separation step is used to isolate the targeted SAF fraction while routing heavier fraction and any other components toward appropriate pathways.

4.4 Full Pathway and Exergy Analysis Results

Aspen Plus was used to simulate the integrated FT unit shown in Figure 4.3. Exergy analysis was then applied to the extended configuration, where the FT section is integrated with the upstream processes from the summer work by connecting the COIN and H2IN streams in Figure 4.3 to the SYNGAS stream in Figure 3.1. This integration

establishes the complete process flowsheet as outlined in Figure 2.1.

With the full pathway defined, exergy analysis was conducted to investigate the effect of introducing recycle loops in the FT section. Figure 4.4 presents the exergy destroyed normalized per tonne of SAF production, assuming SAF is purely C11 alkane. Under these conditions, the exergy destroyed in the FT synthesis unit remains unchanged. However, the reduced need for fresh syngas lowers the throughput in upstream units—particularly the PEM electrolyser—thereby decreasing the overall exergy destruction.

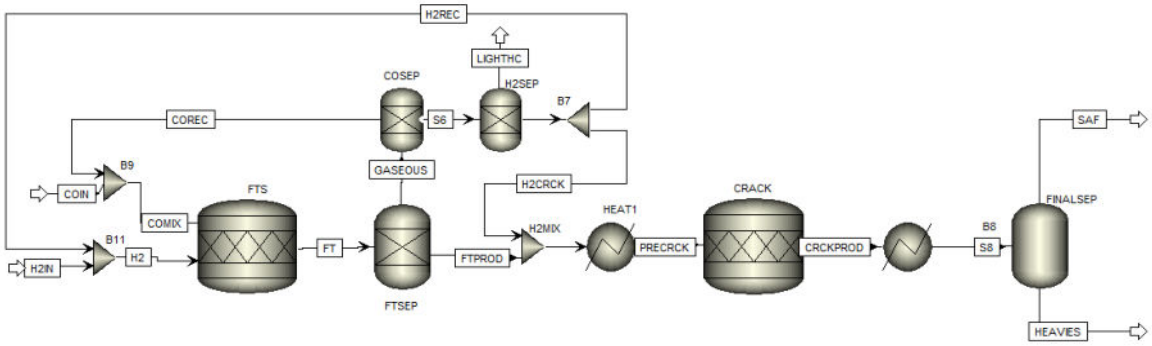


Figure 4.3: Full flowsheet for the FT unit with Syngas Recycling

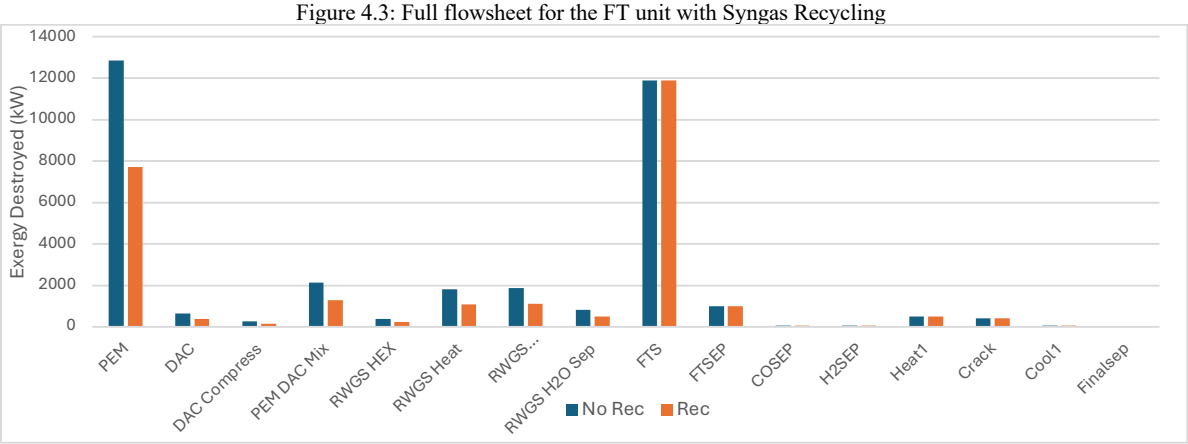


Figure 4.4 Exergy destroyed per tonne of SAF produced of whole integrated process depending on FT recycling

4.5 Comparing Enthalpy and Exergy Balances

In Aspen Plus, enthalpy balances are always conserved, as the software reports heat duties based on the difference between inlet and outlet streams. While this ensures energy conservation is consistently verified, enthalpy balances alone cannot distinguish between high-quality and low-quality energy, nor can they identify irreversibilities. Moreover, determining which streams are "useful" is a subjective choice, further limiting the insights gained from pure enthalpy assessments. Exergy analysis, by contrast, directly quantifies where and how useful work potential is lost, providing a more objective basis for identifying targets for improvement.

4.6 Alternative Fuels: High-Pressure Hydrogen vs SAF

The exergy analysis highlights substantial inefficiencies in SAF production, particularly in the FT synthesis step. The FT process is highly exothermic, meaning that much of the useful work potential (chemical exergy) in the input syngas is lost as heat during the reaction. Moreover, the FT process involves using a high-exergy fuel, such as hydrogen, to create another fuel, inherently compounding inefficiencies. These factors, combined with the extensive downstream processing required to purify SAF from the broader hydrocarbon distribution, result in significant irreversibilities.

Meanwhile, producing high-pressure hydrogen (700 bar) requires only a PEM electrolyser and a compressor, resulting in much lower exergy destruction and process costs. For 1 MW of usable fuel, hydrogen production destroys just 154 kW of exergy, compared to 2024 kW for SAF. This difference shows both the complexity of the SAF pathway and the inherent losses from the FT reactor's exothermic reactions and diverse product distribution. The additional processing required to separate hydrocarbons and achieve the same energy output further increases inefficiencies.

These results demonstrate the thermodynamic trade-offs in SAF production, emphasizing the need for further optimisation of FT synthesis and downstream processing. While SAF remains critical for aviation, the simplicity and efficiency of hydrogen production highlight its potential as an alternative fuel in non-aviation applications.

5. Conclusion, Future Work

This study developed and analysed a complete process for Sustainable Aviation Fuel (SAF) production from air and water, integrating upstream syngas generation with FT synthesis and downstream processing. By applying exergy analysis, the key sources of irreversibility in the process were identified, with the FT reactor and PEM electrolyser emerging as the primary contributors to exergy destruction. The highly exothermic nature of the FT reactions, coupled with the inefficiencies inherent in converting high-exergy hydrogen into hydrocarbons, underlines the thermodynamic trade-offs in SAF production.

Exergy analysis proved invaluable in highlighting these inefficiencies and provided a framework for assessing alternative pathways. A comparison with high-pressure hydrogen production revealed significantly lower exergy destruction and process complexity, demonstrating its potential as a simpler, more efficient fuel in non-aviation applications. These insights indicate the importance of optimising SAF processes, particularly in mitigating the irreversibilities in the FT reactor and downstream processing.

Ultimately, this work illustrates the value of exergy analysis as a diagnostic tool for pinpointing inefficiencies and proposing targeted improvements. By quantifying irreversibilities and evaluating alternative pathways, it provides a foundation for advancing SAF production while informing broader discussions on sustainable fuel strategies.

Future research should focus on refining the comparison between enthalpy and exergy analyses. Developing a more detailed flowsheet with realistic process conditions, such as incorporating pressure drops and other inefficiencies, would allow for a comprehensive enthalpy balance where "useful" and "waste" streams are explicitly classified. Comparing these results with exergetic efficiency could validate whether exergy inherently accounts for such inefficiencies from the outset, further proving its diagnostic value.

For the FT synthesis step, further work is needed to isolate specific sources of exergy destruction, including the highly exothermic reactions, mixing inefficiencies, and suboptimal product distributions. This could guide improvements in reactor design and heat recovery strategies. Additionally, addressing Aspen's incomplete entropy balance, which only considers Gibbs free energy contributions, could provide a more accurate assessment of entropy generation and exergy destruction.

Lastly, developing a standardized exergy-based efficiency metric would enable consistent comparisons across processes and pathways. Such a metric could bridge the gap between enthalpy and exergy analyses, offering both a practical and theoretically grounded framework for evaluating complex systems like SAF production. By addressing these areas, future studies can enhance SAF process optimisation and strengthen the methodologies for evaluating its performance.

References

[1] Afonso, F., et al. (2023). Strategies towards a more sustainable aviation: A systematic review. Progress in Aerospace Sciences, [online] 137(137), p.100878. doi:https://doi.org/10.1016/j.paerosci.2022.100878.

[2] Rojas-Michaga, M.F., et al. (2023). Sustainable aviation fuel (SAF) production through power-to-liquid (PtL): A combined techno-economic and life cycle assessment. Energy Conversion and Management, [online]292,p.117427. doi:https://doi.org/10.1016/j.enconman.2023.117427.

[3] Szargut, J. (2005). *Exergy method : technical and ecological applications*. Southampton ; Boston: Wit Press.

[4] Szargut, J. (1989). Chemical exergies of the elements. *Applied Energy*, 32(4), pp.269–286. doi:https://doi.org/10.1016/0306-2619(89)90016-0.

[5] Szargut, J., Morris, D.R. and Steward, F.R. (1988). *Exergy Analysis of Thermal, Chemical, and Metallurgical Processes*. Chem/Mats-Sci/E.

[6] Gray, T.D. (2019). An introduction to exergy and its evaluation using Aspen Plus. [online] K-state.edu. Available at: https://krex.k-state.edu/items/8fccac48-b083-4180-9a5d-0d14c7397ec8

[7] National Institute of Standards and Technology (1997). *NIST Chemistry WebBook - SRD 69*. [online] Nist.gov. Available at: https://webbook.nist.gov/chemistry/.

[8] NASA (2024). *Carbon Dioxide Concentration | NASA Global Climate Change*. [online] Climate Change: Vital Signs of the Planet. Available at: https://climate.nasa.gov/vital-signs/carbon-dioxide/?intent=121.

[9] Hoseinpoori, S., Pallarès, D., Johnsson, F. et al. A comparative exergy-based assessment of direct air capture technologies. Mitig Adapt Strateg Glob Change 28, 39 (2023). https://doi.org/10.1007/s11027-023-10076-3

[10] Marlair G, Cwiklinski C, Tewarson A (1999) An analysis of some practical methods for estimating heats of combustion in fire safety studies. Proceedings Interflam 1999

[11] S. Shiva Kumar, V. Himabindu, Hydrogen production by PEM water electrolysis – A review, Materials Science for Energy Technologies, Volume 2, Issue 3, 2019, Pages 442-454, ISSN 2589-2991, https://doi.org/10.1016/j.mset.2019.03.002 .

[12] Olivier, P., Bourasseau, C. and Bouamama, B. (2016). Modelling, simulation and analysis of a PEM electrolysis system. *IFAC-PapersOnLine*, 49(12), pp.1014–1019. doi:https://doi.org/10.1016/j.ifacol.2016.07.575.

[13] Daza, Y.A. and Kuhn, J.N. (2016). CO2 conversion by reverse water gas shift catalysis: comparison of catalysts, mechanisms and their consequences for CO2 conversion to liquid fuels. *RSC Advances*, [online] 6(55), pp.49675–49691. doi:https://doi.org/10.1039/C6RA05414E.

[14] Su, X., Yang, X., Zhao, B. and Huang, Y. (2017). Designing of highly selective and high-temperature endurable RWGS heterogeneous catalysts: recent advances and the future directions. *Journal of Energy Chemistry*, 26(5), pp.854–867. doi:https://doi.org/10.1016/j.jechem.2017.07.006.

[15] Prins, M.J., Ptasinski, K.J. and Janssen, F.J.J.G. (2005). Exergetic optimisation of a production process of Fischer–Tropsch fuels from biomass. *Fuel Processing Technology*, 86(4), pp.375–389. doi:https://doi.org/10.1016/j.fuproc.2004.05.008.

[16] ScienceDirect. (n.d.). *Industrial Chemical Process Analysis and Design*. [online] https://www.sciencedirect.com/book/9780081010938/industrial-chemical-process-analysis-and-design.

[17] CHENG, J., HU, P., ELLIS, P., FRENCH, S., KELLY, G. and LOK, C. (2008). A DFT study of the chain growth probability in Fischer–Tropsch synthesis. *Journal of catalysis*, 257(1), pp.221–228. doi:https://doi.org/10.1016/j.jcat.2008.05.006.

[18] Fernandes, F.A.N. and Teles, U.M. (2007). Modeling and optimization of Fischer–Tropsch products hydrocracking. *Fuel Processing Technology*, [online] 88(2), pp.207–214. doi:https://doi.org/10.1016/j.fuproc.2006.09.003.

[19] Dancuart, L.P., de Haan, R. and de Klerk, A. (2007). Processing of Primary Fischer-Tropsch Products. [online] Elsevier, pp.482–532. doi:https://doi.org/10.1016/S0167-2991(04)80463-4.

Exploring Optimal Parameters for Graphene Liquid-Phase Exfoliation Using Taylor-Coutte Device

Sooa Ji
Department of Chemical Engineering, Imperial College London, U.K.

Abstract

Graphene, a two-dimensional (2D) nanomaterial, has garnered significant attention due to its remarkable properties, including a high specific surface area, exceptional tensile strength, and superior electrical and thermal conductivity. These attributes make it invaluable for a broad spectrum of industrial applications. Among the established methods for graphene exfoliation – mechanical, electrochemical, and liquid-phase exfoliation. The present study focuses on optimising liquid-phase exfoliation using a custom-designed high-shear Taylor-Couette device. Sodium cholate (NaCh) in deionised water served as the surfactant-assisted solvent system. Following exfoliation, centrifugation was employed to separate few-layer graphene (FLG) from partially graphite, and the resultant dispersion was diluted and characterised using UV-Vis spectroscopy to determine FLG concentration and the average number of atomic layers. Four distinct graphite precursors (vein graphite, Canadian graphite, expanded graphite, and Chinese graphite flakes) were investigated under varying process conditions of graphite concentration, exfoliation duration, and surfactant concentration to identify parameters that maximise graphene yield. The results revealed that Chinese graphite flakes produced the highest FLG yield among the tested precursors. Moreover, higher initial graphite concentrations correlated with improved yields, while prolonged exfoliation enhanced production efficiency up to a threshold, beyond which diminishing rates of production were observed. An optimal surfactant concentration of 2.5 g/L was established for maximising yield. This study provides a systematic approach to process optimisation, offering valuable insights for enhancing the scalability and efficiency of graphene production via liquid-phase exfoliation.

Keywords: Graphene, Liquid-exfoliation, Taylor-Couette device, Process optimisation

1. Introduction & Background

Carbon is widely distributed on Earth and is a major organic compound component. Carbon atoms have a $2s^2$, $2p^2$ configuration, which can form allotropes and compounds by forming s-p hybridised chemical bonds of sp, sp^2 and sp^3 [1]. Carbon has more than 500 allotropes, including graphite, fullerene, carbon nanotubes, diamond and graphene [2].

Graphene is a two-dimensional nanomaterial in which one carbon atom is bonded to three other carbon atoms in a hexagonal shape in a plane. Graphene has a slight thickness of 0.334 nm but has a large theoretical specific surface area of 2,600 $\mathrm{m^2/g}$ [3], making graphene widely used in energy storage devices such as batteries. In lithium-ion batteries, graphene is used as an anode, which makes the capacity three times higher than that of graphite electrodes. Graphene is highly flexible, so it is widely utilised in wearable electronics. Graphene is 100 times more potent than steel with a tensile strength of 125 GPa [4], and thermal conductivity is 10 times higher than copper with 5×10^3 $\mathrm{W/mK}$ [5].

Moreover, at room temperature, it has electronic mobility of 200,000 $\mathrm{cm^2/Vs}$ due to the band-gap characteristic, which makes electrons move freely between the valence and conduction bands. It also has a high optical transmission of 97.4 % [6], used to make touchscreens and organic light-emitting diodes (OLEDs). Since graphene also has high light absorption, it improves the light absorption capacity of solar cells by bonding graphene to cadmium sulfide (CDS) semiconductors and adding carbon nanotube, increasing the efficiency by 15.2 % [4]. Graphene is a significant nanomaterial in various fields due to its unique and superior properties.

Delocalised π bonds connect the crystal layer of graphite flakes with van der Waals forces, and the binding force is 16.7 $\mathrm{kJ/mol}$ [9]. Carbon atoms form a covalent bond with the sp^2 hybridised orbital with a binding energy of 345 $\mathrm{kJ/mol}$. Since the binding energy between the crystal layers is more than 20 times weaker than the binding energy of the carbon atom and the orbital, the graphene can be exfoliated by separating the layer from the graphite. Through this principle, research to find efficient ways to exfoliate graphene is ongoing.

In 2004, Geim *et al*. discovered a method of exfoliating a graphene sheet from a graphite sheet using the "Sticky tape method" at Manchester University [8]. This mechanical exfoliation method separates graphene from graphite by repeatedly attaching scotch tape and detaching to overcome weak van der Waals forces. This method can produce high-quality graphene but can only be applied at the laboratory scale.

For the electrochemical exfoliation of graphite, oxygen radicals are formed by applying a voltage of 20 V to two graphite rods immersed in an electrolyte composed of 1 % sodium dodecyl sulfate (SDS) with deionised water [11]. The corrosion of the graphite rod electrodes is followed by the insertion of anionic surfactant, which precipitates the sheet of graphene.

Liquid-phase exfoliation is a top-down method of dispersing graphite in a liquid medium and overcoming the weak van der Waals forces using a physical method such as sonication to obtain graphene by separating the layer. The surface energy of graphite is 40 mJ/m^2 [10], and the solvent should have a similar surface energy to graphite, such as N-Methylpyrrolidone (NMP) and Dimethylformamide (DMF), that can offset the bonding energy between the graphite layers.

The objective of this research is to find a way to increase the yield of graphene by experimenting with various precursors in several process conditions through liquid-phase exfoliation.

2. Methods

2.1 Precursors

Vein graphite, Canadian graphite, expanded graphite, and Chinese graphite flake were precursors to exfoliate graphene. The concentration of vein graphite, Canadian graphite, and expanded graphite was fixed at 20 g/L. The concentration of Chinese graphite flake was varied to 20, 100, and 200 g/L to compare the yield variation depending on the graphite concentration.

2.2 Solvent

Flakes tend to adhere to each other due to their cohesiveness, interrupting graphite dispersion in the solvent [13]. To overcome van der Waals forces, the surface energy of the solvent should be similar to the surface energy of graphite to minimise interfacial tension among flakes and facilitate graphene dispersion [13]. DMF (37.1 mJ/m^2) and NMP (40 mJ/m^2) are used as a solvent in liquid-phase exfoliation. However, these organic solvents are expensive and toxic. The research is limited to the effect of NMP on humans; however, Solomon *et al.* experimented with the exposure of NMP, which can cause developmental disability and sensory responsiveness changes in the fetus of mice [14]. James *et al.* categorised DMF as a Substance of Very High Concern (SVHC), which can cause developmental disability and reproductive toxicity [15].

NMP and DMF are substances that encourage the reduction of use in research and industrial sites and the use of safe alternative substances. In this experiment, sodium cholate, which is non-cancerous and can be safely used at an appropriate concentration, was used as an additive at appropriate concentrations to form a stable aqueous dispersing solution. Sodium cholate (NaCh) usually has a surface energy of 30-35 mJ/m^2 and is used to exfoliate layered materials by lowering the surface tension.

This solvent provides electrical and three-dimensional repulsive force to prevent re-aggregation of graphene, thereby increasing dispersion stability. This study used 2 g/L of Sodium cholate solution to exfoliate vein graphite, Canadian graphite and expanded graphite. During the Chinese graphite flake experiment, 1, 2, 2.5, and 4 g/L of sodium cholate were used to observe the change due to the different solvent concentrations. The prepared sodium cholate in each experiment was mixed well with 2 L of ultra-pure deionised water (18 $M\Omega.cm$).

2.3 Taylor-Couette technology

Coleman *et al.* proposed a high-shear rotator mixer, which showed a low yield of 0.1 % [11], and the exfoliation phenomenon was limited to the rotor and stator. Tran *et al*. proposed using the Taylor-Couette device for liquid-phase exfoliation to overcome this [12].

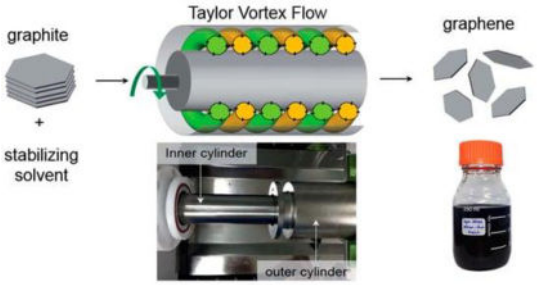


Figure 1 Schematic of exfoliation process and Taylor Vortex Flow
Reproduced with permission from Ref. 12. Copyright 2016 The Royal Society of Chemistry.

The Taylor Colette device consists of two concentric cylinders with stationary outer (stator) and rotating inner (rotor) cylinders. The specifications of the Taylor-Couette device used: outer cylinder diameter = 197 mm, an inner cylinder diameter = 193 mm, and a cylinder length = 280 mm, which was built by the Department of Chemical Engineering at Imperial College London. The flow generated by the rapidly rotating liquid in the small gap of the two coaxial cylinders is called Taylor-Couette flow. When the inner cylinder rotates above the critical rotation speed, the flow becomes unstable, resulting in a secondary flow showing a toroidal-shaped vortex rotating in the opposite direction, known as a Taylor vortex flow shown in Figure 1 [12]. The liquid rotating by these two cylinders creates a high shear. The shear rate can be calculated using the equation below:

$$Re_T = \frac{r_i \omega d}{v} \qquad (1)$$

$$G = 1.45 \frac{\eta_r^{\frac{3}{2}}}{(1-\eta_r)^{\frac{7}{4}}} Re_T^{1.5} \quad (2)$$

$$\gamma = \frac{Gv}{2\pi r_i^2} \quad (3)$$

where Re_T is the Reynolds number for the Taylor Couette flow from Lathrop *et al* [21]. Where reynolds number (Re) < 2000 is a laminar flow, 2000 < Re < 4000 is a transition flow, and 4000 < Re is a turbulent flow. r_i is the inner cylinder radius, ω is the rotational speed in rad.s^{-1}, d is the exfoliation channel width, v is the kinematic viscosity of DI water, G is the dimensionless torque, and $\eta_r = r_i/r_o$. In this process, the van der Waals force is overcome, and the graphite is exfoliated into the solution by graphene due to the secondary flow. The Taylor-Couette device allows the graphene to have high yield and low defect.

2.4 Operational procedures

Prior to the experiment, individual surfactant aqueous solutions were prepared of required concentrations with an overnight mixing to ensure complete dissolution. Figure 2 and Figure 3 shows a photo and schematic, respectively, of the experimental setup used.



Figure 2 Overview of the experimental setup

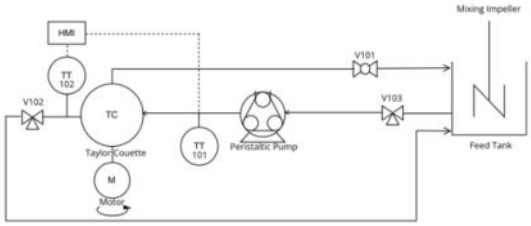


Figure 3 Schematic showing the experimental setup

At the start of the experiment, the device was operated through a software installed on a computer, and the cylinder rotational speed of the Taylor-Couette device could be set via motor-controller connected to this computer. The inner cylinder of the Taylor-Couette device rotates by the connection of the pulley and belt, and the outer cylinder is stationary. The temperature change between the inlet and outlet of the TC device was measured using Pico TC-08 Thermocouple Temperature Data Logger. In this experiment, the appropriate temperature is less than 35 ℃. To prevent overheating, the fan was operated to maintain the appropriate temperature i.e., <70 ℃. When the experiment started, the V-101 was closed, the V-102 was opened, Masterflex 07528-10 L/S peristaltic pump equipped with a 77200-62 pump head was used and operated fixed speed at 250 RPM, and the surfactant was injected into a 2L beaker. When the cylinder rotational speed was gradually increased to reach 400 RPM, the mixer immersed in the surfactant was operated at 450 RPM which provided sufficient agitation without any aeration. The prepared graphite was poured, mixed, and injected into the Taylor-Couette device.

After that, the speed was gradually increased to set the speed suitable for each experiment as shown in Table 1. The exfoliating time, the concentration of precursors, the concentration of surfactant, and the cylinder rotational speed were set for each experiment, and the experiment was carried out. After sufficient exfoliation, the graphene product was extracted back into a 2 L beaker where surfactant was originally contained through a flexible rubber tube when the cylinder rotational speed was 400 RPM.

Table 1 Process conditions

Type of Graphite	Time (hours)	ω_c (RPM)	$C_{graphite}$ (g/L)	C_{NaCh} (g/L)
Vein Graphite 1	3	1000	20	2
Vein Graphite 2	5	1000	20	2
Canadian Graphite	3	1000	20	2
Expanded Graphite	3	1000	20	2
Chinese Graphite Flake 1	3	1000	100	2
Chinese Graphite Flake 2	3	1100	20	2
Chinese Graphite Flake 3	9	800	100	1
Chinese Graphite Flake 4	9	800	100	2.5
Chinese Graphite Flake 5	9	800	100	4

2.5 Shut-down procedure

After removing the whole volume (2 L) of processed product into a Feed Tank (beaker), the cleaning of the system was followed to decontaminate and prepare for the next batch production. The V-101 was opened, closed the V-102, set the pump speed to 350 RPM, set the cylinder rotational speed to 150 RPM, and injected 3.5 L of DI water. Removed the remaining substances in the Taylor-Couette device by opening and closing the primary and outlet valves until the discharged DI water became transparent. When the remaining substances were discharged, the device's speed was changed to 0 RPM, all fluid was removed from the device, and the experiment was finished.

2.6 Centrifugation

The extracted graphene solution was left overnight to leave graphene flakes in the bulk solvent and allow for the graphite particles to settle. After that, 54 mL of graphene solution is extracted from the supernatant with a pipette and placed in six centrifuge tubes. The instrument used in the centrifuge is the Thermo Scientific Sorvall Legend X1R. The weight of the oppositely facing centrifuge tubes positioned in the centrifuge machine were balanced within 0.1 g. Post centrifugation (30 minutes at 1500 rpm) 42 mL of supernatant were collected with the pipette.

2.7 Dilution

Table 2 Dilution factor for each type of experiment

Type of graphite	Dilution Factor
Vein Graphite 1	250
Vein Graphite 2	250
Canadian Graphite	250
Expanded Graphite	50
Chinese Graphite Flake 1	500
Chinese Graphite Flake 2	250
Chinese Graphite Flake 3	500
Chinese Graphite Flake 4	500
Chinese Graphite Flake 5	500

Centrifuged products were collected in a 250 mL vials, which were further diluted with a corresponding surfactant aqueous solution to obtain desired dilution factor. The formula of the dilution factor is

$$M_1V_1 = M_2V_2 \qquad (4)$$

where M_1 is the initial concentration, V_1 is the initial volume, M_2 is the final concentration, and V_2 is the final volume. The diluted solution can be characterised by UV-Vis, which requires a degree of transparency for the analysis. Above table is a showing each appropriate dilution factor.

2.8 Product characterization

UV-Vis-nIR spectrometer (Shimadzu UV-2700) with 3.5ml of quartz cuvettes was used as an instrument. The absorbance of the graphene solution can be observed by light absorbed by the diluted sample at wavelengths of 200 nm to 900 nm. The wavelength of 200 to 400 nm is in the range of ultraviolet (UV), 400 to 800 nm is in the range of visible light (Vis), and 800 to 900nm is in the range of near-infrared (nIR).

When electromagnetic waves irradiate the material, scattering, reflection, absorption, and transmission occur. Absorption occurs when the amount of energy of light irradiated to the material is equal to the energy difference between the ground and excited states of the molecule. The transition of an electron from the ground state to the excited state is called an electron transition. In general, transitions that occur at short wavelengths, such as $\sigma \rightarrow \sigma^*$ and $n \rightarrow \sigma^*$, are not observed, but since $n \rightarrow \pi^*$and $\pi \rightarrow \pi^*$ transitions occur at 200 to 700 nm, they show a strong absorption reaction [16].

A few-layer graphene concentration was determined by applying specified absorbance to the Beer-Lambert Law.

$$A = l\alpha_c C_{FLG} \qquad (5)$$

$$\epsilon = \alpha_c + \sigma_s \qquad (6)$$

A is the absorbance, l is the length of the cuvette length (pathlength), and C_{FLG} is the concentration for few-layered graphene. ϵ is extinction coefficient and σ_s is the scattering coefficient. Backes *et al.* found that the absorbance and extinction coefficients are qualitatively similar [17]. Accordingly, in this experiment, the extinction coefficient to obtain the absorbance was defined and calculated as the absorbance coefficient.

In this experiment, the absorbance at 680 nm was used to calculate the concentration of a few-layer graphene; the absorbance coefficient at this time was 4237 L/g/m (determined by the research group by gravimetric analysis), and the optical length of the cuvette was 0.01m. Average layer number, $<N>$, which can be obtained by the following well-established correlation:

$$< N >= 0.42\lambda_{peak}(nm) - 108 \qquad (7)$$

Where λ_{peak} is wavelength at peak. However, this formula has the problem that it is difficult to find the exact peak point due to the wide range of peaks observed in the experiment, and the position of the peak can be changed depending on the environment by the solvatochromic effect [17].

The intensity ratio of the absorbance at the π–π^* peak to the high wavelength plateau is used to find the exact location of the peak [17].

$$< N1 >= 25 \times \frac{\varepsilon_{550}}{\varepsilon_{max}} - 4.2 \qquad (8)$$

ε_{550} is the absorbance at 550 nm of wavelength and ε_{max} is the maximum absorbance. However, this method also has a problem: solvents and surfactants absorb UV light [17]. As an improvement, the formula changed the intensity ratio to ε550/ε325

$$< N2 >= 35.7 \times \frac{\varepsilon_{550}}{\varepsilon_{325}} - 14.8 \qquad (9)$$

3 Results & Discussion

3.1 Concentration of few-layer graphene

Figure 4 shows the absorbance of all experiments measured by UV-vis according to the wavelength. The π-π* transition of the C=C bond of graphene causes the strong absorption peak seen between 260 and 280 nm [18]. This transition is caused by the delocalization of π electrons present in the overlapping region of the p orbital of sp2 hybridised in defect-free graphene [19]. This structure is the reason graphene has zero bandgap characteristics.

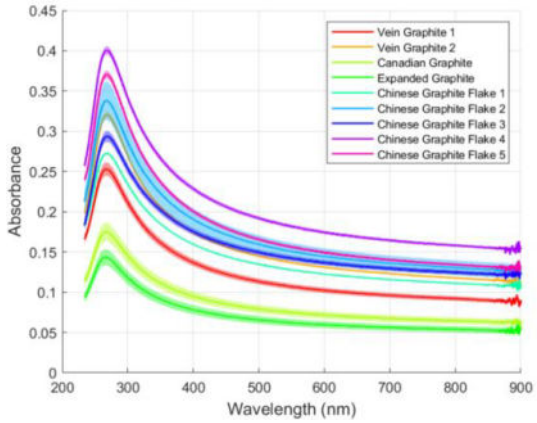


Figure 4 Absorbance intensity spectra for all processed samples including individual standard error

Table 3 Concentration of few-layer graphene

Type of Graphite	C_{FLG} (g/L)
Vein Graphite 1	0.57
Vein Graphite 2	0.74
Canadian Graphite	0.40
Expanded Graphite	0.07
Chinese Graphite Flake 1	1.38
Chinese Graphite Flake 2	0.82
Chinese Graphite Flake 3	1.54
Chinese Graphite Flake 4	1.96
Chinese Graphite Flake 5	1.67

Table 3 shows the value measured by UV-vis, and Equation 4 calculates the concentration of few-layer graphene in the entire experiment. Chinese graphite flake 4 showed the highest yield throughout the experiment.

3.2 Shear rate

The Taylor-Couette device produces a high shear by rotating only the inner cylinder while the outer cylinder is fixed. Based on the Reynolds number, the flow regime is turbulent. Corresponding values are calculated and shown in Table 4, where the shear rate increases as the rotational speed increases, calculated by Equation 3. In all cases, shear rate is greater than the critical shear rate for graphite exfoliation, i.e., $\dot{\gamma}_{crit} = 10^4$ s^{-1}.

Table 4 Shear rates for three different rotational cylinder speeds

Rotational cylinder speed (RPM)	Angular velocity of rotational cylinder ω (rad/s)	Reynolds number for the Taylor Couette flow Re_T	Torque G	Shear rate $\gamma(s^{-1})$
800	83.78	2.19×10^4	4.29×10^9	5.22×10^4
1000	104.72	2.74×10^4	6.04×10^9	7.63×10^4
1100	115.19	3.01×10^4	6.97×10^4	8.80×10^4

3.3 Number of layers

Table 5 Number of atomic layers for each sample

Type of graphite	N_1	N_2
Vein Graphite 1	6.35	6.28
Vein Graphite 2	6.46	6.34
Canadian Graphite	6.46	6.58
Expanded Graphite	6.63	6.95
Chinese Graphite Flake 1	7.48	7.41
Chinese Graphite Flake 2	6.93	6.93
Chinese Graphite Flake 3	7.90	7.99
Chinese Graphite Flake 4	7.15	7.10
Chinese Graphite Flake 5	6.33	6.18

Figure 5 and Table 5 show the number of layers of all experiments obtained according to Equations 8 and 9. In all experiments, the exfoliated graphene is between 2 and 10 layers, indicating that it is a few-layer graphene. In both N1 and N2, Chinese Graphite Flake 3 showed the highest values at 7.9 and 7.99, yet, <*N*> remains consistent between 6 and 8 atomic layers. Figure 6, 7, 8 and 9 show the N1 and N2 of the three experiments under the same conditions. The conformity of N1 is 99. 4% for Vein graphite 1, 98.89 % for Vein graphite 2, 98.26 % for Canadian graphite, 98.42 % for expanded graphite, 98.42 % for Chinese graphite 1, 98.79 % for Chinese graphite 2, 98.25 % for N2, 98.4 % for Vein graphite

2, 98.4 % for Canadian graphite, 97.93 % for expanded graphite, 97.82 % for Chinese graphite 1, and 98.41 % for Chinese graphite 2. The high conformity ensures that the quality of the product is uniform in three batches and has high stability in the process.

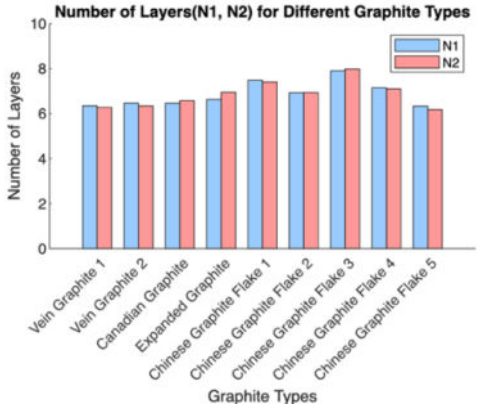


Figure 5 Number of layers (N1, N2) for different graphite types

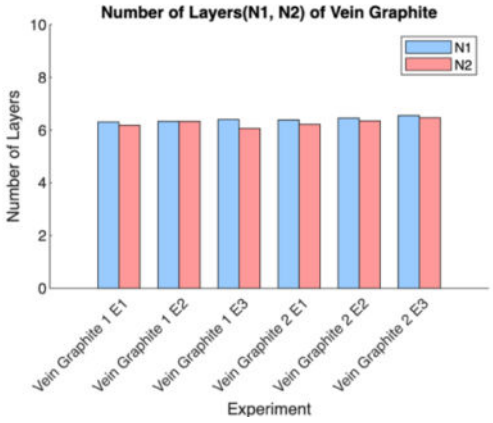


Figure 6 Number of layers (N1, N2) of vein graphite

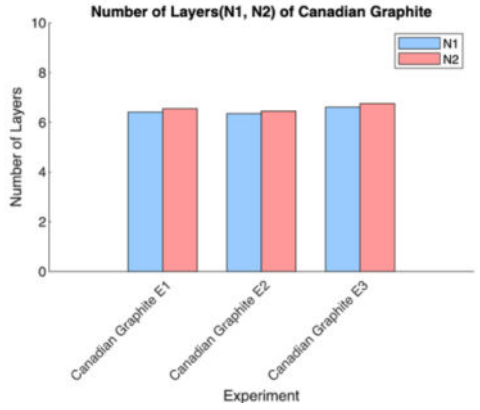


Figure 7 Number of layers (N1, N2) of Canadian graphite

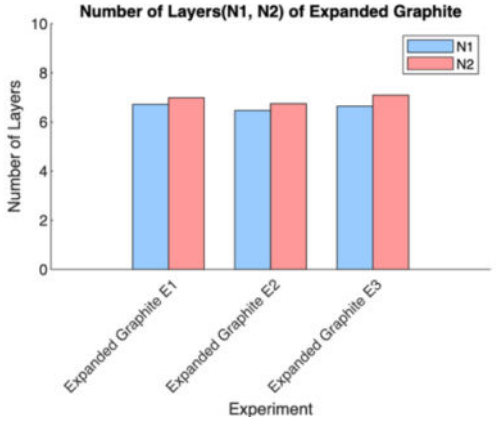


Figure 8 Number of layers (N1, N2) of expanded graphite

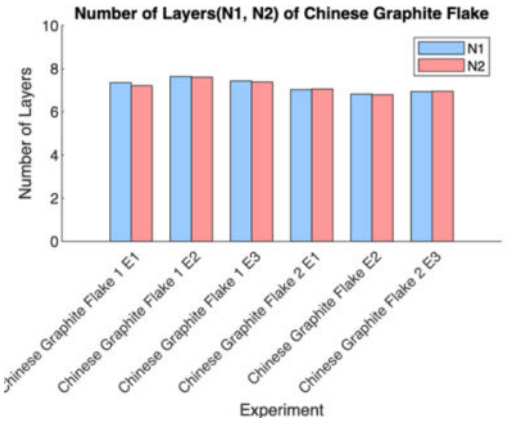


Figure 9 Number of layers (N1, N2) of Chines graphite flake

3.4 Variations depending on time

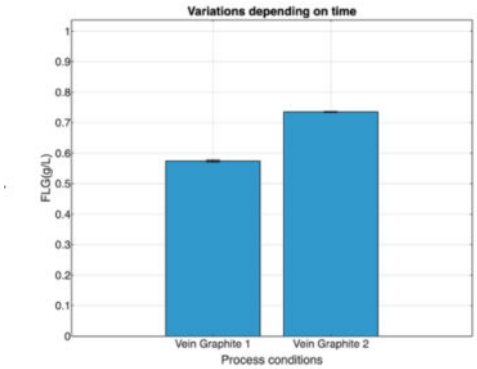


Figure 10 Variations of few-layer graphene concentration depending on time (left: 3 hours; right: 5 hours).

Under identical process conditions, one sample of vein graphite was run for 3 hours, and another was run for 5 hours. The concentration of few-layer graphene was about 1.3 times higher when tested for 5 hours. It can be seen that more graphene is exfoliated by overcoming the weak van der Waals force while performing the exfoliating for a longer time. In Figure 10, when comparing these graphs, it

can be seen that the 9-hour exfoliated Chinese graphite 4, 5, and 6 experiment shows a higher yield than when the Chinese graphite flake is exfoliated for 3 hours (Chinese graphite flake 1). However, the additional 6 hours of exfoliation does not justify the yield of 1.72 % for 9 hours relative to 1.38 % yield for 3 hours. Instead, one could run additional 3 hours batches to achieve higher yield for the same total exfoliation period of 9 hours. This is because the rate of graphene production reduces with exfoliation time i.e., 124.2 g/hr for 3 hrs and 154.8 g/hr for 9 hrs.

3.5 Variations depending on different precursors

After testing the same amount of vein graphite, Canadian graphite, expanded graphite, and Chinese graphite flakes used in the above experiment for 3 hours with a surfactant of 2 g/L, the concentration of Chinese graphite flakes was 2.42 times higher than that of vein graphite, which showed the second highest yield, and 19.71 times higher than that of expanded graphite, which showed the lowest yield.

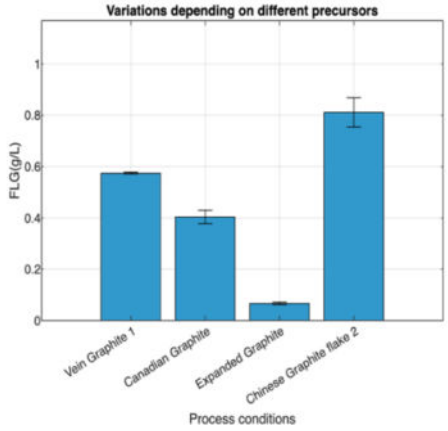


Figure 11 Variations of few-layer graphene concentration depending on precursors

3.6 Variations depending on graphite concentration

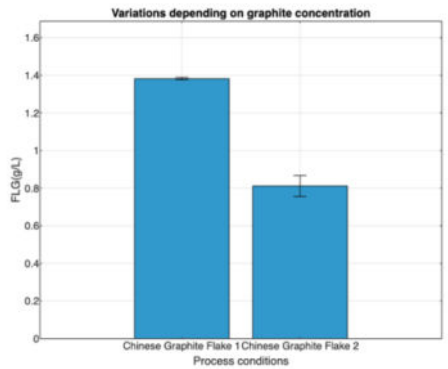


Figure 12 Variations of few-layer graphene concentration depending on graphite concentration (left: 100 g/L; right: 20 g/L).

In the graph above, Chinese graphite flake 1 yields 1.7 times more than Chinese graphite 2. The more significant the amount of initial graphite, the more graphene must be exfoliated.

3.7 Variations depending on NaCh concentration

Chinese graphite flake 4 shows the highest yield. Through this, it can be seen that the high NaCh concentration does not mean the yield is high. In the past, critical micelle concentration (CMC) was thought to affect the nanosheet yield [20]. The CMC of sodium cholate is 3 g/L [23]. If the concentration is higher than that of CMC, a micelle is formed to prevent graphene from being adsorbed to surfactant, and thus, graphene is considered to be low in yield. However, in recent studies, CMC has no impact on the yield of graphene [22]. This is because the micelle is formed regardless of the nanosheet being adsorbed on the surfactant surface. In this study, if the optimum concentration is 2.5 g/L and if smaller than this, the surface is insufficient to form adsorption and higher than this, the increased surfactant narrows the nanosheet gap, causing re-aggregation, resulting in the low yield of a few-layer graphene.

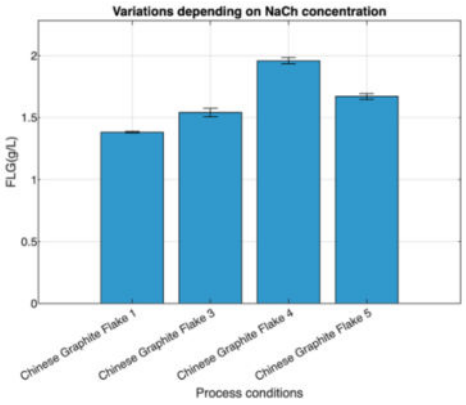


Figure 13 Variations of few-layer graphene concentration depending on NaCh concentration

4 Conclusions

Graphene is a material that is useful in modern industry. Since 2004, research has been steadily conducted to exfoliate graphene from graphite. However, large-scale graphene production is complex, and most methods are complex. This experiment shows a process condition that can increase the yield by proceeding with a simple liquid exfoliation method using a Taylor-Couette device. that enables large-scale graphene production, which can realistically make graphene widely used.

As a result of observation through UV-vis, the highest concentration of new layer graphene of 1.96 was exfoliated in Chinese Graphite Flake 4. This is because the experiment was conducted for a long time at the highest graphite concentration and optimal surfactant concentration. As for the number

of layers, Chinese Graphite Flake 3 was the highest at 7.9 and 7.99. The high conformity of 97 % or more, as shown when the three experiments were tested under the same conditions, increases the stability and reliability of this experiment. In order to exfoliate graphene effectively, the exfoliation time must be increased, but a long-term exfoliated time of 9 hours is unnecessary, considering efficiency. Among the precursors, Chinese graphite flake shows the highest yield; the higher the graphite concentration, the higher the yield. In the case of the surfactant concentration, the optimal concentration was 2.5 g/L. If concentration is higher than the optimal concentration, it formed aggregation and if the concentration is too low, the exfoliation of graphene could not be sufficiently performed.

5 Outlook

In this experiment, liquid exfoliation, which enables the commercialization of graphene, used a relatively eco-friendly and safe sodium chloride solution as a surfactant. However, the original production process uses highly toxic and environmentally hazardous solvents such as NMP and DMF. Suppose sodium chloride is used as a solvent. In that case, the efficiency is reduced to that of NMP and DMF, so finding an eco-friendly and non-toxic solvent is necessary while increasing the exfoliation efficiency. Also, a significant amount of graphene remaining inside the cylinder could be confirmed when cleaning up after the experiment. The graphene remaining inside the cylinder must be reduced in order to increase the graphene yield. In this experiment, the maximum speed was limited to 1,100 RPM for safety, but in case of scale-up, increasing the shear rate by increasing the speed will reduce the amount of graphene remaining.

6. References

1. Yi, X.-W. *et al.* (2022) 'T-carbon: Experiments, properties, potential applications and derivatives', *Nano Today*, 42, p. 101346. doi:10.1016/j.nantod.2021.101346.
2. Hoffmann, R. *et al.* (2016) '*homo citans* and carbon allotropes: For an ethics of Citation', *Angewandte Chemie International Edition*, 55(37), pp. 10962–10976. doi:10.1002/anie.201600655.
3. Ibrahim, A. *et al.* (2021) 'Graphene-based nanocomposites: Synthesis, mechanical properties, and characterizations', *Polymers*, 13(17), p. 2869. doi:10.3390/polym13172869.
4. Mbayachi, V.B. *et al.* (2021) 'Graphene synthesis, characterization and its applications: A Review', *Results in Chemistry*, 3, p. 100163. doi:10.1016/j.rechem.2021.100163.
5. Kim, Keun Soo *et al.* (2009) 'Large-scale pattern growth of graphene films for stretchable transparent electrodes', *Nature*, 457(7230), pp. 706–710. doi:10.1038/nature07719.
6. Blake, P. *et al.* (2008) 'Graphene-based liquid crystal device', *Nano Letters*, 8(6), pp. 1704–1708. doi:10.1021/nl080649i.
7. *Made in Manchester* (no date) *Graphene*. Available at: https://www.graphene.manchester.ac.uk/learn/discovery-of-graphene/
8. Liu, N. *et al.* (2021) 'Graphene synthesis: Method, exfoliation mechanism and large-scale production', *Crystals*, 12(1), p. 25. doi:10.3390/cryst12010025.
9. Saidin, N. *et al.* (2013) 'A*q*-switched thulium-doped fiber laser with a graphene thin film based saturable absorber', *Laser Physics*, 23(11), p. 115102. doi:10.1088/1054-660x/23/11/115102.
10. Stafford, J. *et al.* (2018) 'Towards scale-up of graphene production via nonoxidizing liquid exfoliation methods', *AIChE Journal*, 64(9), pp. 3246–3276. doi:10.1002/aic.16174.
11. Paton, K.R. *et al.* (2014) 'Scalable production of large quantities of defect-free few-layer graphene by shear exfoliation in liquids', *Nature Materials*, 13(6), pp. 624–630. doi:10.1038/nmat3944.
12. Tran, T.S. *et al.* (2016) 'High shear-induced exfoliation of graphite into high quality graphene by Taylor–Couette flow', *RSC Advances*, 6(15), pp. 12003–12008. doi:10.1039/c5ra22273g.
13. Bonaccorso, F. *et al.* (2012) 'Production and processing of graphene and 2d crystals', *Materials Today*, 15(12), pp. 564–589. doi:10.1016/s1369-7021(13)70014-2.
14. Solomon, H.M. *et al*(1995). 1-Methyl-2-pyrrolidone (NMP): reproductive and developmental toxicity study by inhalation in the rat. *Drug and Chemical Toxicology*, [online] 18(4), pp.271–293. doi:https://doi.org/10.3109/01480549509014324.
15. Sherwood, J. *et al* (2024) 'N,N-Dimethyl formamide European restriction demands solvent substitution in research and development,' *ChemSusChem*, 17(8). doi :https://doi.org/10.1002/cssc.202301639.
16. Vogt, C., Wondergem, C.S. and Weckhuysen, B.M. (2023). Ultraviolet-Visible (UV-Vis) Spectroscopy. *Springer handbooks*, pp.237–264.

doi:https://doi.org/10.1007/978-3-031-07125-6_11.
17. Backes, C. *et al* (2016). Spectroscopic metrics allow in situ measurement of mean size and thickness of liquid-exfoliated few-layer graphene nanosheets. *Nanoscale*, 8(7), pp.4311–4323. doi:https://doi.org/10.1039/c5nr08047a.
18. Xiao, D *et al* (2024). A Tape-Wrapping Strategy towards Electrochemical Fabrication of Water-Dispersible Graphene. *Nanomaterials*, 14(9), pp.805–805. doi:https://doi.org/10.3390/nano14090805.
19. Armano, A. and Agnello, S. (2019). Two-Dimensional Carbon: A Review of Synthesis Methods, and Electronic, Optical, and Vibrational Properties of Single-Layer Graphene. *C — Journal of Carbon Research*, 5(4), p.67. doi:https://doi.org/10.3390/c5040067.
20. Lotya, M. *et al* (2009). Liquid Phase Production of Graphene by Exfoliation of Graphite in Surfactant/Water Solutions. *Journal of the American Chemical Society*, 131(10), pp.3611–3620. doi:https://doi.org/10.1021/ja807449u.
21. Lathrop, D.P. *et al* (1992). Transition to shear-driven turbulence in Couette-Taylor flow. *Physical Review A*, 46(10), pp.6390–6405. doi:https://doi.org/10.1103/physreva.46.6390.
22. Wang, S. *et al* (2016). The effect of surfactants and their concentration on the liquid exfoliation of graphene. *RSC Advances*, 6(61), pp.56705–56710. doi:https://doi.org/10.1039/c6ra10933k.
23. Hebling, C.M. *et al* (2008). Sodium Cholate Aggregation and Chiral Recognition of the Probe Molecule (*R*,*S*)-1,1′-Binaphthyl-2,2′-diylhydrogenphosphate (BNDHP) Observed by ^{1}H and ^{31}P NMR Spectroscopy. *Langmuir*, 24(24), pp.13866–13874. doi:https://doi.org/10.1021/la802000x.

Wrinkle Morphology: An Investigation into Plasma Driven Cracking

Giorgos Kkantis and Taran Lal

Department of Chemical Engineering, Imperial College London, U.K.

Abstract

Wrinkled surface patterns, inspired by natural adaptations such as shark denticles and moth eyes, present innovative opportunities for improving industrial processes with enhanced efficiency and environmental sustainability. One way in which these wrinkles can be formed is via the plasma oxidation of an elastomer substrate such as polydimethylsiloxane (PDMS). By applying a strain to PDMS and exposing it to plasma, a rigid oxidised layer is formed, which buckles upon relaxation of the strain, forming a layer of wrinkles on the surface. However, this processed is caveated by the fact that random cracking is often seen in the elastomer surface due to the difference in mechanical properties between the oxidised layer, and the PDMS substrate. This study investigates how the surface morphology of the PDMS exposed to plasma impacts the way in which cracks are formed. By using plasma oxidation and ultraviolet ozonolysis (UVO) wrinkled moulds were created, from which samples were made, ultimately creating fresh, unoxidized, patterned PDMS surfaces. These samples were exposed to varying dosages of plasma, and crack formation was observed. It was concluded that surface morphology does indeed impact crack formation, and a negative correlation was found between crack density and pre-existing surface wrinkle wavelength. This trend was seen in wavelengths of single and tens of micrometres. Furthermore, it was also noted that exposure to plasma significantly decreased the amplitude of wrinkled surfaces that were exposed to it.

1. Introduction

Wrinkled patterns have been observed as an evolutionary solution for species of plants and animals, in order to overcome many of the challenges faced throughout nature, by simply enhancing the efficiency of mechanisms such as movement, visual sensory, traction etc.

The notion of wrinkled surfaces being used to enhance the performance of man-made processes or even provide new alternatives with the focus of environmental sustainability, has stricken the curiosity of many researchers whose objective is to make processes more efficient.

A well-documented example of wrinkles as an adaptation is the shape of the millions of dermal denticles found on the surface of a shark. Wrinkles aid in the shark's swimming by reducing its drag as it flows through water, in other words, sharks are more hydrodynamically efficient with wrinkles, allowing them to be faster when moving toward prey (Fu et al., 2017). In addition to movement, dermal denticles also demonstrate great resistance to biofouling; their wrinkled and diamond shape provides a challenge for bacteria and parasites when attempting to cultivate on the surface of a shark. Due the force of fast-moving water caused by the surface topology of a denticle, microorganisms are simply unable to attach, and even if they do they are quickly washed away (Chien et al., 2020). Modelling shark denticles on the surface of ship hulls has become a new possibility in improving the fuel and operational efficiency of naval voyages, as well as reducing biofouling in the form of barnacles without the need of toxic coatings (Fu et al., 2017). Not only in the sea, but also in hospital surfaces, wrinkled patterns can mitigate the spread of harmful microorganisms without the need of anti-bacterial chemicals, making these scenarios environmentally more sustainable (Mann et al., 2014).

Wrinkles are also exhibited on the surface of the eyes of certain types of Lepidoptera for anti-glaring purposes. More commonly documented in moths, the unique surface topology of their eyes allows light from different angles to enter their eyes, whilst altering the refractive index to prevent the light from reflecting away (Uma Rao et al., 2017). Studies focus on acquiring a wrinkle pattern inspired by this adaptation to improve the efficiency of solar panels by reducing the amount of light reflected off the surface of a solar cell (Uma Rao et al., 2017).

Although the designs of wrinkled surfaces have already been established for certain applications, cheap and precise manufacturing of these patterns remains a challenge, as available techniques such as

photolithography are expensive to operate at high production. One prospect is the use of soft materials with plasma induced surface buckling to create wrinkles, providing a cheaper alternative to the current methods.

However, plasma oxidation causes the formation of surface cracks on the material, becoming a new hurdle for this production path as they reduce wrinkle performance and precision (Glatz and Fery., 2018). Therefore, there is a great need to understand the mechanism of cracking in order to design better production conditions to avoid them. As such, this report focuses on the way in which cracks are formed, and how this is affected by the surface morphology of a soft elastomer substrate known as polydimethylsiloxane (PDMS) when exposed to plasma oxidation.

2. Background

2.1 Wrinkling

As mentioned, PDMS will serve as the model material examined within this report. It is a silicon-based organic polymer that exhibits properties such as optical transparency, hydrophobicity, and non-toxicity (Miranda et al., 2021). However, it was mainly selected for its remarkable elasticity, affordability, and ease of fabrication, making it an ideal candidate for experimental and practical applications (Sales et al., 2022). It is extensively researched in soft matter chemistry, and presents a range of general applications in microfluidics, medicinal technology, and consumer products, such as lubricants and adhesives (Miranda et al., 2021).

One additional property that is highly valued for the study is the ability of PDMS to undergo plasma oxidation. Plasma oxidation is when highly energised oxygen particles collide and break the methyl groups found on the surface of the PDMS, subsequently exposing the silicon backbone, which in turn get oxidised to form hydroxyl groups. This creates a thin, stiff silicate layer (that thickens with plasma exposure time) on the surface of PDMS that is visually glassy and exhibits hydrophilic properties (Waters et al., 2017).

However, when PDMS is oxidised under elastic stress and subsequently relaxed, due to the mechanical instability of the polymer, the surface will buckle. This will in turn form wrinkles, which is ideal for surface patterning. This phenomenon is already well characterised, and it has been concluded that the wavelength, λ, and amplitude, A, of the wrinkles are governed by Equations 1 and 2, respectively (Pellegrino et al, 2019).

$$\lambda = 2\pi h_f \left(\frac{\bar{E}_f}{3\bar{E}_s}\right)^{\frac{1}{3}} \quad (1)$$

$$A = h_f \left(\frac{\varepsilon}{\varepsilon_c} - 1\right)^{\frac{1}{2}} \quad (2)$$

$$\bar{E}_s = \frac{E_{PDMS}}{(1 - v_{PDMS}^2)} \quad (3)$$

$$\varepsilon_c = -\frac{1}{4}\left(\frac{3\bar{E}_s}{\bar{E}_f}\right)^{\frac{2}{3}} \quad (4)$$

In the first two equations, h_f is the thickness of the film formed on the surface during oxidation, E_f is the plane strain modulus, and E_s is the substrate modulus. This value is described by equation 3, where E_{PDMS} and v_{PDMS} are the young's modulus and Poisson's ratio of the polymer respectively (Pellegrino et al, 2019). Additional variables are the strain and critical strain (described in Equation 4) applied on the PDMS, denoted as ε and ε_c, respectively, where ε_c is the minimum strain required for the surface to buckle (Pellegrino et al, 2019). It is important to note that it is only the amplitude and not the wavelength that is impacted by the strain ε, whilst they both depend on the thickness of the film formed h_f. Additionally, the value of h_f has a positive correlation with the oxidation exposure time, i.e., the longer the exposure time, the larger the value of h_f, thus the larger the wavelength and amplitude.

2.2 Cracking

Other than plasma oxidation being used to create wrinkles on the surface of PDMS, another less well-defined phenomenon of cracking also occurs. These cracks perpetuate through the surface of the PDMS in a random manner, with no particular direction. However, it is known that the cracks usually originate at locations on the surface that experience the highest stress intensities, which is why crack formation occurs during wrinkling (Glatz and Fery., 2018). This is because of the formation of the oxidised layer h_f. Due to the large difference in mechanical properties between the oxidised layer and the regular PDMS substrate below, there is a

high stress at the interface between the layer and substrate (Beuth, 1992). This high concentration of stress at the interface is the current theory explaining the formation of cracks when it comes to plasma oxidised wrinkling (Hutchinson and Suo, 1991).

3. Methodology

3.1 Materials & Equipment

3.1.1 Plasma Oxidation Chamber

To create the necessary conditions to oxidise the surface of PDMS sample, a 'FEMTO' vacuum chamber by Diener Electronic was used to carry out this process. This machine works by removing all the present air within the chamber in which the sample is placed, then supplying a small amount of oxygen that is ionised upon activation of the device. This creates what is known as plasma, and it is observed as a light purple glow caused by the emission of UV light by the ionising particles. (Diener electronic GmbH + Co. KG., 2011).

3.1.2 Ultraviolet Ozone Generator (UVO)

Another approach to oxidising the surface of PDMS is through UV ozonolysis that is induced using a 'PSD Pro Series Digital UV Ozone System' (created by Novascan). This device works by utilising UV light to convert oxygen particles into ozone (Novascan Technologies, Inc., 2019), which in turn oxidises the surface of a sample over long exposure times (Berdichevsky et al., 2004). This allows the formation of wrinkles with wavelengths in the order of tens of micrometres, compared to the ones of micrometres or nanoscale wavelengths achieved by plasma oxidation. Furthermore, this method produces wrinkles without the formation of cracks.

3.1.3 Atomic Force Microscopy (AFM)

To physically analyse the surface modifications induced onto samples after plasma oxidation, an 'Innova AFM' (created by Bruker) was used to map out the surface topology of the PDMS. By placing this specialised device into tapping mode, it gently taps and records the height changes on the surface of a sample using a nanoscale tip. Scanning over a specified area, it can produce a 3D image of the surface of a sample, where the wavelength and amplitude of wrinkles can be extracted (Bruker, 2024). This data can then be then correlated back to Equations 1 and 2 to find h_f, as well as analyse any morphological changes in wavelength and amplitude between oxidised and non-oxidised wrinkles.

3.1.4 Fourier Transform Infrared Spectrometer (FTIR)

In order to confirm that the surface chemistry of the PDMS remained unchanged after making the samples, a FTIR spectrometer (from Bruker) was used. This works by passing infrared radiation through a sample and measuring the range of wavelengths in the IR region that are absorbed by the sample. Specific functional groups correspond to specific wavelengths absorbed, resulting in various peaks being formed in the results graph (Mathias, 2022). Analysis of these peaks allows for the surface chemistry to be determined and is discussed in later section of this report.

3.2 Sample Preparation

3.2.1 Wrinkle Formation

Liquid polydimethylsiloxane (PDMS) was made by combining Dow SYLGARD 184 silicone elastomer base with the 184 silicone elastomer curing agent according to a 10:1 ratio on a mass basis. This mixture was stirred using a dropper to distribute the curing agent more evenly, however this simultaneously introduced air bubbles. As such, the mixture was placed into a vacuum to degas it and remove all bubbles, finally providing a clear smooth liquid PDMS. This was then poured onto a setting surface and left in an oven for 4 hours at 75°C to cure into a solid sheet of PDMS.

A strain stage with clamps was set to an initial internal width of 43.0mm measured using metal callipers. Following this, a scalpel was used to cut a rectangle of PDMS out of the cured sheet, with dimensions of 50mm x 30mm. This was placed onto the stain stage, and the clamps were screwed into place to keep the edges of the PDMS securely attached to the stage. Strain was then applied to the PDMS by expanding the strain stage, using the built-in screw system, to an internal width of 51.6mm, ultimately leading to a 20% strain being applied to the sample. Tape was then used to remove any dust from the surface of the PDMS. The strain stage with the sample in it was then put on to a metal plate which was subsequently put into the plasma oxidation chamber.

Once the sample was in the plasma chamber the pump was turned on sealing the chamber, and a vacuum was pulled till a pressure of 0.1 mbar was reached. Following this, oxygen gas was pumped into the chamber and the vacuum pumping continued until the pressure stabilised at 0.22 mbar. At this point the power was set to 99W and the generator was turned on, starting the plasma oxidation of the PDMS surface. At the same time, a timer for 1 minute was started, after which the ventilation button was pressed, stopping the oxidation process, and returning the chamber to atmospheric conditions. It was at this time the sample was able to be removed from the chamber, and the strain relaxed, resulting in the buckling of the oxidised layer and consequently the formation of wrinkles on the surface of the PDMS.

The same procedure outlined above was followed exactly the same way but for plasma oxidation exposure times of 5 minutes and 20 minutes to create wrinkle samples of different wavelengths (still in the order of ones of micrometres). Furthermore, two other samples were made following a similar methodology however with a strain of 50% and using UV-Ozone (UVO) instead of plasma to oxidise the surface. This was done in order to create samples with wavelengths in the order of tens of micrometres. These two samples were exposed to UVO for 90 minutes and 150 minutes respectively, at a distance of 11±1mm away from the UVO bulb source. The increase in strain and exposure time was necessary in order to see the wrinkles appear and to get the larger order wavelength. For trials that were carried out during this experimentation time, no wrinkles were seen using the UVO method with 20% strain.

3.2.2 NOA Mould Creation

Having formed all the wrinkle samples that were desired to be tested, a solid mould needed to be created out of each of them so that multiple trials could be done using the same surface morphology repeatedly. This was done by using Norland Optical Adhesive 81 (NOA), a liquid adhesive that cures solid with exposure to UV light. However, before the moulds could be made an intermediary step was required as exposing PDMS to plasma oxidation changes its surface from hydrophobic to hydrophilic (Tan et al., 2010). As such, to ensure that the mould could easily be removed from the PDMS, the surface was made hydrophobic using octadecyltrichlorosilane (OTS). This was done by pouring a small volume (~5mL) of OTS into a cap and placing that into a chamber with the wrinkled PDMS sample. The chamber was then put under vacuum conditions using a pump, allowing the OTS to form a monolayer coating the PDMS thus making the surface hydrophobic.

The now hydrophobic PDMS was then placed between two pieces of acrylic leaving the wrinkled surface exposed, and NOA was poured onto it. This rig was then place underneath a UV light with energy 50 mW/cm^2, for 1,998 seconds to cure. The rig was then flipped upside down, and the bottom acrylic removed, for the NOA to cure another 999 seconds under the UV, ensuring that the mould was solid. The NOA mould was then removed from the PDMS, thus resulting in a mould complementary to the original. This was repeated for all 5 of the wrinkled PDMS samples made in Section 3.2.1 (3 plasma based, 2 UVO based), thus resulting in 5 different NOA moulds created.

3.2.3 PDMS Sample Preparation for Experimentation

An acrylic tray with 6 separate indentations was prepared so that each of the 5 moulds could be placed into it, along with a flat (non-wrinkled) mould to use as a control. Liquid PDMS was then made according to the method outlined in Section 3.2.1, poured onto the moulds, and left to cure in the oven at 75°C. Once cured, a scalpel was used to cut out a 35mm x 25mm rectangle of PDMS from the moulds which were used as test samples. From there on out, for each round of samples made for experimentation, 4 grams of liquid PDMS was poured into the 35mm x 25mm cavity and then left in the oven to cure. The same 35mm x 25mm sample would then be cut out from each mould and the process repeated, thus yielding the same exact samples to be test for the next set of experiments.

Other than the fact that making the NOA mould allows for the same samples to be tested repeatedly, there are 2 other critical benefits behind this full methodology. Firstly, as the PDMS samples that are cut out from the NOA moulds will have wrinkles complementary to the NOA, they will be exact replicas of the original wrinkled PDMS surfaces the NOA moulds were made from. Secondly, and most importantly, by pouring fresh PDMS onto the

surface of the wrinkled NOA, the result is a wrinkled PDMS surface without the glassy oxidised layer formed from plasma exposure, which is always present when the wrinkles originally form. Therefore, the only difference between plain, flat, unoxidized PDMS and the wrinkled samples that were used in this study was the surface morphology, i.e., the different surface shape cause by the different wavelengths of wrinkles on each sample. To confirm this, the surface chemistry was examined using FTIR spectroscopy on the samples before and after plasma exposure, as well as on plain PDMS. These results are discussed in Section 4.1.

3.3 Experiment Methodology

3.3.1 Sample Wavelength Determination

Through prior studies and experimentation, the formation of random cracks in the surface of PDMS has been regularly documented, thus motivating the focus for this project to investigate if the surface morphology of the PDMS impacts the way in which cracks are formed. Prior to experimentation, the samples created following the method in Section 3.2.3 were imaged with a microscope, along with a microscope calibration slide. This allowed for the wrinkle wavelength corresponding to each sample to be determined, as is outlined in Table 1. This was later cross checked and confirmed using the data retrieved from AFM.

Table 1: Wrinkle wavelength yielded from each oxidation exposure time.

Exposure Time (mins)	Wavelength (µm)
1 (Plasma)	1.3 ± 0.2
5 (Plasma)	2.0 ± 0.1
20 (Plasma)	3.7 ± 0.3
90 (UVO)	30.0 ± 2.0
150 (UVO)	40.0 ± 2.5

As the samples were relatively large (35mm x 25mm), different regions across the surface of the wrinkled PDMS samples were measured throughout the study. There were minor differences in the wavelengths measured of each sample depending on the region examined consequently leading to the uncertainty in the wavelength as illustrated within Table 1.

3.3.2 Surface Shape Impact on Cracking

In order to determine how the wavelength of the surface impacted cracking after plasma oxidation, each PDMS sample was imaged with a microscope (at 10x and 50x) before exposure to plasma. The samples were then put on the centre of a metal platform and into the plasma oxidation chamber where a vacuum was pulled, oxygen gas was introduced, and a pressure of 0.22 mbar was achieved. Following this, the generator was turned on at 99W and the samples were exposed to plasma for one of the following lengths of time depending on the trial that was being carried out: 10, 20, 30, 60 seconds. The chamber was then vented, and the samples taken out and imaged again using the microscope.

Each sample was initially marked on the back using a black marker. This indicated the region that was to be imaged so as to ensure that the same region was always being viewed for each sample. This allows any cracks formed post plasma exposure to be clear to see when compared to the pre-exposure photo, as is illustrated in Figure 1.

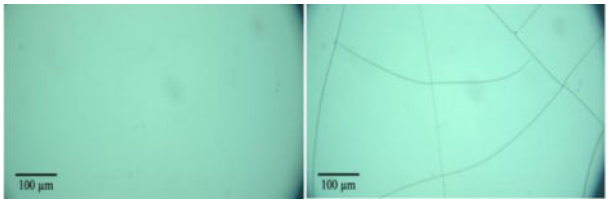


Figure 1: 50X image of plain PDMS before (left) and after (right) exposure to 60s of 99W plasma.

Having obtained images of the cracks formed for each sample after exposure to 99W plasma for varying lengths of time, the crack density was then to be determined. This was done by processing the images using the Fiji software. The post-plasma oxidation images were imported into the software and converted to 8-bit images and grayscaled, after which thresholding was applied. This resulted in the new cracks formed being able to be isolated in the image as shown in Figure 2.



Figure 2: Processed 50x image of post-plasma plain PDMS after thresholding.

After processing all the images following this methodology, it was possible to determine the percentage of black pixels in the processed image using the Fiji software. As these black pixels represented the cracks formed during exposure to plasma oxidation, this percentage was taken as the value for the crack density. This value could then be compared between all samples and trials, which provided the basis for the analysis in the results and discussion section of this report.

3.3.3 Plasma Oxidation Impact on Wrinkle Amplitude

Upon inspection of the images of the PDMS surface before and after plasma exposure, a secondary interesting phenomenon was observed other than the expected cracking. In many of the post-plasma images, the cracks that were present in the mould before exposure to the plasma seemed to fade partially, while new cracks formed in other locations. Furthermore, the wrinkles present had seemed to have lost some of their definition, even at optimal focus on the microscope, in the exact same region. This is illustrated within Figure 3.

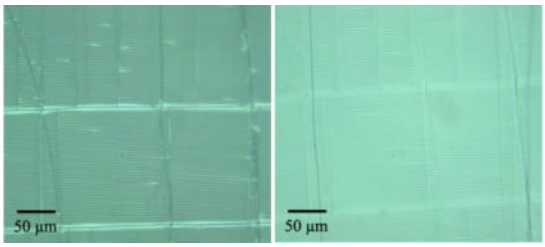


Figure 3: 3.7±0.3µm sample before (left) and after (right) exposure to 60s of 99W plasma.

As such, this prompted a secondary hypothesis that the exposure to plasma was causing a decrease in the amplitude of the wrinkles on the PDMS surface. In order to test this, it was decided that atomic force microscopy (AFM) would be used to map a small region of the wrinkled surface, in order to determine the amplitude before and after plasma exposure.

As the PDMS samples were too large to be used in the atomic force microscope, circular samples of diameter 1cm to be analysed were punched out of the 35mm x 25mm rectangular PDMS. This was done prior to exposure to plasma and imaging the surface with the microscope. A 20µm x 20µm region on the surface of the circular samples was then mapped using AFM in tapping mode, ultimately leading to the determination of the amplitude. The samples were then exposed to plasma following the methodology in Section 3.3.2 and then had their surfaces mapped again using AFM to determine the amplitude after exposure. It is important to note that the 30µm and 40µm wavelength samples were not analysed using AFM as these wrinkle wavelengths were too large to generate clear images. Discussion of the results from this is carried out in Section 4.3.

4. Results and Discussion

4.1 Fourier Transform Infrared Spectroscopy

FTIR was carried out on three different types of samples as mentioned in section 3.2.3: plain unoxidized PDMS, wrinkled unoxidized PDMS taken from the NOA mould, and a plasma oxidised wrinkled PDMS. The full FTIR graph can be seen in Appendix A, however the main results seen from it show that the surface chemistry is exactly the same for plain PDMS and the unoxidized wrinkled PDMS. This can be concluded because the peaks match up exactly on top of each other, indicating that all the same functional groups are present, and nothing has changed. As such, it can be said with certainty that any differences seen in the results using these samples is solely due to the surface morphology. It was also found via FTIR that the plasma exposed sample had a small peak in the region of 3200-3700 cm^{-1} that was not present in the non-plasma exposed samples, indicating the presence of the OH group post exposure. This is in line with the literature and theory discussed in the background, as the silicon backbone is exposed causing it to be oxidised and form hydroxyl groups (Waters et al., 2017).

4.2 Crack Formation with Wavelength Variation

Following the methodology in Section 3.3.2, a crack density was determined for each of the wrinkled PDMS samples, as well as a flat PDMS, upon exposure to varying doses of plasma. These results are illustrated in Figure 4 for the samples in the single micrometre range.

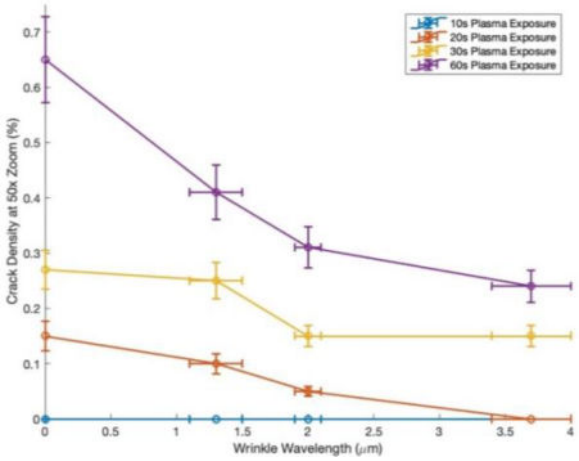


Figure 4: Impact of wrinkle wavelength (ones of μm) on crack density after various plasma exposure times.

Upon inspection of this graph, it can be seen that each point has error bars associated with it, both with regards to the wavelength, and the crack density. The uncertainty in the wavelength has been discussed in Section 3.3.1, and the error bars in the graph reflect this. With regards to the error bars related to crack density, this uncertainty comes from the various trials carried out for each data point. Due to time constraints, two to three trials were carried out for each data point to confirm that the trend seen was repeatable. The error bars represent the variance in the results seen which was roughly up to 18%, 13%, and 12%, for the 20s, 30s, and 60s trials, respectively. As samples with exposure to just 10s of 99W plasma did not show any cracking, there was no uncertainty associated with the crack density for these trails.

Regarding the results from the data illustrated in Figure 4, there are two trends that can clearly be witnessed. Firstly, as the plasma exposure time increases, the crack density increases for all samples, regardless of the wavelength of the pre-existing wrinkle on the surface. When samples were exposed to plasma for just 10s at 99W cracks were not seen at all, whilst when exposed for 60s the crack density was highest for all samples tested. This result is expected and in line with the literature. This is because increasing exposure time increases the thickness of the oxidised film h_f, which results in a higher stress experienced at the interface between the layer and the substrate (Hutchinson and Suo, 1991). Consequently, cracks are more likely to propagate, thus explaining the increased crack density observed at higher exposure times.

The other, more noteworthy result seen from the data in Figure 4, is that as the wavelength of the wrinkles on the surface of the exposed samples increased, the crack density after plasma exposure generally decreased. This same trend was seen across all of the different plasma exposure times. This is interesting to note as it implies that the surface shape/roughness has an impact on how cracks propagate on PDMS when exposed to plasma oxidation. One possible explanation for this is that the larger the wavelength of the wrinkle pre-exposure, the more evenly the stress is distributed across the surface.

The definition of the wavelength λ is the distance between the peaks of two adjacent wrinkles. As such, a larger wavelength sample will have fewer wrinkles within the same distance. This means that there is a more even, smoother curvature then compared with a smaller wavelength wrinkle, which will be more jagged and abrupt, as more wrinkles are fit into the same distance. Consequently, the larger wavelength samples are able to distribute the stress across the surface more evenly, whilst the smaller wavelength samples will be likely to have more, high stress concentration points. This results in less cracking occurring as the surface wavelength of exposed samples increases, as illustrated in Figure 4.

Whilst Figure 4 illustrates a good example of this theory working in the single micrometre scale, results were also analysed in the tens of micrometres scale, through using the samples that were made by UVO. It is important to highlight that it is just the mould for these samples that was made using UVO, to get the larger wrinkles, and that these samples were still plasma exposed to induce cracking following the methodology in Section 3.3.2. These results are shown in a separate plot not just because it is a different scale, but also because 50% rather than 20% strain was applied in order to make the wrinkles. So as to keep variables as controlled as

possible when comparing the data, these results are illustrated separately in Figure 5.

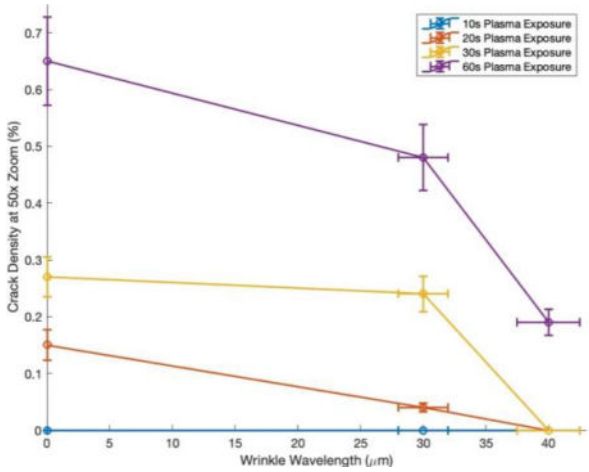


Figure 5: Impact of wrinkle wavelength (tens of μm) on crack density after various plasma exposure times.

As can be seen from Figure 5, all the trends that were seen in the single micron scale, are still observed in the tens of microns scale. Longer exposure time leading to higher crack density, and longer wavelengths decreasing crack density. Furthermore, the source of uncertainty here in the form of the error bars is the same as in the results discussed previously from Figure 4.

While it is clear to see this negative correlation between surface wrinkle wavelength and crack density in both graphs, only 3 and 2 different wavelengths were tested for single and tens of microns scales, respectively. Given more time, more different moulds should be made, and more trials carried out for different wavelengths to confirm this trend is followed. Furthermore, different, more irregular surface morphologies could be experimented with, to see the role this plays in crack formation.

4.3 Plasma Exposure impact on Amplitude

A mentioned in section 3.3.3, non-oxidised wrinkle samples had experienced a decrease in their amplitudes, as a result of plasma exposure. This is illustrated through Figures 6 and 7.

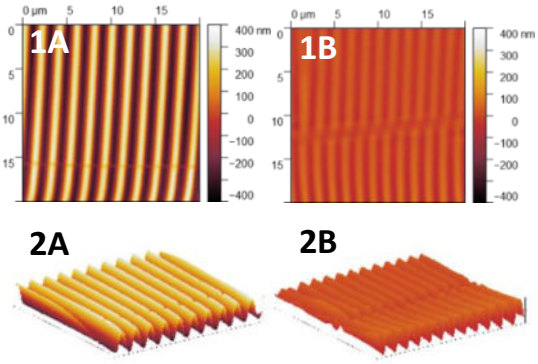


Figure 6: 1A presents the colour topography of a 1.3μm wrinkle before plasma oxidation, with 2A being the 3-Dimensional visualisation of 1A. 1B and 2B are the results after the 1.3μm wrinkle was exposed to plasma for 10 seconds.

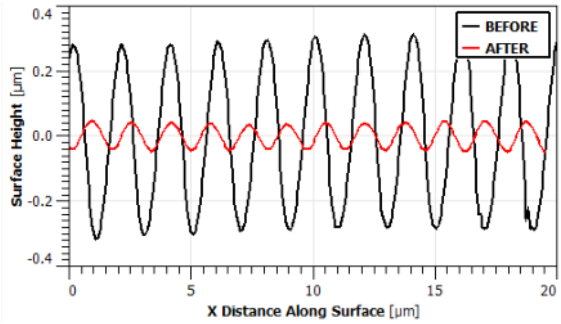


Figure 7: Graphical presentation of the cross-sectional area of a wrinkled surface before and after 10 second exposure time to plasma

As can clearly be seen from the figures, there is a significant drop in the amplitude post plasma exposure, which is denoted in the darker colour change on the wrinkles, this is quantified in Table 2. Additionally, a metric R_a for surface roughness was calculated. This was done by taking the value for the peak height of each wrinkle in a given area and averaging it out by the number of wrinkles accounted for. As expected, the percentage change in R_a closely followed that of the amplitude change, as seen in Table 2.

Table 2: Changes in amplitude and surface roughness of various wrinkled samples after different exposure times to plasma.

Wrinkle Wavelength (µm)	Plasma Exposure Time (s)	Amplitude Change (%)	R_a Change (%)
1.3 ± 0.2	10	-86%	-80%
1.3 ± 0.2	20	-82%	-80%
1.3 ± 0.2	60	-90%	-85%
2.0 ± 0.1	10	-82%	-82%
2.0 ± 0.1	20	-46%	-39%
2.0 ± 0.1	60	-78%	-77%
3.7 ± 0.3	10	-60%	-58%
3.7 ± 0.3	20	-53%	-43%
3.7 ± 0.3	60	-74%	-72%

It can be seen that the 1.3µm wrinkles exhibited an averaged 86% decrease, the 2µm wrinkles an 80% decrease (not including the -46% anomaly), and the 3.7µm a 62% decrease in the wrinkle amplitude. As such, from this sample size, it was observed that as the wavelength of the wrinkle increases, the lower the percentage change in amplitude after plasma.

These results follow closely with a study done by Yang et al. It was posited that due to the formation of the silicate layer during oxidation, as well as other factors such as surface cross-linking, the surface structure of PDMS reorganises into a more orderly formation. This induces some form of tension on the surface that reduces the amplitude of the wrinkles (Yang et al., 2020).

Additional trends that resemble with the results found in Yang et al. can be seen within Table 3, where the wavelengths of the samples seemed to remain significantly unchanged with only minor increases after oxidation (due to the aforementioned reorganisation of the surface). Additionally, it should be noted that there were occasional decreases in wavelength that could have been a result of changing scanning area while using AFM.

Nonetheless, the results found in the current investigation show a great change in the morphology of the samples. This cannot be caused only by changing chemical composition of the surface, but rather proves that smaller wrinkles (with smaller wavelengths and amplitudes) are far more sensitive to surface etching caused by the presence of plasma. There is a greater loss of PDMS fragments, whereas larger wrinkle samples do not experience surface etching to the same degree. However, this requires further experimentation, that due to time constraints was not done, using the variable power output of plasma to find a threshold, if any, where plasma etching begins to take effect.

Table 3: Change wavelength of various wrinkled samples after different exposure times to plasma.

Wrinkle Wavelength (µm)	Plasma Exposure Time (s)	Wavelength Before (µm)	Wavelength After (µm)
1.3 ± 0.2	10	2.0	1.6
1.3 ± 0.2	20	1.2	1.3
1.3 ± 0.2	60	1.3	1.3
2.0 ± 0.1	10	2.0	2.1
2.0 ± 0.1	20	1.9	2.0
2.0 ± 0.1	60	2.0	2.3
3.7 ± 0.3	10	4.0	3.4
3.7 ± 0.3	20	3.5	3.6
3.7 ± 0.3	60	3.4	3.8

5. Conclusion and Outlook

As previously mentioned, achieving an affordable and precise method to produce well-designed micropatterns on surfaces requires to avoid the surface defects, in the forms of cracks, that are created during plasma oxidation.

This study contributes to the understanding of the cracking mechanisms, by revealing that the surface morphology of a wrinkled sample greatly influences the development of cracks during plasma oxidation. It was found that increasing the wavelengths of wrinkle patterns results in a reduction on the crack density on the surface of a sample. To be more precise, stress can be evenly distributed by changing the surface morphology in order to mitigate crack formation. However, further testing should be done on different wavelengths and surface morphologies, to identify accurately wrinkle design parameters or limitations.

Additionally, it was noted that surface etching is much more severe on smaller wrinkle patterns, with smaller wavelengths and amplitudes, compared to that of larger wrinkles. Thus, further experimentation in future work is required to understand how the power output of plasma chamber affects close-to nanoscale patterns to pinpoint a threshold of plasma power, to avoid unwanted surface modifications.

6. Acknowledgements

Great credit would like to be given to Zain Ahmed and the Polymers and Microfluidics Research Group for their supportive guidance throughout the duration of this study.

7. References

Beuth, J. L. (1992). Cracking of thin bonded films in residual tension. *International Journal of Solids and Structures, 29*(13), 1657–1675.

Berdichevsk, Y., Khandurina, J., Guttman, A., & Lo, Y.-H. . (2004). UV/ozone modification of poly(dimethylsiloxane) microfluidic channels. *Sensors and Actuators B: Chemical, 97*(2-3), 402–408. https://doi.org/10.1016/j.snb.2003.09.022

Bruker. (2024). Atomic Force Microscopy FAQs. *Bruker.com*. https://www.bruker.com/es/products-and-solutions/microscopes/materials-afm/faq.html

Chien, H.-W., Chen, X.-Y., Tsai, W.-P., & Lee, M. (2020). Inhibition of biofilm formation by rough shark skin-patterned surfaces. *Colloids and Surfaces B: Biointerfaces, 186,* p.110738. https://doi.org/10.1016/j.colsurfb.2019.110738

Diener electronic GmbH + Co. KG. (2011). PLASMA TECHNOLOGY. https://lampz.tugraz.at/~hadley/semi/ch9/instruments/Diener/Plasmatechnik_eng.pdf

Fu, Y. F., Yuan, C. Q., & Bai, X. Q. (2017). Marine drag reduction of shark skin inspired riblet surfaces. *Biosurface and Biotribology, 3*(1), 11–24. https://doi.org/10.1016/j.bsbt.2017.02.001

Glatz, B. A., & Fery, A. (2018). The influence of plasma treatment on the elasticity of the in situ oxidized gradient layer in PDMS: towards crack-free wrinkling. *Soft Matter, 15*(1), 65–72. https://doi.org/10.1039/c8sm01910j

Hutchinson, J. W., & Suo, Z. (1991). Mixed mode cracking in layered materials. In *Advances in Applied Mechanics* (pp. 63–191). Elsevier.

Mathias, J. (2022). How Does FTIR Work? | Innovatech Labs. *Innovatech Labs*. https://www.innovatechlabs.com/newsroom/672/stuff-works-ftir-analysis/

Mann, E. E., Manna, D., Mettetal, M. R., May, R. M., Dannemiller, E. M., Chung, K. K., Brennan, A. B., & Reddy, S. T. (2014). Surface micropattern limits bacterial contamination. *Antimicrobial Resistance and Infection Control, 3*(1), 28. https://doi.org/10.1186/2047-2994-3-28

Miranda, I., Souza, A., Sousa, P., Ribeiro, J., Castanheira, E. M. S., Lima, R., & Minas, G. (2021). Properties and Applications of PDMS for Biomedical Engineering: A Review. *Journal of Functional Biomaterials*, *13*(1), 2. https://doi.org/10.3390/jfb13010002

Novascan Technologies, Inc. (2019). Novascan PSD-UV, PSDP-UV and PSDP-UVT UV/Ozone Systems. https://micro.fel.cvut.cz/wp-content/uploads/2019/03/Novascan-PSD-1.pdf

Pellegrino, L., Sepideh Khodaparast, & Cabral, J. T. (2019). Orthogonal wave superposition of wrinkled, plasma-oxidised, polydimethylsiloxane surfaces. *Soft Matter, 16*(3), 595–603. https://doi.org/10.1039/c9sm02124h

Sales, F. C. P., Ariati, R. M., Noronha, V. T., & Ribeiro, J. E. (2022). Mechanical Characterization of PDMS with Different Mixing Ratios. *Procedia Structural Integrity*, *37*, 383–388. https://doi.org/10.1016/j.prostr.2022.01.099

Tan, S. H., Nguyen, N. T., Chua, Y. C., & Kang, T. G. (2010). Oxygen plasma treatment for reducing hydrophobicity of a sealed polydimethylsiloxane microchannel. *Biomicrofluidics, 4*(3), 32204. https://doi.org/10.1063/1.3466882

Uma Rao, K., Reddy, C. A., Chaitanya Lakshmi, G., & Vaishnavi, G. (2017). Efficiency Enhancement of the Solar Cells through Subwavelength Scaled Biomimiced Moth Eye Patterning using Lumerical Solutions. *Materials Today: Proceedings, 4*(9), 10407–10411. https://doi.org/10.1016/j.matpr.2017.06.389

Waters, L. J., Finch, C. V., Bhuiyan, A. K. M. M. H., Hemming, K., & Mitchell, J. C. (2017). Effect of plasma surface treatment of poly(dimethylsiloxane) on the permeation of pharmaceutical compounds. *Journal of Pharmaceutical Analysis, 7*(5), 338–342. https://doi.org/10.1016/j.jpha.2017.05.003

Yang, L., Ge, L., Zhou, Q., Jurczak, K. M., & van Rijn, P. (2020). Decoupling the Amplitude and Wavelength of Anisotropic Topography and the Influence on Osteogenic Differentiation of Mesenchymal Stem Cells Using a High-Throughput Screening Approach. *ACS Applied Bio Materials, 3*(6), 3690–3697. https://doi.org/10.1021/acsabm.0c0033

A Hybrid Process for Adsorption-based CO_2 Capture

Bellal Ahadi and Stutya Nagpal

Department of Chemical Engineering, Imperial College London, UK

Abstract

Post-combustion capture through pressure-vacuum swing adsorption (PVSA) systems is a process that has gained growing interest, especially in the backdrop of the energy transition. This interest primarily stems from traditional coal and gas-fired power generation facilities, which view it as a means of mitigating carbon emissions. A PVSA process comprises of four stages which are used to selectively adsorb and remove CO_2 from N_2-rich vapour streams. Previous studies have extensively modelled these processes and performed constrained multi-objective optimisation to determine the candidate operating points which maximise certain key performance indicators, such as CO_2 recovery and purity. However, most works do not account for the inevitable CO_2 liquefaction, which forms an essential part of storage and transport within carbon capture systems on an industrial scale. In addition, little to no research has been conducted on the impact of varying feed composition on the PVSA process. This not only affects process performance, but also imposes further constraints on the feasibility and operational flexibility of the hybrid PVSA-liquefaction process. To this end, a liquefaction system was designed and developed for a standard PVSA equilibrium model, and the overall performance of the combined system was analysed. This was done through varying key decision variables such as pressures, as well as other process parameters like feed composition and adsorbent material, to provide a holistic assessment of this hybrid system.

Keywords: *CO_2 liquefaction, pressure-swing vacuum adsorption, hybrid PVSA-liquefaction*

1. Introduction

During a constantly evolving energy landscape, carbon capture, utilisation and storage (CCUS) systems have emerged as a viable technology to control CO_2 emissions. The United Nations recognises the potential CCUS systems can have in mitigating carbon emissions and estimates suggest that CCUS projects can capture up to 6.3 Gt of CO_2 by 2050 (Unfccc.int., 2022). There is growing interest in these systems for post-combustion capture, specifically in the power sector as they provide a technological hedge to power networks, whilst enhancing flexibility and stable operation (IEA. (n.d.)). Traditional carbon capture technologies often make use of the liquid amine scrubbing technique however, this process can be infeasible as in practice it tends to be highly energy inefficient and generates degraded-solvent waste which requires regeneration (Hack, Maeda and Meier, 2022).

An alternative approach to this involves the use of adsorption techniques such as pressure-vacuum swing adsorption (PVSA). Previous studies on these systems have focused on many factors such as assessing operational flexibility through design space learning (Ward and Pini, 2022), finding optimal operating points based on certain Key Performance Indicators (KPIs) as well as techno-economic assessments of PVSA systems. However, these do not provide many insights into the impact of varying feed conditions to the PVSA process. Furthermore, and perhaps most importantly, the impact of liquefying the CO_2 product, which is an inevitable step for further processing and storage, remains largely unexplored. Hence, the KPIs used for measuring the ability of the system to capture carbon dioxide in previous PVSA works may not provide a true representation of the entire process.

This work focuses on studying the performance of a PVSA system with a liquefaction process directly downstream. The ability of the overall system to capture CO_2 is measured through KPIs similar to the ones used in previous works. These include CO_2 purity, recovery, and Specific Energy consumption (SEN), which is the total amount of energy required by the process per unit mass of CO_2 captured. In this study, an existing equilibrium-based PVSA model was modified and integrated with a liquefaction model to develop a novel hybrid process. This system's performance was analysed by varying parameters like feed concentration, adsorbent material and purge fraction to identify candidate optimums and quantify their respective KPIs.

2. Background

2.1 PVSA Cycle

An equilibrium model is applied to assess the PVSA process, which was developed in an earlier study by Ward, Li and Pini (2023). This model simulates the performance of a multi-stage pressure vacuum swing adsorption system with feed pressurization. These four stages include: high-pressure adsorption (1), forward blowdown (2), reverse (vacuum) evacuation (3), and feed pressurization (4).

As shown in *Fig.1,* the PVSA process is initially fed a gaseous stream of CO_2/N_2 to an adsorption column containing a fixed bed of a solid adsorbent such as Zeolite 13X (Z13X) or Activated Carbon (AC). The feed stream elevates the column pressure to the desired high-pressure level P_H. This is followed by the adsorption of

the CO_2 into the stationary solid phase, which produces a N_2-rich effluent stream and a CO_2-rich adsorbed phase. Once the adsorbent material is saturated with CO_2, the pressure in the column is reduced to an intermediate level P_I, wherein the bulk N_2 gas is expelled from the product end of the column whilst keeping the feed side closed. The product end is then closed off and a vacuum pump is used to reduce pressure from P_I to P_L. During this evacuation stage, a high-purity CO_2 stream is collected from the feed side of the column. Finally, flue gas is once again fed to the column to pressurize it from P_L to P_H and repeat the cyclic process.

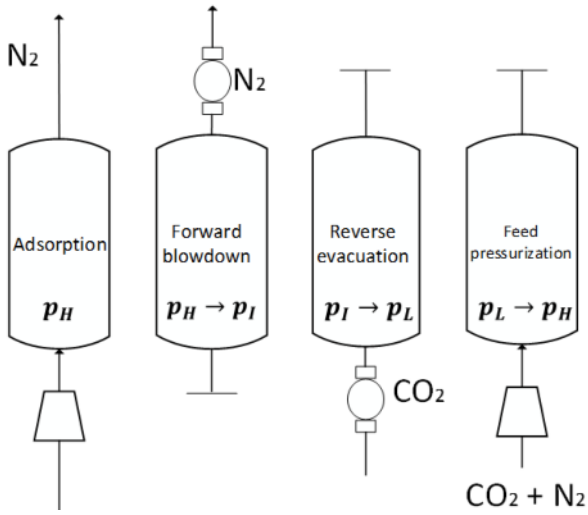


***Figure 1**. The PVSA Process*

In the original study by Ward, Li and Pini (2023), the process performance was analysed through four KPIs: recovery, purity, specific energy consumption and working capacity. Note that since working capacity is not impacted by liquefaction, it was not included as a part of this study. Therefore, the remaining three KPIs are used to compare process performance before and after CO_2 liquefaction which are defined in the original work as follows:

$$Pu_{CO_2}(\%) = 100 \times \frac{n^{evac}_{CO_2,out}}{n^{evac}_{CO_2,out} + n^{evac}_{N_2,out}} \quad (1)$$

$$Re_{CO_2}(\%) = 100 \times \frac{n^{evac}_{CO_2,out}}{n^{pres}_{CO_2,in} + n^{ads}_{CO_2,in}} \quad (2)$$

$$E_T = \frac{E_{ads} + E_{bd} + E_{evac} + E_{pres}}{m^{evac}_{CO_2,out}} \quad (3)$$

where $n^{evac}_{CO_2,out}$ and $n^{evac}_{N_2,out}$ are the molar flows of CO_2 and N_2 removed in the evacuation step respectively, while $n^{ads}_{CO_2,in}$ and $n^{pres}_{CO_2,in}$ refer to the molar flows of CO_2 fed during adsorption and feed pressurization stages respectively. In equation (3), E_{stage} is defined as the total energy consumed by a given stage in the process whilst $m^{evac}_{CO_2,out}$ refers to the total mass of CO_2 extracted during evacuation.

2.2 CO_2 Liquefaction Cycle

Following carbon capture from the PVSA process, the CO_2 rich stream removed from the column can be fed to a liquefaction system. There are many different processes which can be used to liquefy CO_2, with each process having varying degrees of performances. For this work, a commonly used process was implemented for ease of understanding, industrial relevance and simplicity of modelling. Hence, the Linde-Hampson CO_2 liquefaction process was chosen, as demonstrated in Seo et al. (2015). This is also illustrated in *Fig.2.*

Following reverse evacuation in the PVSA system, the gaseous CO_2 rich stream is compressed and then cooled to condense and remove any potential impurities which may damage downstream compressors. These impurities may also accumulate in the system as well as separation column. The CO_2/N_2 vapour stream exiting from the top of the column is mixed with a recycle, followed by compression of this combined stream. This stream is then cooled further in a cooler and heat exchanger before being throttled to reduce the pressure and liquefy the mixture. As CO_2 boils at a significantly higher temperature than N_2, an extremely pure liquid CO_2 stream is obtained, which can be then separated from the remaining gas in a separator. The N_2-rich vapour stream, which contains a substantial amount of CO_2, is then recycled to the heat exchanger to cool down the incoming CO_2/N_2 mixture. Finally, it is purged to prevent N_2 accumulation within the system, with the rest being returned to the mixer to maximise CO_2 recovery.

2.3 Hybrid PVSA-Liquefaction Process

As shown in *Fig.2,* the combined PVSA-liquefaction process consists of a standard Linde-Hampson liquefaction cycle immediately downstream of the evacuation stage. Following the addition of the cycle, the KPIs in *Section 2.1* are redefined to account for the impact of CO_2 liquefaction as follows:

$$Pu_{CO_2}(\%) = 100 \times \frac{n^{liq}_{CO_2,out}}{n^{liq}_{CO_2,out} + n^{liq}_{N_2,out}} \quad (4)$$

$$Re_{CO_2}(\%) = 100 \times \frac{n^{liq}_{CO_2,out}}{n^{pres}_{CO_2,in} + n^{ads}_{CO_2,in}} \quad (5)$$

$$E_T = \frac{E_{ads} + E_{bd} + E_{evac} + E_{pres} + E_{liq}}{m^{liq}_{CO_2,out}} \quad (6)$$

where $n^{liq}_{CO_2,out}$ and $n^{liq}_{N_2,out}$ are the amounts of CO_2 and N_2 liquefied respectively. Equation (3) is redefined to account for the energy consumed by the liquefaction process, E_{liq}. The total mass of CO_2 extracted also changes to $m^{liq}_{CO_2,out}$ post-liquefaction.

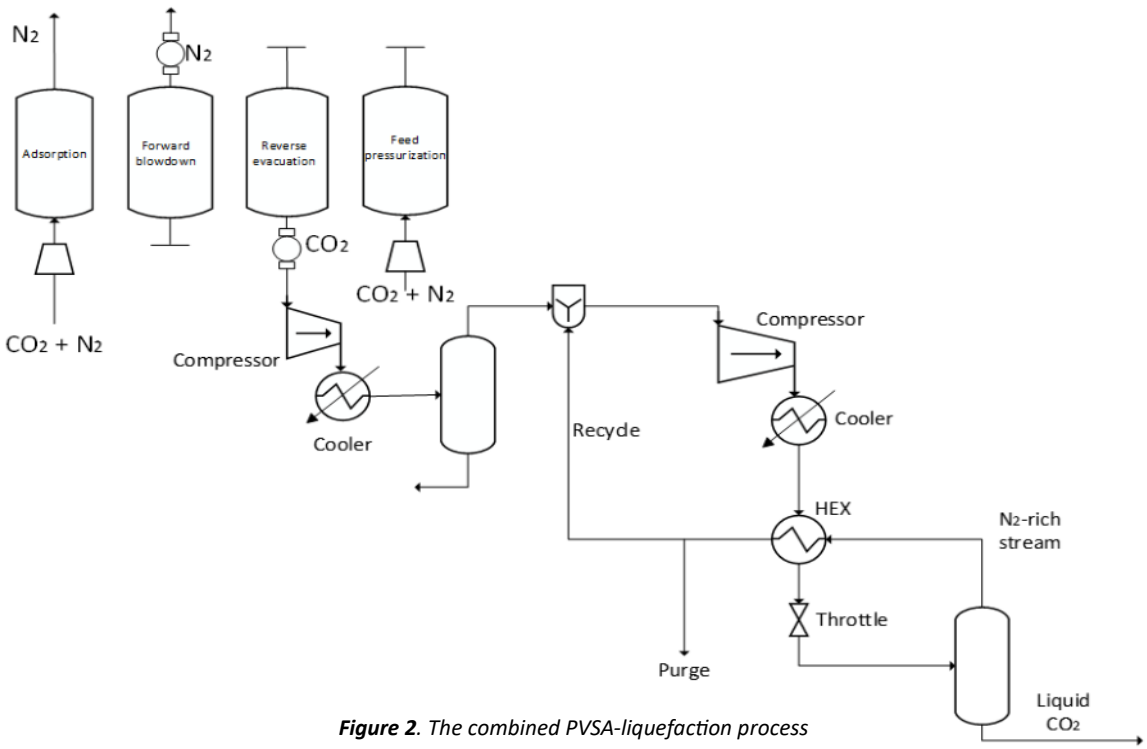


***Figure 2**. The combined PVSA-liquefaction process*

3. Methodology

3.1 PVSA Model

The PVSA model used within this work utilises a modified version of the Dual-adsorbent Batch Adsorber Analogue Model (DBAAM) as developed by Ward, Li and Pini (2023). This DBAAM model describes the separation of a two-component gas-phase mixture (CO_2 and N_2) using a four-step vacuum swing adsorption process, where the cycle is designed to extract the heavy component (CO_2) at high purity during the evacuation phase. The system was simulated using algebraic equations that are governed by a set of mass, momentum and energy balances to define the solid-phase adsorption model. These can be found in the original study by Ward, Li and Pini (2023).

In order to find the optimal points that form the pareto fronts for the relevant KPIs, thousands of simulations were performed at varying system operating pressures. This required a modification of the model whereby operating pressures were chosen randomly within a specified range using the Sobol sampling technique. These ranges are shown in *Table 1* and were chosen to provide a design space large enough to locate where ideal operation occurs within the DBAAM model. It should be noted that since some of the values within the pressures ranges can overlap, any non-physical operating points ($P_I < P_L$ or $P_H < P_L$ or $P_H < P_I$) generated by Sobol sampling were cleaned before being passed into the model.

The model was looped through every randomly generated Sobol point and the simulated results were stored for further processing.

A final modification to the model was required to vary the flue gas composition entering the PVSA system from a fixed value to a variable one. A sensitivity was then performed on the model by varying the feed concentration between 10-20%. Once the model had been edited to meet the requirements of this study, the simulations were performed.

***Table 1**. Operating Pressure Bounds*

Parameter	P_H(bar)	P_I(bar)	P_L(bar)
Lower Bound	1	0.05	0.02
Upper Bound	10	3	0.10

The performance of two adsorbents, Z13X and AC, were investigated in this study. AC is a versatile adsorbent renowned for its industrial applications and was chosen as the benchmark adsorbent for this study. Z13X on the other hand is a promising adsorbent for use in the PVSA carbon capture processes (Mustafa Erguvan and Amini, 2024), hence was chosen to determine the ideal performance of the technology.

Since the original model was developed for a dual-adsorbent system, a weighting factor ($w = \{0,1\}$) was applied to ensure only one set of parameters was used for each set of simulations, corresponding to Z13X or AC. The parameter data for both adsorbents was based off the study from Ward, Li and Pini, (2023).

For each adsorbent, three set of simulations were run with differing flue gas compositions. Most post-

combustion carbon capture processes (particularly those for coal and gas fired plants) operate between feed CO_2 compositions of 3-15%. However, heavy industries like cement and steel production emit larger concentrations of CO_2 between 20-35% (Wang and Song, 2020). Therefore, feed compositions of 10%, 15% and 20% were chosen for sensitivity analyses on the system.

The optimal points generated from the PVSA model were stored and passed to the liquefaction model to determine performance post-liquefaction. This is explained in more detail in *Section 3.3*.

3.2 Liquefaction Model

The CO_2 liquefaction process is modelled using ASPEN HYSYS and is based on the Linde-Hampson system described in *Section 2.2*. As the operating parameters of the liquefaction process and their corresponding effects on the KPIs were not under investigation, values for temperatures and pressures within the system were fixed for the duration of this study. These values are taken from the work by Seo et al. (2015).

After the model was constructed, a sensitivity analysis was performed by varying the CO_2 purity in the liquefaction feed (i.e. the PVSA outlet stream) between 70% - 99%, with a step change of 0.02%.

Although PVSA systems typically produce CO_2 streams exceeding 90% purity, a wider range of purities were inputted into the liquefaction inlet to relax the purity constraint on the PVSA system. This would lead to an increased number of feasible operating points and a larger optimal design space. This is discussed later in the paper, where ideal operating points were found at lower purities.

A limitation with the HYSYS model was its inability to converge the N_2-rich recycle streams for inlet CO_2 purities exceeding 99.08%. This is possibly caused by tolerance errors in the simulation, where the small amounts of N_2 leaving via the purge stream would result in numerical underflow and an accumulation of nitrogen in the system, hence preventing inlet purities greater than 99.08% to be simulated. For this reason, only purities up to 99% were investigated, however such problems can likely be prevented by increasing the nominal inlet flowrate, such that numerical underflow does not occur.

The liquefaction recovery, total compression power and liquid CO_2 outlet purity were measured at points between the range mentioned above to complete the sensitivity analysis. These three variables are required to calculate the post liquefaction KPIs mentioned in *Section 2.3*. It should be noted that while the liquefaction system includes additional energy demands such as pump power, these are several orders of magnitude smaller than the energy required for compression and were therefore ignored from the calculations.

A purge must be included to prevent the accumulation of nitrogen within the system. However, the amount of vapour purged can vary significantly from system to system and can greatly influence the KPIs. Hence, it was decided to also carry out a sensitivity analysis on the purge fraction of the system. These tests were carried out at 5% intervals between the 5-20% range of the recycle flowrate. To determine post-liquefaction KPIs, both post-liquefaction recovery and compressor power were visualised as function of inlet CO_2 purity and the correlations obtained were then used in the final stage of the work to integrate the two systems.

3.3 Hybrid PVSA-Liquefaction Model

The results from the PVSA model were combined to the liquefaction process to calculate the overall KPIs of the hybrid system. This was done by transferring MATLAB data from the PVSA model to the correlations obtained from the ASPEN model. The equations required to calculate these post-liquefaction KPIs were altered from those in *Section 2.3* and are described below:

$$Re_{CO_2}^{HYBRID} = Re_{CO_2}^{PVSA} \times Re_{CO_2}^{LIQ} \quad (7)$$

$$SEN_{CO_2}^{HYBRID} \left[\frac{kWh}{tonne}\right] = SEN_{CO_2}^{PVSA} + SEN_{CO_2}^{LIQ} \quad (8)$$

$$SEN_{CO_2}^{LIQ} = \frac{power_{compressor}}{n_{INLET}^{LIQ} \times purity_{CO_2}^{PVSA} \times Re_{CO_2}^{LIQ} \times M_{CO_2}} \quad (9)$$

where $Re_{CO_2}^{sys}$ and $SEN_{CO_2}^{sys}$ are the recoveries and SENs for the PVSA/liquefaction/hybrid systems respectively. M_{CO_2} is the molar mass of CO_2 and n_{INLET}^{LIQ} is the molar flow into the liquefaction. Both recovery and SEN for the PVSA component are computed by the DBAAM model. The model also calculates the PVSA purity ($purity_{CO_2}^{PVSA}$) alongside the other two KPIs. After the expressions for $Re_{CO_2}^{LIQ}$ and $power_{compressor}$ were found as functions of $purity_{CO_2}^{PVSA}$, the post-liquefaction KPIs were computed. Note that no figures or correlations of post-liquefaction CO_2 purity were developed as all simulated post-liquefaction points reached exceed 99% purity. This is adequate for most uses of liquid CO_2, except food and medicinal applications (Thunder Said Energy, 2024). The computed KPIs are plotted and discussed in *Section 4.*

4. Results and Discussion

4.1 Adsorbent Material

Fig.3 highlights the performances of Zeolite 13x and Activated Carbon as adsorbents at various operating pressures for a fixed flue gas composition of 15%. Given the same constraints of $> 70\%$ purity and $> 80\%$ recovery, Z13X has over 26 times more operating points that meet the applied recovery and purity constraints than AC.

The latter underperforms in comparison to Z13X even after relaxing its recovery constraint from 80% to 60%, demonstrating the scale at which Z13X dominates.

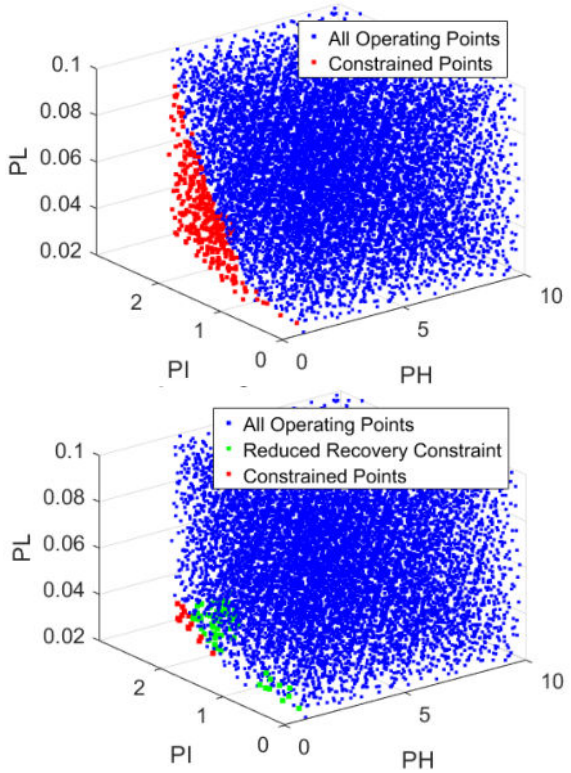


***Figure 3**. Z13X (top) and AC (bottom) optimal design spaces*

This is likely due to the conditions the adsorbents operate under, as Z13X shows higher equilibrium adsorption capacity for CO_2 at lower pressures (Zarghampoor et al., 2017). Z13X also exhibits a lower adsorption capacity for N_2 than AC at the tested pressure range, further strengthening its advantage as an adsorbent. Note that the purity and recovery constraints applied are significantly lower than the recommended standards of 95% and 90% respectively (US DOE. 2024). This reduced recovery constraint is to allow for an increased number of ideal operating points to be investigated, whilst the reduced purity constraint is to later test the hypothesis that high CO_2 purities will be achieved post-liquefaction irrespective of the purity achieved at the end of the PVSA cycle.

A common observation across both adsorbents was the location of the optimal design space. These occur near the lower bounds of the P_H and P_L ranges, and the upper bound of the P_I range. This trend is rather counter-intuitive, as both AC and Z13X have higher equilibrium adsorption capacities at higher pressures (Chue et al., 1995). However, competitive adsorption (Joos, Swisher and Smit, 2013) among other factors can explain this result. Although CO_2 adsorption selectivity is greater than N_2 at lower pressures, at higher pressures the increase in the absolute partial pressure of N_2 is significantly greater than that of CO_2 for a N_2-rich flue gas. Subsequently, the adsorption selectivity of CO_2 decreases, resulting in both lower purities and lower recoveries for both adsorbents.

Another potential reason for diminished performance at higher pressures are the heat effects which occur during the adsorption process. Since adsorption is an exothermic process, large amounts of heat are released when CO_2 adsorbs onto the bed. At higher pressures, this process occurs at a higher rate, resulting in sharp increases in local temperature. Since adsorption capacity decreases with temperature (Budnyak et al., 2020), this could explain the reduced system performance at higher pressures. These heat effects do not occur at lower pressures because adsorption happens more slowly, allowing gradual increases in temperature rather than instantaneous.

Optimal performance occurring at lower P_L and higher P_I is an expected result of the adsorbent. The lower the reverse evacuation pressure, the greater is the amount of the adsorbent that is regenerated as more CO_2 is removed from the pores, thereby increasing the recovery. A higher P_I is expected for improved operation as at lower P_I values, CO_2 losses in the forward blowdown step become increasingly significant, diminishing recovery. Alternatively, a lower P_I would increase CO_2 purity since more nitrogen is removed in the forward blowdown stage. However, this effect is difficult to see in *Fig.3* as a >70% purity constraint has been applied. Increasing this to >97% purity for Z13X, as shown in *Fig.4,* highlights that lower P_I values are preferred at higher purities.

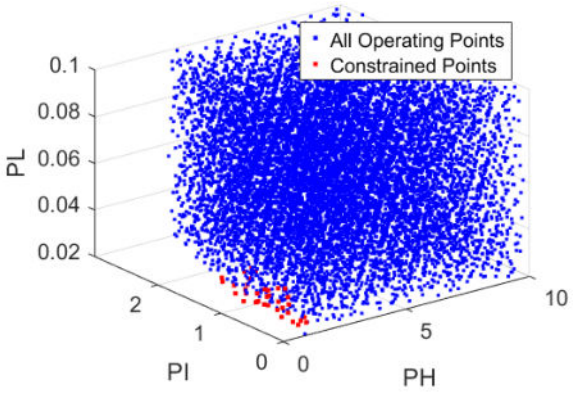


***Figure 4**. Z13X with 97% purity constraint imposed*

4.2 Flue Gas Composition

Fig.5 illustrates the performances of the adsorbents at varying flue gas compositions. It is evident from *Fig.5* that the number of optimal/feasible points increases with increasing flue gas concentration. At 10% CO_2 concentration only 119 optimal operating points exist for Z13X, whereas at 20% there are 621 points. This is an expected result and occurs due to the increased number of points that meet the recovery constraint. This is because post-CO_2 adsorption, the bulk gas phase (N_2-rich) for each feed composition can be assumed to contain similar concentrations of CO_2. This implies that a higher fraction of the CO_2 fed to the column is captured

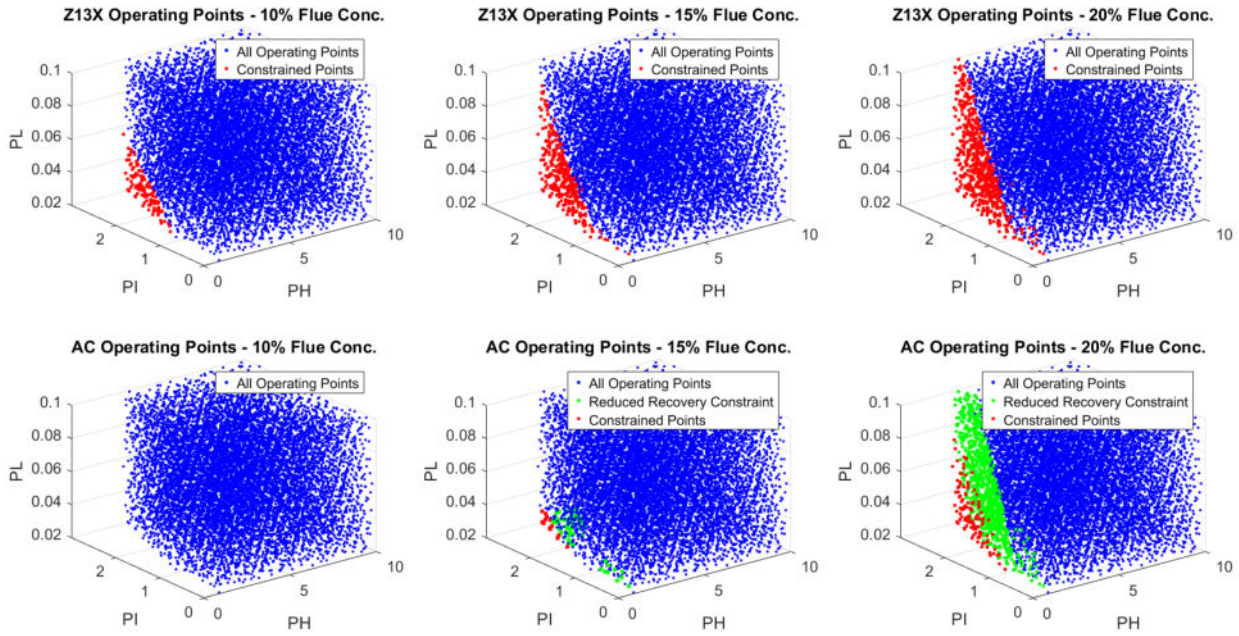


***Figure 5**. Z13X (top) and AC (bottom) performance at varying flue gas compositions*

at higher flue gas compositions enabling higher recoveries. However, it also implies that a similar amount of CO_2 is lost from both systems. Hence, a key takeaway from this test is that changing the flue gas composition drastically impacts the process performance, therefore it is imperative to compare adsorbent performance at equal feed gas concentrations. This is because very high CO_2 concentrations may skew the KPIs (particularly recovery) whilst extremely low concentrations may yield little to no optimal points. A key motivation for this study was the lack of literature on varying inlet CO_2 concentrations, with most studies opting for a fixed 15% concentration. However, this can prove advantageous when performing sensitivity on a given adsorbent material or comparing different adsorbents.

Increasing the flue gas concentration also increases the design space, particularly in the direction of increasing P_I and decreasing P_L which is shown by the emergence of newer points at lower P_L and higher P_I values. Since both adsorbents exhibit similar trends, this observation is independent of adsorbent properties and highlights the inherent nature of the PVSA process to prefer these operating conditions.

4.3 Liquefaction Process

Fig.6 illustrates the pre- and post-liquefaction recoveries for both adsorbents at varying flue gas compositions. Note that the pareto fronts also suggest that higher inlet concentrations contribute to higher recoveries, confirming the observation made in *Section 4.2*. It can be observed that higher flue gas concentrations result in higher purities and therefore increasing feed concentration can improve CO_2 recovery and purity.

The addition of the liquefaction cycle reduces the overall recovery of the hybrid system as shown in *Fig.6*. This is an expected result, due to the additional losses of CO_2 which occur in the purge stream. Unlike AC, Z13X shows consistent recovery reductions of approximately 12-15% post-liquefaction.

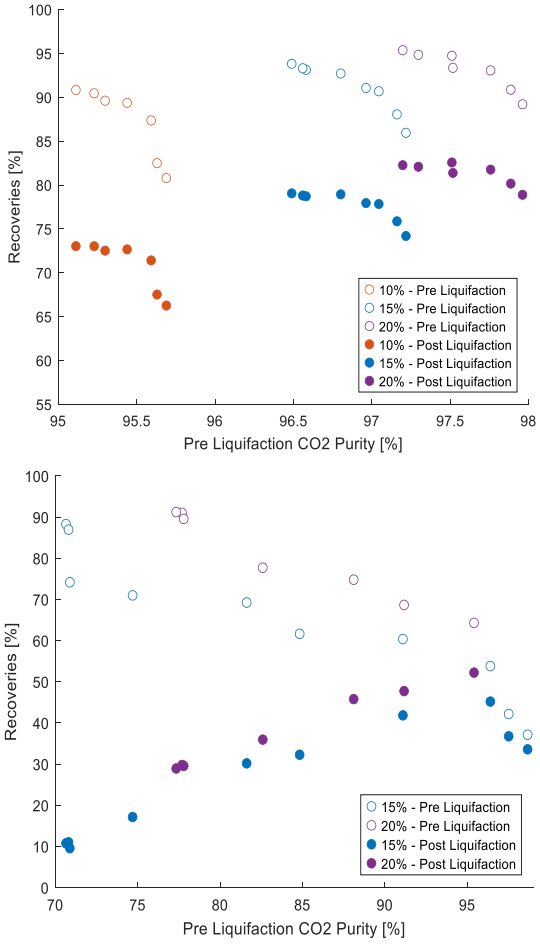


***Figure 6**. Z13X (top) and AC (bottom) Recovery Pareto Fronts*

On the other hand, the reduction in recovery for AC can range between 5-80%, with lower pre-liquefaction CO_2 purities exhibiting poorer performance.

These extreme reductions in recovery can render operation along these points virtually infeasible. Although low purity PVSA outlet streams can yield very high purities post-liquefaction, this comes at the cost of sacrificing recovery and hence, the feasibility of the hybrid system.

The increased losses of CO_2 at lower purities are also explained by the increased N_2 flow into the cycle. If the liquefaction feed is relatively rich in N_2, a large amount of N_2 is retained in the recycle and must be purged to avoid build-up. Larger purge flowrates correspond to greater losses of CO_2, resulting in the extremely low recovery rates. Mitigating strategies can be put in place to reduce these losses such as introducing further downstream multi-stage separation to recover additional CO_2 from the N_2-rich recycle. This technique has been applied in industry when dealing with oxy-fuel combustion, where inlet streams with approximately 70% CO_2 purity are liquefied (Knapik, Kosowski and Stopa., 2018).

Another key result from investigating post-liquefaction pareto fronts is the emergence of a recovery maxima. Both 15% and 20% feed compositions exhibit a maximum for AC in *Fig.6*. Recovery increases with pre-liquefaction purities up to a maximum before starting to decline again. This is caused by increases in recovery in the liquefaction cycle being offset by decreasing recoveries in the PVSA system. These systems integrate to form a maximum recovery, suggesting potential for further optimisation within future works. Note that this trend is also shown by Z13X but cannot be seen in *Fig.6* since lower purities are not included as part of any pareto fronts.

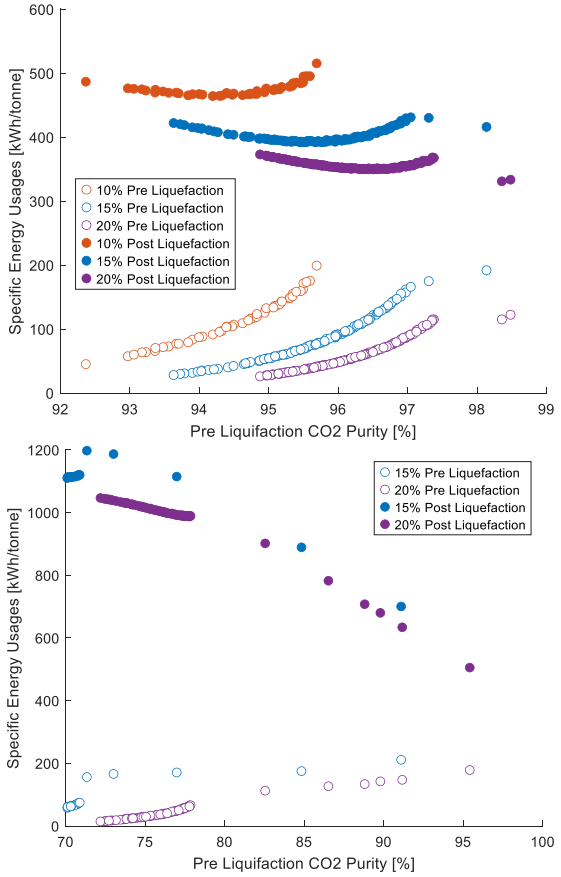


***Figure 7**. Z13X (top) and AC (bottom) SEN Pareto Fronts*

On the other hand, SEN for the hybrid process increases rapidly with the introduction of the liquefaction system as shown in *Fig.7*. This occurs due to the implementation of multiple compressors, which impose additional power requirements on the system. This substantial increase in the energy demand of the combined process highlights the need to integrate liquefaction within existing PVSA models, as neglecting this can present a misleading assessment of the process performance and hence optimal operation.

Additionally, there appears to be a significant disparity between the energy performance of AC and Z13X post-liquefaction. This arises due to the vast differences in purities each adsorbent operates at, with the increase in post-liquefaction SEN being significantly larger at lower purities. This is because at lower purities there is increased recycle within the liquefaction system which imposes further energy demand during compression. Furthermore, at lower purities a lower mass of CO_2 is liquefied which leads to an increased specific energy demand from the mathematical definition of SEN in *Section 3.3*.

Another trend observed in *Fig.7* is that increasing the pre-liquefaction purity increases the specific energy demand on the PVSA system. However, increasing flue gas concentration can help reduce this specific energy usage. This is because at lower flue gas compositions, optimal points lie lower in the P_L range. Pulling a vacuum is an energy intensive process, the greater the vacuum the greater the energy consumed. Since optimal points for lower flue gas compositions operate at lower P_L values, they require stronger vacuums, and thereby consume more energy as illustrated in *Fig.7*.

A key finding shown by Z13X in *Fig.7* are the contrasting trends of pre- and post-liquefaction. As mentioned previously, increasing the pre-liquefaction purity increases the corresponding SEN. This is not the case post-liquefaction where increasing the purity decreases the total SEN. Similar to the recovery maxima shown in *Fig.6* before, the integration of the PVSA system increases SEN with increasing purity but is offset by decreasing SEN with increasing purity within the liquefaction system leading to the emergence of minima. This presents yet another opportunity for further works to optimise the hybrid system and thereby maximise energy efficiency for the combined system.

To conclude, the operation of a hybrid PVSA-liquefaction system is improved with increasing flue gas composition with respect to both recovery and SEN. However, the relationship of these KPIs with pre-liquefaction purity is not so straightforward as observed by the formation of minima/maxima, resulting in the need to locate these stationary points.

4.4 Varying Purge Fraction

As stated previously, the existence of a purge within the CO_2 liquefaction cycle implies that the overall recovery of the system can never equal the recovery predicted by the PVSA model. However, investigating the effects of the purge can help optimise the process and minimise CO_2 losses. *Fig.8* shows the sensitivities carried out on ASPEN for various purge percentages and highlights how increasing the inlet purity improves post-liquefaction recoveries, whilst reducing total compressor power.

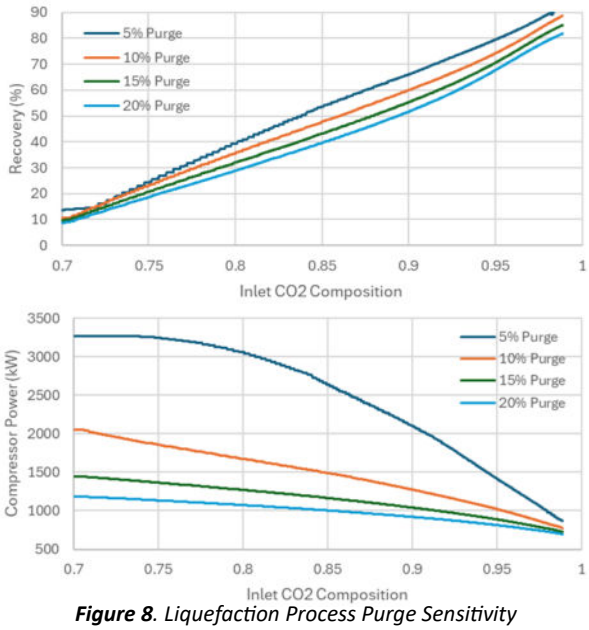


***Figure 8**. Liquefaction Process Purge Sensitivity*

Fig.9 and *Fig.10* show how the recovery pareto fronts change with varying purge fractions given a fixed flue gas composition of 15%. Note that the reduction in recovery occurs just like in *Section 4.3*, however it is the magnitude of the reduction that is of particular interest.

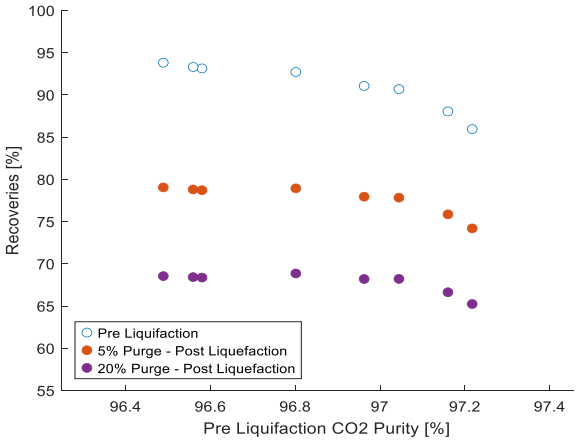


***Figure 9**. Z13X Recovery Pareto Fronts with varying Purge*

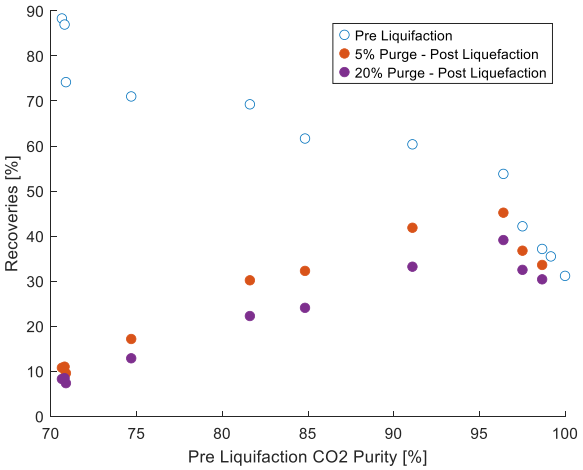


***Figure 10**. Z13X Recovery Pareto Fronts with varying Purge*

For both adsorbents, an increased purge fraction results in a further decrease in post-liquefaction recovery due to the increased losses of CO_2 that occur within the purge. Therefore, to maximise CO_2 recovery it is essential to operate the process at as low purge fractions as possible. On the other hand, increasing the purge fraction can minimise the SEN of the combined process as shown by *Fig.11*. This is because increasing purge fractions reduces the recycle flowrate, resulting in less flow through the compressor which in turn reduces the power demand of the system.

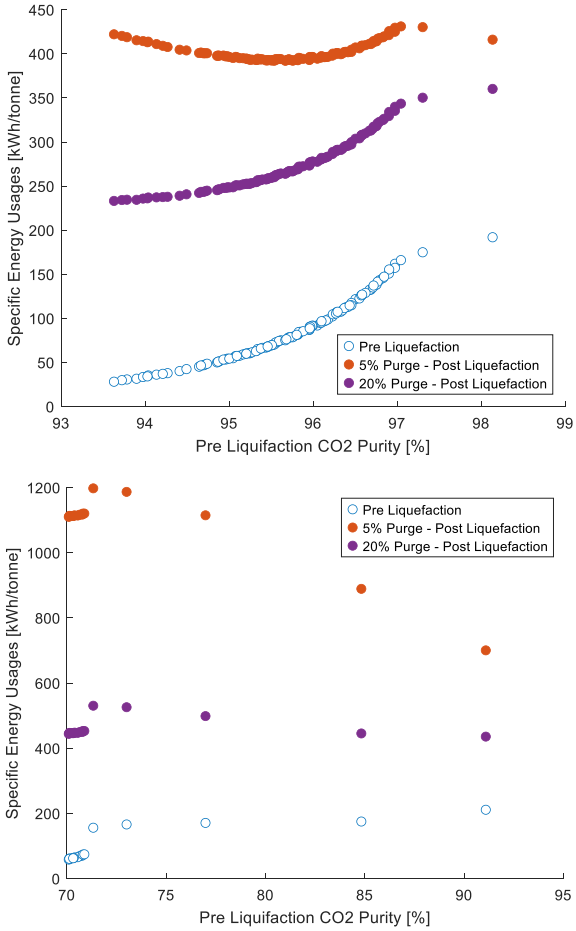


***Figure 11**. Z13X (top) and AC (bottom) SEN Pareto Fronts with varying Purge*

Due to this trade-off between recovery and SEN for varying purge fractions, it is essential to find an optimum between the two when designing an efficient hybrid system. To this end, minimising the energy consumption whilst maximising the recovery of the system could form the subject of future optimisation studies on the purge fraction, particularly since the relationship between SEN and recovery with varying purges remains largely unexplored.

5. Conclusion

The integration of a liquefaction process downstream of the PVSA unit significantly reduces the recovery of the overall hybrid process. It imposes an additional compressor duty, thereby increasing the specific energy consumption of the process. When comparing the performance of the two adsorbents, it was observed that Zeolite 13X outperformed Activated Carbon across all measured KPIs. Increasing feed composition to the adsorption process was also shown to enhance process performance as more CO_2 could be captured allowing higher recoveries and purities to be achieved along with reducing specific energy demand. Furthermore, it was concluded that the purge fraction plays a significant role in the performance of the liquefaction system. Increasing the purge can help reduce energy consumption but at the risk of diminishing the overall recovery of the hybrid process. On the other hand, prioritising recovery by opting for a lower purge ratio can lead to an increased energy demand, particularly from the liquefaction process due to the increasing costs of the additional compressor duties required. Therefore, optimising the purge ratio is critical to balance the trade-off between maximising recovery and minimising the specific energy usage.

6. Outlook

In future works, the results obtained in this study for an equilibrium model can be compared with the performance of a high-fidelity model of the PVSA system combined with a liquefaction cycle. This would account for the dynamic behaviour of the PVSA system (Ward and Pini, 2022) but would also integrate the liquefaction unit.

Furthermore, the liquefaction cycle developed within this work is a simple Linde-Hampson process which allows for simplicity in the implementation of a liquefaction process. However, the performance of the process should also be studied for additional CO_2 liquefaction systems which can include processes such as the pre-cooled Linde Hampson, dual pressure Linde Hampson or even the vapour compression refrigeration systems among others (Aliyon, Mehrpooya and Hajinezhad, 2020).

After studying the processes mentioned above, the optimal liquefaction system can be identified and the operational parameters within this, such as purge fraction and pressure, can be optimised. This is necessary to find a compromise between maximising recovery and minimising the energy usage.

Finally, the hybrid PVSA-liquefaction process can be optimised for operational parameters such as pressures and temperature through the implementation of a black-box optimisation model. This can also involve exploring the potential for recycling the purge towards the beginning of the PVSA process. This is because the purge contains a relatively high concentration of CO_2, despite being rich in N_2 gas. Therefore, mixing the purge stream with the flue gas can produce a feed which has a higher CO_2 composition. This can help increase the performance of the hybrid process by increasing the recovery whilst reducing its specific energy demand.

References

Aliyon, K., Mehrpooya, M. and Hajinezhad, A. (2020). Comparison of different CO2 liquefaction processes and exergoeconomic evaluation of integrated CO2 liquefaction and absorption refrigeration system. *Energy Conversion and Management*, 211, p.112752. doi:https://doi.org/10.1016/j.enconman.2020.112752.

Budnyak, T.M., Błachnio, M., Slabon, A., Jaworski, A., Tertykh, V.A., Deryło-Marczewska, A. and Marczewski, A.W. (2020). Chitosan Deposited onto Fumed Silica Surface as Sustainable Hybrid Biosorbent for Acid Orange 8 Dye Capture: Effect of Temperature in Adsorption Equilibrium and Kinetics. *Journal of physical chemistry. C./Journal of physical chemistry. C*, [online] 124(28), pp.15312–15323. doi:https://doi.org/10.1021/acs.jpcc.0c04205.

Chue, K.T., Kim, J.N., Yoo, Y.J., Cho, S.H. and Yang, R.T. (1995). Comparison of Activated Carbon and Zeolite 13X for CO2 Recovery from Flue Gas by Pressure Swing Adsorption. *Industrial & Engineering Chemistry Research*, 34(2), pp.591–598. doi:https://doi.org/10.1021/ie00041a020.

Hack, J., Maeda, N. and Meier, D.M. (2022). Review on CO2 Capture Using Amine-Functionalized Materials. *ACS Omega*, 7(44), pp.39520–39530. doi:https://doi.org/10.1021/acsomega.2c03385.

IEA. (n.d.). *How carbon capture technologies support the power transition – The role of CCUS in low-carbon power systems – Analysis*. [online] Available at: https://www.iea.org/reports/the-role-of-ccus-in-low-carbon-power-systems/how-carbon-capture-technologies-support-the-power-transition.

Joos, L., Swisher, J.A. and Smit, B. (2013). Molecular Simulation Study of the Competitive Adsorption of H_2O and CO_2 in Zeolite 13X. *Langmuir*, 29(51),

pp.15936–15942. doi:https://doi.org/10.1021/la403824g.

Knapik, E., Kosowski, P. and Stopa, J. (2018). Cryogenic liquefaction and separation of CO2 using nitrogen removal unit cold energy. *Chemical Engineering Research and Design*, [online] 131, pp.66–79. doi:https://doi.org/10.1016/j.cherd.2017.12.027.

Mustafa Erguvan and Amini, S. (2024). PARAMETRIC INVESTIGATION OF CO2 DESORPTION OF ZEOLITE 13X UNDER MICROWAVE CONDITIONS. *Carbon capture science & technology*, 11, pp.100189–100189. doi:https://doi.org/10.1016/j.ccst.2024.100189.

Seo, Y., You, H., Lee, S., Huh, C. and Chang, D. (2015). Evaluation of CO2 liquefaction processes for ship-based carbon capture and storage (CCS) in terms of life cycle cost (LCC) considering availability. *International Journal of Greenhouse Gas Control*, 35, pp.1–12. doi:https://doi.org/10.1016/j.ijggc.2015.01.006.

Thunder Said Energy. (2024). *CCS: what CO2 purity for transport and disposal?* [online] Available at: https://thundersaidenergy.com/downloads/ccs-what-co2-purity-for-transport-and-disposal/.

US DOE. (2024). *Point-Source Carbon Capture: Roles for Artificial Intelligence in Support of FECM RDD&D Priorities.* [online] Available at: https://www.energy.gov/sites/default/files/2024-04/point-source-carbon-capture.pdf

Unfccc.int. (2022). Available at: https://unfccc.int/resource/climateaction2020/tep/thematic-areas/carbon-capture/index.html.

Wang, X. and Song, C. (2020). Carbon Capture From Flue Gas and the Atmosphere: A Perspective. *Frontiers in Energy Research*, 8. doi:https://doi.org/10.3389/fenrg.2020.560849.

Ward, A., Li, K. and Pini, R. (2023). Assessment of dual-adsorbent beds for CO2 capture by equilibrium-based process design. *Separation and Purification Technology*, 319, pp.123990–123990. doi:https://doi.org/10.1016/j.seppur.2023.123990.

Ward, A.S. and Pini, R. (2022). Efficient Bayesian Optimization of Industrial-Scale Pressure-Vacuum Swing Adsorption Processes for CO_2 Capture. *Industrial & Engineering Chemistry Research*, 61(36), pp.13650–13668. doi:https://doi.org/10.1021/acs.iecr.2c02313.

Zarghampoor, M.H., Mozaffarian, M., Soleimani, M. and Takht Ravanchi, M. (2017). Modeling of CO2 Adsorption on Activated Carbon and 13X Zeolite via Vacuum Swing Adsorption. *IOP Conference Series: Materials Science and Engineering*, 206, p.012004. doi:https://doi.org/10.1088/1757-899x/206/1/012004.

High-Temperature Electrolytic Synthesis of Carbon Nanotubes from Molten Li_2CO_3

Tang, Bohui (Robert) and Wijaya, Colleen Handriani
Department of Chemical Engineering, Imperial College London, U.K.

Abstract: Carbon nanotube (CNT) production from molten Li_2CO_3 via high-temperature electrolysis provides a promising and cost-effective approach to carbon capture and utilization (CCU). In this study, various operational parameters were investigated to understand their effects on the Faradic efficiency of CNT electrosynthesis. The parameters investigated included the surface area of the electrodes (ranging from 3.14 to 6.28 cm^2), the mass of the Li_2CO_3 electrolyte (25 g), and the electrolysis time (2 - 4 hours). All experiments were conducted at a temperature of 800 °C. Multiple characterization techniques, including Raman Spectroscopy, X-Ray Diffraction (XRD), Scanning Electron Microscopy and Energy-dispersive X-ray spectroscopy (SEM-EDS), Transmission Electron Microscopy (TEM), and Simultaneous Thermal Analysis (STA), were used to verify the formation of CNTs and elucidate the effects of electrode materials and electrolysis time on their quality. The results indicated that using stainless steel as a cathode with larger electrode surface area, and longer electrolysis duration would enhance both CNT quality and its Faradaic efficiency. Faradaic efficiencies below 100% can be explained using thermodynamic predictions: equilibrium and thermoneutral potential calculations indicated the possibility of CO gas formation under certain conditions. Overall, these findings show the potential of adjusting electrode configurations and operating conditions to optimize CNT production, thereby contributing to the ongoing development of sustainable CO_2 conversion strategies.
Keywords: *Carbon Capture and Utilization, Carbon Nanotubes, Faradaic Efficiency, Thermoneutral Potential*

1. Introduction

Carbon dioxide (CO_2) is a potent greenhouse gas that is the main contributor to global warming. Since the industrial revolution in the 18th century, atmospheric CO_2 concentration has risen by 50%.[1] This has led to extreme weather, ocean acidification, and the disruption of biodiversity and natural habitats. Thus, it is urgent to remove CO_2 from the atmosphere.[2]

The main two methods to remove CO_2 from the atmosphere or limit its release from point sources is carbon capture utilisation and storage (CCUS) and carbon capture utilisation (CCU). CCUS is applied to power plants and industrial facilities and stores CO_2 in geological formations such as depleted oil fields. However, there is a risk of geological leakage. On the other hand, CCU offers a greener approach and transforms captured CO_2 into high economic value products such as polyurethane, methanol, and carbon nanotubes (CNTs). [3] This circular economy aspect makes CCU more beneficial than CCUS as CO_2 is instead stored in stable products with minimal leakage risk.

The main problems faced by CCU are low cost-effectiveness, additional CO_2 emission, and energy consumption during CO_2 conversion. However, carbon nanotubes (CNT) offer a promising alternative, bypassing these issues while providing significant real-world applications such as battery energy storage systems, semiconductor materials in electronics, drug delivery systems, and others.[4] The CNT market also has a compound annual growth rate (CAGR) of 14.2%.[5] It has experienced significant growth due to their exceptional properties such as high mechanical strength, electrical conductivity, and thermal stability.

Traditional CNT production methods such as laser ablation and chemical vapour deposition remain costly, with costs around $114,000 per tonne of CNTs and lack the scalability for large-scale production.[6] A better method to produce CNTs is Molten Salt CO_2 Capture and Electro-Transformation (MSCC ET) which could potentially reduce costs to $700 per tonne.[6] This substantial cost reduction increases the economic viability of large-scale CNT production.

However, the main challenge in producing CNTs through MSCC ET is the difficulty in controlling their quality. Hence, this paper aims to examine how different process conditions impact the quality of CNT products, calculate the faradic efficiency, and optimise it.

2. Background

Molten salt electrolysis is a process that breaks down metal salt electrolytes such as Li_2CO_3, NaCl, rare metal ores, and others into desired components depending on the electrode material used and operating conditions. It is a method for the extraction of important metal commodities and can be used to extract precious metals in battery recycling.[7]

There has been a focus on molten salt electrolysis used to produce carbon materials through the combination of carbon capture technologies. A recent study by Laasonen, et.al (2023) concluded that various carbon nanostructures can be produced via high-temperature molten salt electrolysis of CO_2 in molten Li_2CO_3 at 750 °C. However, metal impurities were present in carbon nanostructures and greatly influenced the carbon morphologies and yield.[8] Another paper by Deqiang et.al (2022) found that optimising the electrolysis temperature, current density, and molten salt structure can impact the synthesis of carbon materials on stainless steel (SS), iron (Fe), and zinc (Zn) cathode.[9] This report focuses specifically on carbon nanotubes production, and how temperature, cathode and anode materials, and electrolysis time affects the product quality and morphology. It is important to note that this electrolysis method only yields multi-walled carbon nanotubes.

The proposed reaction mechanism for reducing captured CO_2 in Li_2CO_3 electrolyte to carbon nanotubes via high temperature electrolysis is shown below: [10]

During electrolysis, the carbonate ion is reduced to carbon at the cathode.

$$CO_3^{2-} + 4e^- \rightarrow C_{CNT} + 3O^{2-} \quad (1)$$

At the same time, the carbonate ions (CO_3^{-}) can be regenerated by the injecting carbon dioxide directly into the electrolyte:

$$CO_2 + O^{2-} \rightarrow {CO_3}^{2-} \quad (2)$$

At the anode, the O^{2-} is oxidised to O_2:

$$2O^{2-} \rightarrow O_2 + 4e^- \quad (3)$$

Hence, the net reaction occurring during electrolysis where solid carbon is produced is:

$$CO_2 \rightarrow C_{\mathrm{CNT}} + O_2 \quad (4)$$

However, in the proposed mechanism, the CNT growth can occur either via 'tip growth' or 'root growth' depending on the catalyst.[11] In tip growth, catalyst particles adsorb carbon and become saturated, leading to nanotube formation that pushes the catalyst upward, keeping it at the nanotube tip. In contrast, for root growth, the catalyst remains anchored to the substrate while carbon precipitates outward, forming nanotubes from the base. In-situ TEM is a technique that allows for the direct observation of dynamic processes within a material as they occur in real-time. Therefore, the exact growth could be determined by using in-situ TEM to observe CNTs nucleation and growth in the future.

The objective of this study was to investigate: (i) the feasibility of producing carbon nanotubes (CNTs) through high-temperature electrolysis of molten Li_2CO_3. (ii) the effects of various operating parameters on the quality and Faradic efficiency of the resulting CNTs , as well as identify potential optimization methods. Throughout the experiments, the temperature was held at 800 °C and the mass of the Li_2CO_3 electrolyte was fixed at 25 g. The primary variables investigated included electrode surface area (3.14 to 6.28 cm^2), electrode material (Stainless steel, Nickel and Iron), and electrolysis time (2 – 4 hours).

3. Methodology

3.1 Reactor setup

The reactor for CNT production, with size 228 mm × 655 mm × 480 mm, consisted of an alumina (aluminium oxide) crucible containing the molten electrolyte, as shown in Figure 1. This crucible is placed within a CARBOLITE GERO 3016 furnace equipped with thermal insulation. Electrode materials were inserted into the crucible for electrochemical reactions, which were powered by the potentiostat (Autolab PGSTAT302N). The measuring and control software NOVA 2.1.7 was used to apply current and measure the electrochemical parameters during the electrolysis process. CO_2 gas was supplied from a CO_2 cylinder via a regulator and mass flow controller and was bubbled into the molten Li_2CO_3 electrolyte. This setup enabled precise control over the reaction temperature, gas flow, and electrochemical parameters, enabling efficient CNT production.

3.2 CNT synthesis process

Firstly, the anode and cathode materials, which were formed from wires, were shaped into spiral forms to increase their surface area. They were then inserted into an alumina holder to prevent breakage during the high-temperature heating process. The assembled electrodes and the weighed lithium carbonate salt were placed together into an alumina crucible.

The electrodes were threaded through FFKM o-rings and the lid holes, and fixed with screws, ensuring they faced each other properly. The crucible was then placed into a split tube furnace, which was closed and clamped.

After ensuring proper ventilation, the furnace was programmed to heat slowly to the set temperature at 4 °C/min. During the heating process, the open circuit potential (OCP) between the anode and the cathode was measured. Upon reaching the target temperature (800 °C), electrolysis was initiated by applying a current through the system (the required current was calculated based on the electrode surface area to ensure a current density of 200 mA/cm^2, which was used in all experiments).

3.3 Post treatment

After the electrolysis process, the solid carbon product remained on the electrodes and inside the crucible. To remove any impurities and excess electrolytes, the product was rinsed with 13% dilute hydrochloric acid. The acid solution was then filtered through a funnel to separate the solid carbon product from the liquid. The filtered solid product was further rinsed with distilled water, dried at 80°C in an oven overnight, collected, weighed, and stored in a vial for characterization. The product weight was used to calculate the Faradaic efficiency.

3.4 Faradic efficiency determination

Faradic efficiency, Φ, is an important metric for assessing how efficiently electrons are utilized in an electrochemical reaction within a system.[12] By quantifying the Faradic efficiency, we can evaluate how effectively the system produces CNTs from lithium carbonate, and identify and mitigate energy losses to improve the overall efficiency. Φ is calculated by the following equation:

$$\Phi/\% = \frac{m_{\mathrm{formed}}}{m_{\mathrm{predicted}}} \times 100 = \frac{nF \times m_{\mathrm{formed}}}{Q} \times 100 \quad (5)$$

Where n is the electron stoichiometry, F is the Faraday constant, Q is the total charged passes in coulombs.

Since the mass of the CNT product is very small (less than 1 gram) and can easily disperse into the atmosphere, a robust protocol needed to be followed to minimize errors. First, the funnel needed to be thoroughly cleaned to remove any residual impurities. The combined mass of the dry filter paper and funnel were recorded beforehand. After the filtration process, the funnel containing the filter paper and CNT product was placed in an oven for drying and was covered with a lid to prevent any CNT product loss. The entire assembly was weighed after drying, and the mass of the lid, filter paper, and funnel was subtracted to obtain the mass of the electrochemically synthesised CNTs. The

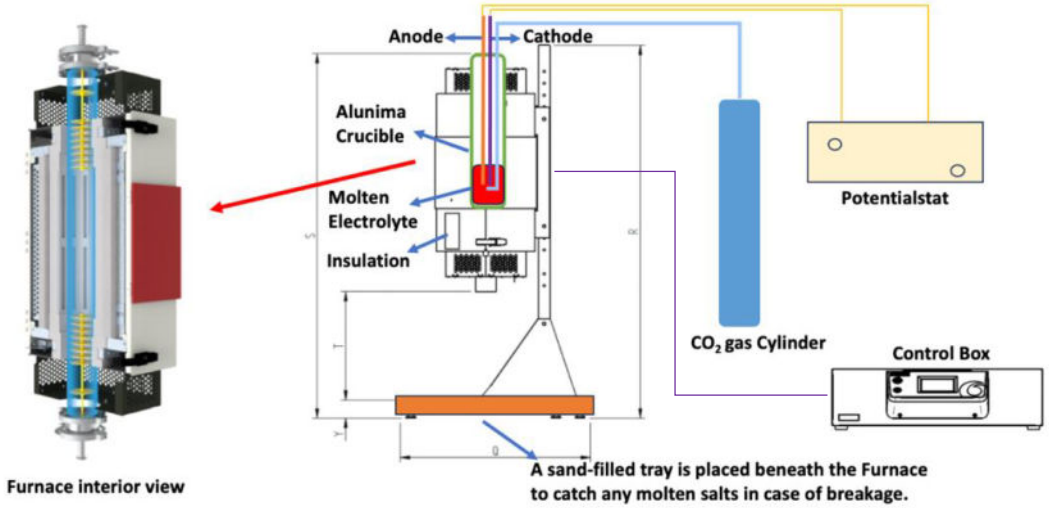


Figure 1. *Experimental set-up & control system*

Faradaic efficiency was then calculated by dividing the actual deposited mass by the theoretical mass.

3.5 Thermoneutral and equilibrium potential modelling

In electrochemistry, the equilibrium potential, E_{eq}, is the potential at which the forward and reverse rates of an electrochemical reaction are equal, meaning no net reaction occurs:[13]

$$E_{\text{eq}} = -\frac{\Delta G}{nF} \quad (6)$$

The thermoneutral potential, E_{th}, is the potential across the electrochemical cell that is necessary not only to drive the cell reaction but also supply the heat required to maintain a constant temperature:[14]

$$E_{\text{th}} = -\frac{\Delta H}{nF} \quad (7)$$

Where the ΔG is the Gibbs free energy change and ΔH is the change in enthalpy.

The change in enthalpy is calculated by the Shomate Equation:[15]

$$\Delta H = H_{298.15} + \left(At + Bt^2 + C\frac{t^3}{3} + D\frac{t^4}{4} - \frac{E}{t} + F - H\right) \quad (8)$$

Where t = temperature (K)/ 1000, the coefficients are different for each component and temperatures and are extracted from NIST website.

The Gibbs free energy is calculated by: [16]

$$\Delta G = \Delta H - T\Delta S \quad (9)$$

The equilibrium potential indicates the minimum theoretical voltage required to drive the electrolysis reaction forward. And by calculating the equilibrium potential of the side reaction, it helps determining which reaction is more favourable in which conditions; Meanwhile, the thermoneutral potential provides information about the heat management; if the applied potential is greater than E_{th}, excess energy is transferred into heat, reducing the overall efficiency. Moreover, it also helps evaluating whether the process is scalable without requiring excessive thermal management.

Raman spectroscopy was employed to characterize the structural properties of the synthesized carbon nanotubes (CNTs). It provides detailed information about the vibrational modes of carbon atoms, which lead to the assessment of defects, disorder, and the degree of graphitization in the CNTs. Raman spectra were obtained using a BRUKER SENTERRA II spectrometer over a spectral range of 200 cm^{-1} to 4000 cm^{-1}.

X-ray Diffraction (XRD) was used to determine the crystallographic structure and phase purity of the produced CNTs. The measurements were performed using a Malvern PANalytical AERIS XRD, the data were collected over a 2θ angle range of 5° to 60°.

3.7 Microscopy techniques

Scanning Electron Microscopy and Energy-Dispersive X-ray Spectroscopy (SEM-EDS) were used to examine the surface morphology and structural features of the CNT samples, as well as to identify and quantify the elemental composition of the observed structures. These analyses were conducted using a GeminiSEM 460.

Transmission electron microscopy (TEM) was employed to investigate the internal structure and crystallinity of individual CNTs. This was carried out with TEM JEOL 2100 Plus electron microscope.

3.8 Simultaneous thermal analysis (STA)

Simultaneous Thermal Analysis (STA) was conducted to monitor the mass changes and thermal events of the CNT samples as a function of temperature at a controlled heating rate. This technique combines Thermogravimetric Analysis (TGA) and Differential Scanning Calorimetry (DSC) to provide comprehensive insights into the thermal stability and phase transitions of the CNTs. The STA was performed using NETZSCH STA 449 at temperatures up to 800 °C.

4. Results & Discussion

A total of 5 experiment data sets were obtained through varying the electrode materials and operating conditions to investigate the impact on the CNT product quality and efficiency. The details are shown in Table 1. Various parameters such as electrode length and electrolysis time were initially planned to be tested, however due to the

***Table 1**. Conducted Experiment details for carbon nanotube production from molten Li_2CO_3*

Experiments	Current Density (mA/cm^2)	Anode	Cathode	Electrode Length(cm)	Temperature (°C)	Electrolysis Time(hr)	Products
EXP1	200	Nickel	Iron	20	800	2.25	CNT
EXP2	200	Stainless Steel	Stainless Steel	10	800	2	CNT
EXP3	200	Nickel	Iron	10	800	2	CNT
EXP4	200	Nickel	Iron	10	800	2	CNT
EXP5	200	Reused Nickel	Reused Iron	10	800	2	CNT

limited project timescale and equipment failures, these could not be fully explored.

In Experiment 1, a nickel anode and iron cathode combination was used as it is known to facilitate CNT production instead of other carbon structures, e.g. carbon fibre, graphene.[17] They are also less expensive than other metal alloys and more economical for future large-scale operations. During Experiment 1, it was noticed that the electrode had dropped into the electrolyte and could not pass high currents at 20 cm electrode length. Furthermore, the potentiostat broke down for operations past 2 hours, thus the electrolysis time was set at 2 hours for the remaining experiments to facilitate direct comparisons of experimental results.

Experiment 2 was conducted using stainless steel as both the anode and cathode material. Stainless steel is also an economical choice and has high corrosion resistance. Its composition includes nickel, iron, chromium, and manganese, which could potentially enhance CNT yield and efficiency, since chromium and manganese may act as catalysts during the reaction.

Electrolysis using a nickel anode and iron cathode were repeated for Experiments 3 and 4 but at an electrode length of 10 cm and electrolysis time of 2 hours. This helped determine the average Faradic efficiency and examine any consistent patterns in the CNT structure produced using nickel anode and iron cathode set-up. The final Experiment 5, the nickel anode and iron cathode from Experiment 3 were reused to evaluate whether reusing electrodes affects the quality of the CNT product and its Faradaic efficiency, while also reducing costs.

Unfortunately, after Experiment 5, the crucible apparatus broke and no further experiments could be conducted.

4.1. Open circuit potential (OCP) measurement

For all experiments, the open circuit potential (OCP) was measured during the heating-up stage of the furnace to provide a baseline of the electrochemical stability of the system without any external electrical influence. This measurement indicates if the state of the electrodes is operationally good, reacting, or degrading.

As shown in Figure 2, before 300 °C, there were significant fluctuations in the potential difference for all the experiments, which were mainly caused by poor ionic conductivity of the Li_2CO_3 salt prior to it melting. In Experiment 2, which utilized stainless steel for both electrodes, the potential difference between the electrodes was approximately zero, which is likely because identical materials were used for both electrodes. Additionally, the unique mechanical and chemical stability of stainless steel over a wide range of temperatures further ensures consistent behaviour.

In contrast, for the other experiments with nickel as the anode and iron as the cathode, there was a gradual increase in the potential difference as the temperature rises from around 300 °C. This is most likely due to the increase in kinetic energy that allows ions in the electrolyte to move more freely and react more readily at the electrode surfaces. In this case, the positive OCP suggests that nickel is more likely to oxidize, leading to increased degradation or corrosion. Meanwhile, at 631K (358 °C), data for nickel from the JANAF tables show a marked increase in heat capacity, indicating a significant transition point.[18] This peak in C_p often signifies a structural or phase stability transition within the material.

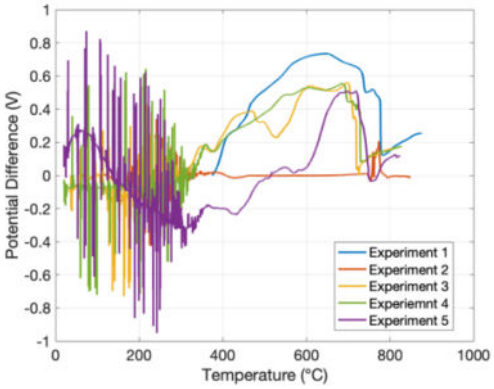


***Figure 2**. Open circuit potential difference of all experiments*

4.2 Potential difference during electrolysis

The cell potential difference was measured during the electrolysis process to monitor the electrochemical reactions occurring at the electrodes and to evaluate the stability of the electrolytic system. As shown in Figure 3, during the first hour of electrolysis, Experiment 2 showed the highest cell potential difference, starting with an initial peak above 2.6 V and stabilizing around 2.5 V. This behaviour is due to the unique mechanical property of stainless steel, especially its chromium-induced passivation layer provides superior corrosion resistance. Experiment 1, which used a 20 cm nickel

anode and iron cathode, started at a slightly lower potential but experienced a gradual decrease over time to about 2.3 V. In contrast, Experiments 3, 4, and 5, which had a similar setup to Experiment 1 but with an electrode length of 10 cm and hence a lower surface area, indicated a more pronounced potential drop to about 1.8 V. This was even more significant in Experiment 5 where reused anodes were employed.

After an hour of electrolysis, there was a significant increase in the potential difference, surging to about 5.0 – 7.0 V for all experiments except Experiment 5 with the reused electrodes, shown in Figure 4. This increase indicates that an oxidation reaction might be occurring at the anode during the electrolysis process. In addition to oxygen evolution happening at the anode, nickel may be oxidized to Ni^{2+}, causing degradation and structural changes in the electrode. This oxidation not only affects the electrode integrity but may also impact the overall electrolysis efficiency. However, in the case of Experiment 5, the reused electrodes were polished before use, resulting in less active material on their surfaces. This polishing likely reduced the amount of nickel available to oxidize, making oxidation less probable in this experiment. As a result, the potential difference did not surge as it did in the other experiments.

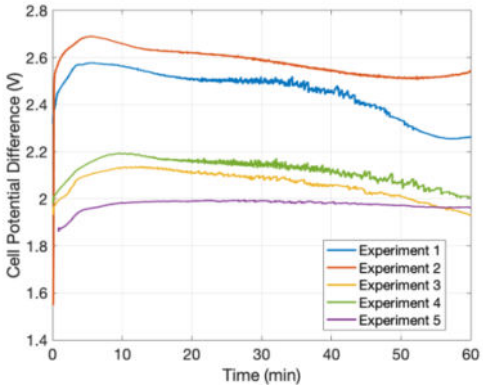


Figure 3. *Cell potential difference of all experiments during the first hour electrolysis*

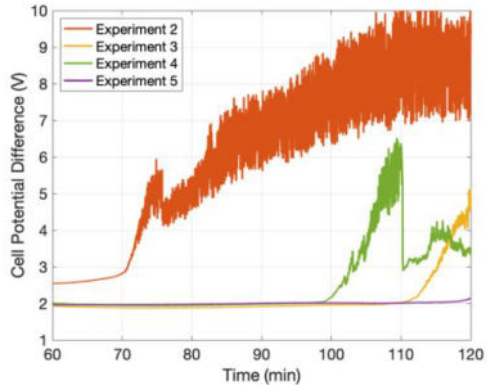


Figure 4. *Cell potential difference of all experiments after 1 hour of electrolysis*

4.3 Faradic efficiency

Table 2. *Faradic Efficiency of all conducted experiments*

N.O	Sample Mass(g)	Faradaic Efficiency (%)
EXP1	0.286	87.14
EXP2	0.094	66.90
EXP3	0.050	35.62
EXP4	0.072	51.50
EXP5	0.104	74.29

The Faradaic efficiency results are shown in Table 2. Experiment 1, which utilized longer electrodes (20 cm) made from nickel and iron and had a slightly longer electrolysis duration (2.25 hours), demonstrated the highest Faradaic efficiency of 87.14%. It is noteworthy that during this experiment, the iron cathode broke and dropped into the electrolyte, which is likely due to the high current applied. In contrast, Experiment 2, with 10 cm long stainless steel electrodes, yielded a moderate efficiency of 66.90%. Experiments 3 and 4 gave relatively lower Faradaic efficiencies, mainly due to measurement issues — some of the CNT product was lost during measurement and balance was not calibrated well, and also because the predicted mass of CNT was pretty low (0.14 g), errors had a significant impact on the result. Experiment 5, which used reused electrodes and incorporated lessons from the previous two experiments, achieved a higher Faradaic efficiency but still less than that of Experiment 1. These results suggest that a larger electrode surface area combined with longer electrolysis time can yield a higher CNT product mass, enhance the effectiveness of the electrochemical process, decrease measurement errors, and result in higher Faradaic efficiency.

Furthermore, another main issue prevents reaching very high Faradaic efficiency is the occurrence of side reactions, particularly the 2-electron pathway:

$$Li_2CO_3 \rightarrow Li_2O + CO + \frac{1}{2}O_2 \qquad (10)$$

During electrolysis, carbon monoxide (CO) is likely to form, which can be confirmed by the calculated thermoneutral and equilibrium potential graph in Figure 5. The graph shows that the thermoneutral potential for CO formation is consistently higher across all temperatures, indicating that this reaction requires more energy compared to the carbon (CNT) formation reaction. Therefore, CO reaction is not desirable as excess electrical energy is converted to heat, potentially reducing the overall system efficiency.

However, the equilibrium potential for CO decreases more steeply than that for carbon as temperature increases, showing that CO formation becomes energetically more favourable at higher temperatures. This trend might explain the difficulty in achieving higher Faradaic efficiencies, as CO formation competes with the desired CNT production.

Our experiments were conducted at 800 °C with a current range of 0.62 to 1.28 A. The cell potential difference was recorded between 1.8 V and 2.6 V during the first hour of electrolysis, as shown in Figure 4. From Figure 5, the equilibrium potentials for the CNT and CO reactions were determined to be 1.32 V and 1.62 V, respectively, while the thermoneutral potentials were 1.58 V and 2.60 V, respectively. This confirms that the CNT formation reaction is thermodynamically more favourable. However, CO gas still probably forms, thereby reducing the Faradaic efficiency.

Theoretically, as indicated by the modelled graph, temperatures below 550 °C and potentials between 1.2 V and 1.6 V significantly favour CNT formation. Under these conditions, the potential lies between the thermoneutral and equilibrium potentials for CNT formation and below the minimum potential required to drive the CO reaction. This maximizes the system's overall efficiency. However, further experiments are needed to confirm this, as lower temperatures might hinder CNT formation or result in poor-quality CNTs.

There are some other methods that should be considered to increase the Faradic efficiency. First, increasing the surface area of the electrodes and increasing electrolysis time might help. However, in this setup, the electrodes in the crucible are prone to breakage due to high currents. Another approach is to add an iron catalyst into the electrolyte to enhance CNT production by providing a more efficient catalytic pathway for carbon deposition.[19]

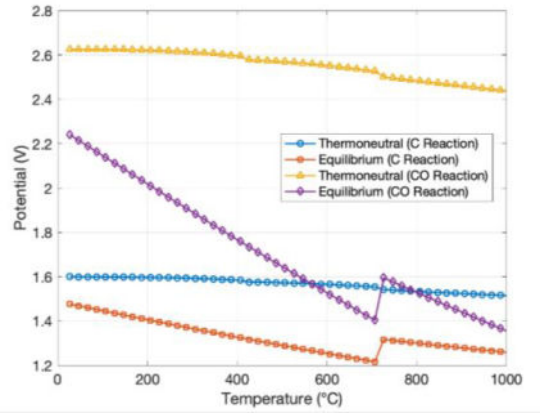


Figure 5. *Modelled equilibrium potential and thermoneutral potential diagram*

4.4 Raman spectroscopy

Raman peaks were observed at 1370 cm^{-1}, 1580 cm^{-1}, and 2700 cm^{-1}, shown in Figure 6. This corresponds reports in the literature, where the first Raman peak at 1370 cm^{-1} in the D band indicates the presence of defects in the CNT products.[10] This can be contributed by the impurities present in the CNTs. The second Raman peak at 1580 cm^{-1} in the G band indicates the stretching motion of sp^2 atom pairs in a long chain. The third Raman peak at 2700 cm^{-1} in the 2D band indicates the multilayer order and crystallinity of the CNT products.

The I_G/I_D ratio is the intensity ratio of the G band peak to the D band peak. A high I_G/I_D ratio indicates fewer defects and a more ordered, graphitic structure, which is that of a high-quality graphene CNT. A low I_G/I_D ratio indicates a higher defect intensity or more disordered structure. All CNT products from the conducted experiments exhibited a higher I_G/I_D than that of the Sigma-Aldrich branded commercial CNT which is 0.980, thus, indicating the high quality of CNT produced electrochemically.

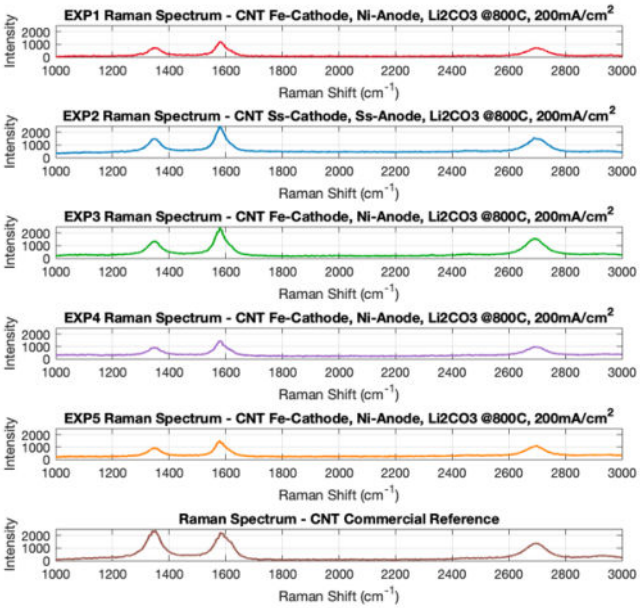


Figure 6. *Raman spectra of all conducted experiments*

Meanwhile, I_{2D}/I_G ratio is the intensity ratio of the 2D band peak to the G band peak. A lower I_{2D}/I_G ratio typically indicates the presence of multi-walled carbon nanotubes (MWCNTs).[9] In our experiment, all samples exhibited a ratio of less than 0.75, indicating multiple-layer CNT has produced.

Table 3*. Raman spectroscopy analysis of various samples*

No	I_G/I_D	I_{2D}/I_G	Properties
1	1.598	0.602	multi-layer
2	1.680	0.611	multi-layer
3	1.847	0.647	multi-layer
4	1.623	0.656	multi-layer
5	1.609	0.746	multi-layer
Commercial	0.980	0.620	multi-layer

4.5 X-ray diffraction(XRD)

X-ray diffraction patterns for all CNT samples are shown in Figure 7. It is noticeable that there are prominent peaks at approximately 18°, 26°, 35°, 37°, 43. According to the literature, the peak at 26° corresponds to the (002) plane in Miller indices, indicating the presence of graphitic carbon structures.[20] The broad, low-intensity peak at 26° suggests the presence of amorphous carbon in the samples and may be masking unresolved smaller peaks. The peak at 18° could be attributed to carbon nanodots, also corresponding to the (002) planes. Significant peaks at 35°, 37°, and 43°

indicate lithium residues and nickel oxide with the lithium peak's intensity varying between samples. Additionally, the peak at 43.9°, corresponding to the (111) plane in Miller indices, may signify the presence of metallic nickel or nickel compounds.

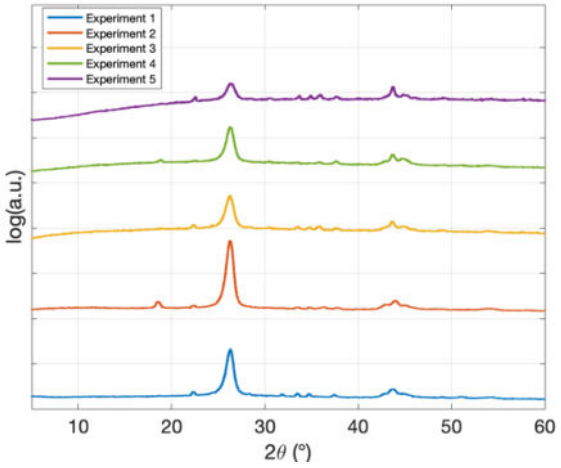


Figure 7. *X-ray diffraction (XRD) of all experiments*

4.6 Scanning electron microscopy(SEM)

Scanning Electron Microscopy (SEM) images from the experiments are shown in Figure 9. It is noticeable that Experiments 1 and 2 produced high-quality CNTs, especially Experiment 2. The images show a high proportion of CNTs that are less clustered, and the nanotubes have comparatively long lengths. Longer and less clustered nanotubes are considered higher quality because they exhibit superior mechanical, electrical, and thermal properties, and are more effective in reinforcing materials and conducting electricity and heat. They also disperse more uniformly in composites, enhancing overall material performance. In contrast, the CNTs produced from Experiments 3, 4, and 5 are very short and heavily clustered together. Additionally, there are shining impurities present in these samples, such as lithium residues and nickel and iron metal particles

4.7 Transmission electron microscopy (TEM)

The quality of the carbon nanotubes (CNTs) was further assessed using Transmission Electron Microscopy (TEM). TEM images, shown in Figure 10, revealed that Experiment 2 produced longer and less clustered CNTs compared to the other experiments. Moreover, it was observed that some metal residues were congested within the hollow tubes, which could be the potential metal catalyst for CNT reaction. At higher magnifications, the diameter of the nanotubes could be measured. For instance, the carbon nanotubes from Experiment 2 had a diameter of approximately 74.56 nm, with the hollow core measuring 16.94 nm. Furthermore, the diffraction patterns could be seen form Figure.10(d), therefore the number of CNT layers could be determined by counting the observed layers in the images.

4.8 Simultaneous thermal analysis (STA)

Simultaneous Thermal Analysis was conducted under an 80% O_2 and 20% N_2 flow to determine the thermal stability and phase transitions of the CNTs for Experiments 1 and 2, as shown in Figure 8. The TGA curve indicates that the main decomposition reaction occurs around 500 °C and concludes at approximately 600 °C. The DSC curve further supports this observation, showing a sharp exothermic peak at around 500 °C, which corresponds to the oxidation and decomposition of carbon materials. After 600 °C, no significant thermal events are detected, indicating that the reaction has concluded. According to the literature, the mass loss prior to this decomposition temperature is attributed to the oxidation of amorphous carbon. The remaining mass (11%) of CNT observed after the TGA process appeared as a dark brown colour.[21],[22] This led to speculation about the presence of iron oxide or some other metal, therefore SEM-EDS was performed to further confirm it.

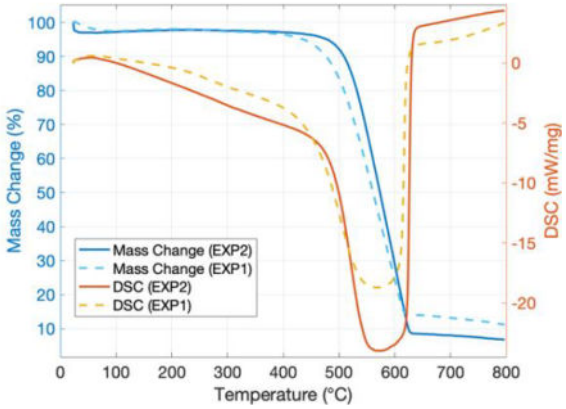


Figure 8. *Simultaneous thermal analysis (STA) of Experiment 2*

EDS characterization was performed on the samples from Experiments 1 and 2 after STA. From the EDS analysis shown in Figure 11, it was found that for Experiment 1, the primary metal substance remaining was iron oxide. According to the literature, iron oxide (Fe_2O_3) is known to act as a catalyst for CNT synthesis; therefore, it is suspected that the metal catalyst was the main substance that retained in the sample after burning.[23] For Experiment 2, the main components present were chromium (Cr), bromine (Br), iron (Fe), and iron oxide (Fe_2O_3). Since these were the only substances remaining after combustion, they are likely residues from the CNT formation catalysts. In this case, Cr and Br might also serve as potential catalysts, which could be explored in future experiments.

Furthermore, it is surprising that the sample for Experiment 2 retained a structure like carbon nanotubes even after combustion, as shown in Figure.12(b). This is likely due to the impurities being congested within the hollow spaces of the CNTs, which preserved the nanotube morphology even after the burning process.

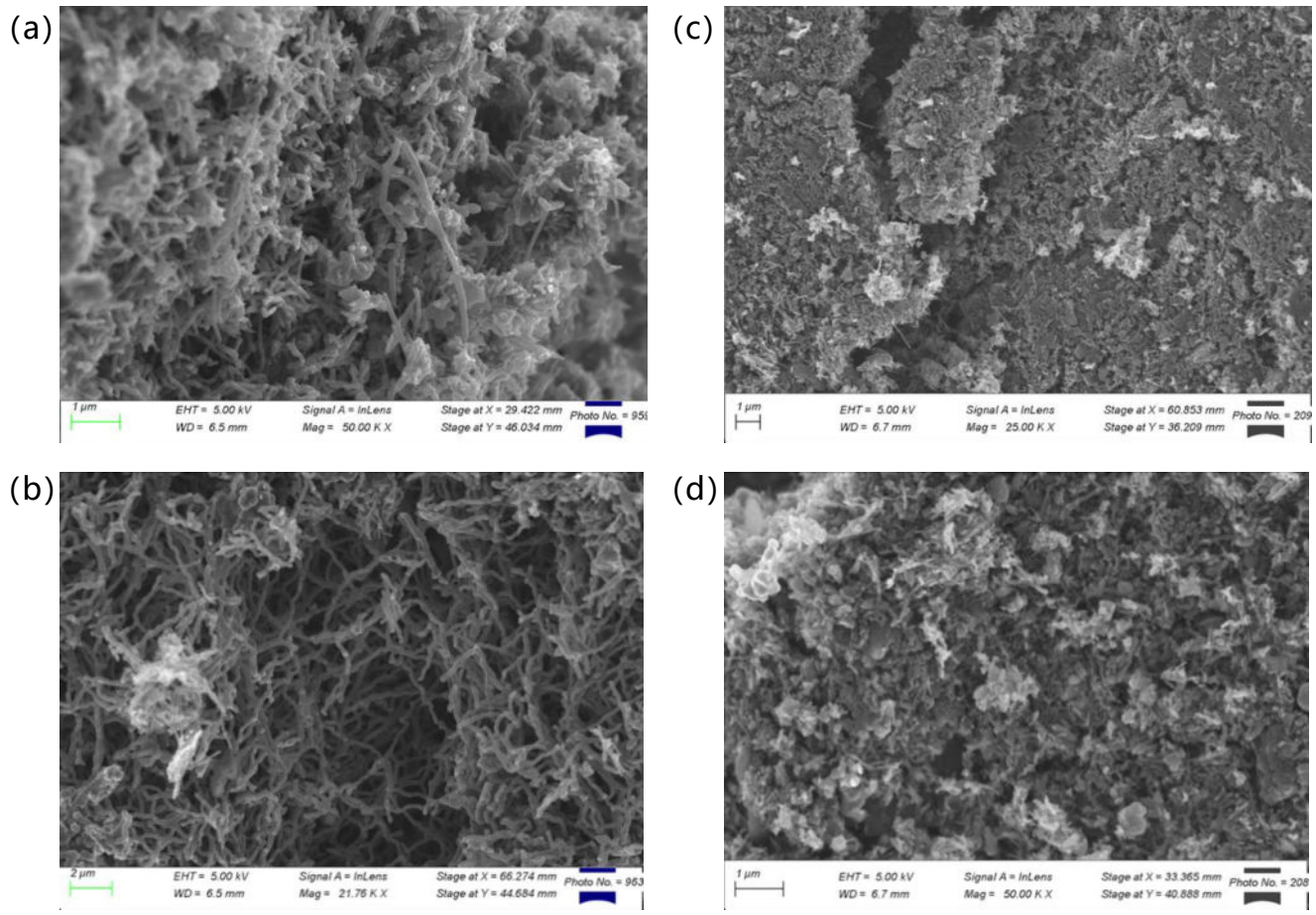


Figure 9. *Scanning electron microscopy (SEM) images of carbon products. (a) SEM image from Experiment 1 (b) SEM image from Experiment 2 (c) SEM image from Experiment 3 (d) SEM image from Experiment 5,. Scale bars indicate 1–2 µm as labeled.*

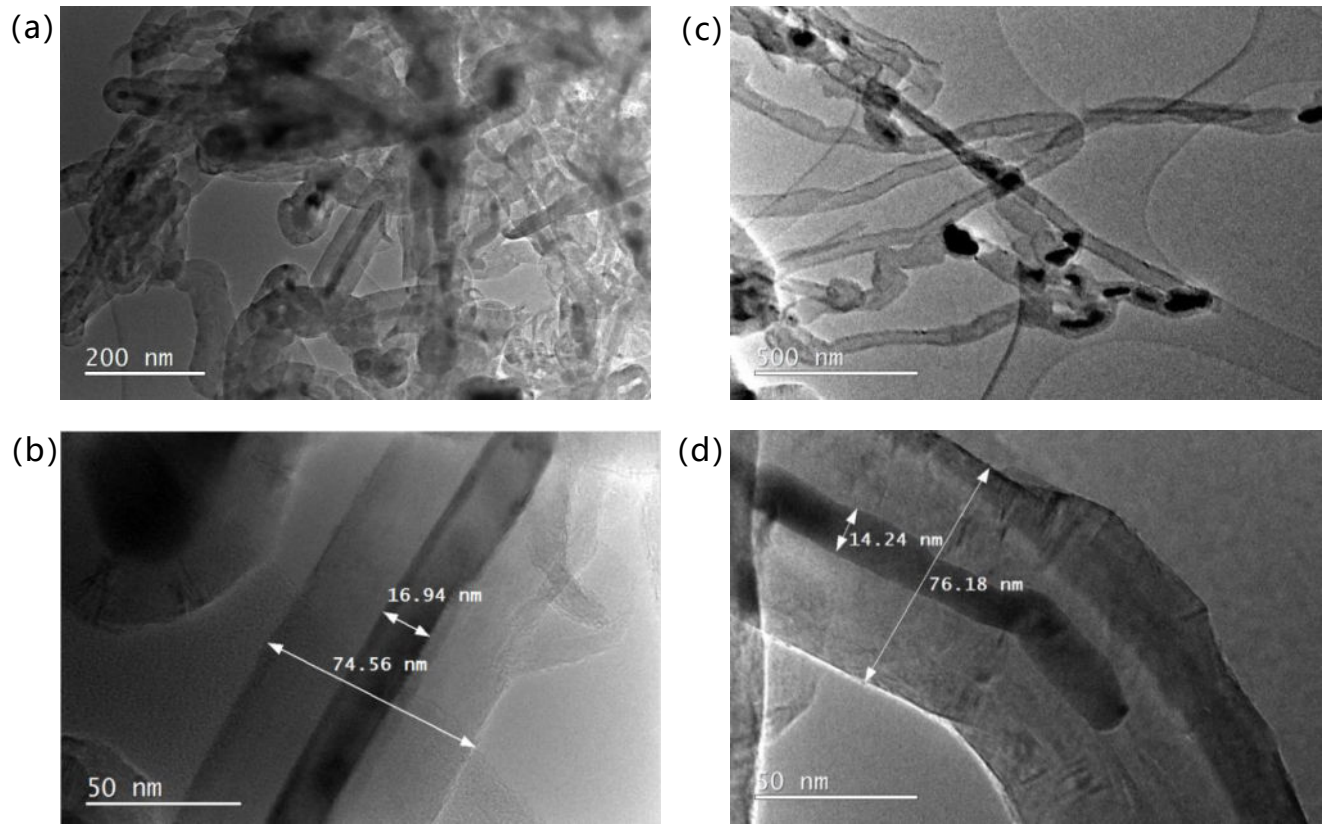


Figure 10. *Transmission electron microscopy (TEM) images of carbon products. (a) TEM image from Experiment 1 at a scale of 200 nm. (b) TEM image from Experiment 1 at a scale of 50 nm. (c) TEM image from Experiment 2 at a scale of 200 nm. (d) SEM image from Experiment 5 at a scale of 50 nm.*

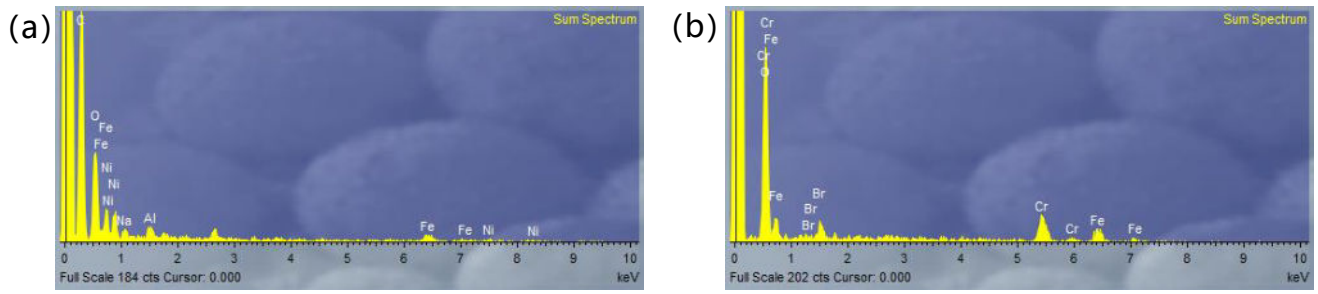


***Figure 11**. Energy Dispersive X-ray Analysis(EDS) of carbon products after STA. (a) SEM image from Experiment 1 (b) SEM image from Experiment 2*

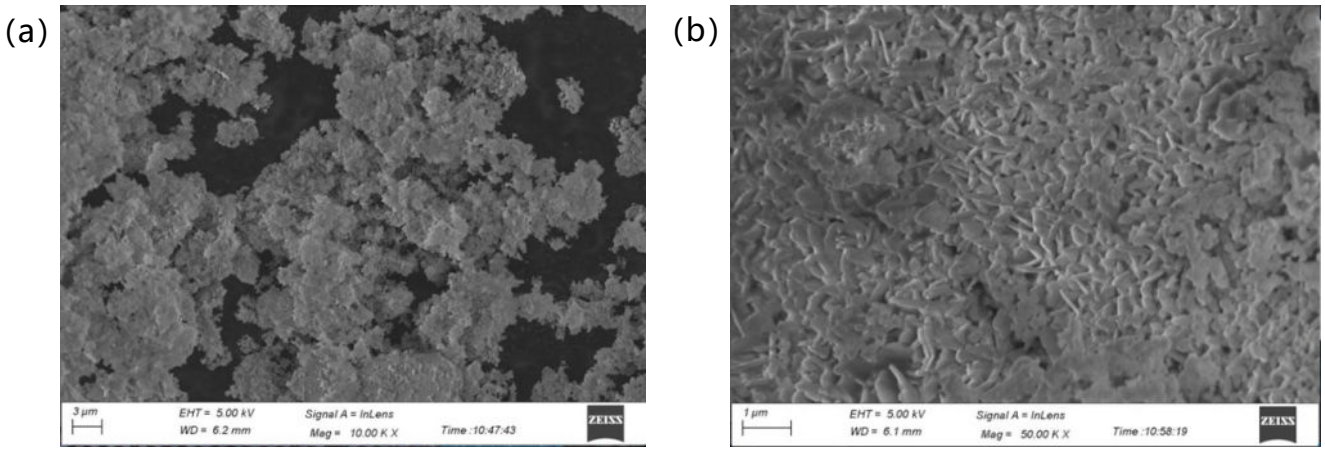


***Figure 12**. Scanning electron microscopy (SEM) images of carbon products after STA. (a) SEM image from Experiment 1 (b) SEM image from Experiment 2*

5. Conclusion

This study successfully demonstrated the production of carbon nanotubes (CNTs) from CO_2 through high-temperature electrolysis in a molten salt electrolyte. By optimizing key parameters, including electrode material, electrode dimensions, and electrolysis duration. we identified conditions that maximize both CNT yield and quality.

Our experiments indicated that larger electrode surfaces and longer electrolysis times would enhance faradaic efficiency. Additionally, thermoneutral potential and equilibrium potential graph modelling suggests the formation of CO gas, indicating the occurrence of side reactions that are critical for optimising faradic efficiency.

A range of characterization techniques were employed to assess the quality of the produced CNTs. Among the conditions tested, the use of stainless-steel electrodes resulted in the highest quality of CNTs.

6. Future work

Having demonstrated the feasibility of producing carbon nanotubes (CNTs) from CO_2 through high-temperature electrolysis in molten salt electrolytes, future work could involve conducting additional experiments to validate and reproduce the results. Due to equipment failure, the number of experiments in this study limited; thus, repeating the experiments would ensure reproducibility and reliability of the findings.

Following that, there are numerous opportunities for improving the electrolysis process. This could involve continuously bubbling CO_2 gas into the electrolyte to regenerate the consumed Li_2CO_3, thus enabling longer electrolysis time. Exploring larger electrode surface areas and extending the electrolysis duration could potentially lead to higher yields and improved product characteristics.

Finally, synthesized CNTs could be tested as electrode materials in lithium-ion batteries. This would provide valuable insights into their practical applications and performance in energy storage devices, opening new avenues for utilizing CNTs produced via this method in large-scale commercial technologies.

Acknowledgements

We would like to thank Zhixu Zhu for his invaluable help throughout this project.

References

1) NASA (2023) *Carbon dioxide concentration.* Available at: https://climate.nasa.gov/vital-signs/carbon-dioxide/ [Accessed 15th August 2023].
2) IPCC Chapter 4 (n.d.) *Global Warming of 1.5 oC.* Available at: https://www.ipcc.ch/sr15/faq/faq-chapter-4/ [Accessed 10th December 2024].
3) IEA (n.d.) *Carbon capture, utilisation and storage - energy system.* Available at: https://www.iea.org/energy-system/carbon-capture-utilisation-and-storage [Accessed 10th December 2024].
4) Patel, D.K. et al. (2020) *Carbon nanotubes-based nanomaterials and their agricultural and biotechnological applications*. Available at:

https://www.mdpi.com/1996-1944/13/7/1679 [Accessed 9th December 2024].
5) Polaris Market Research (n.d.) *Carbon nanotubes market size & growth analysis by 2032*. Available at: https://www.polarismarketresearch.com/industry-analysis/carbon-nanotubes-market [Accessed 9th December 2024].
6) [Imperial College London (n.d.) *Carbon nanotube production from CO2 via high temperature electrolysis*. Available at: https://www.imperial.ac.uk/events/167563/carbon-nanotube-production-from-co2-via-high-temperature-electrolysis/ [Accessed 10th December 2024].
7) Cassir, M., Licht, S., Frangini, S., Brun, C.L., Volkovich, V.A., Souček, P., Kwon, S.W., Griffiths, T.R., Liu, J., Li, H., Hao, W., Xu, X., Hemmes, K., Allen, J.A., Tulloch, J., Choi, S.H., Jiang, C., Yu, B., Li, C., Fan, L. (2016) Molten carbonates for advanced and sustainable energy applications: Part II. review of recent literature. Available at: https://www.sciencedirect.com/science/article/pii/S0360319916306073 [Accessed 9th December 2024].
8) Laasonen, E., Sorvali, M., Ruuskanen, V., Neimela, M. (n.d.) The effect of metal dissolution on carbon production by high-temperature molten salt electrolysis. Available at: https://www.researchgate.net/publication/364551493_The_Effect_of_Metal_Dissolution_on_Carbon_Production_by_High-Temperature_Molten_Salt_Electrolysis [Accessed 10th December 2024].
9) Ji, D., Zhang, F., Qiao, Z., Zhang, J., Wu, H., Wang, G. (2022) Electrochemical synthesis and structural characteristics of new carbon-based materials generated in molten salts. Available at: https://www.mdpi.com/2076-3417/12/19/9923 [Accessed 9th December 2024].
10) Elsevier (2021) Insights into carbon production by CO2 reduction in molten salt electrolysis in coaxial-type reactor. Available at: https://www.sciencedirect.com/science/article/pii/S2213343721019102?via%3Dihub [Accessed 10th December 2024].
11) Raji, K., Sobhan, C.B. (1970) Growth mechanisms in Carbon Nanotube Formation. Available at: https://link.springer.com/referenceworkentry/10.1007/978-3-030-91346-5_70 [Accessed 10th December 2024].
12) [ScienceDirect (n.d.) Faradic efficiency. Available at: https://www.sciencedirect.com/topics/engineering/faradic-efficiency [Accessed 10th December 2024].
13) ScienceDirect (n.d.) Equilibrium potential. Available at: https://www.sciencedirect.com/topics/engineering/equilibrium-potential [Accessed 10th December 2024].
14) Speight, J.G., Lange, N.A. (2017) Lange's Handbook of Chemistry. McGraw-Hill Education.
15) NIST Office of Data and Informatics (n.d.) Lithium carbonate. Available at: https://webbook.nist.gov/cgi/cbook.cgi?ID=C554132&Mask=2#Thermo-Condensed [Accessed 10th December 2024].
16) *ScienceDirect (n.d.) Gibbs Free Energy. Available at: https://www.sciencedirect.com/topics/engineering/gibbs-free-energy [Accessed 10th December 2024]*.
17) Laasonen, E., Sorvali, M., Ruuskanen, V., Niemelä, M., Koiranen, T., Ahola, J., Mäkelä, J.M., Joronen, T. (n.d.) The effect of metal dissolution on carbon production by high-temperature molten salt electrolysis. [Publisher information needed]. [Accessed 10th December 2024].
18) NIST (n.d.) Available at: https://janaf.nist.gov/tables/Ni-001.html [Accessed 10th December 2024].
19) Gong, W., Gu, B., Sirikulbodee, P., Steen, E.V., Cheng, Y., Chernyak, S.A., Bai, J., Zhao, X., Liu, R.-J., Xiong, H., Fang, Y., Deelen, T. van, Rossell, M.D., Egbosiuba, T.C., Bankole, M.T., Chai, S.-P., Kaushal, A., Mohammed, Ih.A., Nariyan, E., Venugopalan, A.T. (2022) Optimization of carbon nanotube growth via response surface methodology for Fischer-Tropsch synthesis over Fe/CNT Catalyst. Available at: https://www.sciencedirect.com/science/article/pii/S0920586122001602 [Accessed 9th December 2024].
20) [ResearchGate (n.d.) Structural quality of carbon nanotube samples. ID/IG band intensity ratios were. Available at: https://www.researchgate.net/figure/Structural-quality-of-carbon-nanotube-samples-ID-IG-band-intensity-ratios-were_fig3_221832801 [Accessed 10th December 2024].
21) [Konstantopoulos, G., Maroulas, P., Dragatogiannis, D.A., Koutsoumpis, S. (2020) The effect of interfacial resistance and crystallinity on heat transfer mechanism in carbon nanotube reinforced polyethylene. Available at: https://www.sciencedirect.com/science/article/pii/S0264127520309564 [Accessed 10th December 2024].
22) [Zhuo, C., Gong, J., Wal, R.L.V., Normand, F.L., Somanathan, T., Radhakrishnan, J.K., Sano, N., Karwa, M., Duy, D.Q., Baddour, C.E., Camilli, L., Herrera, J.E., Alvarez, W.E., Hordy, N., Li, Y., Karimzadeh, N., Cho, B., Freire, L., Veleva, L., Iijima, S. (2014) Oxidative heat treatment of 316L stainless steel for effective catalytic growth of carbon nanotubes. Available at: https://www.sciencedirect.com/science/article/pii/S0169433214012306?via%3Dihub [Accessed 10th December 2024].
23) *Journal of Environmental Chemical Engineering (n.d.) Insights into carbon production by CO2 reduction in molten salt electrolysis in coaxial-type reactor. Available at: https://www.sciencedirect.com/science/article/pii/S2213343721019102 [Accessed 10th December 2024]*.

Combined Effect of Water and Free Fatty Acids on Biodiesel Purity from Transesterification

Pranay Kantaria and Hadi Neseyif

Department of Chemical Engineering, Imperial College London, U.K.

Abstract

Biodiesel is commonly regarded as a sustainable substitute to traditional fossil fuels. However, the combined effect of free fatty acids (FFAs) and water on biodiesel purity from transesterification remains underreported. This project investigated these effects, both independently and in combination, on biodiesel synthesised from rapeseed oil using a sodium methoxide catalyst. Experiments were conducted by first varying the FFA and water content of rapeseed oil from 0 to 1% wt. individually, followed by factorial combinations of 0.2%, 0.5%, and 0.8% wt. for both impurities. In the individual case, product purity displayed enhanced sensitivity to FFA concentration, as predicted by literature, while in combination, it was water that had the greater impact. This was attributed to the mass transfer limitations of water in the oil phase, which were reduced in the presence of FFAs due to emulsion formation. These results emphasise the importance of considering multivariable effects and provide a foundation for further studies into emulsion formation and solubility.

Keywords: Biodiesel, Free Fatty Acids, Water Content, Mass Transfer, Emulsion.

1. Introduction

Mitigating the effects of climate change through the reduction of greenhouse gas emissions is one of the most pressing challenges of our time. Dependence on fossil fuels has long been recognised as a primary contributor to emissions. Consequently, extensive research has been committed to alternative fuels in recent decades.

An area of promising research is biodiesel, a mixture of fatty acid alkyl esters greater than 96.5%, produced through the transesterification of triglycerides from fats and oils (Eze et al., 2018). Biodiesel's significance lies in its chemical similarities with traditional diesel, which enable seamless integration into current energy infrastructure, such as diesel engines. This removes the need for expensive infrastructure modifications, which are the primary financial barrier in the transition away from fossil fuels (Huang et al., 2012). Biodiesel has definite advantages as a renewable alternative to traditional fuels. However, production faces several challenges, notably the issue of feedstock availability which conflicts with demand for the food industry (Koizumi, 2015). Biodiesel research has traditionally been conducted on high-yield feedstocks, such as soy or palm. However, these products are in high demand for food worldwide, resulting in conflict between food and energy needs. Research has shown that tying these needs together economically creates a dependency, where fluctuations in energy markets can create widespread food insecurity (Reza Miri et al., 2017).

Consequently, it is imperative to explore alternative feedstocks, such as non-edible vegetable oils or feedstocks that people rely on less for food, such as rapeseed oil. However, these feedstocks often display higher impurity levels when compared to those used in the food industry, as food-grade vegetable oils must meet rigorous standards (Reza Miri et al., 2017). Impurities, such as free fatty acids (FFAs) and water, are known to inhibit catalyst performance in transesterification, reducing the purity of biodiesel produced (Pullen &

Saeed, 2015). Despite this, the behaviour of FFAs and water, especially when present together, is not fully explored (Pullen & Saeed, 2015). Additionally, it is imperative to determine if feedstocks with these impurities can still be economically viable. This research paper first investigates the individual effect of water and FFA on biodiesel purity in section 4.1, before analysing the combined effect of both impurities in section 4.2. It then goes on to briefly discuss the effect of the mixing regime in section 4.3, before conducting an economic analysis on the cost of impurities in section 4.4.

2. Background

Transesterification is a chemical reaction where an ester exchanges an organic group with an alcohol. The reaction can proceed with simple reactant mixing; however, a strong acid or base catalyst is routinely used to increase the reaction rate. Research has evaluated base catalysts to be more efficient; hence, only they will be discussed going forward (Meher et al., 2006). The general reaction for transesterification is shown in Figure 1, where R groups represent hydrocarbon chains.



Figure 1: General Transesterification Reaction. Adapted from Meher et al. (2006).

Triglycerides are lipid molecules found in fats and oils that contain three fatty acid chains connected to a glycerol backbone by ester linkages. Biodiesel synthesis typically proceeds with the swapping of the glycerol backbone for the alkyl groups on three alcohol molecules. In the case of methanol, the alkyl group is a methyl group, resulting in the formation of three fatty acid methyl esters (FAMEs) and a glycerol byproduct. The fatty acid chains vary in length, and the resulting FAME mixture constitutes biodiesel. Triglyceride transesterification with methanol (methanolysis) is shown in Figure 2, where R groups represent the variable-length fatty acid chains.

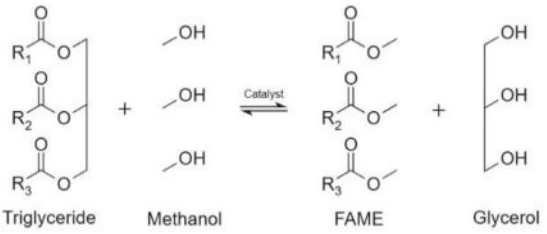


Figure 2: Methanolysis of Triglycerides. Adapted from Meher et al. (2006).

The reaction proceeds in a stepwise manner, with a triglyceride first being converted to a diglyceride, then to a monoglyceride, before a final conversion to glycerol. At each stage, this process consumes one molecule of methanol to produce one molecule of FAME. The reaction steps are described in Figure 3.

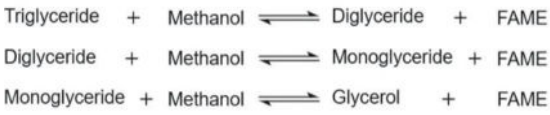


Figure 3: Reaction Steps for Methanolysis of Triglycerides. Adapted from Meher et al. (2006).

Each step of the methanolysis reaction proceeds via nucleophilic substitution. The initial step is the formation of methoxide ions from the base catalyst. This involves either the reaction of hydroxide ions with methanol when using hydroxide catalysts, or the direct dissociation of methoxide ions with methoxide catalysts (Jr et al., 2018). Methoxide formation is described in Figure 4.

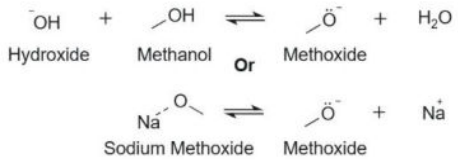


Figure 4: Methoxide Nucleophile Formation. Adapted from Pullen and Saeed (2015).

The generated methoxide ions first attack the carbon centre of the ester, breaking the CO double bond, forming a tetrahedral intermediate. The negatively charged oxygen on the intermediate then reforms the CO double bond, cleaving the bond with the glycerol backbone in the process, forming FAME and a deprotonated

diglyceride. The deprotonated diglyceride then scavenges a proton from a molecule of methanol, regenerating the methoxide nucleophile. This mechanism for the first stage of methanolysis is shown in Figure 5.

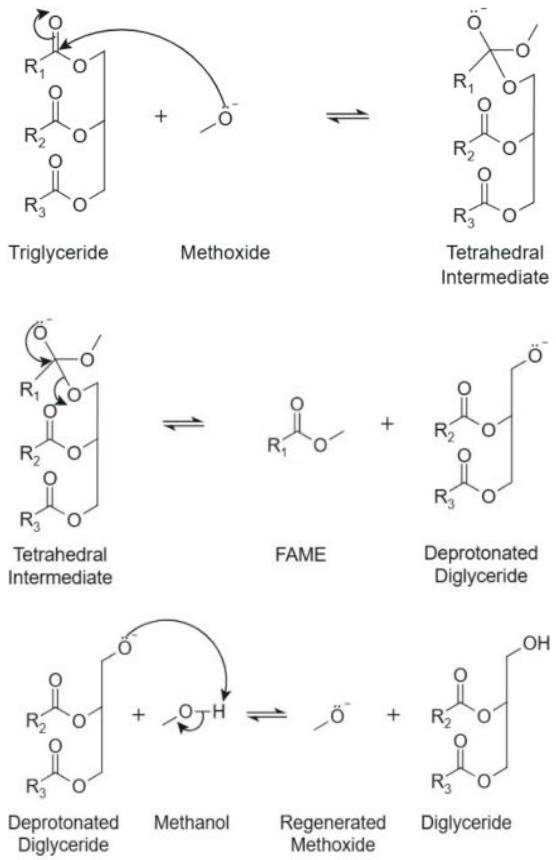


Figure 5: Reaction Mechanism for Methanolysis of Triglycerides. Adapted from Jr et al. (2018).

However, several side reactions can also occur during biodiesel synthesis, particularly if FFAs or water are present. A common side reaction is saponification – the reaction of FFA with the base catalyst to produce soap. Saponification is detrimental to biodiesel synthesis, as it consumes the catalyst so has the potential to slow the reaction significantly if the FFA concentration is large. Furthermore, saponification produces a layer of soap which is partially miscible with both the product layer and the by-product layer, making downstream processing difficult (Mandari & Devarai, 2022). The saponification reaction is shown in Figure 6.

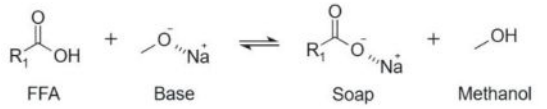


Figure 6: Saponification Reaction. Adapted from Pullen and Saeed (2015).

Another side reaction is the hydrolysis of triglycerides to fatty acids, which can occur when water is present in the system. Triglycerides are broken down into FFAs and diglycerides, as shown in Figure 7. The generated FFAs can then undergo saponification.

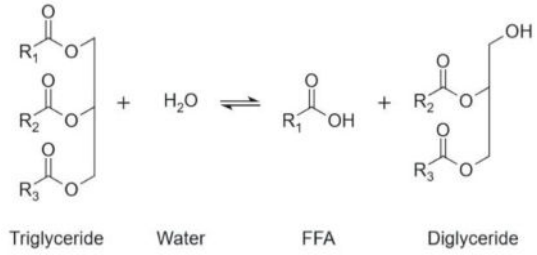


Figure 7: Hydrolysis of Triglycerides. Adapted from Pullen and Saeed (2015).

Water can also result in the hydrolysis of the FAME product back into FFAs and methanol. This typically isn't an issue at the synthesis stage but can cause challenges in long-term biodiesel storage.

Saponification and hydrolysis are vital to understand, as the need for non-edible feedstocks makes the use of feedstocks with higher levels of impurities necessary. Upstream processing is possible, but costly so determining if these feedstocks are viable for biodiesel synthesis is important.

Furthermore, both saponification and hydrolysis are significant reactions, even for feedstocks that contain very little FFA or water. This is because, when common alkali catalysts, such as sodium hydroxide, are used, they produce water by the mechanism shown in Figure 4. This generated water can then undergo hydrolysis to generate FFAs for saponification. Typically, in industrial production, these alkali catalysts are used due to their low cost and availability, so both side reactions are always relevant for biodiesel synthesis (Mandari & Devarai, 2022).

Optimising biodiesel synthesis has been the focus of extensive research looking at a range of variables, including catalysts, temperature, reaction time, and mixing rate. Exploratory work into kinetics by Freedman et al. (1986) investigated the effect of catalyst type and the molar ratio of methanol to triglyceride in attempts to fit kinetic models. The study found that sodium methoxide and a molar ratio of 6:1 of oil to methanol were sufficient to produce high-purity FAMEs at low temperatures, hence these conditions were used in this study. Since then, research has looked at optimising different parts of the process, from reaction conditions to specific feedstocks and catalysts. The individual effects of FFA and water content have also been the subject of several reports, such as those by Meher et al. (2006) and Pullen and Saeed (2015).

Despite this, studies that focus on the combined interaction of water and FFA content – particularly when investigated with a methoxide catalyst (which does not produce its own water during the reaction) – are underreported. Therefore, the focus of this report was an investigation into the combined effect of water and FFA impurities with a methoxide catalyst.

3. Methodology

3.1. Materials

Table 1: Materials Used in Study.

Chemical	Purity	Supplier
Rapeseed Oil	99.99%	Oillia
Oleic Acid	>99.00%	Sigma Aldrich
Distilled Water	99.99%	Imperial College London
Sodium Methoxide	100.00%	Acros Organics
Chloroform	99.80%	Sigma Aldrich
Methanol	>99.99%	Sigma Aldrich

3.2. Experimental Method

For preliminary experimentation, a methodology based on Pullen and Saeed (2015) was devised and adapted. First, 20.000g rapeseed oil was weighed into three 200ml beakers. To one of the beakers, 0.200g DI water was added to form a 1% wt. water solution. To another of the beakers, 0.200g of oleic acid was added to form a 1% wt. FFA solution. The third beaker was left untouched to act as a control. To prepare the catalyst, 0.400g sodium methoxide was dissolved in 4.424g methanol to produce a solution that would contain 2% wt. dry catalyst to oil, while maintaining the 6:1 methanol to oil molar ratio mentioned previously. Three such catalyst solutions were prepared then added to each of the prepared oil samples. The solutions were then mixed using magnetic stirrers at 800rpm at 60°C on hotplates for 30 minutes. After the time had elapsed, the solutions were poured into separating funnels and left to settle overnight.

After separation, a sample of the top layer of biodiesel was taken and prepared for NMR. From the sample, 15-25mg was weighed out into microcentrifuge tubes before 0.7ml of chloroform was added by pipette. The solution was mixed, then glass pipetted into NMR tubes for analysis. The preliminary round of experiments was conducted to test and adjust the experimental method, as well as determine the upper bounds of FFA and water concentration.

Due to logistical challenges, the experiment was then repeated using centrifuge tubes for the separating step, and the results compared to those using glass separating funnels. As no noticeable difference between the results could be detected, future rounds of experimentation were conducted using this method.

For the first round of experiments, samples ranging from 0-1% wt. FFA and 0-1% wt. water were prepared using 0.2% wt. increments. A batch solution of 12.000g sodium methoxide and 132.707g methanol was mixed up to reduce inaccuracies from

independently mixing the catalyst solution for each sample. To each, 4.824g of the batch catalyst was added to again give 2% wt. dry catalyst to oil. The mixing, settling and NMR methods proceeded using the same method as prior. This was then repeated for another round of oil samples to increase the amount of data to assist in identifying any outliers and anomalies.

For the second round of experiments, samples containing water and FFA together were prepared. The samples were prepared by varying both FFA and water concentration using a factorial design for 0.2%, 0.5%, and 0.8% wt. resulting in 9 samples total. Another batch of catalyst was prepared and again 4.824g of this solution was added to each oil sample and the mixing, settling and NMR methods were conducted as prior. The experimental method was repeated for another round of samples to increase the amount of data for analysis.

It was observed that during the mixing, some of the stirring bars in the samples behaved in a chaotic fashion resulting in an unstable mixing regime. It was uncertain how this affected the reaction so it was decided that the samples where this was observed should be repeated twice more to determine the effect of the chaotic mixing.

3.3. Analytical Method

Figure 8 shows the different proton environments of the reactants, intermediates, and products, highlighting the methoxy group in green and the α-carbonyl groups in red. The first carbon in the fatty acid chain has been shown for clarity. Additional proton environments for triglycerides, diglycerides and monoglycerides are also shown in blue.

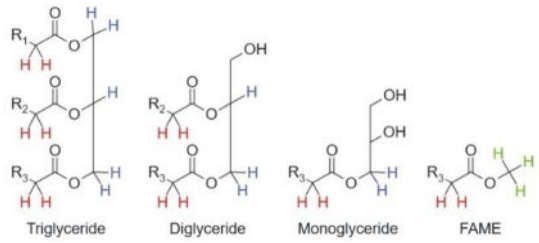


Figure 8: Hydrogen Environments in Reactants, Intermediates and Products.

Each of these proton environments produce peaks in an NMR spectrum. The proton NMR spectra from the experiments were analysed using *MestreNova* software, with calibration performed by comparing the chloroform peak at 7.26 ppm in the spectrum to the software's standard reference value. The key peaks were identified, with the methoxy group located 3.7 ppm and the α-carbonyl group located 2.3 ppm, as shown in Figure 9.

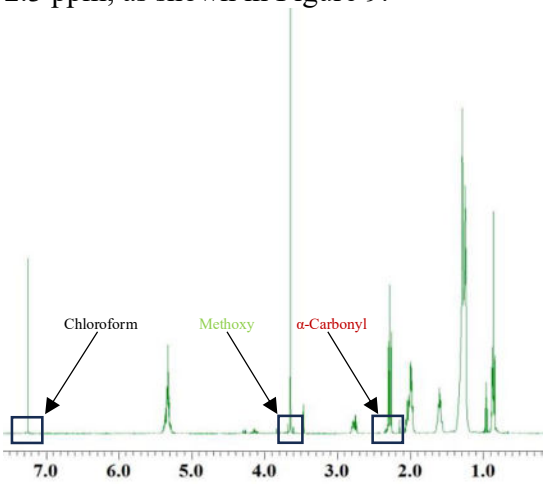


Figure 9: Proton NMR Spectrum Showing Relevant Peaks (Units of ppm).

These peaks were integrated with a consistent width of 15 Hz on either side of the peak's maximum. It can be noted from Figure 8 that only FAME contains the methoxy group and that the ratio of protons between the methoxy and α-carbonyl groups in FAME is 3:2. Hence, the purity of the FAME in the product was determined by calculating the ratio of the areas of the methoxy peak to the α-carbonyl peak on the NMR spectra, normalised by the number of hydrogens, and calculating the deviation from the ideal 3:2 ratio. This calculation is summarised in Equation 1.

$$FAME\ Purity\ (\%) = \frac{2 \times \int normalised\ 3.7\ \text{ppm}}{3 \times \int 2.3\ ppm} \times 100$$

Equation 1

From this method, the FAME purity of the biodiesel product was determined for each of the oil samples, and an average FAME purity for each mixture was determined.

3.4. Economic Analysis

The economic analysis was conducted to assess the impact of feedstock impurities on the biodiesel production process. The total costs of the process were estimated using 2024 market prices for the reactants, products, and utilities required for the reaction. The quantities of reactants and products were determined based on molar ratios provided in Figures 2, 6, and 7. Starting with an initial feedstock of 1000 tonnes of rapeseed oil, the amounts of biodiesel and glycerol produced were calculated based on stoichiometry and with experimental purity data. Utility and labour costs for operating a plant of this capacity were estimated and incorporated into a detailed profit and loss analysis where biodiesel and glycerol were sold as products (Leung et al., 2010).

Three calculations were computed: one using pure rapeseed oil to establish baseline yields, another using the price of used oil with FFA impurities from restaurants, and a third using the same used oil but accounting for yield loss due to water impurities instead of FFA. These analyses quantified the effects of impurities on biodiesel production. The 2024 market prices used are shown in Table 2.

Table 2: Prices of Chemicals (ECHEMI, 2024).

Chemical	Price $/ MT
Rapeseed Oil	800
Used Oil	272
Methanol	343
Sodium Methoxide	2500
Biodiesel	1300
Glycerol	908

4. Results and Discussion

4.1. Individual Effect of Impurities

The initial stage of the experiment focused on assessing the impact of changing the concentrations of FFA (oleic acid) and water on biodiesel purity. The data from the first round of experiments is displayed in Figure 10.

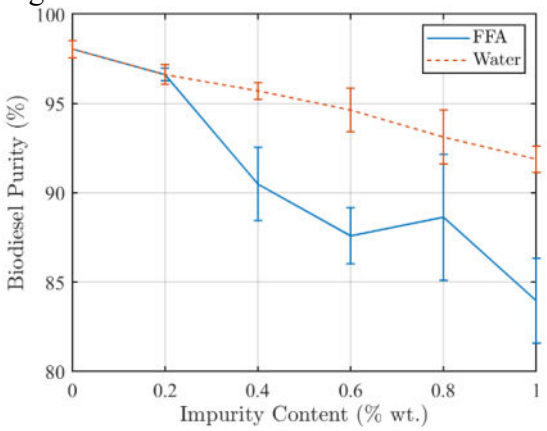


Figure 10: Effect of Impurity Concentration on Biodiesel Purity.

The data shows that increasing the concentration of either FFA or water decreased biodiesel purity, with FFA having a greater impact than water at the same concentrations. This trend was expected based on research by Meher et al. (2006), Pullen and Saeed (2015), and Eze et al. (2018).

As expected, the greatest product purity of 98% was achieved when no additional water or FFA was present. The inability to achieve 100% purity arose from the fact that even the control oil sample contained impurities, which could undergo the saponification and hydrolysis reactions.

The addition of FFA resulted in biodiesel purity decreasing due to saponification. This was because the catalyst was consumed by reaction with the FFA to produce soap and methanol, which reduced catalyst availability for transesterification. Consequently, the concentration of reaction intermediates, such as diglycerides and monoglycerides in the product increased, reducing biodiesel purity (Ma & Hanna, 1999). At higher concentrations, this effect became more noticeable, with the purity dropping to 84% at 1% wt. FFA. Furthermore, FFA created emulsions between glycerol and biodiesel layers, which would complicate separation during downstream processing (Gerpen, 2005).

In contrast, varying the water concentration had a less significant impact on biodiesel purity. Water caused hydrolysis of triglycerides, generating FFAs and glycerol, as described in Figure 7. This reduced the availability of triglycerides for transesterification, while the resulting FFAs further consumed catalyst through saponification. Additionally, water could have reacted with sodium methoxide, forming an equilibrium with sodium hydroxide and methanol, as shown in Figure 4. Sodium hydroxide is a less effective catalyst and more prone to saponification, which could also be responsible for the reduced biodiesel purity (Eze et al., 2018).

The greater impact of FFA compared to water was attributed to mass transfer limitations. Triglycerides are non-polar, while water is highly polar, resulting in two immiscible phases: one phase containing the triglycerides, and the other containing the water. Due to water's poor solubility in the oil phase, diffusion across the phase boundary was low. Therefore, hydrolysis primarily occurred at the phase interface, significantly limiting the reaction rate. In contrast, FFAs, with their long fatty acid chains, were miscible with triglycerides and interacted freely with the methoxide catalyst, allowing saponification to occur readily.

This can also be seen on the NMR spectra where water produced a sharp peak at 2.14ppm as displayed in Figure 11.

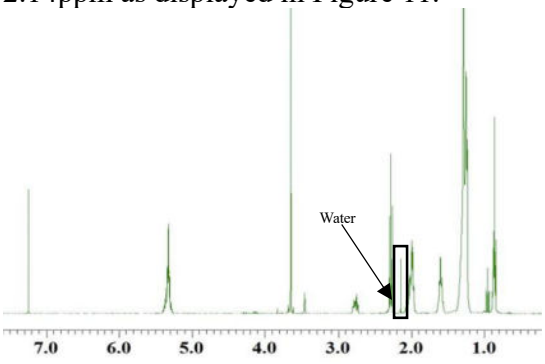


Figure 11: NMR Spectrum Showing Water Peak (Units of ppm).

This peak was present in all of the samples from the first round of experimentation, even when no water was added – likely due to water in the starting oil or in the chloroform used to prepare the NMR samples. However, the peak grew noticeably when the water content increased, indicating a higher level of water in the biodiesel. The presence of water at all indicated that the rate of hydrolysis was much slower than transesterification, since the water was unable to react with the triglycerides. This again indicates the mass transfer limitations of the hydrolysis reaction.

It is also significant for downstream processing, as the water must be removed before long-term storage, since water is known to break FAME back down into FFAs and methanol over prolonged time periods.

4.2. Combined Effect of Water and FFA

The second part of the experiment investigated the combined effect of water and FFA impurities on the biodiesel purity. The results are displayed in Figure 12.

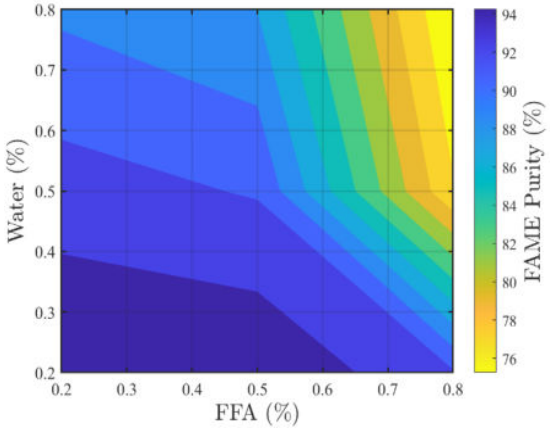


Figure 12: Combined Effect of FFA and Water Impurities on Biodiesel Purity.

The data in Figure 12 shows that, under the combined effect of both water and FFA, the system was more sensitive to water. This was evidenced by the fact that, at 0.5% wt. water and 0.2% wt. FFA, the biodiesel purity was 93.24%, while at 0.2% wt. water and 0.5% wt. FFA, the biodiesel was significantly purer at 95.92%. Similarly, at 0.2% wt. water and 0.8% wt. FFA, the purity was still 92.59%, while for 0.8% wt. water and 0.2% wt. FFA, the purity dropped

to 90.09%. This contrasted the results from the individual experiments where FFA concentration was more significant.

This can be explained by the properties of the soap produced from saponification. It can be seen in Figure 6 that surfactant molecules have a polar head containing oxygen atoms, in addition to long, non-polar fatty acid tails. Such molecules are termed amphiphilic, meaning they have an affinity to both polar solvents and non-polar solvents. In the context of the experiment, this allowed them to distribute between the oil and methanol phases, creating a layer between the two solvents known as an emulsifier. Emulsifiers reduce interfacial surface tension, stabilising droplets of one phase within the other, and preventing coalescence.

During the reaction, the soap produced from the saponification of FFA allowed a stable emulsion to form between the polar and non-polar phases, enabling more effective mixing. When only water was present, mass transfer across phases limited the rate of hydrolysis due to water's poor solubility in the oil phase. However, when FFA was added, an emulsion formed, improving the mass transfer efficiency of the process, allowing water to pass into the oil phase with less restriction and undergo hydrolysis with the triglycerides. Consequently, when both FFA and water were freely able to react, it was the water that had the greater impact likely due to the combined effect of triglyceride consumption and FFA formation.

Additionally, the water peaks in the NMR spectra described in Figure 11 were not visible for any of the mixed samples from the second round of experimentation. This further indicates that the rate of hydrolysis was greater when FFA was present, allowing all the water in the oil layer to react. Consequently, no water was left in the biodiesel, resulting in the absence of the NMR peak. This is significant for biodiesel storage as it indicates that a little FFA in the feedstock may actually be preferable to ensure all of the water reacts, removing the need for additional processing to remove moisture for long term storage.

4.3. Mixing Results

In total, five of the samples displayed chaotic mixing rather than the smooth mixing regime that stirring bars usually exhibit. The reason for this was likely due to a loss in magnetism of those specific stirring bars. However the chaotic mixing could be attributed to the changing viscosity as the reaction proceeded since triglycerides have a higher viscosity than the FAME product, due to their long fatty acid chains. The five experiments were repeated twice more, and the average biodiesel purity was determined for both types of stirring. The results are displayed in Table 3.

Table 3: Mixing Regime Results.

Sample	Chaotic Mixing	Normal Mixing
0.4% FFA	88.99	93.45
0.6% FFA	84.46	70.11
1% FFA	79.20	88.69
0.8% FFA, 0.5% Water	88.49	83.33
0.8% FFA, 0.8% Water	82.01	70.11

Of the five samples tested, two displayed improved biodiesel purity while three showed reduced purity. This was due to the chaotic nature of the mixing – in some cases, it created a more optimal mixing regime for the reaction, while in others it hindered the reaction or promoted hydrolysis or saponification. Consequently, the exact impact of the different mixing was difficult to determine and would require additional research to draw meaningful conclusions.

4.4. Economic Results

The data from the economic analysis of impurities is displayed in Table 4.

Table 4: Economic Analysis with 1000 MT of Feedstock.

Scenario	Pure Oil ($)	Oil and FFA ($)	Oil and Water ($)
Revenue	1,366,817	1,157,610	1,297,082
Cost	1,005,460	477,199	477,199
Profit	361,357	680,411	819,883

These results highlight the impact of feedstock impurities on biodiesel production. Pure feedstock produced the purest biodiesel and the highest revenue but was less economically viable due to its high cost. Impure feedstock, such as the oil demonstrated in the experimental method, was cheaper but yielded biodiesel with lower purity.

FFA impurities had a greater negative impact on revenue and profitability than water impurities, emphasising the importance of pre-treatment to remove FFAs. The findings are significant as in reality, used oil conversion would be lower due to higher impurity levels and therefore cause much lower revenues. Overall, the table shows how impure feedstocks can still be economically viable, however even small impurities can greatly affect biodiesel yield and profitability. A key assumption in this analysis was that separation costs were not included, which would reduce the profitability of the impure products further.

5. Conclusions and Outlook

The focus of this research was to investigate the combined effect of FFA and water on biodiesel purity. The findings showed that when FFA and water were tested individually, FFA had a more significant impact on biodiesel purity, as discussed in literature. However, when both impurities were present, water had the more significant effect on biodiesel purity. This was due to emulsion formation which reduced the mass transfer limitations of water in the oil phase.

This conclusion highlights the increasing importance of removing water from feedstocks, a conclusion often overlooked by literature which cites FFAs as the most significant impurity (Pullen & Saeed, 2015). Despite this, the economic analysis shows how using impure feedstocks can still be profitable which is a welcome conclusion given the need for biodiesel production to move away from edible feedstocks.

Additionally, the absence of water in the biodiesel product when FFA was present indicates that some FFA impurity in the feedstock may actually be beneficial. Since water has a hydrolysing effect on FAME in long-term storage, a reaction that produces biodiesel with no water is economically beneficial as it removes the need for additional drying stages.

The study was unfortunately limited by time constraints and resources however the preliminary findings were promising. Going forward, future studies should look to expand the data set of the combined impurity experiments by looking at a wider range of process conditions, feedstocks and catalysts to determine if the results of this study are reproducible. Additionally, investigating the effect of variables such as FFA and catalyst type on emulsion formation and attempting to model the effect of emulsions on reaction kinetics would be a promising avenue of research.

The specific effect of temperature on the solubility of water in the oil phase is also under-researched. For example, Leung et al. (2010) pointed out that heating the oil prior to the mixing can increase the reaction rate of transesterification however the effect on hydrolysis is unknown.

Finally, it is suggested to further investigate the effect of the mixing regime on biodiesel purity, especially focusing on how viscosity changes as the reaction progresses and how this can affect the mixing regime.

References

ECHEMI. (2024). *ECHEMI Price Database* https://www.echemi.com/price-database.html

Eze, V. C., Phan, A. N., & Harvey, A. P. (2018). Intensified one-step biodiesel production from high water and free fatty acid waste cooking oils. *Fuel*, *220*, 567-574. https://doi.org/10.1016/j.fuel.2018.02.050

Freedman, B., Butterfield, R. O., & Pryde, E. H. (1986). Transesterification kinetics of soybean oil 1. *Journal of the American Oil Chemists' Society*, *63*(10), 1375-1380. https://doi.org/10.1007/BF02679606

Gerpen, J. V. (2005). Biodiesel processing and production. *Fuel Processing Technology*, *86*(10), 1097-1107. https://doi.org/10.1016/j.fuproc.2004.11.005

Huang, D., Zhou, H., & Lin, L. (2012). Biodiesel: an Alternative to Conventional Fuel. *Energy Procedia*, *16*, 1874-1885. https://doi.org/10.1016/j.egypro.2012.01.287

Jr, J., Mendonça, A., Campos, D., Mapele, R., Barra, C., Bauerfeldt, G., & Tubino, M. (2018). Biodiesel Synthesis: Influence of Alkaline Catalysts in Methanol-Oil Dispersion. *Journal of the Brazilian Chemical Society*, *30*. https://doi.org/10.21577/0103-5053.20180183

Koizumi, T. (2015). Biofuels and food security. *Renewable and Sustainable Energy Reviews*, *52*, 829-841. https://doi.org/10.1016/j.rser.2015.06.041

Leung, D. Y. C., Wu, X., & Leung, M. K. H. (2010). A review on biodiesel production using catalyzed transesterification. *Applied Energy*, *87*(4), 1083-1095. https://doi.org/10.1016/j.apenergy.2009.10.006

Ma, F., & Hanna, M. A. (1999). Biodiesel Production: a Review. *Bioresource Technology*, *70*(1), 1-15. https://doi.org/10.1016/S0960-8524(99)00025-5

Mandari, V., & Devarai, S. K. (2022). Biodiesel Production Using Homogeneous, Heterogeneous, and Enzyme Catalysts via Transesterification and Esterification Reactions: a Critical Review. *BioEnergy Research*, *15*(2), 935-961. https://doi.org/10.1007/s12155-021-10333-w

Meher, L. C., Vidya Sagar, D., & Naik, S. N. (2006). Technical aspects of biodiesel production by transesterification—a review. *Renewable and Sustainable Energy Reviews*, *10*(3), 248-268. https://doi.org/10.1016/j.rser.2004.09.002

Pullen, J., & Saeed, K. (2015). Investigation of the factors affecting the progress of base-catalyzed transesterification of rapeseed oil to biodiesel FAME. *Fuel Processing Technology*, *130*, 127-135. https://doi.org/10.1016/j.fuproc.2014.09.013

Reza Miri, S. M., Mousavi Seyedi, S. R., & Ghobadian, B. (2017). Effects of biodiesel fuel synthesized from non-edible rapeseed oil on performance and emission variables of diesel engines. *Journal of Cleaner Production*, *142*, 3798-3808. https://doi.org/10.1016/j.jclepro.2016.10.082

Haemodynamic Analysis in Abdominal Aortic Aneurysms using Computational Fluid Dynamics

Danyar Arif and Kelvin Zhou
Department of Chemical Engineering, Imperial College London, U.K.

Abstract Using computational fluid dynamics (CFD), this study examines the haemodynamic properties and the impact of key modelling assumptions within abdominal aortic aneurysms (AAAs). ANSYS was used to create and analyse AAA models with a patient-specific geometry. The study included a mesh sensitivity analysis, comparison of Newtonian and non-Newtonian blood models, and the effect of different inlet boundary conditions, which were instrumental in conducting this study. With significant variations in wall shear stress (WSS) and velocity profiles amongst models, the results showed that realistic boundary conditions and finer mesh densities drastically altered haemodynamic predictions. These results imply that more sophisticated CFD models offer important insights into AAA behaviour, potentially leading to more precise risk stratification than traditional methods and CFD models currently used in medical practice.

Introduction

An abdominal aortic aneurysm (AAA) is a life-threatening condition that affects millions of people worldwide. It remains asymptomatic during most of its development but ruptures can lead to fatal consequences. In the U.S alone, AAA affects an estimated 1.1 million people[1] and rupture is responsible for over 5,000 deaths annually. Tragically, nearly half of the patients with a ruptured AAA do not survive before reaching the hospital[2], with an overall mortality rate of 90%[3].

Current clinical practice primarily uses aneurysm diameter to assess rupture risk and typically surgical intervention would occur when the diameter of the aneurysm had reached a mean average of 58.3 mm[4]. However, relying on diameter alone is often unreliable, as some aneurysms rupture at smaller diameters, while others do not rupture even when larger[5]. This highlights the need for more precise risk assessment methods that go beyond size.

Recent studies suggest that haemodynamics within the aneurysm, specifically wall shear stress (WSS) and flow dynamics, play a critical role in aortic expansion and rupture risk[6][7][8][9]. Computational fluid dynamics (CFD) has emerged as a powerful tool for simulating blood flow and analysing these haemodynamic factors in patient-specific AAA models and their impact on arterial wall stability[10]. This study focuses on examining the impact of key modelling assumptions commonly used in computational analysis of AAAs. Factors such as different viscosity models, inlet velocity profiles and rotational flow affect the simulations and therefore will have an impact on hemodynamic results. By addressing these assumptions, this work aims to provide valuable insights into their influence on simulated outcomes and improve the reliability of future studies, enabling researchers to design more accurate computational models for assessing AAA behaviour. Such models could be hugely valuable in predicting disease progression, independent of traditional clinical predictors like aneurysm diameter.

The CFD simulations were conducted using a patient-specific AAA geometry, and key metrics such as velocity contours and WSS will be analysed. The results, combined with clinical data from AAA patients, will provide guidance on the specific assumptions required to ensure physiologically accurate CFD models.

Background

In current reports of using CFD to simulate AAAs, it is a common assumption that blood is a Newtonian fluid (typically of constant viscosity of 0.0035 Pas)[11], regardless of the shear rate. However, it is known that blood is not a Newtonian fluid[12], and has shear thinning properties[13], as opposed to viscoelasticity and yield stress.

The assumption for blood viscosity was investigated by testing whether using different models had significant implications for CFD simulations from oversimplifying the flow behaviour and failing to capture critical haemodynamic variations. Non-Newtonian models, such as the Carreau and Power Law models, have been developed to better represent the shear-thinning characteristics of blood. Although rare, studies have shown that these non-Newtonian models provide significantly different predictions of WSS and flow dynamics in AAAs, whereas parameters such as time-averaged wall shear stress (TAWSS), and oscillatory shear index (OSI) showed similarities to the Newtonian model[14]. By capturing these complex flow behaviours, these models could potentially improve the accuracy of rupture risk assessments compared to Newtonian approximations.

Furthermore, most CFD studies on AAAs still rely on the assumptions of flat inlet velocity profiles developing into parabolic flow despite the fact blood flow is typically parabolic, leading to inaccurate predictions. Previous studies have highlighted the importance of using more realistic boundary conditions and non-Newtonian models, however, many of these studies have focused on a single set of parameters for non-Newtonian models or have not compared the variability in results caused by different velocity profiles and non-Newtonian models.

Although CFD simulations are incredibly insightful, there are lots of potential issues that may arise. Due to the fact that they require vast amounts of computational power and time, issues that may arise when trying to further navigate models include convergence issues, high resolution meshes, uncertainty in turbulence modelling and more[15].

Given the limitations of current CFD practices, this study aims to bridge the gap by investigating the role of non-Newtonian haemodynamics in to provide insight on more accurate models. Specifically, it explores how swirling flow components and non-Newtonian models affect flow patterns and compare the predictive accuracy of different non-Newtonian models. However, in this study only models that exhibit shear thinning properties, the Quemada and Carreau-Yasuda models[16] were chosen.

Methodology

This section details the CFD methods used to analyse blood flow in AAAs. All simulations presented used ANSYS CFX as the courier, using a unique hybrid finite-element/finite-volume method to obtain solutions.

Blood is frequently treated as a continuous medium in computational models, despite its complex makeup of cells, plasma, and proteins. At the scale of these simulations, typically millimetre to centimetre ranges, the influence of individual blood cells, becomes negligible. Therefore, it was reasonable to assume that blood behaved as a continuous fluid when modelling overall flow behaviour in larger arteries. This assumption was particularly useful for solving the Navier-Stokes equations, which relied on the continuum hypothesis and provided a practical way to model blood flow dynamics. Through the use of the Navier-Stokes equations, CFD can model complex interactions between blood flow and the artery walls, offering detailed insights into flow patterns and WSS throughout the aortic artery[17].

$$Continuity\ equation:$$
$$\nabla \cdot \boldsymbol{u} = 0 \quad \text{(Eq.1)}$$

$$Momentum\ Equation:$$
$$\rho\left(\frac{\partial \boldsymbol{u}}{\partial t} + \boldsymbol{u} \cdot \boldsymbol{\nabla u}\right) = -\boldsymbol{\nabla} \mathrm{p} + \nabla \cdot \boldsymbol{\tau} + \rho \mathbf{F} \quad \text{(Eq.2)}$$

where ρ is the density, μ is the blood viscosity, t is the time, $\boldsymbol{u}$ is the fluid velocity vector, p is the pressure, $\boldsymbol{\tau}$ is the viscous stress tensor and $\mathbf{F}$ is the force vector.

Geometry

The AAA geometry used throughout this research project was created from CT scans from a single patient with AAA from St. Mary's Hospital. The model inlet was cut proximal to the abdominal branches, removing the healthy ascending aorta, aortic arch, and proximal thoracic descending aorta for computational efficiency. The geometry then extended through the aneurysm region and beyond the first bifurcation point, resulting in six outlet arteries: the celiac artery, superior mesenteric artery, right and left renal arteries, and right and left iliac arteries. The 3D geometry was premade and provided a framework for testing different blood flow conditions and their effects on the aneurysm haemodynamics in a typical AAA patient.

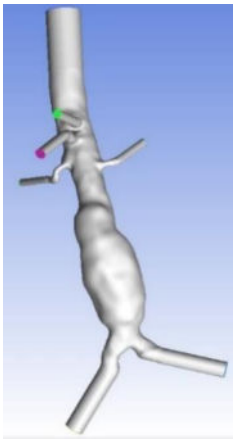


Figure 1.1 Image from ANSYS ICEM of the AAA patient specific geometry used for the project

Fabricating the meshing

Once the geometry was set, the spatial domain was discretized into many individual cells using meshing in ICEM. This was important for solving the Navier Stokes equations, to ensure that there were enough points to iterate between for accurate predictions. With insufficient data points, the model cannot converge to the correct solution. On the contrary, too many discretized cells and the computational processing time becomes more taxing. Therefore, a mesh that strikes a balance between accuracy and processing time is pivotal in any CFD study.

Due to the non-ideal nature of the complex patient specific geometry, an unstructured mesh had to be implemented with the use of densities for local refinement. Densities were applied to sharp curves and bends (including where outlet branches were joined the main artery body). This was done to define meshing parameters in those areas that would override global parameters.

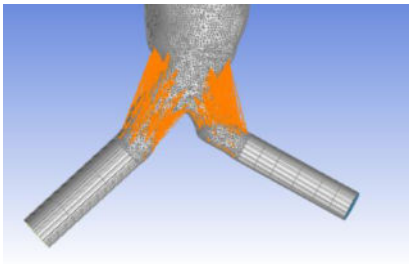


Figure 1.2 Image from ANSYS ICEM of the lower end of the AAA displaying the densities created for local refinement

Once completed, the full mesh was constructed. The first step computed a robust Octree mesh with a constant element size throughout the spatial domain as a base mesh. To overcome quality issues, a quick Delauny mesh was laid on top of the existing mesh to smooth out any abrupt volume transitions between elements. Lastly, a 10-layer prism mesh was created from the geometry wall to improve accuracy of the simulations close to the borders. The final mesh quality was deemed acceptable if more than 99% of cells had a quality greater than 0.3. In addition, the mesh expansion factor was deemed acceptable if it was all below 30.

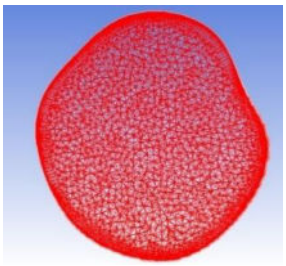


Figure 2.1 Cross section plane showing meshing structure from mesh with 4 million elements chosen

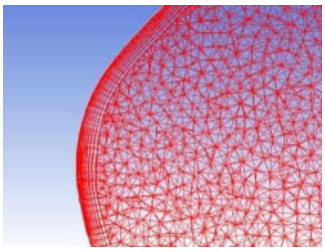


Figure 2.2 Magnified version of **Figure 2.1**

Meshing Sensitivity Analysis

A mesh sensitivity analysis was utilised to find an optimal balance between computational time and accuracy, Multiple meshes of varying element sizes were made consisting of approximately 2, 3, 4 and 5 million elements. This was done to ensure that CFD results were independent of the mesh density and that the discretized number of elements were minimised and therefore had the lowest computational time. These tests were carried out

under these assumptions to simplify the model: blood is Newtonian with a viscosity of 0.0035 Pas, a flat inlet velocity profile and outlet boundary conditions of 0 Pa.

The meshes of different elements were prepared for CFD simulations using CFX in ANSYS. The analysis was carried out at peak systole, and a post-processing context file was created so each simulation was analysed at the exact same points to reduce uncertainty. Values were taken from each mesh and compared against each other, with the criteria that a mesh should be accepted as optimal when its convergence to the next mesh was less than 10% for all key parameters.

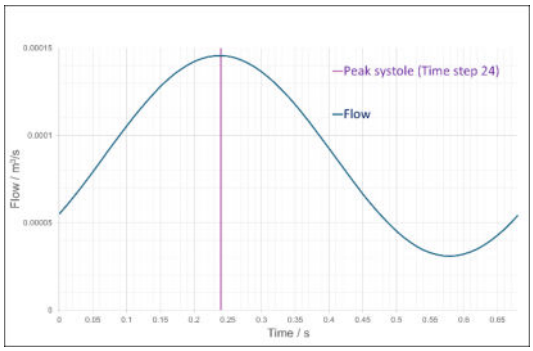


Figure 3 Inlet flow as a function of time, demonstrating a pulsatile flow with peak systole every 0.24 seconds

Effects of implementing non-Newtonian model

After a mesh was confirmed, several changes were made to investigate the effects implementing non-Newtonian models opposed to a Newtonian model.

All outlet boundary conditions which were used in the meshing sensitivity analysis (0 Pa) were changed to a 3-element Windkessel model, tuned in accordance with Pirola et al for the most accurate results[18]. Several studies found that using outlet boundary conditions with 0 Pa failed to reproduce any anatomically realistic flow[19][20][21], whilst using the Windkessel model would closely mirror the systemic response of the cardiovascular system and values consistent with physiological conditions[19]. It is widely used to simulate the resistance and compliance characteristics of the distal vasculature. Utilising patient-specific information in the form of brachial cuff pressure and heart rate readings, the proximal resistance (R1), distal resistance (R2) and compliance (C) values were calculated, displayed below for each artery.

Table 1 Windkessel parameters used for each outlet

Artery	R1 (kg/m^4s)	R2 (kg/m^4s)	C (kg/m^4s)
Celiac	2.61E+08	2.65E+07	2.65E+07
Left iliac	5.92E+07	3.09E+08	3.09E+08
Left renal	3.83E+08	1.83E+07	1.83E+07
Right iliac	4.05E+07	2.25E+08	2.25E+08
Right renal	4.17E+08	1.51E+07	1.51E+07
Superior mesenteric	1.70E+08	2.80E+07	2.80E+07

Throughout the project, all models used the boundary type of 'opening' for all arteries. The use of a pulsatile flow can lead to backflow, however when rendering the boundary type as an opening rather than an outlet, it allows the arteries to become an inlet when necessary.

To ensure simulations were more realistic, it was essential to negate the Newtonian blood assumption. As blood is a shear thinning fluid, and in this case does not exhibit yield stress, the use of the Cross, Herschel-Bulkley and others can be confidently eliminated[16]. The Power Law model, although is a shear thinning model, was determined to be uncharacteristic as with lower shear stresses, would be unrepresentative due to its tendency to inflate the viscosity. One of the most common models for a non-Newtonian model of blood is the Carreau-Yasuda model[22]. This was implemented instead of the Carreau model due to a smoother change in viscosity with shear rate due to the extra flexibility in the α parameter. In addition, the Quemada model was chosen due to it being well documented as well as a robust model [23].

The effects of using non-Newtonian models in CFD were explored, assessing whether any substantial differences were observed. Two variations of the Carreau-Yasuda model were implemented, Carreau-Yasuda 1 and Carreau-Yasuda 2, each with distinct parameter sets. This approach was adopted due to the wide range of parameterisations reported in the literature, suggesting slight variations in model configuration could lead to significantly different results. Carreau-Yasuda 1&2 were based off previous models of blood in aortic [25] and cerebral[26] aneurysms, respectively, and utilising this the potential variability can be evaluated to ensure a comprehensive analysis of the model's performance.

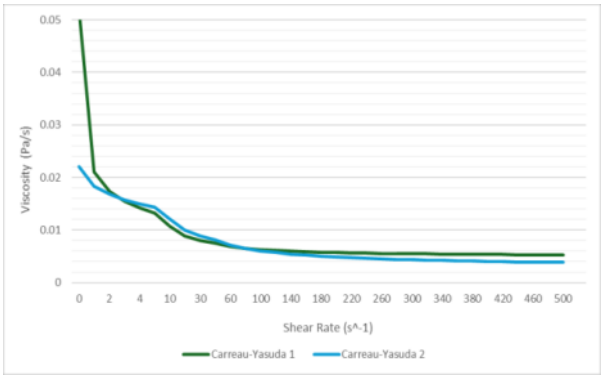


Figure 4 Viscosity as a function of shear rate from $0-500\ s^{-1}$ (range of shear rate in healthy arterial systems)[24] for Carreau-Yasuda 1 & 2

The inlet boundary conditions were fixed as a flat velocity profile for all 4 models, Newtonian, Quemada, and Carreau-Yasuda 1 and 2, for which the results were then compared.

$$Carreau-Yasuda\ equation:$$
$$\mu = \mu_\infty + \frac{\mu_0 - \mu_\infty}{(1+(\lambda\dot{\gamma})^\alpha)^{\frac{1-n}{\alpha}}} \quad (Eq.3)$$

where μ is the blood viscosity, μ_∞ is the infinite-shear viscosity, μ_0 is the zero-shear viscosity and $\dot{\gamma}$ is the shear rate[25]. The empirical parameters follow, where λ is the relaxation time, α is the transition parameter, n is the power index.

Table 2.1 Parameter values for both Carreau-Yasuda models implemented

Parameters	Carreau-Yasuda 1[25]	Carreau-Yasuda 2[26]
μ_∞	0.00476 Pa s	0.0022 Pa
μ_0	0.0519 Pa s	0.022 Pa
λ	0.438 s	0.110 s
$\dot{\gamma}$	Determined during simulation (Ansys CFX)	Determined during simulation (Ansys CFX)
α	0.409	0.64
n	0.191	0.392

$$Quemada\ equation:$$
$$\mu = \mu_P\left(\frac{k_0 + k_\infty\left(\frac{\dot{\gamma}}{\dot{\gamma}_c}\right)^{\frac{1}{2}}}{1 + k_\infty\left(\frac{\dot{\gamma}}{\dot{\gamma}_c}\right)^{\frac{1}{2}}}\frac{\phi}{2}\right)^{-2} \quad (Eq.4)$$

where μ is the blood viscosity, μ_P is the plasma viscosity, $\dot{\gamma}$ is the shear rate, $\dot{\gamma}_c$ is the critical shear rate, k_0 is the intrinsic viscosity at zero shear rate and k_∞ is the intrinsic viscosity at infinite shear rate[27].

Table 2.2 Parameter values for Quemada model implemented

Parameters	Quemada[28]
μ_P	0.0012 Pa s
k_0	4.33
k_∞	2.07
$\dot{\gamma}$	Determined during simulation (Ansys CFX)
$\dot{\gamma}_c$	1.88 s^{-1}
ϕ	0.345

Manipulating inlet boundary conditions

Furthermore, it is rare to see the implantation of both non-Newtonian models and a non-flat velocity profile. By changing the inlet boundary conditions with all other assumptions being fixed, any subsequential change was assumed to be influenced by the varying inlet boundary conditions. This allowed for investigation of the impact on results associated with varying the inlet profiles.

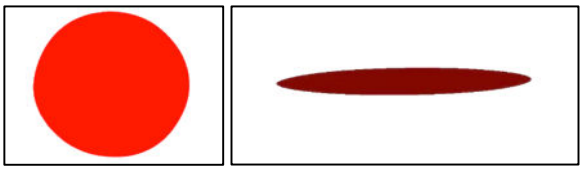


Figure 5 Flat inlet velocity profile top view (left), side view (right)

The inlet boundary conditions were manipulated to simulate parabolic and skewed parabolic velocity profiles at the inlet using an in-house Python script. The inlet velocity plane was generated for three different scenarios (a normal parabolic profile and parabolic skewed left and right profiles respectively). The skewed velocity profiles were adopted to account for the variations in aortic arch curvature observed among patients, which influence how blood flows into the abdominal aorta and create distinct velocity patterns. Evaluating how these differences affect flow dynamics and WSS was essential for understanding their impact on the accuracy and predictions of the simulations.

The first step was to generate the desired inlet velocity plane. This was then mapped to the patient specific geometry, aligning with the aortic flow direction. These conditions were then coded to the CFX solver to use these inlet velocity profiles for the simulation at the correct time step at peak systole. Lastly, the final Python script would generate .CSV files of XYZ velocities on the plane at different points. Using the Carreau-Yasuda 1 model, 4 different velocity profiles at the inlet were compared, the basis of a flat velocity profile, a standard parabolic velocity profile, and 2 skewed parabolic profiles at a flow jet angle (FJA) of 27° and -27°.

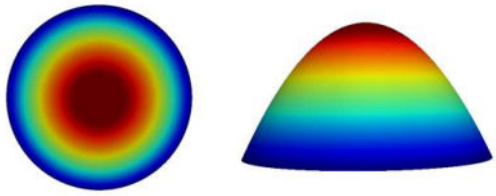


Figure 6.1 Standard parabolic velocity profile implemented

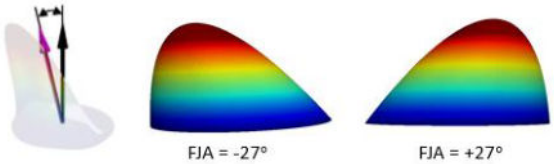
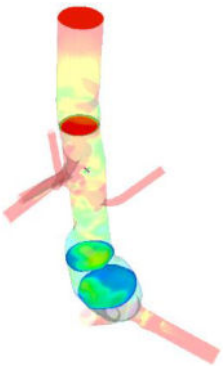


Figure 6.2 Flow jet angle of skewed parabolic velocity profiles

Analysing results

All completed simulations were processed using ANSYS EnSight with the use of contexts to ensure all values retrieved of each plane and section were of the exact same location, to ensure accurate comparisons. This context included three cross sectional plane cuts (one close to the inlet, one towards the start of the aneurysm and one at the apex of the aneurysm), and calculations of several parameters across the whole geometry and within these planes. These parameters included: maximum velocity, maximum spatial mean velocity (the maximum of the average velocity of each node in a specified plane or section), maximum WSS and maximum TAWSS. Velocities in different directions were acknowledged to be of interest when deemed necessary for each model, hence the spatial mean velocity for the inlet and all 3 planes in each axis (X,Y,Z) were also analysed.



The positions of the planes discussed can be seen on the left.

Figure 7 Image captured from ANYS CFX showing positions of the planes in the geometry analysed within the AAA

On top of comparing quantitative results, qualitative analysis of the streamlines in the aneurysm, plane contours of the WSS and velocities were also compared to further visualise the results.

Results

Sensitivity analysis

As mesh elements increase, accuracy increases up to a cut-off point where it converges to a value. The task here was to determine at which mesh density the simulation converged, whilst minimising excessive use of computational resources. The number of elements in each mesh were as follows: Mesh 1 – 1,903,246. Mesh 2 – 2,849,240. Mesh 3 – 4,046,111. Mesh 4 – 5,162,630.

Table 3 Convergence between meshes to determine ideal mesh

Mesh Comparison	Number of key parameters that converged to < 10%	Number of parameters that didn't converge to <10%
Mesh 1/2	14	3
Mesh 2/3	11	6
Mesh 3/4	17	0

Table 3 revealed that all key haemodynamic parameters converged to less than 10% difference between mesh 3 and 4. Moreover, all parameters achieved less than 5% difference with the exception of plane 2 maximum WSS, which still had a low 9% difference. Results also show that mesh density had significant effects on WSS and TAWSS, specifically within planes 2 and 3 in the aneurysm and negligible effects on velocity related parameters. Ultimately, mesh 3 resulted in an ideal balance of maximising accuracy and minimising computational time.

Comparing Newtonian & non-Newtonian models

Flow Patterns

The maximum velocity and spatial mean velocity across all planes showed minimal differences between the models, ranging from 0.008% to 12%. Particularly plane 1 showed virtually no difference in values in both velocity parameters between all 4 models. Plane 2 showed slight yet still minimal differences in both velocity parameters, whereas differences in values were more pronounced in plane 3, notably where the aneurysm was located.

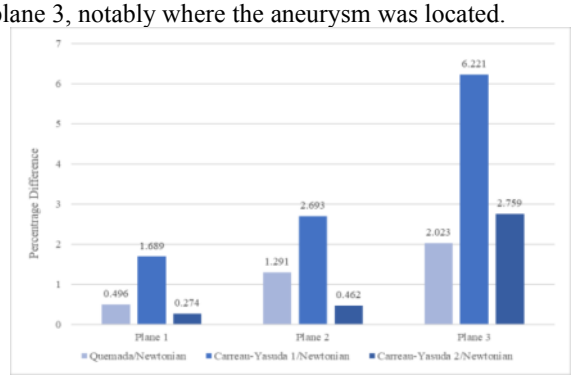


Figure 8.1 Percentage difference in **maximum velocity** between non-Newtonian and Newtonian models

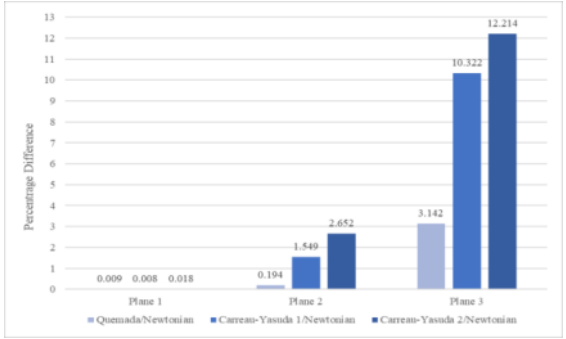


Figure 8.2 Percentage difference in **maximum spatial mean velocity** between non-Newtonian and Newtonian models

When comparing the non-Newtonian models to the Newtonian model, the Carreau-Yasuda models produced slightly higher velocity variations compared to the others, however had lower spatial mean velocities overall. Particularly the Carreau-Yasuda 2 model showed the largest difference of 12%, further reflecting the localised effects of shear-thinning properties.

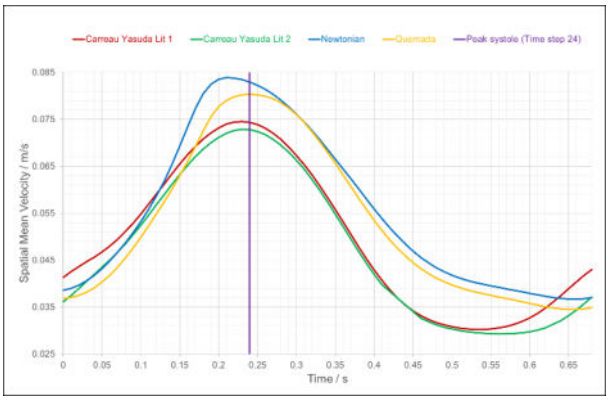


Figure 9.1 Plane 3 **spatial mean velocity** as a function of time between all 4 non-Newtonian models

Meanwhile, the maximum velocity showed the largest difference of up to 6% in plane 3 with the Carreau Yasuda 1 model, with the remaining values having minimal differences between 0.27% to 2.76%.

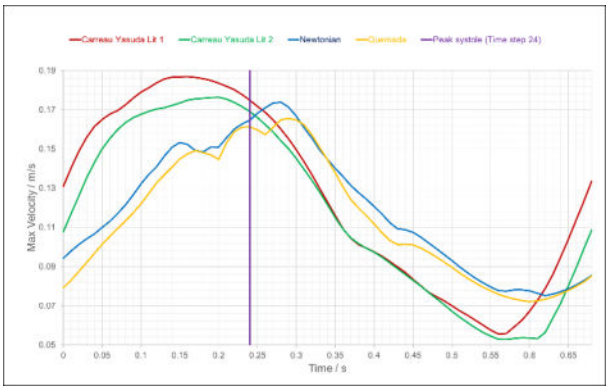


Figure 9.2 Plane 3 **maximum velocity** as a function of time between all 4 non-Newtonian models

Wall Shear Stress (WSS)

The maximum WSS and TAWSS revealed more substantial differences between the models against the Newtonian model over all 3 planes, particularly with the Carreau-Yasuda models. Variations ranged from 4% to 140% compared to the base Newtonian model, with the most significant differences observed in plane 3 once again, the apex of the aneurysm. The Newtonian model generally overestimated WSS in areas of high shear rates and underestimated it in low-shear regions. Whilst all models projected similar values in planes 1 and 2, plane 3 showed important changes. In plane 3, the Quemada model had minimal variations unlike the Carreau-Yasuda models in results, reflecting the importance of different non-Newtonian models displaying vastly different viscosity behaviours with areas of abnormal flow patterns (the aneurysm). The atypical section of the geometry would thus cause the shear rate to behave differently, hence causing the variations in viscosity.

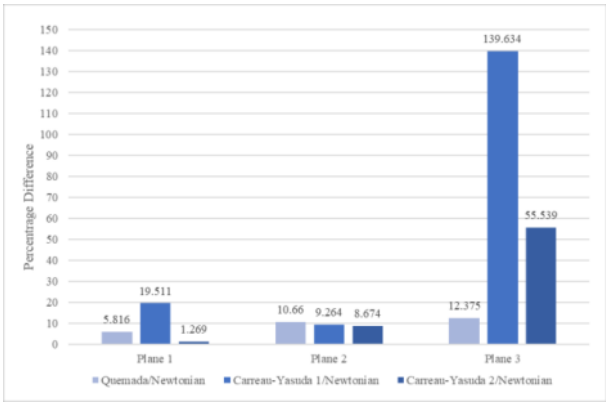


Figure 10.1 Percentage difference in **maximum WSS** between non-Newtonian and Newtonian models

The Carreau-Yasuda models had notable differences in the maximum WSS, ranging from 1.2% to 139.6% in variety, whilst the Quemada model only showed a difference of up to 12.4%. In particular, plane 3 showed the largest difference in variety of 139.6% in the Carreau-Yasuda 1 model, and 55.5% in the Carreau-Yasuda 2 model.

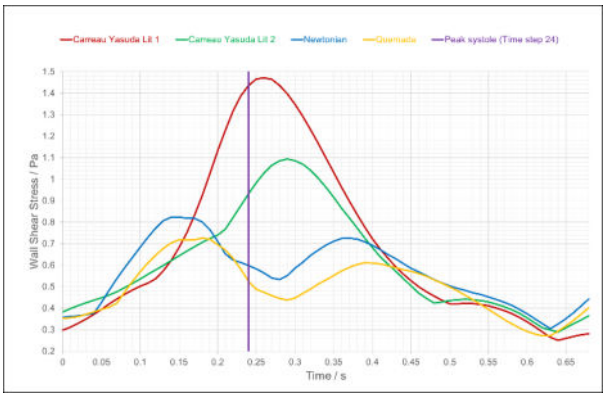


Figure 10.2 Plane 3 **maximum WSS** as a function of time between all 4 non-Newtonian models

The maximum TAWSS exhibited similar trends, with the largest differences found when compared the Newtonian model in plane 3 in the Carreau-Yasuda 1 model and Carreau-Yasuda 2 model at 50.9% and 24.7% respectively. The results varied from 4.6% to 50.89% difference, giving a lower variance than the maximum WSS. Notably, there appeared an anomaly of 21% difference in the Carreau-Yasuda 1 model in plane 1, potentially caused by the model being tuned in a different way to Carreau-Yasuda 2.

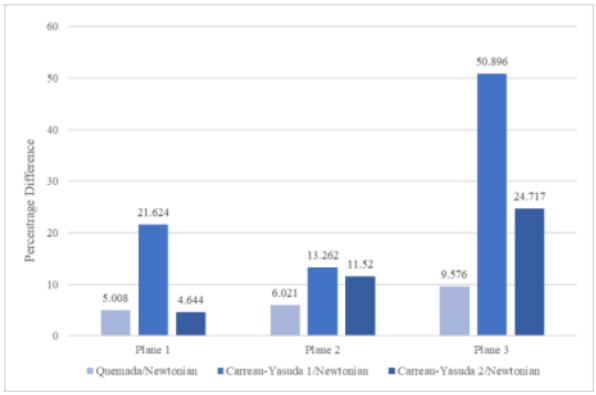


Figure 10.3 Percentage difference in **maximum TAWSS** between non-Newtonian and Newtonian models

Figure 11 below portrays the TAWSS on the abdominal aortic wall. It was evident that the Newtonian and Quemada models generally had a lower TAWSS over the geometry compared to the Carreau-Yasuda models.

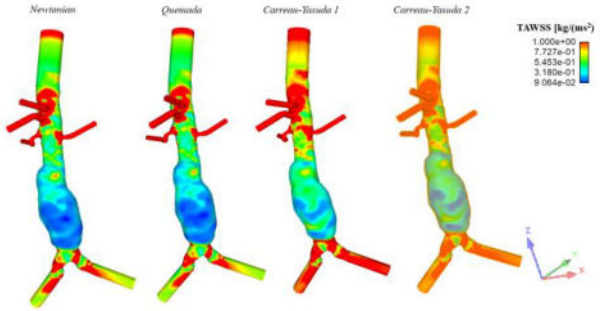


Figure 11 TAWSS contours of geometry wall of Newtonian model (left), Quemada model (left centre), Carreau-Yasuda 1 (right centre), Carreau-Yasuda 2 (right)

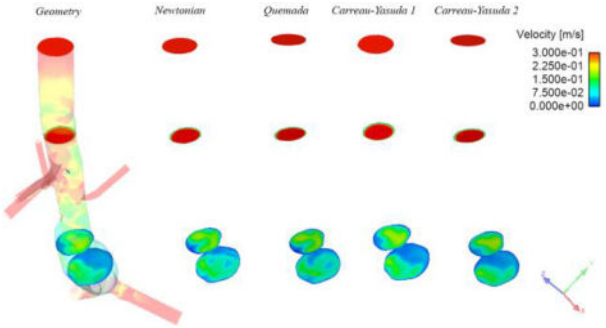


Figure 12 Geometry (left) & **Velocity** contours of the inlet, planes 1,2,3 (top to bottom) of Newtonian model (2nd left), Quemada model (3rd left), Carreau-Yasuda 1 model (4th left), Carreau-Yasuda 2 model (right)

Overall, using different models to represent the behaviour of the bloods viscosity had a significant role on the impact of computed values throughout the entire geometry, most notably in the aneurysm where low shear rates are more evident. Although differences in max velocities were not significant as discussed in Figure 8.1, the velocity contours shown above in Figure 12 shows the velocity profile was drastically different, increasing in variation from plane 2 to plane 3. The Newtonian and Quemada models seem to show a more centralised profile with the maximum velocity tending towards the core, whilst the Carreau-Yasuda models had a skewed profile towards the posterior, with virtually no flow occurring in the anterior.

<u>Comparing inlet boundary conditions</u>

This section presents the effects of different inlet boundary conditions on haemodynamic properties within the AAA. Four velocity profiles were taken into consideration in this analysis: a flat profile, a parabolic profile, a parabolic profile skewed left by 27° and a parabolic profile skewed right by 27°, all being implemented into the Carreau-Yasuda 1 model for further testing. Key parameters such as WSS, velocity components and flow patterns were evaluated qualitatively and quantitatively.

It was evident that maximum velocity achieved was not affected significantly, however, there was some effect on this from different velocity profiles. The parabolic profile displayed the highest overall velocity of 1.304 m/s with the lowest from skewed 2 of 1.289 m/s. This change was small

but showed that the effect of swirling had an effect on velocity properties. More interestingly was the direction of swirling. While skewed 2 displayed the lowest velocity, skewed 1 had a velocity of 1.303 m/s, which was almost the same as a normal parabolic profile. The flat profile had a velocity of 1.299 m/s. This showed that the direction in which the parabolic profile was skewed can influence velocity profiles to behave closer to a normal parabolic or flat profile.

However, this changes when looking at velocities within the specific planes. In plane 1, the skewed inlet boundary conditions showed a considerably higher maximum velocity at peak systole, 0.547 m/s in skewed 1 and 0.537 m/s in skewed 2, a considerable difference to 0.420 m/s in the flat profile and 0.462 m/s in the normal parabolic profile, highlighting the effects of swirling components close to the inlet.

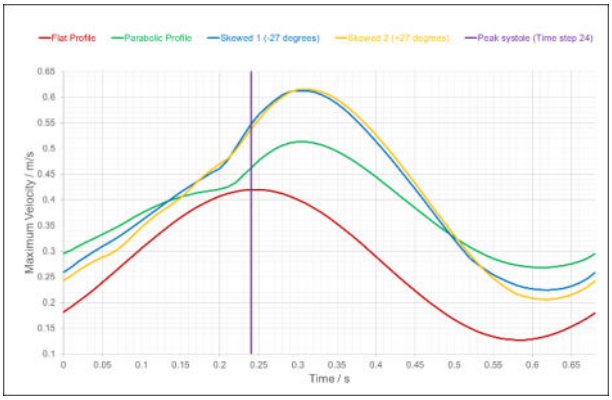


Figure 13.1 Plane 1 **maximum velocity** as a function of time between all 4 varying inlet boundary condition models

From the maximum velocity profiles plotted against time seen below in plane 3, it was demonstrated that over time, they behaved in vastly different ways across the 4 inlet boundary conditions. This demonstrated that as flow progressed through the aneurysm towards the middle, the effects of different inlet profiles became more prominent with positions of the apex varying from before and after peak systole.

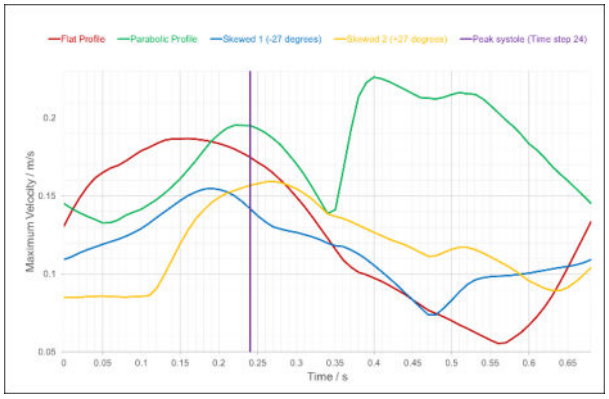


Figure 13.2 Plane 3 **maximum velocity** as a function of time between all 4 varying inlet boundary condition models

Within the aneurysm, the maximum WSS and TAWSS displayed higher values in a flat profile compared to other profiles. Among the other profiles, a normal parabolic profile displayed the lowest values, with skewed 2 displaying the highest of the three. Figure 14 shows the TAWSS contours and its variation in location maximum and intensity in the aneurysm between different profiles.

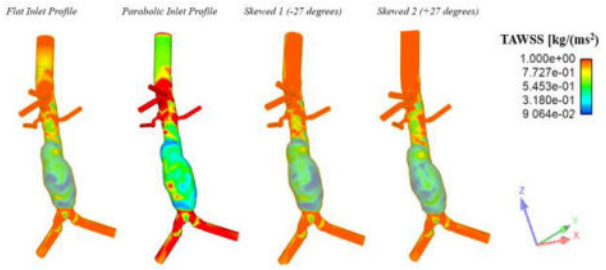


Figure 14 Images of the aneurysm section, highlighting TAWSS for flat (left), parabolic (left centre), skewed 1 (right centre) and skewed 2 (right) inlet velocity profiles

When looking at the spatial mean velocities in X, Y and Z directions across the planes, the flat profile and parabolic profile displayed extremely similar values. The two skewed profiles however, displayed varying results, most of which were different to flat and normal parabolic profiles. These more prominent differences in skewed velocity profiles indicated the effects of swirling on the directions of the velocities in the planes. It was shown that the most significant impact came from the parabolic profile and the skewed profiles, which displayed relatively even vectors but skewed slightly. The flat profile showed unevenness similar to the parabolic profile but much less pronounced.

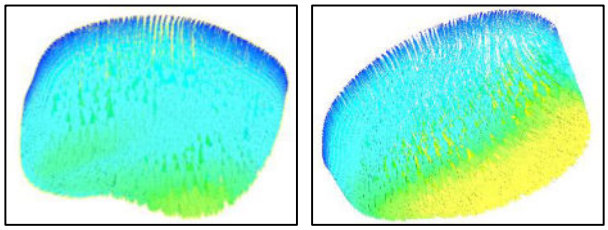


Figure 15 Images of velocity vector arrows in plane 2 for skewed 1 (left) and skewed 2 (right) inlet velocity profiles

Shown below are flow streamlines visualised within the aneurysm; the flat profile exhibited the most uniform velocity distribution within the aneurysm. The parabolic profile exhibited a similar distribution but with a concentration in the centre and some recirculation. Streamlines from the skewed profiles displayed recirculation and flow asymmetry, emphasising the impact of swirling.

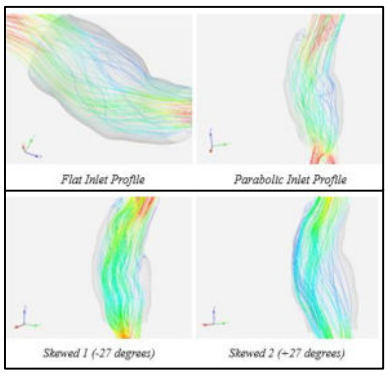


Figure 16 Images showing flow streamlines within the aneurysm for flat (top left), parabolic (top right), skewed 1 (bottom left) and skewed 2 (bottom right) inlet velocity profiles

Discussion

Sensitivity analysis

The mesh sensitivity analysis highlighted the importance of determining a feasible balance between number of elements and accuracy to keep computational taxes low. Key haemodynamic parameters such as maximum velocity, maximum spatial mean velocity and flow showed strong convergence across all mesh densities with less than 3% difference across meshes in the planes. The only exception was maximum velocity in plane 3 with differences of 9.6% across mesh 1 and 2, and 10.2% across plane 2 and 3. This converged to a much lower threshold with a percentage difference of 5.1% across mesh 3 and 4. This was significant as plane 3 was in the apex of the aneurysm and could be considered the most important location to have strong convergence for the purpose of predicting aneurysm rupture in a medical environment.

On the other hand, the parameters that showed greatest mesh density dependence were the maximum WSS and TAWSS in plane 2 and 3, which both sat within the aneurysm, and was vital to habitat good convergence, since the goal was to improve the current ways in which haemodynamic properties were analysed within AAA's. With respect to the convergence between mesh 1 and 2, although in plane 2 these values converged to an acceptable threshold of less than 5%, the percentage difference in plane 3 was 70.8% and 18.6% for the maximum WSS and TAWSS, respectively. These differences were simply too large to accept this convergence value. When the mesh density was increased and meshes 3 and 4 were compared, it was found that the percentage difference dropped significantly. All significant haemodynamic parameters appeared to have converged here with 5% or less difference. The only exception was the maximum WSS in plane 2 with a 9.6% difference. With the importance of keeping computational time as low as possible, mesh 3 with approximately 4 million elements was selected to proceed with further analysis of viscosity models and inlet velocity profiles. This decision was taken since all key haemodynamic parameters no longer displayed largely significant dependence on mesh density from mesh 3 onwards.

This also highlighted the point of using finer meshes with more elements and higher density to model areas with more complex flow dynamics, such as within the aneurysm, to eliminate errors arising from mesh density dependence. Interestingly, for studies focusing solely on velocity related haemodynamic properties, ignoring shear stress in the wall, a coarser mesh would be acceptable with fewer elements. This was shown through the low percentage difference across velocity related properties across meshes. Furthermore, investigating whether different meshes would be necessary for non-Newtonian models would be beneficial for further accuracy.

Overall, the results emphasise the importance of mesh refinement, especially the need for local refinement closer to the wall and high gradient areas such as the sharp curves where branches join the main artery body. This was particularly important for capturing critical haemodynamic properties such as WSS and TAWSS. This analysis also highlighted the importance of sensitivity analysis to ensure that CFD simulations were both accurate and computationally efficient, mitigating effects of mesh density on results.

Comparing Newtonian & non-Newtonian models

The results demonstrated notable differences between Newtonian and non-Newtonian models, particularly in the WSS metrics, underscoring the critical importance of considering blood's rheological properties in AAA simulations. These findings provided insights into the limitations of Newtonian assumptions and the implications of shear-thinning behaviour on haemodynamic predictions.

Flow Patterns

The relatively small variations in velocity parameters across models (0.008% to 12%) suggested that the Newtonian model performed adequately in capturing linear flow characteristics. However, the pronounced differences observed in plane 3, particularly for the Carreau-Yasuda models, indicated that non-Newtonian effects became more significant in regions with complex flow, such as the aneurysm bulge.

The skewed velocity profiles predicted by the Carreau-Yasuda models highlighted the possibility of redistributing flow toward the posterior wall. This aligned with clinical observations of flow asymmetries within aneurysms and could have significant implications when identifying regions of stagnation or recirculation that promote thrombus formation[29]. The Newtonian and Quemada models, with their more centralized flow profiles, may have underestimated these effects, potentially leading to less accurate predictions. The Carreau-Yasuda models' ability to reflect the additional localized effects from viscosity changes enhanced their utility in simulating regions with higher shear stress, potentially more accurate of the flow behaviour in the aneurysm.

Wall Shear Stress (WSS)

The substantial differences in WSS metrics, ranging from 4% to 140%, highlighted the sensitivity of these parameters to the assumed rheological model. These differences were clinically significant, as WSS is a critical predictor of aneurysm progression and rupture risk[6][7][8][9]. The Newtonian model's tendency to overestimate WSS in high-shear regions and underestimate it in low-shear regions limits its accuracy in capturing the haemodynamics of aneurysms.

The Carreau-Yasuda models, particularly Carreau-Yasuda 1, exhibited the highest WSS values in plane 3, indicating their ability to capture the effects of reduced viscosity under low shear rates. Elevated WSS near the aneurysm wall, as predicted by the Carreau-Yasuda models, suggests areas of potential wall degradation, while lower WSS regions may correspond to zones of thrombus formation[30]. The differences in maximum TAWSS, with Carreau-Yasuda 1 showing variations up to 50.9%, further emphasised the importance of accurately modelling blood rheology. While the Carreau-Yasuda models offer greater flexibility in capturing shear-thinning behaviour, the large

deviations they predicted in certain regions, such as plane 3, warrant a closer examination of whether these models might be overpredicting WSS under specific flow conditions. The anomaly observed in plane 1 for the Carreau-Yasuda 1 model, which arose in an otherwise steady flow region, highlighted the potential for non-Newtonian effects to introduce unexpected variations. This could stem from the sharpness of the transition between low and high shear rates, controlled by parameters like α and n. However, the Carreau-Yasuda 2 model did not exhibit any anomalous difference, thus showing the importance of further exploring setting the Carreau-Yasuda parameters to fit the behaviour of blood more accurately, displaying that even the same model with differently tuned parameters results in drastic changes.

In contrast, the Quemada model, while less sensitive to shear-thinning effects, produced more consistent results across all planes, as it may be avoiding over-amplifying viscosity transitions. The Quemada model is used extensively to model rheology and is often the best fit hemodynamically[31], however, the model and its ability to accurately capture the complex flow dynamics as seen in plane 3, where the aneurysm was located, may be misleading. The Quemada model showed less pronounced variations in WSS compared to the Carreau-Yasuda models, which could mean it underestimated WSS in regions of high flow disturbance.

Clinical and Computational Implications

The contrasting results between the models underscored the challenges of selecting the most appropriate non-Newtonian representation for AAA simulations. The Carreau-Yasuda models' ability to capture sharp gradients and localized effects makes them potentially more accurate for assessing rupture risks and identifying areas prone to thrombus formation. However, the possibility of overprediction, particularly in low-shear regions, must be considered. The Quemada model, with its simpler and less dynamic response, might be more suitable for applications where computational efficiency and consistency are prioritized. However, this comes at the cost of potentially underestimating critical haemodynamic features in regions of disturbed flow. These differences suggest that the choice of model should be fine-tuned for the specific objectives of the simulation. For example, detailed rupture risk assessment might benefit from the Carreau-Yasuda models, while broader studies of general flow patterns could rely on the Quemada model.

Future Directions

Future research should aim to validate these predictions of against experimental or clinical data, such as 4D-flow MRI measurements, to determine whether the higher WSS values predicted by the Carreau-Yasuda models are physiologically accurate or represent overpredictions. Similarly, the Quemada model's underprediction of WSS in plane 3 should be assessed to ensure it does not overlook critical haemodynamic variations.

Adjusting and standardizing parameters, particularly for the Carreau-Yasuda models, could help mitigate the risk of overprediction. For instance, a sensitivity analysis of parameters like α could provide guidelines for tuning these models to better match clinical observations.

Comparing inlet boundary conditions

The results revealed significant effects on haemodynamic properties with varying inlet boundary conditions, particularly within the aneurysm. Skewed inlet profiles, designed to mimic physiological effects of swirling, were found to alter flow dynamics more than flat and normal parabolic profiles. These differences were most pronounced when looking at the plane 3 maximum velocity over time and the asymmetry in flow patterns from the streamlines in the aneurysm as well as the vector arrows in planes. Although differences were also present when looking at plane 1, close to the inlet, these effects were diminished when looking at the profiles over time.

WSS showed moderate sensitivity to the inlet boundary conditions. Whilst maximum WSS in the aneurysm showed slight differences in values, when looking at the contour across the aneurysm for TAWSS, the positions and intensities varied. Differences in contours between the unhealthy aneurysm and the healthy section were also more pronounced with skewed profiles. This is significant for predicting areas at higher risk of rupture.

Spatial mean velocities in the X, Y and Z directions displayed minimal variations between flat and normal parabolic profiles, indicating that the average flow dynamics were not heavily influenced by changing to a normal parabolic profile. However, slight differences arose when changing to a skewed profile, which displayed asymmetry particularly at the inlet, where swirling effects are most prominent. This highlighted the potential benefits of skewed profiles in capturing localised flow patterns.

The differences observed emphasise the importance of incorporating more realistic velocity profiles into CFD simulation for modelling flow within AAA's. Skewed profiles that introduced physiological swirling effects, provided a more detailed representation of realistic flow dynamics that could have had significant impacts on analysis rupture risk. However, it is important to understand that these differences do not necessarily indicate increased accuracy. Further analysis would be required to determine whether variations in haemodynamic properties in these CFD simulations translate into a more reliable method of predicting AAA rupture across all patients. To fully evaluate the impact of inlet boundary conditions, additional studies would be needed to validate these CFD predictions against clinical data. Moreover, analysing a wider range of patient specific AAA geometries would aid in generalising these findings and assessing the robustness of the impact of using different flow profiles.

Future Directions

Future research could also involve modelling these profiles across different non-Newtonian and Newtonian models with several patient specific geometries. In addition to this, more advanced techniques of modelling could be explored, such as other complex inlet velocity profiles and several skewed angles in different directions.

This would provide a more comprehensive understanding of AAA haemodynamic properties and the reliability of using CFD simulations across different patients.

Conclusion & Outlook

This study highlighted the value of tuning computational fluid dynamics (CFD) to further develop understanding of the complex haemodynamics of abdominal aortic aneurysms (AAAs). The results underscored the importance of selecting appropriate modelling assumptions, such as viscosity models and inlet boundary conditions, to enhance the accuracy and reliability of simulations. The findings demonstrated that finer mesh densities, particularly with complex geometries, were necessary to ensure convergence of critical parameters, including WSS and velocity related haemodynamic properties, when modelling physiological systems.

In addition, the comparison of Newtonian and non-Newtonian blood models revealed significant differences in WSS predictions, emphasising the limitations of assuming constant viscosity. Non-Newtonian models, particularly the Carreau-Yasuda models, amplified shear-thinning effects, showing potential in identifying high-risk regions prone to rupture in medical practices. However, further experimental validation is necessary to confirm whether these predictions align with physiological reality.

Moreover, further study of varying inlet boundary conditions demonstrated that skewed velocity profiles, introducing more realistic flow dynamics[32], alter flow asymmetry and WSS distributions within AAA's, particularly within the aneurysm. These findings point to the potential benefits of incorporating realistic boundary conditions in CFD models to capture haemodynamic abnormalities more effectively.

Future work should aim to validate these computational findings against clinical data and expand the analysis to multiple patient geometries. Exploring advanced modelling techniques, such as more sophisticated boundary conditions and fine tuning the parameters in each non-Newtonian model could further enhance the predictive accuracy of CFD simulations. Ultimately, this research lays the groundwork for leveraging CFD as a powerful tool to forecast AAA rupture risks and how to effectively manage them in clinical settings.

Acknowledgements

The authors would like to express our sincere gratitude to Dr. Chloe H. Armour for her invaluable guidance and support throughout this study. We extend our heartfelt thanks to Kaihong Wang for his technical support.

References

1. D. Yamanouchi, International Journal of Molecular Sciences. 24, 7125 (2023).
2. E. S. Wise, K. M. Hocking, C. M. Brophy, Journal of Vascular Surgery. 62, 8–15 (2015).
3. K. C. Kent, New England Journal of Medicine. 371, 2101–2108 (2014).
4. A. Karthikesalingam et al., New England Journal of Medicine. 375, 2051–2059 (2016).
5. K. Ouriel et al., 15, 12–20 (1992).
6. M. Shojima et al., Stroke. 35, 2500–2505 (2004).
7. H. Meng, V. M. Tutino, J. Xiang, A. Siddiqui, American Journal of Neuroradiology. 35, 1254–1262 (2013).
8. G. Zhou, Y. Zhu, Y. Yin, M. Su, M. Li, Scientific Reports. 7 (2017), doi:https://doi.org/10.1038/s41598-017-05886-w.
9. A. J. Boyd, D. C. S. Kuhn, R. J. Lozowy, G. P. Kulbisky, Journal of Vascular Surgery. 63, 1613–1619 (2016).
10. A. Tziotziou et al., Atherosclerosis. 387, 117387 (2023).
11. A. D. Caballero, S. Laín, Computer Methods in Biomechanics and Biomedical Engineering. 18, 1200–1216 (2014).
12. J. Y. Moon, D. C. Suh, Y. S. Lee, Y. W. Kim, J. S. Lee, Neurointervention. 9, 1 (2014).
13. J. A. Leo, M. J. Simmonds, S. Sabapathy, Experimental Physiology. 105, 244–257 (2019).
14. N. Sarkar, S. D. Sharma, S. Chakraborty, S. Roy, Computers & Fluids. 279, 106337 (2024).
15. Quadco Engineering, Advantages and Disadvantages of CFD | www.quadco.engineering
16. T. Sochi, Non-Newtonian Rheology in Blood Circulation. arXiv: Fluid Dynamics (2017), (available at https://www.semanticscholar.org/paper/Non-Newtonian-Rheology-in-Blood-Circulation-Sochi/f481947bc6b36dc0ca8b7d8dde87cc8e0f1235c1/figure/1).
17. A. Manta, K. Tzirakis, Annals of vascular surgery. 110, 63–81 (2024).
18. S. Pirola et al., IEEE Transactions on Biomedical Engineering. 66, 3411–3419 (2019).
19. S. Pirola et al., APL bioengineering. 2 (2018), doi:https://doi.org/10.1063/1.5011960.
20. C. H. Armour et al., Biomechanics and Modeling in Mechanobiology. 20, 481–490 (2020).
21. S. Pirola et al., Journal of Biomechanics. 60, 15–21 (2017).
22. F. L. Macrae, F. Swieringa, J. W. M. Heemskerk, R. A. S. Ariëns, Blood Advances. 5, 3468–3477 (2021).
23. H. Liu et al., Frontiers in Physiology. 12 (2021), doi:https://doi.org/10.3389/fphys.2021.718540.
24. K. Sriram, M. Intaglietta, D. M. Tartakovsky, Microcirculation. 21, 628–639 (2014).
25. H. E. Salman, B. Ramazanli, M. M. Yavuz, H. C. Yalcin, Frontiers in Bioengineering and Biotechnology. 7 (2019), doi:https://doi.org/10.3389/fbioe.2019.00111.
26. J. Bernsdorf, D. Wang, Computers & Mathematics with Applications. 58, 1024–1029 (2009).
27. D. Quemada, Biorheology. 18, 501–516 (1981).
28. Z. Mimouni, Open Journal of Biophysics. 06, 29–33 (2016).29. E. Weigang et al., European Journal of Cardio-Thoracic Surgery. 34, 11–16 (2008).
30. J. J. Paszkowiak, A. Dardik, Vascular and Endovascular Surgery. 37, 47–57 (2003).
31. A. Fuchs, N. Berg, L. Fuchs, Lisa Prahl Wittberg, Bioengineering. 10, 1240–1240 (2023).
32. S. Madhavan, E. M. C. Kemmerling, BioMedical Engineering OnLine. 17 (2018), doi:https://doi.org/10.1186/s12938-018-0497-1.

The Flexural Strength and Strength Development of Concrete with Carbon Addition

Thomas Dolbel and Yongbin Cho
Department of Chemical Engineering, Imperial College London, UK

Abstract

As the global demand of hydrogen as a renewable energy source is increasing, the potential of hydrogen production through methane pyrolysis, which generates pyrolytic carbon as a byproduct, is highlighted. This study focuses on the application of modified carbon in the concrete industry, which potentially reduces CO_2 emissions, focusing on the impact on flexural strength and strength development. The cement in concrete were replaced in different proportions with modified carbon additives, milled, air-milled, and nitric acid-milled coke, and unmodified carbon to test the flexural strength of carbon-enhanced concrete. These concretes were then analysed at various curing intervals (1, 3, 7, and 28 days) to determine the trends of strength development. Results show that replacing cement with modified carbon additives significantly improves flexural strength especially at 5% replacement. Milled and air-milled carbon demonstrated the best performance regarding flexural strength, reaching up to 40% of blank concrete. Furthermore, the strength development trends provide insight into the surface properties and hydration trends with usage of S_u and k_t, determined as KPIs through this study. These findings emphasize the importance of using carbon-enhanced concrete in real-world applications, benefiting in reduction of CO_2 emissions as well as enhanced performance, leading to a greener concrete industry with improved quality.

1. Introduction

1.1 Background

The demand of hydrogen has been increasing over the past decade since it is key to clean energy transition. The global demand of hydrogen was 97 Mt in 2024 and is expected to surpass 100 Mt in 2024, highlighting the importance of it as a renewable energy source. (IEA, 2024) Compared to the traditional methods of producing hydrogen, methane pyrolysis is an eco-friendly (Sánchez-Bastardo, et al., 2021) and cost-effective (Vine, 2024) technique, producing pyrolytic carbon instead of CO_2 as a byproduct. Thus, the utilisation of pyrolytic carbon has become a very important issue.

Pyrolytic carbon is known for its flexible applications such as in agriculture and battery manufacturing. (Hou, et al., 2023) (Mlonka-Medrala, et al., 2021). This study focuses on the industrial applications of pyrolytic carbon in the concrete industry. As previous studies have shown (Salazar, 2024), rather than using the natural carbon itself, modifications on the carbon additives are a necessary process, such as milling and thermal modification since they have demonstrated the ability to enhance mechanical and physical properties of concrete, such as compressive strength and workability, due to its small particle size and functional groups.

Meanwhile, the concrete industry, which is one of the largest industries globally, is estimated to produce 30 billion tons of concrete and 4.1 billion tons of cement annually (Statista, 2024) (Potter, 2023). A large amount of cement can be replaced with carbon, if carbon enhanced concrete is used. Also, cement production is a process that emits a large amount of CO_2, 8% of the world's emissions according to (Purton, 2024). The replacement of a portion of this cement with carbon, produced in a way that reduces CO_2 emissions, will allow for an improvement in the carbon footprint of concrete. By utilizing modified carbon, a dual strategy is available since it enables a cleaner production of hydrogen as well as cement.

Building up on the previous findings, the need for further research was recommended on another important aspect of concrete performance, flexural strength and strength development which is the main focus of this study. Flexural strength indicates the material's ability to resist bending stresses and cracking under load (Sowmya, 2020), which is critical for ensuring the stability and durability in real world applications. Testing the flexural strength will evaluate the impact of carbon additives and the corresponding resistances of the carbon-enhanced concrete to tensile stresses.

Furthermore, understanding how strength develops over the curing period of concrete, which is 28 days (ACI Committee 209, 1997), is vital since it may influence construction schedules and other various aspects in the industry. This indicates that strength development is crucial in concrete

evaluation. Since modified carbon additives can significantly alter hydration dynamics, aggregate bonding and matrix structure (Salazar, 2024), it is important to analyse the effects of both compressive and flexural strength during the curing period.

This study mainly focuses on investigating the performance of concrete with carbon additives with diverse properties by examining flexural and compressive strength across the curing period in intervals of 1, 3, 7, and 28 days. It aims to and explore the role of carbon surface modifications on the flexural strength of concrete, analyse the rate of strength development and identify potential trade-offs between compressive and flexural properties. Through these findings, the research seeks to provide deeper understanding of carbon-enhanced concrete, offering solutions for sustainable and mechanically optimized materials, therefore meeting the need of modern-day construction industry.

2. Methods

2.1 Overview

There are two main components to carry out for this project, casting the concrete bars, and then testing their strength using various methods. 5 different types of concrete were made, those being bars with no cement replacement, and bars with different methods of treating carbon replacing cement.

In order to cast the concrete bars that are required for this project we used cuboid moulds, with the dimensions of 20mm/20mm/80mm, shown in *Table 1*.

Table 1. Inner Dimensions of the Moulds

Length/mm	Width/mm	Depth/mm
80	20	20

These were then filled with the wet concrete mixture and labelled, then covered in a wet paper towel (Weerheijm, 2013) and left to set for 24 hours, until solid enough to remove from the moulds and be kept in a high moisture environment.

After the required curing time, these bars were then tested using a 3-point flexural test (ASTM International, n.d.) and after this test the halves were used to complete a compressive strength test.

2.2 Materials

To make the concrete, the following materials are needed: Kiln dried sand, cement, water and the carbon used for the cement replacement. These are unmodified coke, milled coke, air treated milled coke and nitric acid treated milled coke.

The kiln dried sand needed to be dried again in the lab, as it absorbed moisture from the environment, so was left in an oven at 150°C for 3 hours.

The different types of modified coke are produced in various ways, with all of them using milled coke as a base, which is coke that is milled using Fritsch mortar grinder pulverisette 2 for 4 hours. This is then either heated in an oven for 3.5 hours to air treat it or added to a reflux with 10% nitric acid at 90°C for 16 hours to treat it.

Regular coke is used instead of pyrolytic carbon for this study, due to the relatively large amounts needed to produce the bars, and the number of tests carried out. Previous work by (Salazar, 2024) shows that the difference between using these two types of carbon is limited, and coke will work as a suitable substitute.

2.3 Casting the Concrete Bars: Detailed Method

To cast the concrete bars a ratio of 3:1:0.6 of sand, cement and water respectively. The quantity used to make three samples per carbon replacement was 180g/60g/36g, with different percentages of the 60g of cement being replaced with carbon. The replacement percentages used are 2.5/5/7.5/10% of carbon, based on the findings from (Salazar, 2024). These components are then mixed for 2 minutes using an electric handheld mixer, to ensure that they are well mixed, and then casted into bar moulds of size 20mm/20mm/80mm.

This casting is carried out by filling the moulds a third at a time, tamping and stamping the mould on the bench for each third to eliminate any formation of bubbles. Once filled, the moulds are then smoothed out and labelled, and then covered in wet paper towel to ensure a high moisture environment to ensure a good cure. (Weerheijm, 2013) After one day the samples had hardened enough to be removed from the moulds and then fully wrapped in a wet paper towel to ensure the whole bar is in a high moisture environment to enhance hydration. These are then removed and tested after the allocated number of days.

For the 28-day test the full range of cement replacement percentages are used, while for the 1-, 3-, and 7-day tests only 5% replacement is used. Shown in *Table 2* below. This percentage replacement is chosen as it has the greatest increase in strength, both flexural and compressive, as can be

seen in the *Results* and from (Salazar, 2024) respectively.

2.4 Testing the Concrete Bars: Detailed Method

The Lloyd EZ 50 is used with the NEXYGEN software to configure the tests. 2 different load cells are used, with 1kN being used for the weaker bars, 1~7 days, to allow for a greater accuracy, while the 50kN load cell is used for the stronger flexural tests and the compressive tests. The load cell is moved down at 2mm/min and the tests are stopped at the point the samples break.

After the bars cured for the allocated time, they were then marked at points 10, 40 and 70mm along the bar to give markers to line up with the points on the 3-point test, as 10mm clearance either side of the points is necessary (ASTM International, n.d.). These are then tested using a 1kN load cell for the 1-, 3- and 7-day bars, and 50kN cell on the 28-day bars, due to their greater strength. The maximum load required to break the bars is then measured.

This process leaves 2 pieces that can then be used to carry out a compressive strength test, using the same equipment with an increased load, 50kN, as a greater load will be required to break the samples compressively. The compressive strength is found to allow for a measure of how the strength develops, according to *Equation 3*.

These were all completed three times for each sample type.

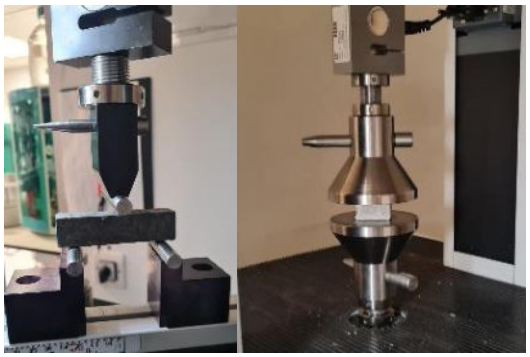


Figure 1. Different Configurations of EZ 50

2.5 Analysis Methods

The forces required to break the samples are converted to flexural and compressive strength, using *Equation 1&2* (Callister & Rethwisch, 2018). These are then compared against a bar with no carbon addition, to see how the modified carbon changes the strength, and rate of strength increase of the concrete.

$$S_f = \frac{3FL}{2bd^2} \quad (1)$$

$$S_c = \frac{F}{A} \quad (2)$$

Where S_f is flexural strength, S_c is compressive strength, F is the maximum load, L is the length of the bar, b is the breadth of the bar, d is the width of the bar and A is the cross-sectional area of the side bearing the load. The average of the 3 tests per replacement will be taken, and a standard deviation will be found for each set.

Table 2. Different Curing Times and Percentage Replacements

Curing Time/Days	Percentage Replacements Studied
1	5%
3	5%
7	5%
28	2.5%, 5%, 7.5%, 10%

The 28-day tests will be compared to find how the different carbon types affect the flexural strength of the concrete, while the 1-, 3-, 7- and 28-day tests at 5% cement replacement will be used to find if the rate of strength gain is affected by the different cement replacements. The tests on different curing times will also show the carbon additions will change the relationship between flexural and compressive strength (ACI).

The rate of strength development will be compared to the function, *Equation 3*, found by (Tank & Carino, 1991).

$$S_c = \frac{S_u k_t t}{1 + k_t t} \quad (3)$$

With S_u (MPa) being the ultimate compressive strength after an infinite period, k_t (day^{-1}) being the rate constant of strengthening at a constant temperature and t being the number of days that curing took place. This equation is then linearised, *Equation 4*, and a least squares regression over the experimental data is used to find S_u and k_t.

$$\frac{1}{S_c} = \frac{1}{S_u} + \frac{1}{S_u k_t} \cdot \frac{1}{t} \quad (4)$$

To complete the least-squares regression on *Equation 4* the average of the S_c values found for each sample, and the number of days that it was cured for, were used. This provided values for $\frac{1}{S_u k_t}$ and $\frac{1}{S_u}$ as the gradient and intercept respectively. These were then used to find S_u and k_t and used to fit

Equation 3 as a trendline on the plots in the results section.

3. Results

3.1 28-Day Tests

The results of the 28-day tests are shown in *Figure.2*, with the dashed line denoting the flexural strength of the blank concrete and the error bars showing 1 standard deviation between the flexural strength of each sample for a different composition. The high standard deviations of the 2.5 and 5% unmodified and the 10% air milled can be explained by the raw data, both unmodified data set had a missing sample, due to equipment issues. This causes a greater standard deviation, and the 10% air milled had one anomalous result that is half the strength of the others in the data set. There is a clear trend that shows that the addition of carbon to replace cement give the bars an increased flexural strength, with the smaller quantities of carbon are the most effective.

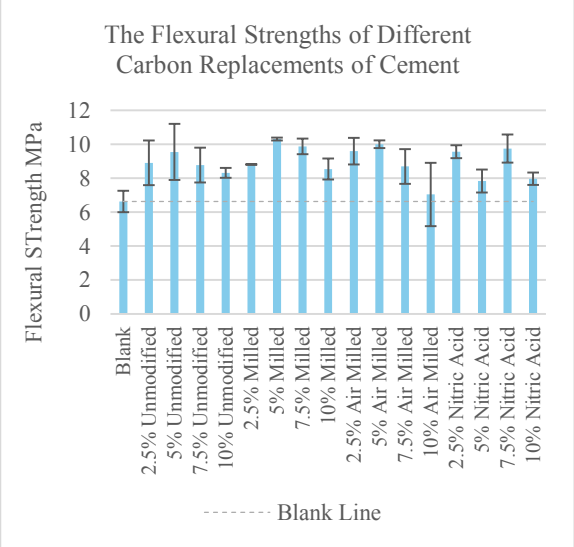


Figure 2. The Flexural Strength After 28 Days

The 2.5% and 5% replacements have a good strength advantage of between 15-40%. The milled and air milled are the most effective of the carbon types, with 5% of each type have flexural strengths of 10.3MPa and 10.0MPa respectively. This is considerably higher than the 6.62MPa flexural strength that the blank concrete had. The standard deviation of each of these results are only 0.08 and 0.22 for milled and air milled respectively, indicating that these results are reliable.

All of the modified carbon tests show a considerably higher strength than the blank concrete, however the nitric acid treated tests do not follow the same trend as the rest. The 5% replacement being considerably lower than both the 2.5% and 7.5% at 7.8MPa, compared to 9.6MPa and 9.7MPa for 2.5% and 7.5% respectively. This is very different to the other tests, and the standard deviations for the nitric acid tests are reasonable, so it is not easily explained by anomalous data.

3.2 Strength Development

3.2.1 Compressive Strength Development

To analyse the strength development, a graph was made using *Equations 1, 2 and 3.*

For compressive strength, as shown in *Figure 3* the trendline produced accurately fits with the data obtained through the experiment. The 7-day data were ignored which will later be explained in the *Discussion* section.

According to the figure above, all the cement replacement types show an increasing trend in compressive strength over time, which supports the curing behaviour of concrete.

Early-Stage Strength (1~7 Days)

The blank starts with a strength of 5MPa, one of the lowest starting strengths, however, experiences a high growth rate over the next 6 days. The unmodified carbon concrete has a similar strength at 1 day and experiences a relatively low growth rate across the next days only reaching around 19MPa after 7 days.

Milled carbon exhibits the highest strength at the 1-day curing time, with a strength around 9MPa. Despite this high early strength, the rate of growth for the milled carbon is quite low, bringing the strength more in line with the others at 7 days. The

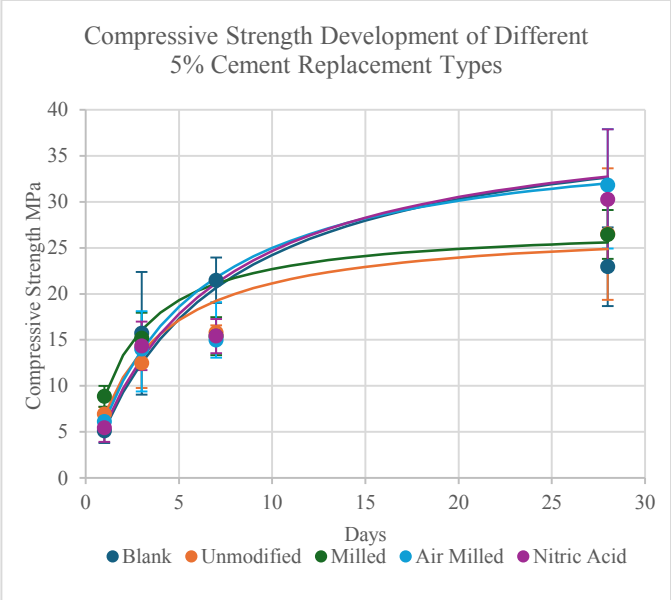


Figure 3. Compressive Strength Development Over 28 Days.

air milled carbon addition has a similar 1-day strength to the others, around 6MPa, but it has a high growth towards the 3-day cure, and then very low growth to 7 days, which will be later covered in *Discussion*. The nitric acid treatment shows a similar early-stage growth to the air milled, passing 20MPa after 7 days, following the trendline.

Final Strength (28 Days)

The blank reaches 23MPa after 28 days, the lowest strength found after 28 days. The unmodified carbon and milled carbon both have 28-day strengths of 26.5MPa, giving them similar characteristics for compressive strength. The air milled has the highest strength at 28 days, at 31.8MPa, considerably stronger than the previous modifications. Finally, the nitric acid modifications show a similar, but slightly low strength than air milled.

Ultimate Strength (S_u)

Table 3. 28-Day Compressive Strength Compared to Ultimate Strength

	Strength at 28-days/MPa	Strength at t=inf/MPa
Blank	23.0	32.3
Unmodified	26.5	27.6
Milled	26.5	27.5
Air Milled	31.8	38.0
Nitric Acid	30.3	40.0

As shown in the table above the ultimate strength, S_u, value of the compressive strength always exceeds the strength at 28 days. This indicates that a maximum compressive strength value will exist after a certain period after 28 days, thus supporting the equations used to produce the trendlines based on the data. This table used is excluding the 7-day data as stated before, the reason for which will be shown in *Discussion*.

3.2.2 Flexural Strength Development

The flexural strength development of the bars follows a similar trend to the compressive strength development, as shown in *Figure 4*, however the ultimate flexural strength is underestimated. The rate of strengthening for the different samples is similar for a low number of days, but over the whole 28 days all of the additions exhibit a greater strength development. The milled and air milled show the greatest strength improvement, as is shown in *Figure 4* where these are seen as the strongest after 28 days.

Early-stage strength (1~7 Days)

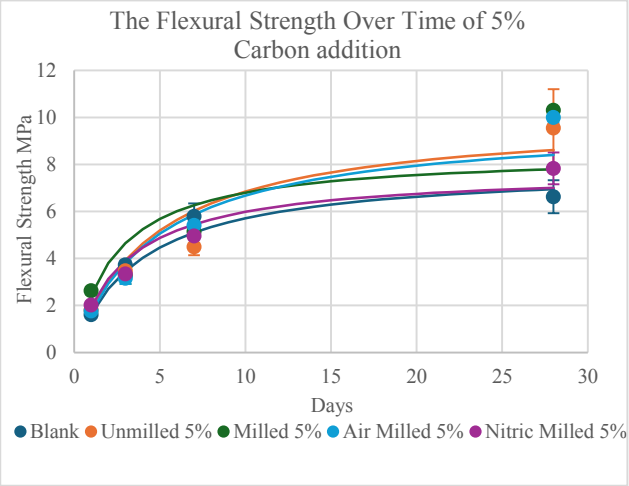


Figure 4. Flexural Strength Development Over 28 Days

The strength at 1-day for the blank is just under 2MPa, which is similar to the other modifications, with a lower growth rate up to 7 days. The milled carbon modification has the highest 1-day strength, at 2.5MPa, but with a low early-stage growth. The air milled and nitric milled samples both have very similar development, with the air milled gradually overtaking the nitric milled samples in strength.

Final Strength (28 Days)

The strength after 28 days shows a more accurate measure of the strength. The air milled, milled and unmodified are the strongest of the modifications, and the nitric acid and blank samples lag behind.

Metric for Comparison

The 7-day data is much more appropriate with the trend of the flexural strength, however as it is being discounted for the compressive strength it will not be used for correlations related to the flexural strength either.

The comparison between the flexural strength and compressive strength of the carbon, *Figure 5*, is a useful way to see the relative flexural strength, compared to what is expected. *Figure 5* shows that the unmodified and milled carbon have a slightly improved flexural strength than what is expected of the blank. Despite the high flexural strength when air milled carbon is added the percentage of the compressive strength is not much higher than that of the blank.

This will be a much more useful benchmark for flexural strength than S_u and k_t, as the fit is much better.

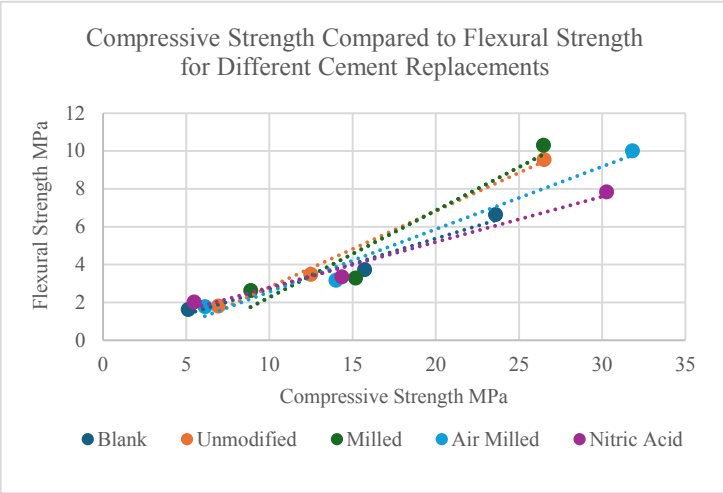


Figure 5. Compressive Strength Compared to the Flexural Strength

4. Discussion

4.1 28-Day Tests

The 28-day tests have clearly shown that with the carbon replacement of cement the flexural strength of the concrete is increased. Despite this, after the 5% replacement point the strength begins to decrease. This is in agreement with previous studies completed by (Salazar, 2024) where the compressive strength of the concrete was found to be improved by the addition of modified carbon, with the greatest increase in strength being for 2.5% and 5% replacement. This increase is likely due to an increase in active groups found in the carbon. According to data from, (Nikolis, 2024) the addition of carbon, modified and unmodified add additional C-O functional groups to the concrete mixture. This means that the drop in strength after 5% replacement is due to these functional groups not being suitable to replace the cement fully, just providing new sites for reaction when there is a high quantity of cement already. The milled, air milled, and nitric acid milled modifications have the greatest number of C-O groups, so these would be expected to, and have, the greatest strength. Despite this the nitric acid milled is considerably lower than the aforementioned modifications.

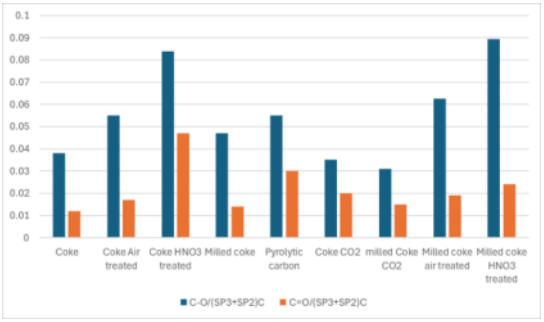


Figure 6. Prevalence of Different Active Sites for the Carbon Modifications, Courtesy of Dimitrios Nikolis (Nikolis, 2024)

The 5% replacement with both milled and air milled carbon were found to have the highest strength of the modifications, this is similar to what (Salazar, 2024) found for the compressive strength with these modifications. However, in the case of flexural strength the nitric acid milled coke provided a much worse improvement in strength, while the milled carbon provided a greater improvement. This could be due to the milled carbon filling more pores within the concrete, which as found by (Birchall, et al., 1981) filling a greater number of small pores for the same volume as a larger particle would give a higher flexural strength. The particle size distribution of the nitric acid milled coke could be studied to find if there is a difference in size that could contribute to this lower flexural strength.

4.2 Strength Development

Although the compressive strength development observed in this study show an overall increasing trend, it clearly shows differences between early and late-stage performance for each type of carbon modification. It demonstrates the role of hydration dynamics, particle characteristics and chemical modifications.

The trends are closely related to the hydration behaviours shown in the Isometric DSC data (Nikolis, 2024) shown in the graphs below, where cumulative hydration energy was calculated over 53 hours. The data found by (Nikolis, 2024)shows how the high the rate of reaction is over 53 hours. Whilst not necessarily the useful hydration, as this data lines up with the strengths that were found at 3-day curing time, (72 hours), it can be suggested that an increased hydration energy shows a high rate of early strength development.

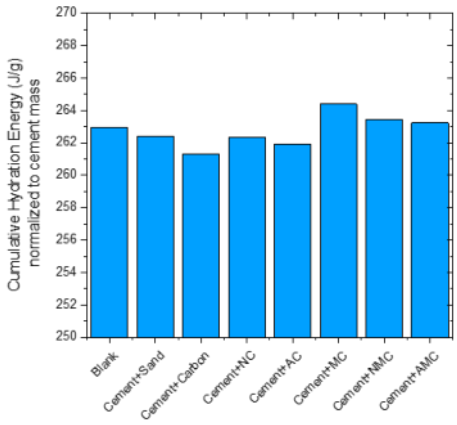


Figure 7. Hydration Energies of Different Carbon Modifications, Courtesy of Dimitrios Nikolis (Nikolis, 2024)

Blank achieved an intermediate hydration energy, thus resulting in intermediate rate of growth. untreated coke shows lowest hydration energy among the modified carbons, these line up with the blank and unmodified coke having the lowest strengths at 3 days. The milled coke having the highest hydration energy over this time corroborates the milled coke having the highest strength at 3 days. Air milled and nitric acid milled coke both have similar hydration energy, which again, lines up with their similar strengths at 3 days, being just below the milled coke. Based on the findings above, a clear correlation between the accumulation of hydration energy and the early-stage development is established.

The higher strengths, and hydration energies for the different milled samples suggests that the increased surface area gives a greater hydration energy, and therefore strength growth. This also supports the hypothesis that the modifications introduce a larger particle size, as the air milled and nitric milled samples have a lower hydration energy. Nitric milled coke having a higher hydration energy than air milled is likely due to a higher number of C-O active sites, as shown in *Figure 7*, despite the hypothesised larger particle size.

4.3 7-Day Data Exclusion

The graph for the compressive strength excludes the 7-day compressive strength values. The original chart including the 7-day data is shown below

Table 4. Strength at 28 Days Compared to Ultimate Strength, With 7-Day Data Included

	Strength at 28 days/MPa	Strength at t=inf/MPa
Blank	26.2	40.5
Unmodified	26.5	24.5
Milled	26.5	23.4
Air Milled	31.8	30.2
Nitric Acid	30.3	33.8

From the *Table 4*, unmodified, milled and air milled exhibits S_u values lower than the strength at 28 days which is infeasible data since the ultimate strength value should be higher than the 28-day strength value. This means that the inclusion of the 7-day compressive strength data does not correctly represent the ultimate strength equation, thus not fitting the trendline produced. This results in limited insights in the early-stage development. This is also supported by the very low growth between 3 and 7 days as shown in *Table 5*, where other examples, like (Salazar, 2024) and (Tank & Carino, 1991) show the strength should still have a high growth.

Table 5. Difference Between Compressive Strength at 3 and 7 Days

Sample	S_{3Days}/MPa	S_{7Days}/MPa	Difference/MPa
Blank	15.7	22.6	6.9
Unmodified	12.5	16.6	4.1
Milled	15.2	16.2	1.0
Air milled	14.0	15.9	1.9
Nitric Milled	14.3	16.3	2

The differences in *Table 5* show that the 7 days for the blank has a very high growth rate, unmodified is a reasonable increase and then the other 3 have a very low growth.

The issues with the 7-day data may be caused by an unexpected or different trend that is happening in the concrete. This would have to be different for the blank than for the carbon included samples, as the blank has an unexpectedly high strength, and the carbons have a low strength. It is more likely that there was uneven mixing within the concrete, or a worse curing time. There is also a possibility that the curing temperature was different for the sample, as they were cured at room temperature, which can fluctuate over time. The main takeaway is that more repetitions of the 7-day samples should be carried

out, and then these inconsistencies could be investigated.

Until this, excluding this data from the KPIs is necessary, as either the data is incorrect, or the model from (Tank & Carino, 1991) nor does it fit the carbon addition concrete.

4.4 KPIs

The experimental data of the compressive strength development excluding the 7-day data fitted well with *Equation 4*, which reflects the parameters S_u and Kt as key performance indicators.

According to the ACI 318 the flexural strength is usually between 10%-15% of the compressive strength (Kępniak & Woyciehowski, 2016)so this is the basis that will be used to compare how good the flexural properties of the different concrete types.

The previously mentioned additions with the greatest flexural properties, milled, air milled and unmodified, have the greatest gradients in *Figure 3*. This corresponds to the flexural strength being a greater percentage of the compressive strength. The overall percentages are shown in *Table 6*.

Table 6. Percentage of Compressive Strength at 28 days

Replacement Type	Percentage of Sc
Blank	28.1%
5% Unmodified	36.0%
5% Milled	39.0%
5% Air Milled	31.4%
5% Nitric Acid Milled	25.9%

The percentage of compressive strength gives insight into how appropriate for flexural applications each modification is. This supports the suggestion that the nitric milled carbon is a less appropriate addition for flexural strength. These percentages are considerably higher than the estimation from the ACI 318 (Kępniak & Woyciehowski, 2016). This will be due to different ratios and materials used to make the concrete.

The KPIs, key performance indicators, of the strength development of concrete are the ultimate strength. S_u, the rate constant, k_t and what comparison between the compressive strength and flexural strength. S_u denotes how high the strength can get, k_t the rate at which it strengthens, the higher the better. The comparison between flexural strength and compressive strength shows whether a material is predisposed to having better flexural strength than would be typical.

Table 7. KPIs for Compressive Strength for Different Modifications

Sample	S_u /MPa	k_t/day^{-1}
Blank	32.3	0.195
5% Unmodified	27.6	0.329
5% Milled	27.5	0.468
5% Air Milled	38.0	0.192
5% Nitric Acid Milled	40.0	0.160

The ultimate strength is highest for the air milled and nitric milled, with lines up well with the expectation that they are strong compressively, from the data in *Figure 3* and the findings from (Salazar, 2024). The lower values of S_u for the unmodified and milled samples compared to the blank, is likely due to the reduced C-O bonds as previously mentioned. Despite this, they have a greater strength at 28 days than the blank, due to the far greater rate of strength growth, shown by the k_t values.

The k_t values of unmodified and milled carbon samples are much higher than the others, the milled value being higher can be explained by the DSC data in *Figure 7*. The greater reaction rate for the first 3-days is shown with the higher k_t value. The unmodified sample is more difficult to explain, however it may be due to the low strength it can reach, see the S_u value, *Table 7*, causing a similar numerical increase in strength to have a high rate constant.

The linear regression to find the S_u and k_t values was completed over only 3 data points, so there is potential for errors in the unmodified results, see the high standard deviation shown in the error bars for *Figure 3*. This is another factor that could cause an exaggeration of the potential of the unmodified coke addition.

5. Conclusions

5.1 Flexural Strength

With the results and discussion above, the most effective modification for maximising flexural strength at 28 days is the milled coke. This is due to both its high flexural strength, and the ease of modification. As it is just as strong as the air milled coke in terms of flexural strength, and only requires one step to modify it.

If compressive strength is also taken into account, as it would be in most cases, the air milled comes out as the most effective. This is due to having a high flexural and compressive strength, with a minor compromise in terms of flexural strength. The slightly more difficult modification is not a big issue in this case, as air treating the milled coke is very simple, just heating in the presence of air. This could be useful in construction applications, specifically roads, bridges and beams (Atlas Fibre, 2023). Using a higher strength concrete could reduce the amount of reinforcement required to support tensile loads, and therefore also reduce the cost of reinforced concrete.

5.2 Strength Growth

The KPIs however suggest which modifications are the most effective. The air milled and nitric milled give the highest ultimate strength, and therefore would be the most suitable for use. The high strength development of the milled and unmodified could have uses where short term strength is required. It is possible.

Despite a low rate constant for the air milled and nitric acid milled, the higher ultimate strength will also contribute to improve increasing the rate of strength gain, so they still have acceptable strength growth.

The discrepancy with the 7-day data may suggest that there is a different trend than the one suggested by (Tank & Carino, 1991), as the addition of carbon may impact the growth rate.

6. Outlook and Recommendations

6.1 Improvements to experimental approach

The experimental approach for the project could be improved with a good labelling system for the bars. During the project, the high moisture environment caused markings on the bars to wash off, so use of a waterproof labelling system, such as laminated labels could be implemented.

6.2 Further Research

For future research in the subject, it is recommended that the 7-day samples get repeated, along with carrying out the tests for other curing times such as 14 days. This would allow for the trendlines to be verified, and the least squares regression to be found more accurately.

If, upon repetition of the 7-day samples, the values are found to be consistent with the findings in this report a further analysis of the strengthening over time should be carried out. To do this a study of many different curing times could be carried out, to find an empirical trend.

A particle size analysis of the modified carbon could also be carried out, to find if the reduced strength of the nitric acid treated coke addition is due agglomeration, as suggested in *Discussion*. Furthermore, the pores in the concrete could be investigated, using microscopy to see whether this flexural strength change is due to these pores.

Additionally, research into other effects of these modifications could be carried out. For example, the effects on hydrophobicity or workability could be investigated.

7. Acknowledgements

The authors would like to thank Dimitrios Nikolis for his guidance and support during the project.

References

ACI Committee 209, 1997. *Prediction of Creep, Shrinkage,,* s.l.: ACI.

ASTM International, n.d. *Standard Test Method for Flexural Strength of Concrete,* s.l.: ASTM International.

Atlas Fibre, 2023. *Understanding Flexural Strength: Guide to Flexural Strength in Materials,* s.l.: Atlas Fibre.

Birchall, J. D., Howard, A. J. & Kendall, K., 1981. Flexural Strength and Porosity of Cements. *Nature,* Volume 289, pp. 388-390.

Callister, W. D. J. & Rethwisch, D. G., 2018. *Materials Science and Engineering : an Introduction.* 10th Edition ed. s.l.:Hoboken, NJ; Wiley.

Hou, C. et al., 2023. Rigid-flexible double coating silicon oxide composed of pitch pyrolytic. *Advanced Composites and Hybrid Materials,* 6(143).

IEA, 2024. *Global Hydrogen Review, Hydrogen Demand,* s.l.: IEA.

Kępniak, M. & Woyciehowski, P., 2016. The Statistical Analysis of Relation Between Compressive and Tensile/Flexural Strength of High Performance Concrete. *Archives of Chemical Engineering,* 62(4).

Mlonka-Medrala, A. et al., 2021. Pyrolysis of agricultural waste biomass towards production of gas fuel and high-quality char: Experimental and numerical investigations. *Fuel,* Volume 296.

Nikolis, D., 2024. *Previous Work Completed In The Lab,* s.l.: s.n.

Potter, B., 2023. *There Will Soon Be More Concrete Than Biomass on Earth,* s.l.: Heatmap.

Purton, M., 2024. *Cement is a big problem for the environment. Here's how to make it more sustainable,* s.l.: World Economic Forum.

Salazar, J. P., 2024. *Evaluation of the Effects of Coke and Surface-Modified Coke as a Cement and Sand,* London: Department of Chemical Engineering, Imperial College London, U.K..

Sánchez-Bastardo, N., Schlögl, R. & Ruland, H., 2021. Methane Pyrolysis for Zero-Emission Hydrogen Production: A Potential Bridge Technology from Fossil Fuels to a Renewable and Sustainable Hydrogen Economy. *Industrial & Engineering Chemistry Research,* 60(32).

Sowmya, C. B., 2020. *Flexural Strength of Concrete,* s.l.: CivilEngineering.org.

Statista, 2024. *Production of cement in the United States from 2010 to 2023,* s.l.: Statista.

Tank, R. C. & Carino, N. J., 1991. Rate Constant Functions for Strength Development of Concrete. *ACI Materials Journal,* 88(1), pp. 74-83.

Vine, D., 2024. *Methane pyrolysis: the case for cleaner hydrogen with existing infrastructure,* s.l.: C2ES.

Weerheijm, J., 2013. *Understanding the Tensile Properties of Concrete,* s.l.: Woodhead Publishing.

Pluronic F-127 Hydrogels for Delivery of Lipid Nanoparticles Containing mRNA

Daniel Penn and Sarah Hyndman
Department of Chemical Engineering, Imperial College London, U.K.

Abstract Messenger RNA (mRNA) delivery has attracted significant attention due to its diverse range of theraputic applications, yet its efficient and targeted delivery remains a challenge. This work explores a novel approach, combining lipid nanoparticles (LNPs) with a Pluronic F-127 hydrogel to form a lipogel for injectable and localized delivery. The gels were evaluated based on their gelation temperature, particle size distribution, strength, and transfection efficiency. A number of Pluronic-based conjugated hydrogels were synthesized, incorporating bovine serum albumin (BSA), hyaluronic acid (HA), and sucrose to modulate gelation temperature. It was found that BSA and HA increase the gelation temperature, while sucrose had a less significant opposite effect. Synthesizing the gel in phosphate-buffered saline (PBS) or tris-buffered saline (tris) significantly reduced the gelation temperature compared to water. Dynamic light scattering (DLS) revealed that conjugated hydrogels exibited a greater degree of polydispersity, while pure Pluronic hydrogels displayed two size populations (7nm and 40nm). Upon incorporation of LNPs to form a Pluronic-LNP lipogel, a single peak was attained, with an average particle size of 80-90nm. Rheology measurements showed that the lipogels had a suitable strength, at around 3500-4200 Pa, and were stronger than pure Pluronic hydrogels. Pluronic F-127 with C24 ionizable lipid-based LNPs was able to achieve a transfection efficiency 54% of a lipofectamine control. The results demonstrate potential of lipogels as a targeted mRNA delivery vehicle, however, further characterisation and testing is needed.

Keywords – Pluronic F-127, Lipogel, Hydrogel, Lipid Nanoparticles, Injectable, mRNA, Drug Delivery

Introduction

Hydrogels have been investigated as a method of drug delivery due to their biocompatibility and tuneable properties, allowing for controlled drug release (Zhong et al., 2023). They consist of a network of natural or synthetic polymers which swell in water and are formed through either physical or chemical crosslinking (Desfrançois, Auzély and Texier, 2018). They allow for either continuous release through manipulating properties such as pore size distribution, molecular weight, and crosslinking density, or stimuli-based release upon a change in pH, radiation, or the presence of enzymes (Zhong et al., 2023).

There are over 100 available hydrogel products that have been approved by the FDA (Clegg et al., 2024), with applications in RNA delivery for wound healing, cancer treatment, as well as bone and cardiac repair (Zhong et al., 2023). The structure of hydrogels allow for targeted application of a drug, enabling it to be retained where it is needed (Zhong et al., 2023).

For a hydrogel to be injectable, it should be a liquid as it is injected, and then form a gel in situ around body temperature (37°C). They must also be non-toxic, biocompatible, and have good mechanical properties to withstand deformation in the body (Alonso et al., 2021). One such hydrogel of interest for injectable drug delivery systems is Pluronic F-127, due to its stability, thermo-reversibility, and gelation at body-temperature (Jaquilin P J et al., 2022). Furthermore, the preparation of Pluronic hydrogels does not require toxic initiators or radiation, as seen in the synthesis of other synthetic hydrogels, and instead can be prepared through simple dissolution in cold water.

Messenger RNA (mRNA) is emerging as a new method to treat and prevent disease, having a wide range of therapeutic applications, including cancer treatment, vaccines, and protein replacement therapy (Hou et al., 2021). However, the delivery of mRNA faces significant barriers, as its large size and charge makes it difficult to permeate the cell membrane and escape endosomes to reach the cytoplasm (Dowdy, 2017). Furthermore, mRNA is prone to degradation in vivo by enzymes and the immune system (Hou et al., 2021). Therefore, to improve the transfection efficiency while exploiting the advantageous properties of hydrogels, one can incorporate a secondary carrier inside the hydrogel matrix, such as a lipid nanoparticle.

Lipid nanoparticles are another novel method of drug delivery, with the ability to encapsulate RNA through electrostatic interactions (Hou et al., 2021). They are comprised of a single layer of surfactants surrounding a lipophilic core (Desfrançois, Auzély and Texier, 2018) and can range in size between 40-1000nm (Tenchov et al., 2021). They can contain lipids in various proportions, typically including cationic and ionizable lipids, as well as polyethylene glycol, cholesterol, and phospholipids (Hou et al., 2021). The components in LNPs are FDA and EMA-approved due to their high biocompatibility (Desfrançois, Auzély and Texier, 2018), with LNPs being used in both the Moderna and Pfizer-BioNtech COVID-19 vaccines (Silva-Pilipich et al., 2024).

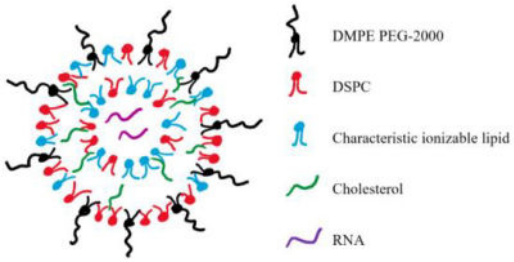


Figure 1. LNP Structure

LNPs complement the features of hydrogels, as they can protect RNA from degradation and facilitate endosomal escape (Hou et al., 2021). Therefore, a system combining LNPs into a Pluronic F-127 hydrogel, forming a lipogel, may be a viable method of mRNA delivery that combines both the tuneable and targeted drug delivery properties of hydrogels with the transfection efficiency of LNPs.

Background

This investigation evaluates a novel injectable drug delivery system comprised of LNPs in Pluronic F-127 hydrogel, aiming to expand the scope of therapeutic applications of mRNA.

Pluronic F-127 is a tri block copolymer with a poly(propylene oxide) (PPO) core, flanked on either side by poly(ethylene oxide) (PEO) tails (Jaquilin P J et al., 2022). It is soluble in cold water due to hydrogen bond formation between water molecules and the hydrophilic PEO segments, but when heated, the interactions between the hydrophobic PPO chains begin to dominate and result in the formation of micelles (Jaquilin P J et al., 2022). These micelles are formed with the hydrophobic chain oriented inwards, and the hydrophilic chain oriented outwards (Figure 2). As the temperature continues to rise, the micelles assemble into a regular packed structure, forming a gel (Jaquilin P J et al., 2022).

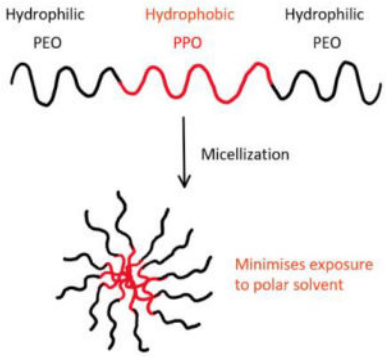


Figure 2. Pluronic structure and micellization

Pluronic can also be physically mixed or chemically crosslinked with other components to form a conjugated gel with modified mechanical properties. In this investigation, conjugated hydrogels were formed using bovine serum albumin (BSA), sucrose, and hyaluronic acid (HA).

BSA is a biocompatible hydrophilic protein that can be easily incorporated into Pluronic to form mixed micelles of a larger size, thus influencing diffusion rates of encapsulated drugs, and can be used to modulate gelation temperature (Hirun, Kraisit and Soontaranon, 2023; Zhou et al., 2011). Sucrose was used as it has been shown to help stabilize and maintain efficacy of lipid nanoparticles in storage (Kafetzis et al., 2023). Finally, HA was used as it has been shown to improve the strength of Pluronic hydrogels through its involvement in inter-micellar packing (Jung et al., 2017). All of these components can be combined into Pluronic hydrogel simply through physical mixing, meaning that RNA-loaded LNPs can also be included within the system without risk of degradation during the synthesis process.

Four LNPs were incorporated into Pluronic hydrogel in this study (Table 1), each with different ionizable lipids. These are key components for cellular delivery as they become protonated in the lower pH environment within endosomes, which can destabilize the membrane and allow for release of mRNA (Hou et al., 2021). C12-200 and Dlin-MC3-DMA were selected as these are commonly used in LNP formulation and are considered benchmark lipids (Hajj et al., 2020), with MC3 lipids being used in the first FDA approved siRNA drug (Hou et al., 2021). 306-O12B was selected as this has been identified by studies to be a promising lipid for RNA delivery, demonstrating higher efficiency than Dlin-MC3-DMA (Qiu et al., 2021). Finally, C24 is a more novel lipid with high stability and multiprotic properties, which may facilitate better endosomal escape (Alishetty et al., 2021).

Previous research has seldom incorporated LNPs into Pluronic hydrogel. In cases where they have been used, this has primarily been studied for non-injectable systems (Hao et al., 2014; Almeida et al., 2016; Yu et al., 2018; Din et al., 2015). Furthermore, research using Pluronic to deliver mRNA is limited, and is more frequently studied using small interfering RNA (siRNA). Kadekar et al. (2021) delivered siRNA by conjugating it with Pluronic using a disulfide linkage, and Lan et al. (2021), incorporated glycogen-encapsulated siRNA for wound healing. However, there is no research into the use of Pluronic hydrogels to deliver LNPs containing large mRNA molecules.

In this investigation, the concentration of Pluronic in water, phosphate-buffered saline (PBS), and tris-buffered saline (tris) was optimized to achieve physiological gelation temperatures (37°C). Then, Pluronic was combined with additives including sucrose, BSA, and HA. The gels which attained

desired gelation temperature through visual tests were sized using DLS. Finally, the strengths of the gels were measured, and a cell transfection was performed.

Materials and Methods

1. Materials

Pluronic F-127 (12600Da), sucrose, and BSA were purchased from Sigma-Aldrich, Italy. A Milli-Q water purification system was used to produce pure water. Dulbecco's Phosphate Buffered Saline (PBS) and Tris-buffered saline (tris) were obtained from Thermo Fisher Scientific. Hyaluronic acid sodium salt (200-400kDa) was purchased from Glentham Life Sciences. For the purposes of initial DLS, rabies saRNA was used. For the transfection, fLuc mRNA was used.

2. Synthesis of Pluronic hydrogel via 'cold method'

To synthesize the Pluronic hydrogels, the 'cold method' from Schmolka (1972) was followed. Pluronic was slowly added to a beaker of solvent (water, PBS, or tris) held in an ice bath, with gentle mixing until the solution became clear and homogenous. Once the Pluronic fully dissolved, additional components were added through physical mixing to form a conjugated hydrogel (0.5-2 %w/v BSA, 5-10 %w/v sucrose, and 0.5-1.5 %w/v HA).

To incorporate LNPs into the hydrogels to form lipogels, a 40%w/v Pluronic hydrogel in PBS was first synthesized. Then, a measured volume of LNP in PBS was added and mixed into the hydrogel in an ice bath, such that the desired final concentration of both Pluronic and LNP was achieved.

3. Synthesis of LNPs

LNPs were prepared using the standard method outlined by Barbieri et al. (2024). All LNPs were prepared by mixing a characteristic ionizable lipid (C12-200, C24, 306-O12B or Dlin-MC3-DMA) with a helper lipid (DSPC), cholesterol and DMPE-PEG2000. The molar ratios of these components and the final RNA to lipid ratios are shown in Table 1.

Stock solutions of all of the lipid components were prepared in absolute ethanol and mixed in excess relative to the RNA. RNA solutions were prepared in an excess of sodium acetate and sodium chloride buffer at a pH of 5.5. The LNPs were then formulated in an Ignite microfluidics device (Precision NanoSystems Incorporated) with a flowrate of 8mL/min and a flowrate ratio of RNA:lipid of 3:1. The formulation was then concentrated in PBS using a VivaSpin 6 centrifugal concentrator at 4000rpm and 18°C until the desired concentration of RNA was achieved.

4. Gelation temperature measurement

To measure the gelation temperature of the hydrogel, a sample was placed inside a Benchmark Myblock Mini Dry Bath, and the temperature was slowly increased from 25-37°C. The sample was inverted at every 1°C degree increment, and the temperature at which the gel remained in a solid state when inverted was recorded as the gelation temperature (Figure 3). The time taken to un-gel at room temperature was also recorded. The target range for gelation temperature was chosen to be 30-37°C, such that it gels at body temperature, but not at room temperature.

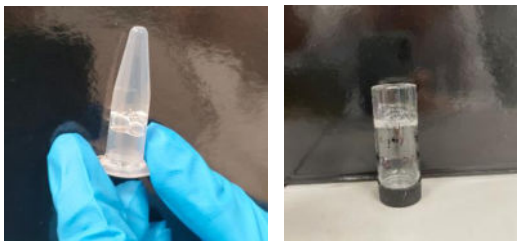


Figure 3. Inversion test on Pluronic hydrogel

5. Dynamic Light Scattering

Particle size, polydispersity index (PDI), and zeta potential were measured using a Malvern Zetasizer. The hydrogel or lipogel sample was added to a DTS1070 cuvette at a 1:100 dilution using PBS. Measurements were run at 25°C. For lipogel samples, a concentration of 10µg/ml of rabies saRNA inside LNPs was used with varying concentrations of Pluronic. This was selected as doses of 5µg of saRNA have been successful in clinical trials (Schmidt and Schnierle, 2023), and the typical vaccine volume is 0.5ml (Oxford Vaccine Group, 2022).

Table 1. Composition of lipid nanoparticles

LNP	mol% Ionizable Lipid	mol% DSPC	mol% Cholesterol	mol% DMPE-PEG2000	mRNA/lipid (w/w) ratio
C12-200	35	16	46.5	2.5	0.024
C24	50	10	38.5	1.5	0.038
306-O12B	50	10	38.5	1.5	0.030
Dlin-MC3-DMA	50	10	38.5	1.5	0.045

6. Rheology

The procedure from Shmool et al. (2022) was followed to measure the strength and loss moduli of the gels. A TA-Discovery HR-1 hybrid rheometer was used with a 40mm Peltier steel plate. The strain was set to 1% and the angular frequency to 1 radian per second. Lipogel samples of 1.3ml were used for testing, containing 18%w/v Pluronic and a fLuc mRNA concentration of 0.75µg/ml.

7. RNA Transfection Efficiency

The transfection was carried out using the standard method outlined by Barbieri et al. (2024). HEK 293T/17 cells were seeded in 96-well plates at 50×10^3 cells per well. The next day, a sample of 18%w/v Pluronic with 10µg/ml of RNA inside LNPs was added such that a total of 100ng of fLuc mRNA was added to each well. The transfection ran for 24 hours. 50µL of medium was removed from each well and replaced by 50µL Bright-Glo™ luciferin substrate. Five minutes later, 100µL of each well was transferred to white plates, and a FLUOstar Omega plate reader was used to measure luminescence intensity.

Results and Discussion

1. Gelation temperature

Figure 4 shows the effect of the concentration of Pluronic and the solvent choice on the gelation temperature of the hydrogel. It was found that increasing the concentration of Pluronic resulted in a decrease in the gelation temperature, which is in agreeance with previous studies (Jung et al., 2017). As more Pluronic is dissolved in the solution, the micellization more readily occurs, and less heat is needed to facilitate the formation of the micellar lattice.

Water, PBS, and tris were investigated as solvents for Pluronic hydrogel synthesis. Natural water is a poor solvent for storing RNA, as deviations in pH and the presence of ions catalyse hydrolysis, leading to degradation (Chheda et al., 2023). Therefore, it would be advantageous for the excipient in an RNA delivery system to also function as a buffer.

Generally, the ideal storage buffer for RNA is Tris with ethylenediaminetetraacetic acid (EDTA), to chelate with free calcium ions (Chheda et al., 2023). This is typically kept at -80°C, to slow down hydrolysis, enzymatic or other degradation reactions to a standstill (Chheda et al., 2023).

However, it was found that Tris buffer gelled at far too low temperatures to be injectable. PBS had a gelation temperature much closer to that of water, and was chosen as the ideal candidate to control pH in an injectable gel. This agrees with previous studies that found that increasing the ionic strength of a solution of Pluronic hydrogel lowered its gelation temperature (De Neve et al. 2017).

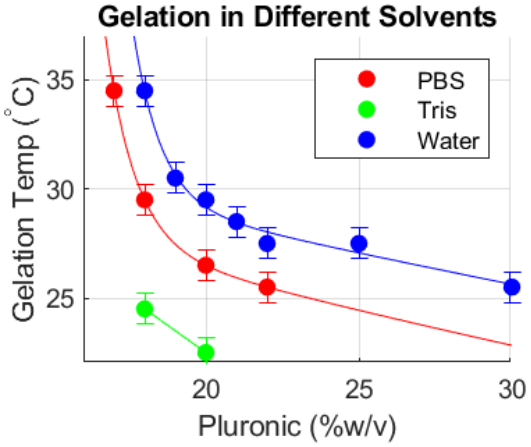


Figure 4. Gelation temperature of Pluronic in various solvents

Table 2. Time to un-gel for water and PBS hydrogels

%w/v Pluronic	Time to revert in water (min)	Time to revert in PBS (min)
18	13	2
19	14	2
20	11	4

The time taken for the hydrogels to un-gel was also recorded (Table 2). It was noticed that PBS derived hydrogels reverted to liquid much faster than their water-based counterparts. Adding additional components into the hydrogel also decreased the time to revert, though the solvent was a bigger factor.

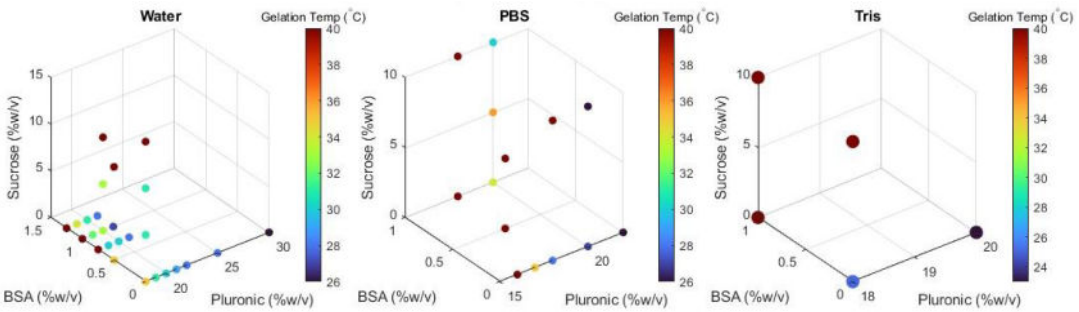


Figure 5. Gelation temperature of conjugated gels

Table 3. 19%w/v Pluronic hydrogels in water

%w/v BSA	Gelation temperature (°C)	Time to revert (min)
0	30	11
0.75	30	10
1	32	10
1.25	34	6

Table 4. 18%w/v Pluronic hydrogels in PBS

%w/v HA	Gelation temperature (°C)	Time to revert (min)
0	30	<1:00
0.5	36	3:30
1	37	1:30
1.5	39	1:30

Figure 5 shows the gelation temperatures of various conjugate hydrogels containing Pluronic, sucrose, and BSA in the three solvents. It was found that BSA generally increased the gelation temperature (Table 3), agreeing with previous research by Hirun, Kraisit and Soontaranon (2023), which they hypothesize is due to steric hindrance imposed on the micellization by the BSA. However, BSA was found to precipitate in some samples after a week, posing storage concerns.

HA was also found to increase the gelation temperature of Pluronic hydrogels (Table 4), which is in agreeance with literature by Ramya et al. (2019). They state this may be because the HA affects the micellization process and hinders the formation of a regular lattice. Interestingly, this contrasts research by Balan, Precupas and Matei (2023), which found that using high molecular weight HA (1500–2200kDa) increases the gelation temperature, as the longer chains fill the space between Pluronic micelles, reducing the interactions between Pluronic and water, thus promoting micellar packing. Thus, the molecular weight of HA may be a variable that can be manipulated to control the impact on gelation temperature.

Sucrose had a less noticeable reducing effect on the gelation temperature up until a critical point of 15%w/v where it prevented gelation. This is likely due to the extremely hydrophilic nature of sucrose, with a stronger affinity to water than that of Pluronic. The sucrose therefore dehydrates the PEO-PPO interface, promoting aggregation of hydrophobic PPO chains and in turn micellization (Sarkar et al., 2013).

2. Dynamic Light Scattering

Figure 6 shows the particle size distribution of a 20%w/v Pluronic hydrogel in water. Two peaks were found at around 7nm and 40nm. This is in agreeance with the findings from Constantinou et al. (2022), which found that unimers of Pluronic had a size of 9nm, and micelles of 30nm. They also found that the endothermic micellization process occurs around 12°C, thus explaining the presence of two peaks at the measurement temperature of 25°C.

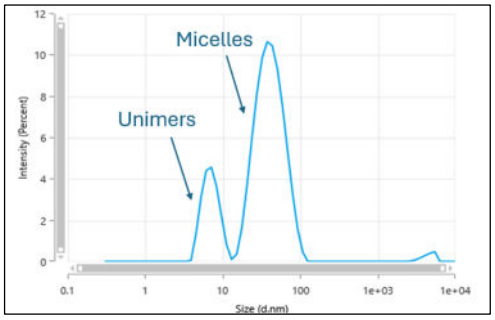


Figure 6. Pluronic gel characteristic size in water

Table 5 shows the particle size and zeta potential of various conjugated Pluronic gels. It was found that pure Pluronic gels had the lowest average particle size and lowest polydispersity index, while mixed gels saw an increase in both. A single peak with low PDI is indicative of low aggregation, such that particles are evenly distributed, which is ideal for delivery and reduced immunogenicity (FDA, 2014).

The FDA terms particles in the 2–10-micron range as subvisible, noting that they are much more likely to cause an immune response (FDA, 2014). As conjugated Pluronic gels with BSA, sucrose and hyaluronic acid formed aggregates within this range and so were discarded. Thus, only the pure Pluronic hydrogels were combined with LNPs for further investigation.

Table 5. Particle size and zeta potential of various gel formulations in PBS

Gel Components (%w/v)	No. of peaks	Average Particle Size (nm)	Polydispersity Index (PDI)	Zeta Potential (mV)
20% Pluronic	1	7.44	0.187	-14.09
20% Pluronic 1% BSA	4	43.30	0.294	-12.50
18% Pluronic 1% HA	3	24.72	0.704	-9.46
20% Pluronic 1% BSA 5% Sucrose	2	11.67	0.233	-8.58

Table 6 shows two lipogels and their particle size, PDI, and zeta potential. The LNPs themselves have a size between 80 and 160 nm and a near zero zeta potential. The concentration of Pluronic was re-adjusted when incorporated with LNPs to achieve sufficient gelation temperature.

It was found that the zeta potential of the gels increased slightly after incorporating LNPs. The size

of lipogels exhibited broad single peaks (Figure 7) in some experiments, and sharper double peaks in others (Figure 8). This is likely an artifact of the mixing technique, as while all samples were vortexed to thoroughly mix the LNP and hydrogel, mixing durations were not consistent which may be an important factor in preventing aggregation. In the future a low shear mixing technique should be explored. Nevertheless, promising candidates presenting a single peak were identified and investigated further in rheology and transfection tests.

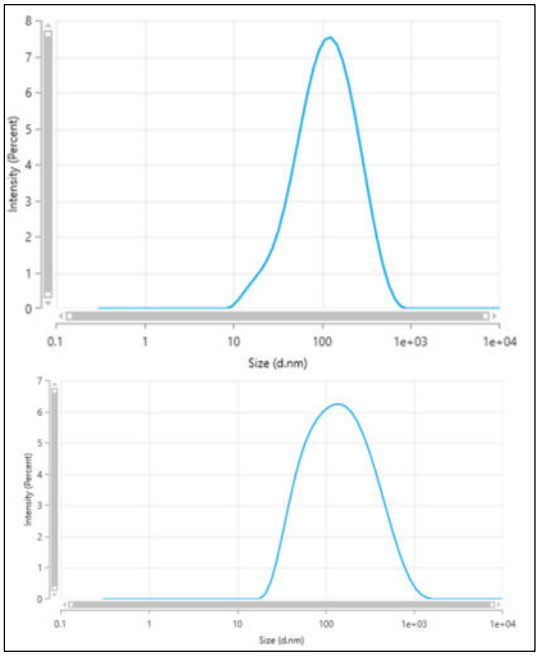


Figure 7. Particle size of Pluronic hydrogel with MC3 LNPs (top) and C24 LNPs (bottom) displaying a single peak

Table 6. Particle size and zeta potential of lipogel formulations

Gel components	Average Particle Size (nm)	Polydispersity Index (PDI)	Zeta Potential (mV)
20% Pluronic MC3 LNP	80.61	0.422	-2.00
16% Pluronic C24 LNP	89.09	0.508	-10.20

The ability of the system to recover after gelation was also studied. A sample of MC3-based LNP lipogel was measured prior to gelation, then heated to 31 °C, allowed to un-gel at room temperature, then was tested again. Finally, a third measurement was taken 30 minutes later. Figure 8 shows that while immediately after gelling and un-gelling, some aggregates of larger sizes were present, these began to dissociate and revert to the pre-gelation state after 30 minutes. This demonstrates that the hydrogel remains thermo-reversible, even when incorporated with LNPs, thus supporting the system's overall stability in structure.

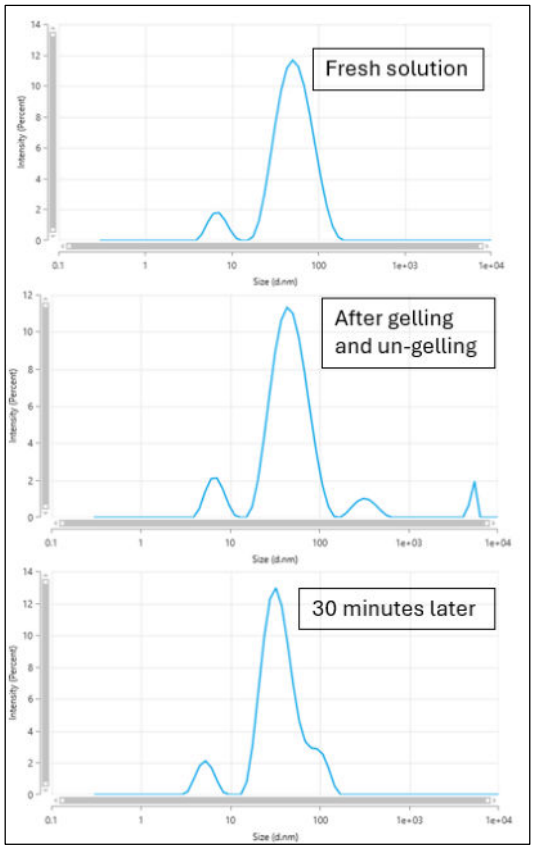


Figure 8. Particle size of 20%w/v Pluronic gel containing MC3-based lipid nanoparticles

3. Rheology

The strengths of the lipogels were measured to evaluate the suitability of their mechanical properties for injection and stability in the body. The storage modulus (G') represents the elastic storage of energy and is used to represent the strength of a material, while the loss modulus (G") reflects energy dissipation. Figure 9 shows the storage and loss moduli of 18% Pluronic hydrogel, while Figure 10 compares lipogels with 4 different LNPs.

It was found that pure 18%w/v Pluronic hydrogel achieved a strength of 1123 Pa at 37°C. The storage modulus is lower than that found in literature by Shriky et al. (2020), who found that the strength depends on concentration and reached around 3000 Pa for a 15 w/w% Pluronic hydrogel. This difference may be explained by the differences in solvent, as they used water as opposed to PBS, which may have increased the strength of their gel.

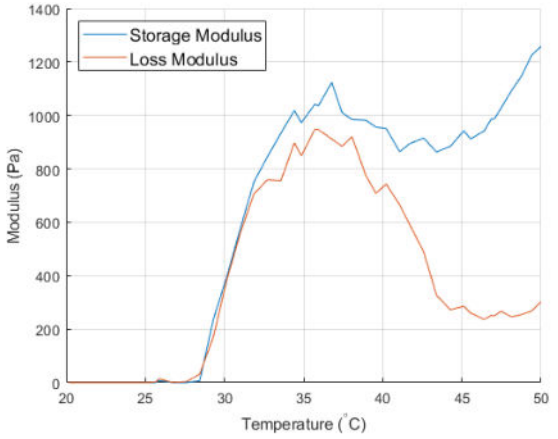


Figure 9. Storage and loss modulus of 18%w/v Pluronic hydrogel

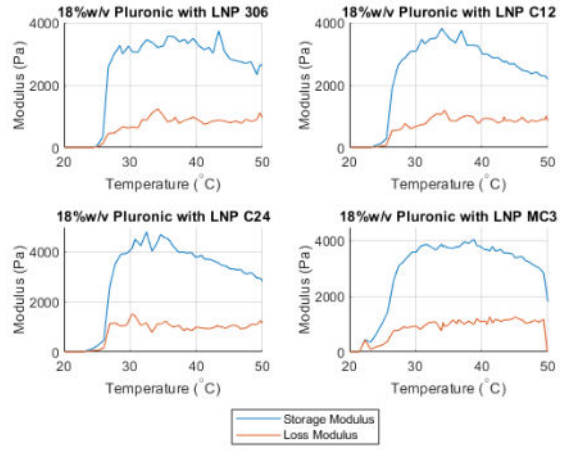


Figure 10. Storage and loss modulus of 18%w/v Pluronic hydrogel with lipid nanoparticles

The value of G' peaks and plateaus for all of the lipogels at around 30°C, consistent with visual observations of gel formation at this temperature. This corresponds to the point where the solid gels are fully formed. In contrast, the hydrogel does not exhibit a plateau in G', and the values of G' and G" are much closer, indicating significant viscous dissipation similar to a liquid. This explains why the hydrogel with PBS reverts quickly as the solid gel phase remains weak.

Table 7. Maximum storage modulus of Pluronic hydrogels

Hydrogel	Storage modulus at 37°C (G', Pa)
18% Pluronic	1123
18% Pluronic, MC3	3768
18% Pluronic, C12	3753
18% Pluronic, C24	4235
18% Pluronic, 306	3520

The final gel strengths at body temperature are found in Table 7. All LNPs improved the storage modulus of the hydrogel by over 3-fold. This may arise from favorable interactions and entanglement between the structurally analogous PEG chains on the LNP exterior and PEO chains on the micelle exterior. The introduction of LNPs into the hydrogel would then prevent Pluronic micelles slipping past one another, forming a stronger solid.

In a review of injectable hydrogels, Overstreet et al. (2012) proposed that 1000 Pa is low for a storage modulus, but achieving 4000 Pa is suitable. Therefore, the C24-based lipogel shows promising rheological properties.

4. RNA Cell Transfection Efficiency

The relative luminescence of transfection for 4 different lipogels can be seen in Figure 11. The transfection reagent lipofectamine was used to delivery fLuc mRNA as a positive control and Pluronic hydrogel loaded with fLuc without LNP encapsulation was used as a negative control.

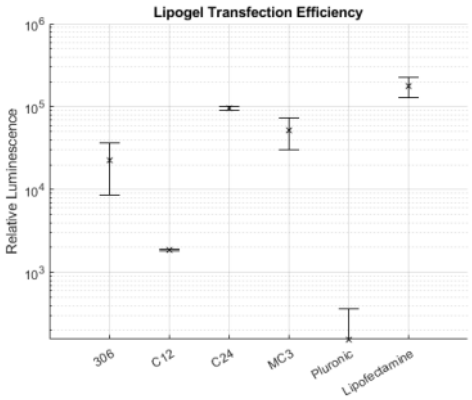


Figure 11. Relative luminescence from cell transfection using lipogels to deliver fluc RNA

All lipogels were able to transfect, with C24 and DLin-MC3 showing particularly high transfection efficiencies. C12-200 showed consistently poor transfection efficiency while 306-O12B showed fair transfection efficiencies but over a broad range.

Structural diversity between the different ionizable lipids leads to different protonation mechanisms, which cause large differences in endosomal escape and observed transfection efficiency. This supported by Alishetty et al. (2021), who found that C24 undergoes multiple protonation stages that allow for a more efficient delivery of cargo into cells, explaining why it had the highest transfection efficiency of all lipogels.

The poor transfection results from the C12-200 LNP agrees with the findings of Barbieri et al. (2024), who investigated the transfection efficiency of LNPs alone, and found that MC3 resulted in a similar level of expression to lipofectamine, whereas C12 had a relative luminescence an order of magnitude lower. More research is needed to understand whether the different ionizable lipids affect interactions between

LNPs and Pluronic micelles, in addition to the differences in cellular uptake and release abilities.

The C24-based lipogel achieved a transfection efficiency of 54% respective to the positive lipofectamine control. Barbieri et al. (2024) found that MC3 LNPs alone achieved a transfection efficiency of 65% of the lipofectamine control. Thus, in comparison, lipogels have a slightly reduced transfection efficiency. Nevertheless, considering the benefits of hydrogels for localized delivery, a sufficiently high transfection efficiency was obtained to demonstrate the potential of lipogels for mRNA delivery.

Conclusion

In this work an injectable lipogel comprised of Pluronic F-127 loaded with LNP was developed and proven to successfully deliver mRNA to mammalian cells.

Increasing the concentration of Pluronic in the hydrogel was found to lower gelation temperature and a minimum concentration of 16%w/v Pluronic was required to facilitate gelation. Both PBS and water-based Pluronic hydrogels had suitable gelation temperatures to be injectable (30-37°C), however using tris as a solvent resulted in far too low gelation temperatures. The solvent also had a large impact on the time for gels to revert, with buffers reverting 5 times faster than pure water. Adding BSA to the hydrogels increased the gelation temperature but posed precipitation issues over time. Sucrose was found to have a minor reducing effect on gelation temperature up to 15%w/v where it prevented gelation. HA was found to have an increasing effect on gelation temperature, though this may heavily depend on the chain length of HA, which needs further exploration.

DLS of pure Pluronic yielded two distinct peaks around 7nm and 40nm, with a polydispersity index of 0.19. Mixing BSA, sucrose or HA with the Pluronic resulted in additional peaks, indicating the formation of aggregates. Furthermore, the measurements of mixed gels also had reproducibility issues, indicating that the gels formed were non-uniform and less stable. The incorporation of LNPs into Pluronic hydrogel formed a single, but much broader peak compared to pure Pluronic. Notably, 16%w/v Pluronic with C24 derived LNP had a single peak, with an average particle size of 89nm and a polydispersity index of 0.51. 20%w/v Pluronic with MC3 derived LNP also had a single peak, with an average particle size of 81nm and a polydispersity index of 0.42.

Rheology showed that pure 18%w/v Pluronic hydrogel exhibited a storage modulus of 1123 Pa at 37°C. Incorporating LNPs into the gel improved this strength by more than three-fold for every LNP tested. The strongest lipogel was that of 18%w/v Pluronic with a C24-derived LNP with a storage modulus of 4235 Pa at 37°C.

All four LNPs were capable of cell transfection in a 18%w/v Pluronic hydrogel, but the C24-derived LNP was consistently the best, with an average relative luminescence 54% that of the positive lipofectamine control.

Outlook

One limitation of the experiment was that due to the equipment available, DLS was only run at 25°C. In the future it would be beneficial to run tests up to 37°C (body temperature) to get a better understanding of the size of the aggregates formed in situ, which is important from an immunogenicity perspective. Furthermore, to observe the internal structure of the lipogels and the degree of aggregation, it would be beneficial to lyophilize the samples and image them using a scanning electron microscope or perform fluorescence microscopy.

One important drawback of an injectable hydrogel is that it does not set immediately, resulting in an initial burst release (Huang et al. 2001). This phenomenon may be minimised by rapid gelation and slow diffusion of bulky LNPs, but the release profile of mRNA is difficult to observe. If the lipogels were loaded with different cargo, such as antibodies, this could be measured using an ELISA assay.

Further characterisation of Pluronic-LNP lipogels should be conducted including investigating factors such as the swelling, injectability, syringeability, and storage and degradation properties (Alonso et al., 2021). After demonstrating the viability of the system through in vitro testing and characterisation, then animal studied can be conducted.

Acknowledgments

We would like to thank Dr Paul McKay, Gloria Nzeako, Xu Liu and Maungo Poomore for their support in our research project.

References

Alishetty, S., Carrasco, M., Alameh, M.-G., Paige, M., Said, H., Wright, L.R., Narayanan, A., Alem, F., Hernandez, K.R., Gillevet, P.M., Soliman, O.Y., Hicks, P., Manzoni, T.B., Bates, P., Stephens-Shields, A.J., Cleveland, T.E., Grishaev, A., Weissman, D. and Buschmann, M.D. (2021). Novel lipid nanoparticle provides potent SARS-CoV-2 mRNA vaccine at low dose with low local reactogenicity, high thermostability and limited systemic biodistribution. *Research Square*. doi:https://doi.org/10.21203/rs.3.rs-

798453/v1.

Almeida, H., Lobão, P., Frigerio, C., Fonseca, J., Silva, R., Quaresma, P., Lobo, J.M.S. and Amaral, M.H. (2016). Development of mucoadhesive and thermosensitive eyedrops to improve the ophthalmic bioavailability of ibuprofen. *Journal of Drug Delivery Science and Technology*, 35, pp.69–80. doi:https://doi.org/10.1016/j.jddst.2016.04.010.

Alonso, J.M., Andrade del Olmo, J., Perez Gonzalez, R. and Saez-Martinez, V. (2021). Injectable Hydrogels: From Laboratory to Industrialization. *Polymers*, 13(4), p.650. doi:https://doi.org/10.3390/polym13040650.

Balan, G.-A., Precupas, A. and Matei, I. (2023). Gelation Behaviour of Pluronic F127/Polysaccharide Systems Revealed via Thioflavin T Fluorescence. *Gels*, [online] 9(12), p.939. doi:https://doi.org/10.3390/gels9120939.

Barbieri, B.D., Peeler, D.J., Samnuan, K., Day, S., Hu, K., Sallah, H.J., Tregoning, J.S., McKay, P.F. and Shattock, R.J. (2024). The role of helper lipids in optimising nanoparticle formulations of self-amplifying RNA. *Journal of Controlled Release*, 374, pp.280–292. doi:https://doi.org/10.1016/j.jconrel.2024.08.016.

Chheda, U., Selvi Pradeepan, Esposito, E.A., Strezsak, S.R., Fernandez-Delgado, O. and Kranz, J.K. (2023). Factors affecting stability of RNA – temperature, length, concentration, pH, and buffering species. *Journal of Pharmaceutical Sciences*, 113(2). doi:https://doi.org/10.1016/j.xphs.2023.11.023.

Clegg, J.R., Kolade Adebowale, Zhao, Z. and Samir Mitragotri (2024). Hydrogels in the clinic: An update. *Bioengineering & Translational Medicine*, 9(6). doi:https://doi.org/10.1002/btm2.10680.

Constantinou, A.P., Nele, V., Doutch, J., Correia, J.S., Moiseev, R., Cihova, M., Gaboriau, D., Krell, J., Khutoryanskiy, V.V., Stevens, M.M. and Georgiou, T.K. (2022). Investigation of the Thermogelation of a Promising Biocompatible ABC Triblock Terpolymer and Its Comparison with Pluronic F127. *Macromolecules*, 55(5), pp.1783–1799. doi:https://doi.org/10.1021/acs.macromol.1c02123.

De Neve, L., York, M., Dickens, J., Leys, J., Meurs, G., Sinnaeve, D. and Van der Meeren, P. (2017). Molecular structure and ionic strength both affect the micellization and solubilization behavior of PEO-PPO-PEO surfactants. Colloids and Surfaces A: Physicochemical and Engineering Aspects, [online] 536, pp.172–179. doi:https://doi.org/10.1016/j.colsurfa.2017.07.023.

Desfrançois, C., Auzély, R. and Texier, I. (2018). Lipid Nanoparticles and Their Hydrogel Composites for Drug Delivery: A Review. *Pharmaceuticals*, 11(4), p.118. doi:https://doi.org/10.3390/ph11040118.

Din, F. ud, Mustapha, O., Kim, D.W., Rashid, R., Park, J.H., Choi, J.Y., Ku, S.K., Yong, C.S., Kim, J.O. and Choi, H.-G. (2015). Novel dual-reverse thermosensitive solid lipid nanoparticle-loaded hydrogel for rectal administration of flurbiprofen with improved bioavailability and reduced initial burst effect. *European Journal of Pharmaceutics and Biopharmaceutics*, 94, pp.64–72. doi:https://doi.org/10.1016/j.ejpb.2015.04.019.

Dowdy, S.F. (2017). Overcoming cellular barriers for RNA therapeutics. *Nature Biotechnology*, 35(3), pp.222–229. doi:https://doi.org/10.1038/nbt.3802.

Hajj, K.A., Melamed, J.R., Chaudhary, N., Lamson, N.G., Ball, R.L., Yerneni, S.S. and Whitehead, K.A. (2020). A Potent Branched-Tail Lipid Nanoparticle Enables Multiplexed mRNA Delivery and Gene Editing In Vivo. Nano Letters, 20(7), pp.5167–5175. doi:https://doi.org/10.1021/acs.nanolett.0c00596.

Hao, J., Wang, X., Bi, Y., Teng, Y., Wang, J., Li, F., Li, Q., Zhang, J., Guo, F. and Liu, J. (2014). Fabrication of a composite system combining solid lipid nanoparticles and thermosensitive hydrogel for challenging ophthalmic drug delivery. *Colloids and Surfaces B: Biointerfaces*, 114, pp.111–120. doi:https://doi.org/10.1016/j.colsurfb.2013.09.059.

Hirun, N., Kraisit, P. and Soontaranon, S. (2023). Role of Bovine Serum Albumin Addition in Micellization and Gel Formation of Poloxamer 407. *Polymers*, 15(11), p.2465. doi:https://doi.org/10.3390/polym15112465.

Hou, X., Zaks, T., Langer, R. and Dong, Y. (2021). Lipid nanoparticles for mRNA delivery. *Nature Reviews Materials*, [online] 6(6), pp.1078–1094. doi:https://doi.org/10.1038/s41578-021-00358-0.

Huang, X. and Brazel, C.S. (2001). On the importance and mechanisms of burst release in matrix-controlled drug delivery systems. Journal of Controlled Release, 73(2-3), pp.121–136. doi:https://doi.org/10.1016/s0168-3659(01)00248-6.

Jaquilin P J, R., Oluwafemi, O.S., Thomas, S. and Oyedeji, A.O. (2022). Recent advances in drug delivery nanocarriers incorporated in temperature-sensitive Pluronic F-127–A critical review. *Journal of Drug Delivery Science and Technology*, 72, p.103390. doi:https://doi.org/10.1016/j.jddst.2022.103390.

Jung, Y., Park, W., Park, H., Lee, D.-K. and Na, K. (2017). Thermo-sensitive injectable hydrogel based on the physical mixing of hyaluronic acid and Pluronic F-127 for sustained NSAID delivery. *Carbohydrate Polymers*, [online] 156, pp.403–408. doi:https://doi.org/10.1016/j.carbpol.2016.08.068.

Kadekar, S., Nawale, G.N., Rangasami, V.K., Le Joncour, V., Laakkonen, P., Hilborn, J., Varghese, O.P. and Oommen, O.P. (2021). Redox responsive Pluronic micelle mediated delivery of functional siRNA: a modular nano-assembly for targeted delivery. Biomaterials Science, 9(11). doi:https://doi.org/10.1039/d1bm00428j.

Kafetzis, K.N., Papalamprou, N., McNulty, E., Thong, K.X., Sato, Y., Mironov, A., Purohit, A., Welsby, P.J., Harashima, H., Yu-Wai-Man, C. and Tagalakis, A.D. (2023). The Effect of Cryoprotectants and Storage Conditions on the Transfection Efficiency, Stability, and Safety of Lipid-Based Nanoparticles for mRNA and DNA Delivery. *Advanced Healthcare Materials*, 12(18), pp.2203022–2203022. doi:https://doi.org/10.1002/adhm.202203022.

Lan, B., Zhang, L., Yang, L., Wu, J., Li, N., Pan, C., Wang, X., Zeng, L., Yan, L., Yang, C. and Ren, M. (2021). Sustained delivery of MMP-9 siRNA via thermosensitive hydrogel accelerates diabetic wound healing. Journal of Nanobiotechnology, 19(1). doi:https://doi.org/10.1186/s12951-021-00869-6.

Overstreet, D.J., Dutta, D., Stabenfeldt, S.E. and Vernon, B.L. (2012). Injectable hydrogels. *Journal of Polymer Science Part B: Polymer Physics*, 50(13), pp.881–903. doi:https://doi.org/10.1002/polb.23081.

Oxford Vaccine Group (2022). *Vaccine ingredients*. [online] vaccineknowledge.ox.ac.uk. Available at: https://vaccineknowledge.ox.ac.uk/vaccine-ingredients#General-information.

Qiu, M., Glass, Z., Chen, J., Haas, M., Jin, X., Zhao, X., Rui, X., Ye, Z., Li, Y., Zhang, F. and Xu, Q. (2021). Lipid nanoparticle-mediated codelivery of Cas9 mRNA and single-guide RNA achieves liver-specific in vivo genome editing of Angptl3. *Proceedings of the National Academy of Sciences*, 118(10), p.e2020401118. doi:https://doi.org/10.1073/pnas.2020401118.

Ramya, K.A., Kodavaty, J., Dorishetty, P., Setti, M. and Deshpande, A.P. (2019). Characterizing the yielding processes in pluronic-hyaluronic acid thermoreversible gelling systems using oscillatory rheology. *Journal of Rheology*, 63(2), pp.215–228. doi:https://doi.org/10.1122/1.5045073.

Sarkar, B., Ravi, V. and Paschalis Alexandridis (2013). Micellization of amphiphilic block copolymers in binary and ternary solvent mixtures. Journal of Colloid and Interface Science, 390(1), pp.137–146. doi:https://doi.org/10.1016/j.jcis.2012.09.028.

Schmidt, C. and Schnierle, B.S. (2023). Self-Amplifying RNA Vaccine Candidates: Alternative Platforms for mRNA Vaccine Development. *Pathogens*, 12(1), p.138. doi:https://doi.org/10.3390/pathogens12010138.

Schmolka, I.R. (1972). Artificial skin I. Preparation and properties of pluronic F-127 gels for treatment of burns. *Journal of Biomedical Materials Research*, 6(6), pp.571–582. doi:https://doi.org/10.1002/jbm.820060609.

Shmool, T.A., Constantinou, A.P., Jirkas, A., Zhao, C., Georgiou, T.K. and Hallett, J.P. (2022). Next generation strategy for tuning the thermoresponsive properties of micellar and hydrogel drug delivery vehicles using ionic liquids. *Polymer Chemistry*, 13(16), pp.2340–2350. doi:https://doi.org/10.1039/d2py00053a.

Shriky, B., Kelly, A., Isreb, M., Babenko, M., Mahmoudi, N., Rogers, S., Shebanova, O., Snow, T. and Gough, T. (2020). Pluronic F127 thermosensitive injectable smart hydrogels for controlled drug delivery system development. *Journal of Colloid and Interface Science*, [online] 565, pp.119–130. doi:https://doi.org/10.1016/j.jcis.2019.12.096.

Silva-Pilipich, N., Beloki, U., Salaberry, L. and Smerdou, C. (2024). Self-Amplifying RNA: A Second Revolution of mRNA Vaccines against COVID-19. *Vaccines*, [online] 12(3), p.318. doi:https://doi.org/10.3390/vaccines12030318.

Tenchov, R., Bird, R., Curtze, A.E. and Zhou, Q. (2021). Lipid Nanoparticles—From Liposomes to mRNA Vaccine Delivery, a Landscape of Research Diversity and Advancement. *ACS Nano*, [online] 15(11). doi:https://doi.org/10.1021/acsnano.1c04996.

U.S. Department of Health and Human Services Food and Drug Administration (FDA), Center for Drug Evaluation and Research (CDER) and Center for Biologics Evaluation and Research (CBER) (2014). Guidance for Industry: Immunogenicity Assessment for Therapeutic Protein Products. [online] FDA. Available at: https://www.fda.gov/media/85017/download.

Yu, Y., Feng, R., Yu, S., Li, J., Wang, Y., Song, Y., Yang, X., Pan, W. and Li, S. (2018). Nanostructured lipid carrier-based pH and temperature dual-responsive hydrogel composed of carboxymethyl chitosan and poloxamer for drug delivery. *International Journal of Biological Macromolecules*, 114, pp.462–469. doi:https://doi.org/10.1016/j.ijbiomac.2018.03.117.

Zhong, R., Talebian, S., Mendes, B.B., Wallace, G., Langer, R., Conde, J. and Shi, J. (2023). Hydrogels for RNA delivery. *Nature Materials*, [online] pp.1–14. doi:https://doi.org/10.1038/s41563-023-01472-w.

Zhou, Q., Zhang, Z., Chen, T., Guo, X. and Zhou, S. (2011). Preparation and characterization of thermosensitive pluronic F127-b-poly(ε-caprolactone) mixed micelles. *Colloids and Surfaces B: Biointerfaces*, 86(1), pp.45–57. doi:https://doi.org/10.1016/j.colsurfb.2011.03.013.

Surfactant Optimized Macroporous Polystyrene for Water Treatment Applications

Weng Kit Leong and Zewen Liu
Department of Chemical Engineering, Imperial College London, U.K.

Abstract
Clean water scarcity is a critical global issue, whilst industrial wastewater is a big contributor to potential pollution to global water sources, requiring innovative approaches for effective treatment and safe disposal. Breath figure formation on macroporous polystyrene (PS) particles offers a promising solution to the problem due to their tuneable surface morphology and good adsorption properties. This study explores the effects of surfactant (AOT) concentration on tuning the structures of breath figures formed on acoustically levitated polymer droplet surfaces. Solutions consisting of 8 w/w% PS and varying concentrations of AOT (0 – 0.05 w/w%) were dried in an acoustic levitator whilst controlling humidity and temperature. The resulting particles are analysed by Scanning Electron Microscopy, surface pore sizes and nearest neighbour data are analysed by ImageJ Voronoi function, while drying mechanism quantified through normalised equivalent radius over time. The results revealed that higher concentrations of surfactant past 0.03 w/w% lead to increasingly disordered pore structures, with self- assembled breath figures up to 0.02 w/w% AOT. These findings verify that interfacial tension governed by surfactant concentration plays a critical role in breath figure formation. The results also point to an optimal concentration of AOT added where pore sizes increase thereby increasing the surface area available for adsorption whilst maintaining regularity of the breath figures formed. The findings gathered by this study could guide to the optimisation of spray drying processes for scalable macroporous polystyrene particles production for water treatment applications.

Introduction

Access to clean water is a critical challenge to human being, regardless of scale. Data from 73 countries compiled by the United Nations Water Organization reveals that only 60% of industrial wastewater was safely treated.[1] With the increasing threat of water stress globally, new methods for water treatment should be explored to allow to treat water from various sources to enable safe disposal into the environment or human-related consumption. [2]

Polystyrene is a synthetic polymer used in various applications due to its versatility and lightweight nature. For wastewater treatment purposes, polystyrene can be processed into different forms, this includes and is not limited to Ion-exchange Resins, Filtration Membranes,[3] Microfibres,[4] and Solid Foams.[5] Good adsorption properties for these materials have been reported. In this study macroporous polystyrene resins are the main application of focus, since they have good adsorption characteristics and can be produced via spray drying. Therefore, further studies on the pore distribution and analysis are crucial to design and optimise polystyrene adsorption performance.

On various polymer-based films and particles, micro-scale structures called breath figures can be induced to form. These are porous structures that are created on the surface of the particle under certain conditions. Furthermore, these breath figures (BFs) can be designed to form ordered arrangements. Most of the current research on this topic aims to elucidate the behaviour and formation of breath figures on polymer-based films.

Breath figures have been explored for uses in a wide variety of fields. Within the biological field, these include the creation of cell scaffolding with breath figures used as the method for forming ordered surface morphologies, and the fabrication of porous membranes for bio-separation.[6] Furthermore, using breath figures as a method for drug delivery[7] has been explored. This underscores the importance of investigating tunability of surface morphologies using breath figures.

A novel method of producing micro-scale structures on polymer particles has been explored.[8] This study utilises acoustic levitation to suspend droplets of polymer solution enabling the formation of micro-scale particles with breath figures. As mentioned, a common method in industry to produce particles is via spray drying. However, spray drying proves to be a challenging method to analyse the dynamics of a singular particle or droplet due to a multitude of reasons. These challenges arise from factors like temperature variations within the spray drying column, heated air flows as disturbances which could affect the motion of droplet, also the small sizes of the particles lead to difficulties in closer analysis. Lastly, the transient nature of the spray drying is another factor making the analysis difficult.

To overcome these limitations, acoustic levitation can be employed as an alternative method to large scale spray drying. By levitating a liquid droplet with an acoustic field, and allowing it to dry in a controlled system, one can accurately observe the drying mechanism and conduct quantitative analysis. This method provides better control over various experimental parameters, which allows the isolation of the effect of surfactant addition to the polymer mixture.

Previous research done [9] has shown that surfactants are able to influence the formation of ordered breath figures. This suggests that interfacial tension plays a

crucial role in their structure. However, this study focuses breath figure formation on films, leaving a question unanswered for whether these observations can be directly translated to acoustically levitated particles or not. As the dynamics and conditions under which breath figures form on an acoustically levitated particle span a wider range compared to films.

This paper aims to investigate the impact of varying the interfacial tension, through the addition of surfactants on the formation of breath figures onto the levitated particle surfaces. The findings of this study have great potential to engineer spray drying processes, enabling large scale production of porous polystyrene particles for wastewater treatment applications.

Background

Polymer drying in any form is a complex process which is affected by a multitude of variables, but in this paper the focus will be on acoustically levitated particles forming from a droplet of a solution.

Acoustic levitation is the process of using standing acoustic waves to trap an object, such as a polymer droplet. Acoustic waves exert forces on the object that counteracts the force of gravity. When those forces are balanced, the object will levitate in space. The TinyLev acoustic levitator, was used to achieve acoustic levitation for the results presented in this paper.

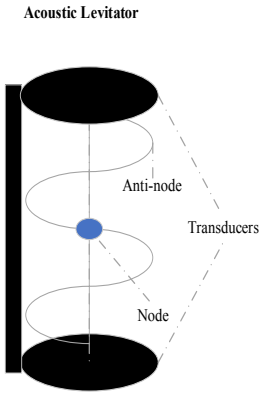


Figure 1. Schematic of an acoustic levitator

The acoustic waves generated by the transducers create a standing wave within the levitator. The arrangement of the transducers creates a focusing effect that helps to stabilise the levitated particle at the central node of the standing wave in all axes. As shown in Figure 1, there are multiple nodes but the particle is placed within the central node due to the focusing effect. The acoustic field can be modelled by the Gork'ov potential, of which the negative derivative can be taken to model the acoustic radiation pressure exerted on a spherical particle within the field.[10][11] Key variables that influence the force exerted on the particle and the stability of its levitation are the type of transducer used, and the voltage applied to the transducers.

The importance on controlling the force exerted on the particle arises from the effect the acoustic radiation force has on the droplet. This force causes a liquid droplet's shape to diverge from sphericity. In general, increasing voltage elongates the levitated droplet and an increase in eccentricity and aspect ratio is observed.[11] This ultimately highlights the importance of keeping voltage applied to the transducers constant through all experiments.

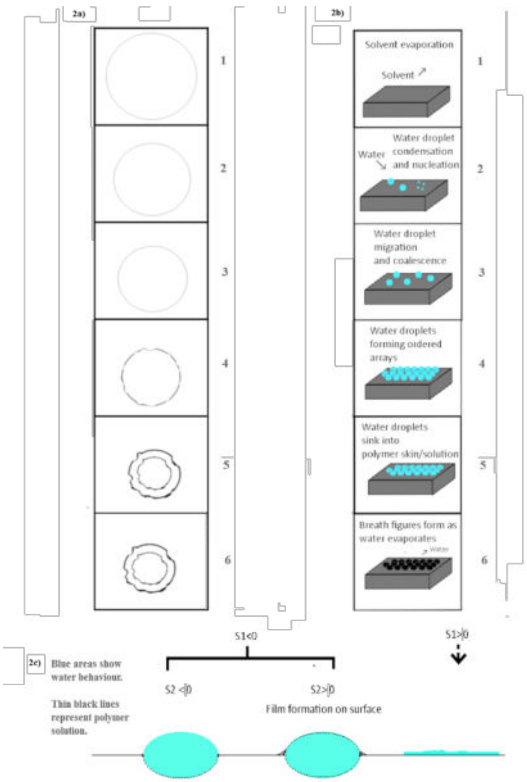


Figure 2: Diagrammatic representations of the following:
a) Polymer droplet drying process
b) Spreading coefficient effect on film formation on water droplet surface
c) Breath figure formation process

As solutions containing polymers within a solvent dry under acoustic levitation, they follow the following stages. This is visually shown in Figure 2a.

1. The droplet is assumed to be homogeneous and well mixed. This is when the droplet is first placed at the node in the acoustic levitator.
2. The droplet begins to dry, and the drying is assumed to be primarily controlled via Fickian diffusion
3. Droplet continues to dry, reducing in size.
4. A gel-like polymer rich skin forms, which impedes the diffusion of solvent from within the droplet to the boundary with air.[8][12]

5. The skin layer thickens as the concentration of the polymer within the droplet increases due to solvent evaporation. A cavity forms.
6. Kinetic arrest occurs where the droplet size does not change further. This occurs when the droplet reaches the glass transition temperature, which marks the gel like skin transitioning into a vitreous state.

The temperature at this transition occurs varies with concentration of the droplet. The concentration at which this temperature reaches the ambient temperature is referred to as the Berghmans point.[13]

During the process of polymer drying, the evaporation of the solvent causes the droplet to experience heat losses resulting in a decrease in droplet temperature (Figure 2c. step 1).[14] If the droplet's surface temperature decreases below the dew point, condensation of water on the droplet surface occurs (Figure 2c. step 2).[15] At this junction, the interfacial tension between water and the polymer solution plays a big role in determining whether breath figures are formed. Interfacial tensions can be defined as a spreading coefficient, S_1, describes the behaviour of the condensed water droplets on the polymer surface[16][17]. At this junction, the interfacial tension between water and the polymer solution plays a big role in determining whether breath figures are formed. Interfacial tensions can be defined as a spreading coefficient, S_1, which describes the behaviour of the condensed water droplets on the polymer surface.[15][16] S_1 is a function of three interfacial tensions, defined by the following equation.[17][18]

$$S_1 = \gamma_{sv} - \gamma_{sl} - \gamma_{lv} \text{ (Eqn.1)}$$

γ denotes the interfacial tension and the subscripts sv, sl, and lv denote the polymer solution/vapour, solution/water, and water/vapour interfaces respectively. For partial wetting to occur, which is the prerequisite for breath figure formation, S_1 must be less than zero. The water droplets can then form structured arrays through self-assembly. Marangoni convection, both soluto-capillary and thermo-capillary in nature, have been used to explain the spontaneous self-assembly process. This has been used to explain how the water droplets migrate on the surface of the polymer droplet and how they can be pulled downwards into the droplet (Figure 2b. steps 3 and 4).[16][18] However, due to the complex drying process and the interplay between the interfacial forces, changing temperatures and concentration gradients, the definite driving force behind self-assembly is still debated.

After this, the water droplets sink into the array that forms a porous structure (Figure 2b. step 5). As the droplet warms up, the water droplets evaporate leaving behind the porous structure, creating breath figures (Figure 2b. step 6).

A study by Fukuhira et al. (2009)[9] discusses the further effects of the interfacial tension between water droplets and the polymer solution on the formation of ordered breath figures. This concludes that higher interfacial tension stabilises water droplets as the solvent evaporates from the polymer droplet surface, leading to ordered breath figures.

The behaviour of water droplets condensing on the surface has been theorised by Tian et al. (2006) to be also stabilised by the formation of the priorly mentioned skin,[19] assisting the arrests of the droplets and prevents coalescence. Excessive coalescence of water droplets results in disordered arrays with a large variation in pore sizes. In this way, the formation of ordered breath figures can be said to be affected by both the solvent and solute chosen. In addition, there are other factors affecting the formation, structure and uniformity of breath figures, These include relative humidity,[20] temperature,[21] concentration of polymer solution,[22] airflow,[23] and pressure.[24] Furthermore, interfacial tension plays a crucial role in the formation of the final breath figure structure. This is via the formation of a polymer layer on encapsulating the droplet, known as water droplet encapsulation. This typically occur as the water droplet sink into the polymer droplet surface to a depth where the become fully encapsulated.

Based on the work by Ma et al. (2011), the formation of the final observable surface structure of the polymer particle after drying can be explained by the encapsulated droplets bursting. This "bursting hypothesis" posits that the water droplet that is encapsulated may not initially evaporate during encapsulation and. After a period of time, due to the droplet temperature staying low. Once the temperature increases back to ambient conditions, the water droplet then evaporates and bursts through the polymer layer, creating a smaller hole.[25]

Further work by J. David Smith et al. (2013) defines another spreading coefficient S_2 described by the following equation.[26]

$$S_2 = \gamma_{lv} - \gamma_{sl} - \gamma_{sv} \text{ (Eqn.2)}$$

According to that study, if $S_2 > 0$, the polymer solution will coat the surface of the water droplet which has not fully sunk into the surface. Once the solvent within the polymer solution coat evaporates, a polymer layer is deposited on the water droplet.

Pivoting back to the purpose of this paper, most of the studies referenced in the section about polymer

droplet drying under acoustic levitation do not mention breath figures or seek to investigate their formation and self-assembly on the levitated particles. Likewise, other studies referenced about breath figure formation and self-assembly focus on polymer films created through various methods such as drop casting and spin casting. Recent work by[8] married the two topics together to investigate the conditions under which breath figures can form with acoustically levitated polymer solution droplets. The findings show that whilst varying polymer concentration, ordered breath figures can form at a lower relative humidity for levitated polymer droplets versus polymer films. At polymer concentrations of 10% w-w and higher, polymer films at the same conditions could not form ordered breath figures at all relative humidities in stark contrast to acoustically levitated particles.

In light of these findings and the previous work by Fukuhira et al. (2009),[9] the primary effect of surface tension on ordered breath figure formation with acoustically levitated particles was chosen to be investigated. This is due to the hypothesis that the different stresses on a particle floating akin to spray drying may have cascading effects on the formation of breath figures.

Methods

The acoustic levitator is constructed using TinyLev schematics,[10] with MCUST10P40 B07RO transducers, 10mm in diameter and operating at 40kHz. Images were recorded by a Basler acA2000 -165uc Camera, connected via Pylon View software. aligned horizontally next to the levitator to focus on the node of the levitator where droplets are placed. An additional LED light was mounted opposite to the camera as the main light source. All the apparatus mentioned above is placed inside a humidity chamber to maintain stability of the node and kept at designated RH value. Polymer used for the experiment is Poly(styrene) ($M_w = 100{,}000$, $Viscosity = 2.0 - 2.4\ cP$, BDH Chemicals Ltd Poole England), solvents Dichloromethane (1.322 g/cm^3 at 20℃, VWR Chemicals), surfactants sodium bis(2-ethylhexyl) sulfosuccinate ($M_w = 444.56\ g/mol$, $T_{melt} = 173 - 179$℃, Aldrich).

From reported literature, at 40% relative humidity, constant temperature of 21℃ , 8% (weight by weight) of polystyrene in the solution, resulted dried polymer sample will exhibit self-assembled breath figures.[8] Therefore, those conditions will be kept at constant throughout this experiment. The controlled variable is therefore the concentration of surfactants in the polymer solution (weight by weight percentage) from 0.005% – 0.05%. The low concentration range is desired to prevent surfactant micelle formation. This would add another layer of complexity to the structure of the polymer particle which was determined to be unwanted.

Solutions were prepared in 30 ml vials and put on tuber roller for thorough mixing. Before the addition of the solution droplet to the acoustic field. The well-mixed solution droplet is then deposited into the acoustic field via a syringe and 33 ga needle. The drying process takes roughly 10 minutes for the samples to be completely dried. Samples are then collected.. The recording is then halted, and images exported to ImageJ for further analysis. Before doing Scanning Electron Microscopy (SEM) (Zeiss Auriga Crossbeam), samples are mounted on carbon tape and coated with gold. The resulting SEM images are analysed by ImageJ. Voronoi analysis (BioVoxxel Toolbox plug-in) is used to perform nearest neighbour analysis of the breath figure pores.

$$R_{eq}(t)\begin{cases}(R_0 - R^*)\left(1 - \frac{t}{\tau_1}\right)^{\alpha_1} + R^* & if\ t < \tau_1 \\ (R^* - R_\infty)\left(1 - \frac{t - \tau_1}{\tau_2 - \tau_1}\right)^{\alpha_2} + R_\infty & if\ \tau_1 < t < \tau_2 \\ R_\infty & if\ t > \tau_2\end{cases}$$

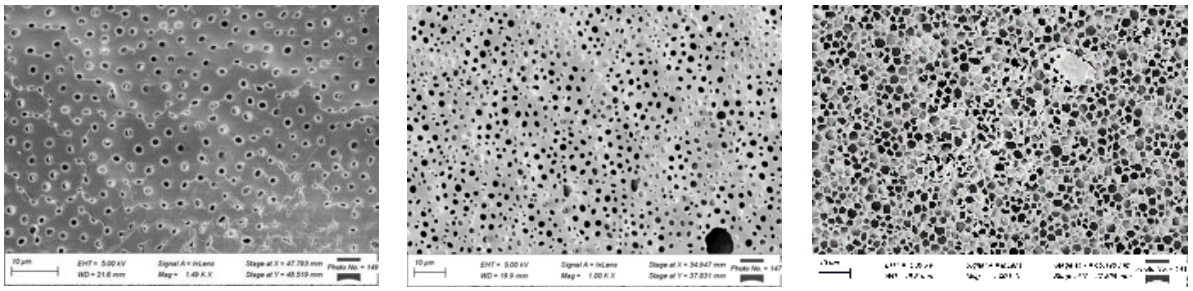
(Eqn.3)

The equation above shows the empirical drying profiles related to polymer droplet drying mechanism. By utilising ImageJ, the images of the drying process were analysed to fit an ellipse to each droplet. The volume equivalent radius of each ellipse was then calculated and tracked over time. To account for variations in initial droplet size, the equivalent radius was normalised against its initial value. The curves were then fitted by separating the drying process into two stages. The first stage is qualitatively equal to steps 1-3 as described in the drying section on page 2, and the second stage equal to steps 4-6. The time (t) at which the drying process transitions form the first to the second stage was determined, coinciding with and equivalent radius of R^*. This marks the initiation of the kinetic arrest period, characterised by the formation of a polymer gel-like skin. The second stage concluded when kinetic arrest occurs, at this point the equivalent radius R_∞ remains constant as the particle stops shrinking. While this is a simplified representation of the complex drying process, this model provides parameters for comparing drying behaviours across various surfactant concentrations.

SEM images of pores were then analysed to look at the pore coverage for a given area. The pores were identified by thresholding the image to create a black and white image which only highlights the pores. This was then processed using the Analyse Particles ImageJ function allowing the identification of the number of pores, and the area of the pores. The pore count and area of each pore was then calculated through this function, then allowed for the derivation of the percentage area coverage of the pores.

Results

a) 0%AOT+8%PS+DCM b) 0.005%AOT+8%PS+DCM c) 0.01%AOT+8%PS+DCM



d) 0.02%AOT+8%PS+DCM e) 0.03%AOT+8%PS+DCM f) 0.05%AOT+8%PS+DCM

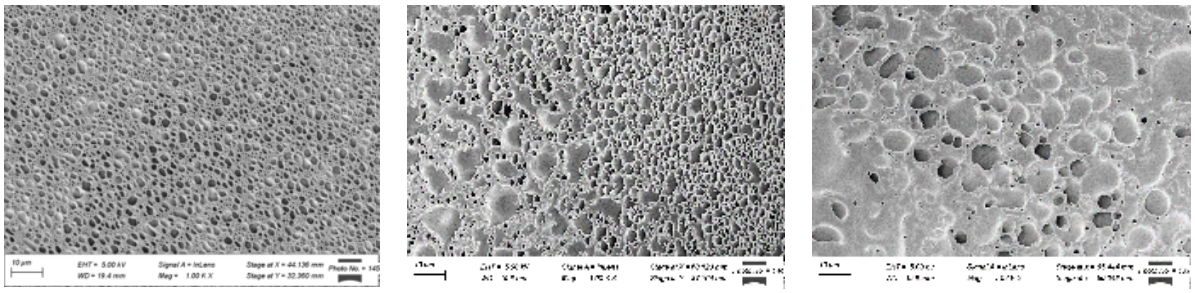
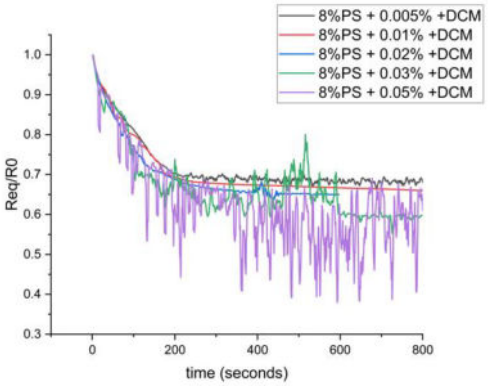


Figure 3 Scanning Electron Microscopy (SEM) results of different surfactant added samples (all samples prepared in 40%RH, 21℃ and constant polystyrene concentration). These images characterise the surface morphology of polystyrene (PS) samples prepared with varying surfactant (AOT) concentrations in the solvent, dichloromethane. The subfigures arranged from a) to f) represent increasing surfactant concentrations from 0 to 0.05 w/w%.

As one can observe in Figure 3, at lower concentrations (a-c), highly ordered pores are observed, this indicates self-assembled breath figure formation. As surfactant concentration increases to 0.02 w/w% (c), pores become smaller and more compact. At 0.03% and higher w/w% concentrations of AOT (d-f), pore ordering diminishes, resulting in irregular or anomalous surface morphologies.

a)



b)

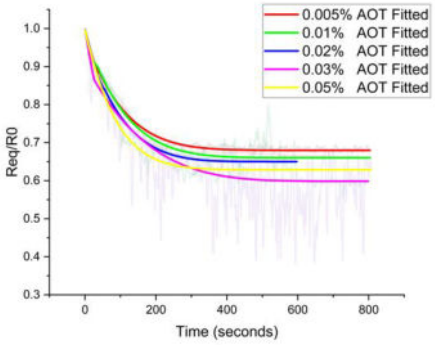


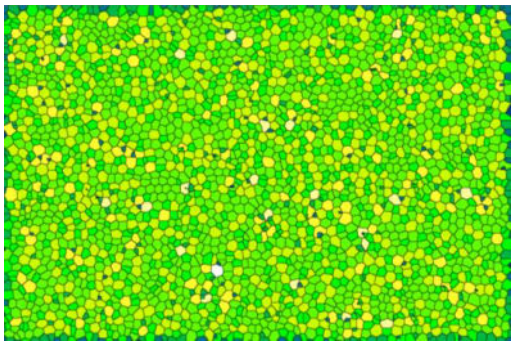
Figure 4 a) Drying curve of different surfactant concentrated samples x axis Req/R0 is the equivalent particle radius normalised by its original radius. b) Drying curve fitted, first stage fitted with eqn_ as the first stage (pre skin formation), eqn_ as the second stage of drying (post skin formation).

Drying curves were analysed to reveal the effects of surfactant concentrations on the particle drying rates and final sizes. Figure 4 illustrates these curves for various tested surfactant concentrations. Figure 4 a) presents the normalised equivalent radius $(R_{eq})/(R_0)$ showing how droplet sizes decrease over time. The curves indicates that as the concentration of surfactant increases, the drying process becomes increasingly chaotic, especially in the later stages of drying, indicating increasingly chaotic behaviour.

Furthermore, Figure 4 b) displays the fitted curves based on the data from a), highlighting the radius transitioning

to a plateau at the end of drying where the equivalent radius stabilises. The final radius values appear to decrease with increasing surfactant concentration, except the 0.05 w/w% one. Across all concentrations, lower surfactant samples (i.e. 0.005 w/w%) exhibit smooth drying profiles, while higher ones ultimately lead to noisier results. These observations highlight clear differences in drying behaviours across different surfactant concentrations, especially in the consistency and final sizes of the particles.

a) BioVoxxel toolbox neighbouring analysis



b) Neighbouring distribution of each pore.

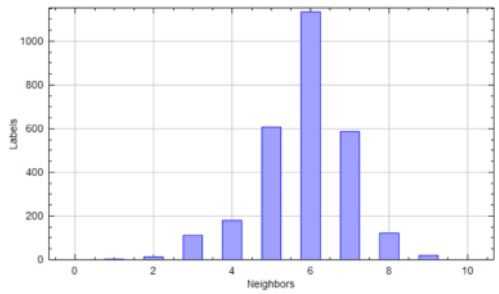
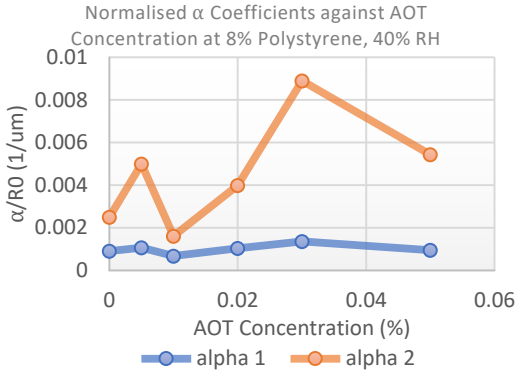


Figure 5 Voronoi neighbouring analysis of 0.01%AOT + 8%PS + DCM sample. Analysis on other concentrations can be found in Supplementary Information Section 1

ImageJ with BioVoxxel toolbox plug-ins is used to analyse the nearest neighbours of each pore from the particle surface. The results shown in Figure 5 presents the results for a sample containing 0.01% surfactant concentration with 8% PS in DCM. Figure 5a) illustrates a Voronoi diagram visualising the pore structure assembly, where each pore is represented by a polygon, and the neighbouring relationships are highlighted. The pores appear to be densely packed with consistent polygon shapes. Whereby Figure 5 b) shows the distribution of the number of neighbouring pores of each polygon. The histogram reveals a peak at six neighbours. This indicates that the most pores are surrounded by six others, with some variability observed on either side of the peak. This agrees with existing literature from O'Connell et al. (2023).[8]

a)



b)

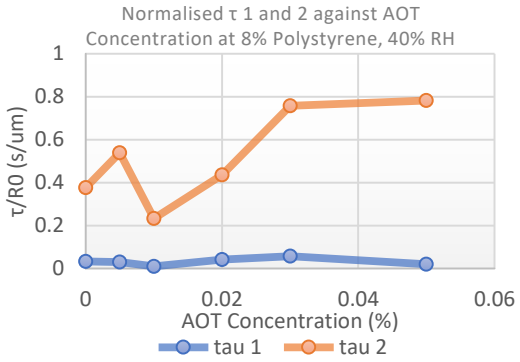


Figure 6 Key drying profile parameters normalised by initial radius (R_0) and characterised against surfactant concentrations.

As aforementioned in the background section, the key parameters needed to characterise the drying profile are α_1, α_2, τ_1, and τ_2. Those parameters for various surfactant concentrations are normalised by the initial droplet radius and presented in Figure 6. a) shows the exponents α_1 and α_2, which empirically represent the drying rates for the pre-skin and post-skin stages, respectively. Both parameters demonstrate non-linear trends, with α_1 having slight fluctuations at all concentrations and peaking at 0.03 w/w%. Whereas α_2 exhibits a steeper change and irregularity at intermediate concentrations.

Figure 6 b) talks about the timescales $\tau_1$and τ_2 associated with drying. τ_1 (time where skin formation occurs) appears to remain relatively low and minimal variations for most concentrations, since the skin forms rapidly after the droplet is placed in the acoustic field. However, $\tau_2$increases significantly, plateauing at higher surfactant concentrations. These results highlight the distinct trends across parameters and surfactant levels.

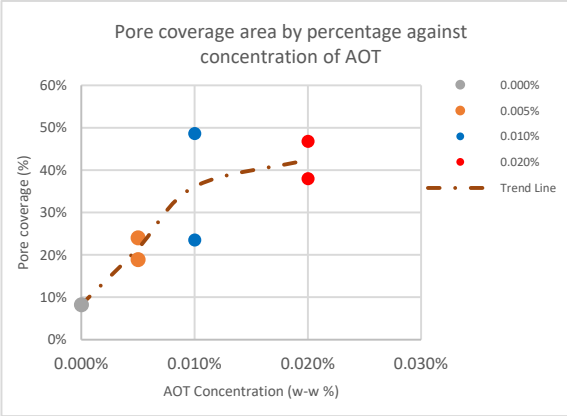


Figure 7 Pore coverage % against concentration change

Figure 7 analyses the SEM images of varied surfactant concentrations; the data is then presented in area coverage by breath figure pores against change in surfactant concentrations. There is a trend upwards, where the total surface level pore coverage area increases as AOT concentration is increased. However, at 0.03% AOT concentration and above, surface level pore coverage is no longer regular enough to conduct analysis.

Discussion

Given the aim of this research paper is to check for the tunability of breath figures by varying interfacial tension on acoustically levitated particles, the first result to be discussed will be the pore coverage area as this will give an indication of overall surface area.

Looking at the trend of pore coverage area both qualitatively and quantitatively, increasing AOT concentrations up to 0.02% yielded increased pore coverage as a function of area. Qualitatively, the increasing visibility of subsurface structures within Figure 1b) lends credence to the bursting hypothesis as described by Ma et al. (2011),[27] which would mean that the underlying structures observed are formed from the initial array of water droplets, and the surface level structures are formed could be formed from the droplets sinking and then bursting through the polymer surface which covered the sunk droplets. This would mean the encapsulation of the water droplet array, and that the bursting hypothesis or something akin to it occurred.

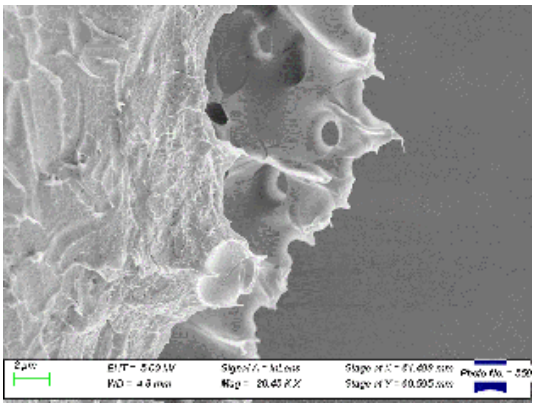


Figure 8 SEM image of a cross-section of polymer particle with composition of 8% w-w polystyrene, as seen with the interior of the particle in the foreground of the image.

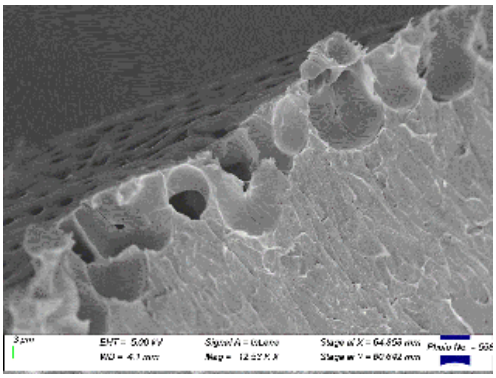


Figure 9 SEM image of the cross-section of the particle, with the skeleton of the water droplet array visible.

This is further supported by SEM images taken of a cross-section of a polymer particle, as shown in figures 8 and 9. This shows the substructure of a polystyrene particle at 8% w-w concentration, with no AOT added. The cross-section in Figure 9 seems to show the existence of spherical structures underneath the surface level pores. Due to the formation of sub-surface structures, and the spherical structures seen in Figure 9, there is evidence for encapsulation theory being the method by which the final surface breath figures are formed. However, it currently is difficult to ascertain what exactly the water droplets are doing after deposition.

Something to note in comparison with past literature from Fukuhira et al. (2009) is the number of surface level pores per water droplet.[9] For the SEM images of 8% polystyrene droplets with 0.01% w-w AOT (Figure 2c.), there seem to be multiple surface level pores per underlying individual sub-surface pore from a single droplet. This was not interpreted as multiple surface level pores per droplet by Fukuhira et al. (2009), [9] but instead just irregularly shaped sub-surface pores still each with a singular surface level pore. However, with Figure 2c. and Figure 8, there is evidence to suggest multiple holes can form per sub-surface pore. The reason for why there are multiple surface level pores per sub surface breath figure link to discussions in the following paragraphs on the formation of polymer films on the water droplet surface, and the effects of interfacial tension. However, the exact process and whether those theories can be validated are still questions to be answered.

Firstly, to discuss the role of surfactants, an explanation of what interfacial tension is must be mentioned. Interfacial tension governs the behaviour of water droplets which condense, and the behaviour of the polymer solution surface in contact with both air and the water droplets. It can be defined as the force required to deform an interface by a given length or the energy required to increase the area of the interface. Lower interfacial tensions result in less energy required to deform the interface. As such with the addition of a surfactant such as AOT, interfacial tensions are lowered resulting in easier deformation of the interfaces. For water droplets, a decrease in interfacial tension of the interface with the polymer solution will result in the droplets having a decreased ability maintain a globular shape. Once $S_1 > 0$,

the water droplets will no longer be able to hold their shape form a film of water that covers the interface instead. The film of water may not just be formed from a single droplet of water, but from multiple droplets undergoing coalescence.

The process of droplet sinking deforms the solution/water interface. With lower interfacial tensions from increasing surfactant concentrations, the ability for the droplets to sink may be increased. This will result in larger pores as more water may be able to sink into the polymer surface. This can be observed in the SEM images with larger AOT concentrations forming larger surface pores and more apparent sub-surface structures. With regards to the bursting hypothesis, the bursting of the droplet through a polymer/air interface with lower interfacial tensions may result in larger pores due to easier deformation of the surface. This easier deformation may allow for water to burst through a larger area of the polymer surface covering the droplet, resulting in multiple larger pores. This is seen in the progression of SEM images of droplets with 0% AOT to 0.01% AOT. At 0.02% and above, it could be the case that the interfacial tension is so low as to allow for the entirety of the polymer skin covering the droplets to be burst through due to decreasing interfacial tension.

Looking at the SEM images taken of particles with 0.03% and 0.05% AOT concentration w-w, the structure of the breath figures observed, although not ordered, support the importance of interfacial tension for water droplet coalescence on the surface of the particle during the drying process, as the non-spherical or elliptical shape of the breath figures suggest that the interfacial tension was too low for the water droplets to maintain their sphericity. What can be observed in those pictures is thereby due to the low interfacial tension resulting in the droplets deforming and forming flatter, less spherical droplets, which result in larger, flatter breath figures with lack of order or regularity. This is in agreement with the work by Fukuhira et al. (2009)[9] as AOT has been found to have a HLB (hydrophilic-lipophilic balance) value of 10.5. This is a measure of how hydrophobic or hydrophilic a surfactant is. With a value of 10.5, AOT can be said to be within the middle of the scale. According to a trend which can be observed in their measurements of interfacial tension against polymer concentration, as the HLB of a surfactant increases, the interfacial tension of solutions containing that surfactant decreases.[9]

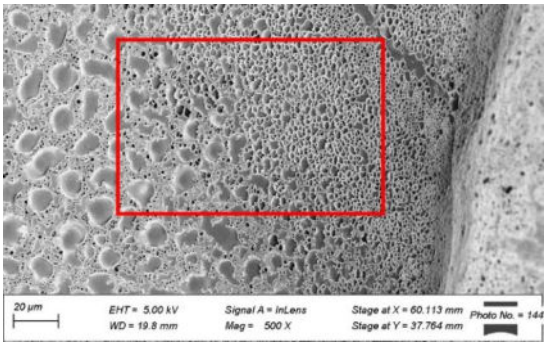


Figure 10: Zoomed out SEM image taken of polystyrene particle from a solution of 8% Polystyrene, 0.03% AOT in DCM.

Something of additional interest is how Figure 3 shows a gradient in the regularity and pore size of breath figure formation across the image. Coincidentally, this SEM image was taken near the edge of the droplet as shown in Figure 10, the red rectangle indicated the area shown in Figure 3e. As the particle curves in the right-most section of Figure 10, there are likely some additional mechanical stresses acting on the entire surface seen in Figure 10. This suggests that mechanical deformation and stresses may impact breath figure formation. This poses an additional complication in attempting to extend studies of breath figure formation with acoustic levitation to spray drying applications, as the mechanical forces experienced by spray dried droplets may be greatly different to droplets in an acoustic levitator.

Another possible explanation of the net-like surface structure shown in Figure 3c could be the fulfilment of the fulfilment of the conditions where the spreading coefficient S_2 is greater than zero. This will result in the polymer solution climbing around the water droplet and forming a thin film. This occurs when a high surface energy droplet is placed on a low surface level liquid phase. As the surfactant concentration increases, the surface energy of the polymer concentration is expected to decrease, meaning that is becomes more likely for the thin polymer film to form. However, this potential interaction will be interacting with the effect changing surfactant concentrations has on the spreading coefficient S_1. According to the bursting hypothesis, the surface level pores could be formed from the evaporation of water bursting through the thin film that forms, resulting in small holes. Tying into this, a further potential explanation would be the thinning of the polymer film that forms on the surface of the droplet. This might be the cause of the reduction in surface level polystyrene coverage and the thinner surface level pore wall thickness as observed. This results in the trends seen with pore percentage area coverage in Figure 7. However, the thinner wall might not be able to fully encapsulate the droplet when dried, meaning that instead of the water droplet bursting out, the surface level polymer film already forms the net like structure before the water evaporates allowing for less hindered water evaporation. This could be a product of the mechanical stresses induced on the polymer as part of the drying process

The validity of the theories discussed above is a question of whether the surface level pores or structure forms before or after water evaporation, and that is a question which cannot be answered with this current report and could be an area of further research. If they form before the water evaporates, that supports the theory of polymer film formation on the surface of the droplet. If they form after, then the theory of water droplet bursting after encapsulation through sinking is supported. Something that is also be occurring is the migration of surfactant molecules into the water droplets, however, the timescale over which that can occur and whether it is thermodynamically viable is yet to be known.

Furthermore, according to Kralova and Sjöblom (2008), [28] emulsifying agents with HLB values between 8 and 16 (of which AOT fulfils those conditions), can form oil-water emulsions. Perhaps this can be extended to organic phase/water emulsions suggesting that the AOT and polystyrene are able to dissolve partially into the water droplet. This may then decrease the interfacial tension for water/air and water/polymer interfaces further. This might alter the behaviour of the water droplets, which might explain the irregular breath figures observed.

Circling back to drying kinetics, surfactant concentration does not seem to generally affect the first stage of drying which is assumed to be controlled by Fickian diffusion. This is due to the alpha 1 and tau 1 results generally staying constant across the varied AOT w-w concentrations used. This would suggest that the initial stage of drying is not generally affected by interfacial tension, as this could be prior to water droplet condensation.

In that process, the final surface level breath figures seen can be attributed to the deposition of that film. According to the bursting hypothesis, the surface level pores could be formed from the evaporation of water bursting through the thin film that forms, resulting in small holes. Tying into this, a further potential explanation would be the thinning of the polymer film that forms on the surface of the droplet. This might be the cause of the reduction in surface level polystyrene coverage and the thinner surface level pore wall thickness as observed. This results in the trends seen with pore percentage area coverage in Figure 7. However, the thinner wall might not be able to fully encapsulate the droplet when dried, meaning that instead of the water droplet bursting out, the surface level polymer film already forms the net like structure before the water evaporates allowing for less hindered water evaporation. Essentially, this is a question of whether the surface level structure forms before or after water evaporation, and that is a question which cannot be answered with this current report and could be an area of further research. This could be a product of the mechanical stresses induced on the polymer as part of the drying process.

It can be suggested that there is a trend for tau 2 and alpha 2, with increasing AOT concentrations causing both to increase, with a marked increase at 0.03%. This can potentially be explained again interfacial tension and using the spreading coefficient S_1 . As mentioned previously, with increasing surfactant concentration, the interfacial tension of the solution decreases, which would mean at a certain point where $S_1 > 0$, only a film of water is able to form (Figure 2b.). This would be supported by the idea that since a surfactant is being used to influence surface tension, γ_{sv} and γ_{sl} will be the parameters affected assuming that a negligible amount of AOT dissolved in the water droplet. Furthermore, γ_{sl} will likely be decreased more than γ_{sv} due to the AOT's role as a surfactant with water. As such as γ_{sl} may decrease more than γ_{sv}, S_1 will increase as surfactant concentration increases. Now if this film formation occurs, given dichloromethane's immiscibility with water, dichloromethane molecules may not only be hindered by the formation of the gel-like polymer skin as described in Figure 2a., step 4. They may also be hindered by transport through a thin film of water, which decreases the rate of evaporation of dichloromethane. This would be significant due to the immiscibility of dichloromethane with water. This would then result in an increase in final time to kinetic arrest, as is observed, and a decrease in the hyperbolicity of the second stage of the drying curves, which is also observed.

Linking back to the goal of applying these findings to water treatment applications, adsorption is influenced by the total surface area available for pollutants to adsorb onto. As such maximising the total surface area will allow for better adsorption properties. Although the porous structure of 8% polystyrene particles with zero% AOT show the most highly ordered breath figures or surface pores, the pore sizes increase with increasing surfactant concentration as shown in Figure 7. However, this effect is only useful up until a certain point at which very irregular breath figures form with a reduction in total coverage due to the excessive coalescence of water droplets into a film.

Conclusions

In conclusion, the results of this study can be said to agree with past literature studying the role of interfacial tension on breath figure formation.

The results from this study also lend credence to the bursting hypothesis for the formation of breath figures and shows some key areas for potential research on the sub-surface level structures underneath surface level pores observed. These results may also show the potential for the use of surfactants improving the performance of polymer-based particles for use in adsorption in water treatment. This stems from the tunability in pore size and coverage area as interfacial tension is decreased via the addition of surfactant. As such, these findings suggest that there may be an ideal concentration where pore sizes of breath figures obtained are increased with addition of surfactant, but regularity of the breath figures are still maintained.

This would entail investigating the drying kinetics and transport of solvents through both the gel-like polymer skin formed during the drying process, and a thin water layer. This might allow for better understanding of the drying process when interfacial tensions are low, and for finding an expanded range of conditions and polymer-surfactant-solvent systems with

Further investigation into the bursting hypothesis can also shed light on the formation of multiple surface level pores per water droplet. In this vein, further looking into the sinking process of water droplets into the polymer surface may shed light on whether the surface level pore structures are formed from a polymer film coating water

droplets, from the bursting of the water droplet, or from the mechanical stresses acting on the polymer film.

Acknowledgements
We would like to express gratitude to Donina, Liva and Angheliu, Andru who have been of immense help. Further appreciation is extended to the Polymers and Microfluidics research group for research advice provided.

References

1. *Progress on Wastewater Treatment – 2024 Update*. (n.d.). UN-Water. Retrieved 11 December 2024, from https://www.unwater.org/publications/progress-wastewater-treatment-2024-update
2. *Water Stress: A Global Problem That's Getting Worse | Council on Foreign Relations*. (n.d.). Retrieved 11 December 2024, from https://www.cfr.org/backgrounder/water-stress-global-problem-thats-getting-worse
3. Ghosh, A., Roy, A., & Ghosh, A. K. (2023). Recycling of Commonly Used Waste Plastics to Fabricate Membranes for Filtration Applications. In H. Ismail, S. S. M., & I. R. A. (Eds.), *Recycled Polymer Blends and Composites: Processing, Properties, and Applications* (pp. 347–362). Springer International Publishing. https://doi.org/10.1007/978-3-031-37046-5_17
4. El-Samak, A. A., Ponnamma, D., Hassan, M. K., Adham, S., Karim, A., Ammar, A., Alser, M., Shurbaji, S., Eltai, N. O., & Al-Maadeed, M. A. A. (2021). Multifunctional Oil Absorption with Macroporous Polystyrene Fibers Incorporating Silver-Doped ZnO. *ACS Omega*, *6*(12), 8081–8093. https://doi.org/10.1021/acsomega.0c05683
5. Zhang, N., Zhong, S., Chen, T., Zhou, Y., & Jiang, W. (2017). Emulsion-derived hierarchically porous polystyrene solid foam for oil removal from aqueous environment. *RSC Advances*, *7*(37), 22946–22953. https://doi.org/10.1039/C7RA02953E
6. Martínez-Campos, E., Elzein, T., Bejjani, A., García-Granda, M. J., Santos-Coquillat, A., Ramos, V., Muñoz-Bonilla, A., & Rodríguez-Hernández, J. (2016). Toward Cell Selective Surfaces: Cell Adhesion and Proliferation on Breath Figures with Antifouling Surface Chemistry. *ACS Applied Materials & Interfaces*, *8*(10), 6344–6353. https://doi.org/10.1021/acsami.5b12832
7. Shi, B., Wei, H., Miu, H., Li, L., Zhu, J., Huo, Y., Xu, L., & Sun, W. (2024). Fabrication of micro-porous polymeric coating with dynamic drug-eluting property on plastic biliary stent for antiproliferative treatment. *Colloid and Interface Science Communications*, *62*, 100801. https://doi.org/10.1016/j.colcom.2024.100801
8. O'Connell, R. A., Sharratt, W. N., & Cabral, J. T. (2023). Breath Figure Assembly on Evaporating Polymer Solution Droplets in Levitation. *Physical Review Letters*, *131*(21), 218101. https://doi.org/10.1103/PhysRevLett.131.218101
9. Fukuhira, Y., Yabu, H., Ijiro, K., & Shimomura, M. (2009). Interfacial tension governs the formation of self-organized honeycomb-patterned polymer films. *Soft Matter*, *5*(10), 2037–2041. https://doi.org/10.1039/B821183C
10. Marzo, A., Barnes, A., & Drinkwater, B. W. (2017). TinyLev: A multi-emitter single-axis acoustic levitator. *Review of Scientific Instruments*, *88*(8), 085105. https://doi.org/10.1063/1.4989995
11. Vashi, A., Yadav, A. S., Nguyen, N.-T., & Sreejith, K. R. (2023). Parametric analysis of acoustically levitated droplet for potential microgravity application. *Applied Acoustics*, *213*, 109624. https://doi.org/10.1016/j.apacoust.2023.109624
12. de Gennes, P. G. (2002). Solvent evaporation of spin cast films: "Crust" effects. *The European Physical Journal E*, *7*(1), 31–34. https://doi.org/10.1140/epje/i200101169
13. (PDF) Thermoresponsive Gels. (2024). *ResearchGate*. https://doi.org/10.3390/gels3010004
14. Niimura, Y., & Hasegawa, K. (2019). Evaporation of droplet in mid-air: Pure and binary droplets in single-axis acoustic levitator. *PLoS ONE*, *14*(2), e0212074. https://doi.org/10.1371/journal.pone.0212074
15. *Breath Figure: A Nature-Inspired Preparation Method for Ordered Porous Films | Chemical Reviews*. (n.d.). Retrieved 11 December 2024, from https://pubs.acs.org/doi/10.1021/acs.chemrev.5b00069
16. Rodríguez-Hernández, J., & Bormashenko, E. (2020). Breath-Figures Formation: Physical Aspects. In J. Rodríguez-Hernández & E. Bormashenko (Eds.), *Breath Figures: Mechanisms of Multi-scale Patterning and Strategies for Fabrication and Applications of Microstructured Functional Porous Surfaces* (pp. 13–49). Springer International Publishing. https://doi.org/10.1007/978-3-030-51136-4_2
17. *Predicting Wetting Behavior from Initial Spreading Coefficients | Langmuir*. (n.d.). Retrieved 11 December 2024, from https://pubs.acs.org/doi/10.1021/la001668u
18. *Multiple interfaces in self-assembled breath figures—Chemical Communications (RSC Publishing)*. (n.d.). Retrieved 11 December 2024, from https://pubs.rsc.org/en/content/articlelanding/2014/cc/c3cc49826c/unauth
19. Tian, Y., Jiao, Q., Ding, H., Shi, Y., & Liu, B. (2006). The formation of honeycomb structure in polyphenylene oxide films. *Polymer*, *47*(11), 3866–3873. https://doi.org/10.1016/j.polymer.2006.03.081
20. Maruyama, N., Karthaus, O., Ijiro, K., Shimomura, M., Koito, T., Nishimura, S., Sawadaishi, T., Nishi, N., & Tokura, S. (1998). Mesoscopic pattern formation of nanostructured polymer assemblies. *Supramolecular Science*, *5*(3), 331–336. https://doi.org/10.1016/S0968-5677(98)00027-3
21. *On the Nature of the Breath Figures Self-Assembly in Evaporated Polymer Solutions: Revisiting Physical Factors Governing the Patterning—Bormashenko—2012—Macromolecular Chemistry and Physics—Wiley Online Library*. (n.d.). Retrieved 11 December 2024, from https://onlinelibrary.wiley.com/doi/full/10.1002/macp.201200272?saml_referrer=
22. Huh, M., Jung, M.-H., Park, Y. S., Kang, T.-B., Nah, C., Russell, R. A., Holden, P. J., & Yun, S. I. (2012). Fabrication of honeycomb-structured porous films from poly(3-hydroxybutyrate) and poly(3-hydroxybutyrate-co-3-hydroxyvalerate) via the breath figures method. *Polymer Engineering & Science*, *52*(4), 920–926. https://doi.org/10.1002/pen.22161
23. Srinivasarao, M., Collings, D., Philips, A., & Patel, S. (2001). Three-Dimensionally Ordered Array of Air Bubbles in a Polymer Film. *Science*, *292*(5514), 79–83. https://doi.org/10.1126/science.1057887
24. Muñoz-Bonilla, A., Fernández-García, M., & Rodríguez-Hernández, J. (2014). Towards hierarchically ordered functional porous polymeric surfaces prepared by the breath figures approach. *Progress in Polymer Science*, *39*(3), 510–554. https://doi.org/10.1016/j.progpolymsci.2013.08.006
25. Ma, H., Tian, Y., & Wang, X. (2011). In situ optical microscopy observation of the growth and rearrangement behavior of surface holes in the breath figure process. *Polymer*, *52*(2), 489–496. https://doi.org/10.1016/j.polymer.2010.11.031
26. *Droplet mobility on lubricant-impregnated surfaces—Soft Matter (RSC Publishing)*. (n.d.). Retrieved 11 December 2024, from https://pubs.rsc.org/en/content/articlelanding/2013/sm/c2sm27032c
27. Ma, H., Tian, Y., & Wang, X. (2011). In situ optical microscopy observation of the growth and rearrangement behavior of surface holes in the breath figure process. *Polymer*, *52*(2), 489–496. https://doi.org/10.1016/j.polymer.2010.11.031
28. Kralova, I., & Sjöblom, J. (2009). Surfactants Used in Food Industry: A Review. *Journal of Dispersion Science and Technology*, *30*(9), 1363–1383. https://doi.org/10.1080/01932690902735561

Transfer Learning for Continuous Biopharmaceutical Manufacturing

Saarah Ahmed and Hana Ruzickova
Department of Chemical Engineering, Imperial College London, U.K.

Abstract The biopharmaceutical sector, and particularly the monoclonal antibody (mAb) production subsector, is experiencing rapid growth. To meet demand and product quality targets, it is becoming increasingly desirable to transition from using Chinese Hamster Ovary (CHO) cells in traditional fed-batch bioreactors to continuous bioprocessing. This transition is complicated by the time-consuming and expensive nature of experiments and the inherent complexity of cell metabolism; effective modelling approaches can alleviate this. Machine learning approaches, specifically recurrent neural networks (RNNs), are a promising tool for the time-series prediction of metabolites. To address the limited availability of continuous process data, this project introduces a transfer learning methodology: pre-training an RNN on abundant fed-batch data for real-time predictions of process variables, then fine-tuning the model to continuous data. Although the methodology was successfully implemented, the transfer learning itself faced challenges due to limitations in data quantity and quality. The findings indicate a direction for refining transfer learning approaches in biopharmaceutical applications.

1. Introduction

Biopharmaceuticals is a high-growth sector with a global market of $1.5 trillion [1]. Monoclonal antibodies (mAbs), the largest class of biopharmaceuticals, contribute to this growth as therapeutic proteins, produced in engineered cell lines to target specific antigens on cells. They are used as effective personalized treatments for a broad range of diseases including cancers, inflammatory diseases, and infectious diseases like COVID-19. The preferred manufacturing method for these proteins is using mammalian expression systems, the primary choice being Chinese Hamster Ovary (CHO) cells as they are simple to cultivate on a large-scale and tolerate changes to external conditions such as temperature [2].

Currently, fed-batch bioreactors are the most common upstream process for mAb production, offering high volumetric productivity with relatively low operational complexity [3]. However, continuous bioprocessing methods, such as perfusion systems, are gaining popularity. The market for these methods was estimated to be worth over $218 million in 2023 and is projected to grow to over $559 million by 2028. This is thanks to advantages including reduced downtime between batches, smaller facility footprints, and greater flexibility [4]. The stable environment and shorter retention times improve consistency of product quality. Additionally, perfusion systems achieve higher viable cell concentrations of up to 200 million cells/mL and mAb titers of up to 27g/L compared to 26 million cell/mL and 13g/L in fed-batch operation [5] [6]. These factors make continuous bioprocessing methods well-suited to meet the growing demand for mAb production.

However, transitioning from fed-batch to continuous manufacturing poses challenges. Optimizing processes during development requires a deep understanding of system dynamics and real time control strategies. This is time-consuming and resource intensive as experiments in continuous systems often operate for over a month and require large volumes of media [7]. Thus, modelling is useful in process development as it reduces experimental burden. Furthermore, the CHO cell metabolism consists of multiple reaction pathways with nonlinear interactions between correlated variables [8], which are difficult to observe directly inside the cells without killing them. Thus, cell metabolism is often inferred by the time progression of metabolite concentrations. To determine these without excessive experiments, mechanistic models, composed of differential equations derived from first principles, can be used. These models are useful, but computationally expensive and heavily reliant on simplifying assumptions. In contrast, data-driven models such as machine learning models require less knowledge of the biological background and can capture nonlinear relationships between variables without underlying assumptions. They can also be readily adapted, making them ideal for real time applications such as predicting metabolite concentrations to refine control strategies [9]. An example of a machine learning technique used in this field is reinforcement learning, which has been introduced for the optimization of nutrient feeding [8]. Artificial neural networks (ANNs) are another example, which have previously been used to describe cell culture processes in mammalian cells [10]. A particularly useful subclass of ANNs is recurrent neural networks (RNNs), as they can capture temporal dependencies of process variables.

A further consideration is that data-driven models require large amounts of data under a wide range of experimental conditions to produce reliable results [9]. Although there is currently abundant fed-batch data, continuous datasets remain limited. To address this, the use of transfer learning is proposed. This is a machine learning approach that uses a model trained on one task as the starting point for a target model and is most applicable when there is a limited supply of target data [11]. The study of the application of transfer learning to the continuous bioprocessing transition has so far been limited. Its implementation is therefore the objective of this study, with the goal of fully training an RNN on fed-batch data and then fine-tuning it to continuous data.

The resulting model should output predictions for subsequent time steps to describe the cell culture progression.

Initially, the model will aid in the development of continuous processes and could later be integrated into manufacturing, enabling real time predictions to enhance operational efficiency and product quality.

2. Background

The background section will highlight the key metabolites in CHO cell production and differences between fed-batch and continuous operation to showcase the underlying patterns the model is expected to learn. This will be followed by details of the machine learning concepts used in formulating the model.

2.1. CHO Cell Metabolism

Monoclonal antibody production in CHO cells involves an initial exponential growth phase followed by a stationary phase when cell growth is approximately equal to cell decay and there is high mAb production. The key metabolites involved in both phases are glucose, glutamine, ammonia, and lactate. Glucose and glutamine are consumed by the cells as carbon and nitrogen sources respectively, and so should be monitored to schedule feed intervals or flowrates. Ammonia and lactate are inhibitory by-products which can also be consumed when feed nutrients are depleted [12]. Predicting these metabolite concentrations is therefore essential to ensure optimal cell growth, leading to higher mAb production. This is why they feature in this study amongst the process variables that the model learned to predict.

2.2. Fed-Batch and Continuous Systems

Fed-batch operation consists of a base medium used to facilitate initial cell growth, and a feed medium added at intervals to prevent nutrient depletion and extend the cell growth phase [3]. The most common continuous system, meanwhile, is a perfusion system where media is continuously fed into the bioreactor and a cell free harvest is removed. Viable and dead cells are retained in the reactor and a bleed stream is used to maintain the cell concentration [6]. The different modes of operation result in fed batch data being more transient due to feed intervals whereas continuous systems typically reach a steady state as the reactor environment is kept constant. This is why they were initially considered as different tasks, with the hope that transfer learning would enable a model to both leverage common features and adapt predictions to a steady-state mode.

2.3. Recurrent Neural Networks (RNNs)

Recurrent neural networks (RNNs) are designed for processing sequential data, like time-series data. They consist of an input layer, one or more hidden layers, and an output layer. The hidden layers incorporate recurrent connections and are often referred to as RNN layers.

An RNN layer is made up of RNN cells, each of which processes one time step of input data while maintaining a hidden state. The hidden state, passed between cells, acts as memory, enabling the RNN layer to use information from previous time steps while processing the current input [13].

The hidden state computation within each cell is given by Equation 1 below:

$$h_t = g(W_{ih}x_t + W_{hh}h_{t-1} + b_h) \quad (1)$$

and the output is calculated as shown in Equation 2:

$$y_t = W_{ho}h_t + b_o \quad (2)$$

where $t \in \mathbb{Z}_+$ is the time step index, and x, y, and h are inputs, outputs, and hidden states; b_j are bias vectors and W_{jl} are weight matrices, with $j, l \in \{i, h, o\}$ denoting the input (i), hidden (h), and output (o) layers. The function $g(\cdot)$ is a standard activation function such as tanh, sigmoid, or ReLU [14] [15].

The weights and biases are parameters that the RNN learns from training data by minimizing the predictive error, as quantified by a loss function. The gradients of the loss function with respect to the parameters are estimated, then used by an optimization algorithm to update the parameters [13].

Training with standard RNN cells often encounters the problem of vanishing or exploding gradients, where the gradients used to update the model parameters become excessively small or large.

To address these challenges, Long Short-Term Memory (LSTM) cells were developed, improving the accuracy and stability of RNNs [13] [14]. The LSTM cell structure is mathematically formulated as follows:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (3)$$
$$f_t = \sigma\left(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f\right) \quad (4)$$
$$c_t = f_t c_{t-1} + i_t \tanh\left(W_{xc}x_t + W_{hc}h_{t-1} + b_c\right) \quad (5)$$
$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad (6)$$
$$h_t = o_t \tanh(c_t) \quad (7)$$

Gate values are denoted by i (input gate), f (forget gate), and o (output gate), while c refers to the cell state. These are also used as indices for the weight matrices W and bias vectors b. $\sigma(\cdot)$ denotes a logistic sigmoidal function [15].

The LSTM cell is the most used type of cell for transfer learning applications [16], and the one implemented in all final model iterations in this study.

2.4. K-Fold Cross-Validation

RNNs are often trained and evaluated using separate datasets. However, a single data split can lead to biased performance estimates, especially when the dataset is small. K-fold cross-validation is a common technique used to address this problem and was employed in this study. It entails dividing the dataset into k subsets, or "folds". The model is trained on $k-1$ folds and validated on the remaining one. The process is repeated k times, with each fold serving as the validation set once, and the final performance metric is computed as an average over all k iterations [17].

2.5. Transfer Learning

Transfer learning entails using the knowledge gained from a source task on a target task. This approach is particularly useful when the target dataset is small or

lacks the diversity needed to train a robust model from scratch [18].

In model-based time series analysis, the predominant transfer learning technique is pre-training followed by fine-tuning. Here, the parameters (weights and biases) of a model pre-trained on source data are used to initialize a new model that is fine-tuned on target data. Fine-tuning involves controlled training, often with reduced learning rates. A special form of it is partial freezing, where certain layers or neurons in the model are kept unchanged (frozen) during training. Frozen layers are typically initialized with source parameters, while non-frozen layers may be initialized with source parameters or trained from scratch [16]. The method employed in this study leverages a combination of these approaches.

3. Methodology

The objectives of this study were (1) to develop models trained on fed-batch data capable of making one-step-ahead predictions of key process variables, and (2) to implement transfer learning approaches using continuous process data.

During the early stages of the study, a model with standard RNN cells was trialed. However, it encountered gradient instability issues, as described in Section 2.3, which prompted a transition to LSTM cells. As the study progressed, the possibility of training a model on a combined set of fed-batch and continuous data emerged. Based on these insights, the models presented in this study were developed in the following sequence:

1. Two fed-batch models were developed, one with a single LSTM layer (1-layer model) and one with two LSTM layers (2-layer model).
2. Two transfer learning models were developed, one using the 1-layer fed-batch model as a base, and the other using the 2-layer fed-batch model as a base.
3. A combined model, trained on both fed-batch and continuous data, was developed.

This structure guided the data preparation, model training, hyperparameter tuning, and evaluation processes detailed in the following sections.

3.1. Data Description and Preparation

The dataset used for this study was collected from six process runs of a CHO cell culture. Measurements were taken for the following variables:

- Viable cell density (X_v).
- Dead cell density (X_d).
- Monoclonal antibody (mAb) concentration.
- Ammonia (Amm), glucose (Glc), lactate (Lac), and glutamine (Gln) concentrations.

Four of the runs were fed-batch process runs, each spanning approximately 10 days. Measurements were taken at uneven intervals of between 4 and 30 hours, yielding 15 samples for each run. The remaining two runs were continuous process runs, one spanning 30 days and the other 40 days. Measurement intervals ranged from 9 to 90 hours, yielding 24 and 28 samples.

Table 1. Overview of data splits for different models.

Model type	Training Set		Test Set
	Training subset	Validation subset	
Fed-batch model	• 3 fed-batch runs	N/A	• 1 fed-batch run
Transfer learning model	• 1 continuous run	• 1 continuous run	N/A
Combined model	• 2 fed-batch runs • 1 continuous run	• 1 fed-batch run	• 1 fed-batch run • 1 continuous run

The dataset was used to construct training and test sets for the three model types, as summarized in Table 1.

Initially, model training was attempted using real-world scale data, but the model failed to learn effectively due to significant differences in magnitude among process variables. To address this, data normalization was implemented. Specifically, the training and test sets for each model were standardized using the mean and standard deviation of the training set, as this approach proved more effective than alternative techniques like min-max scaling. To prepare the sequences for processing, all sequences were padded to match the longest sequence in the dataset (28 samples, from the 40-day continuous run).

Due to the requirement of using one continuous sequence for training and another for validation during hyperparameter optimization and early stopping, constructing an entirely independent test set for the transfer learning models was not feasible. Future studies should prioritize obtaining sufficient continuous process data to ensure the availability of a truly unseen test set.

3.2. Model Architecture

All models were implemented in Python 3.12 using the LSTM class from PyTorch.

The input to each model was a batch of sequences, where each sequence represented data from a single fed-batch or continuous process run. As described earlier, sequences were padded, and the last time step was removed to align with the shape of the target outputs. Each sequence contained 8 features: Δt_{real}, the time difference between the current and last time step, and the 7 measured variables as listed in Section 3.1. Δt_{real} was included to account for variable measurement intervals, enabling the model to learn dependencies in real time.

The architecture consisted of one or two recurrent LSTM layers and a linear layer mapping the LSTM layer outputs to the final output dimensions. Weights in all layers were initialized using Xavier initialization, one of the most common techniques for LSTMs. Xavier initialization controls the variance of weight initialization based on the input and output dimensions of the layer. It therefore does not require the tuning of a distribution range unlike alternatives such as random initialization and decreases the probability of vanishing or exploding gradients [19].

The model produced a batch of sequences as outputs, where each sequence contained predictions one step ahead of the corresponding input sequence.

3.3. Model Training

The training of all models aimed to minimize the prediction error for the 7 process variables, as quantified by the Mean Squared Error (MSE) loss function, which is the standard for regression tasks [20]. A stochastic gradient descent (SGD) optimizer was used for all models as in initial runs it resulted in much faster convergence than adaptive algorithms. Masks were applied to ignore padding during the loss calculation, and L2-regularization was employed across all models to mitigate overfitting. This was particularly important given the limited size of the datasets.

3.4. Hyperparameter Tuning

The model hyperparameters were tuned to optimize prediction performance. The focus was on the number of LSTM layers, hidden size (number of features in the hidden state), optimizer learning rate, regularization strength, and batch size. All hyperparameter tuning was carried out using the Optuna hyperparameter optimization framework, which was selected for the efficiency of its sampling and pruning mechanisms, the ease of set-up, and its seamless integration with PyTorch [21].

The optimal hyperparameters obtained for the models for which hyperparameter optimization was implemented are summarized in Table 2.

3.5. Model-Specific Details

3.5.1. Fed-Batch Models

Fed-batch models were used as the baseline for this study, with their development workflow illustrated in Fig. 1.

K-fold cross-validation was integrated into the Optuna objective function to ensure robust model evaluation. Each of the three training sequences was treated as a fold. Early stopping was employed to terminate training on a fold if the validation loss did not improve for 10 consecutive intervals of 5 epochs. This higher patience value was informed by initial observations of transient fluctuations in loss during training. The output from the objective function to be minimized was the average validation loss across all folds.

An initial optimization of the number of layers identified a 1-layer LSTM model as optimal. To allow the implementation of transfer learning techniques requiring multiple layers, subsequent optimization was constrained to models with more than one layer. This resulted in the 2-layer model.

For the final re-training phase, all three training sequences were used, and the number of training epochs was determined by averaging the termination epochs from the three cross-validation folds.

Table 2. Overview of optimal hyperparameters found for different models.

Hyperparameter	Model		
	1-layer fed-batch	2-layer fed-batch	TL model with 2-layer base
Hidden size	104	91	N/A
Batch size	2	1	N/A
Learning rate	0.0033	0.0020	0.000010
Regularization strength	0.000090	0.0000032	0.00000011

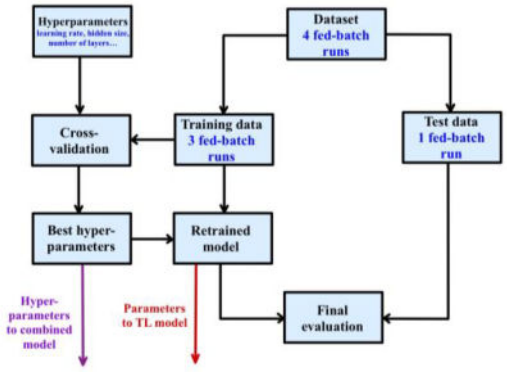


Fig. 1. Model development workflow for fed-batch models. Adapted from [17].

3.5.2. Transfer Learning Models

The development workflow for the transfer learning models is depicted in Fig. 2.

As elaborated on in Section 3.1, there was no independent test set. However, the training and validation sets were kept separate throughout the development process, including during hyperparameter optimization. This meant that results on the validation set could be used as a preliminary measure of model performance.

Two transfer learning approaches were implemented:

- 1-layer fed-batch model as base: A 2-layer LSTM model was created, where the first layer was initialized with the weights from the pre-trained 1-layer fed-batch model and frozen. The second layer was initialized using the Xavier initialization method and trained on continuous data.

- 2-layer fed-batch model as base: A 2-layer LSTM model was initialized with weights from the pre-trained 2-layer fed-batch model. The first layer was frozen, while the second layer was fine-tuned on continuous data.

3.5.3. Combined Model

The development workflow for the combined model, shown in Fig. 3, closely followed the process used for the fed-batch models. Due to time constraints, the hyperparameters from the 1-layer fed-batch model were reused rather than being specifically optimized for the combined model. Additionally, instead of training for a fixed number of epochs, early stopping was implemented by using a fed-batch sequence from the training set as a validation set.

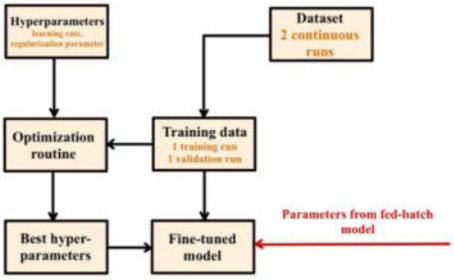


Fig. 2. Model development workflow for transfer models.

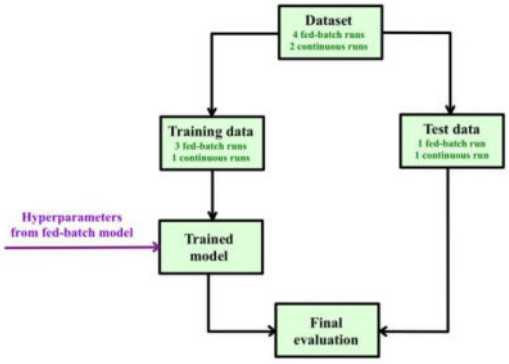


Fig. 3. Model development workflow for combined model.

3.6. Model Evaluation

All models were evaluated using three metrics: the root-mean-squared error on normalized data (nRMSE), and the coefficient of determination (R^2) and mean absolute percentage error (MAPE) on inverse-normalized data. nRMSE was obtained by taking the square root of the validation loss (nMSE) in the final model evaluation stage. It provides an assessment of overall model performance across all process variables, with the normalized scale ensuring that variables with larger magnitudes did not disproportionately influence the metric. R^2 and MAPE were calculated on inverse-normalized data to offer real-world insights into the model's predictive accuracy. They were calculated for each process variable individually and then averaged.

4. Results

The results section evaluates the models' performance across different datasets. It begins with an assessment of model performance on unseen fed-batch data to establish baseline predictive capabilities. This is followed by an evaluation of model performance on unseen continuous data, including an assessment of the effectiveness of the two transfer learning approaches.

4.1. Performance on Unseen Fed-Batch Data

The two fed-batch models were first evaluated on an unseen fed-batch test sequence. It was important to establish this baseline before using the models for transfer learning as without demonstrating learned process features, their parameters would not be useful for transfer. For comparison, the combined model was also tested on the fed-batch data, with all results presented in Fig. 4.

The 1-layer fed-batch model successfully captured some of the general trends, such as the increasing dead cell density. However, it struggled with accurately scaling parameters like ammonia and lactate, likely due to the training data insufficiently covering a range of scales for these parameters. The 2-layer fed-batch model performed worse, as its additional parameters required more data for effective training, favoring the simpler 1-layer architecture. Interestingly, the combined model performed similarly to the 1-layer fed-batch model, demonstrating that the inclusion of continuous data in the training set did not degrade its performance on unseen fed-batch data.

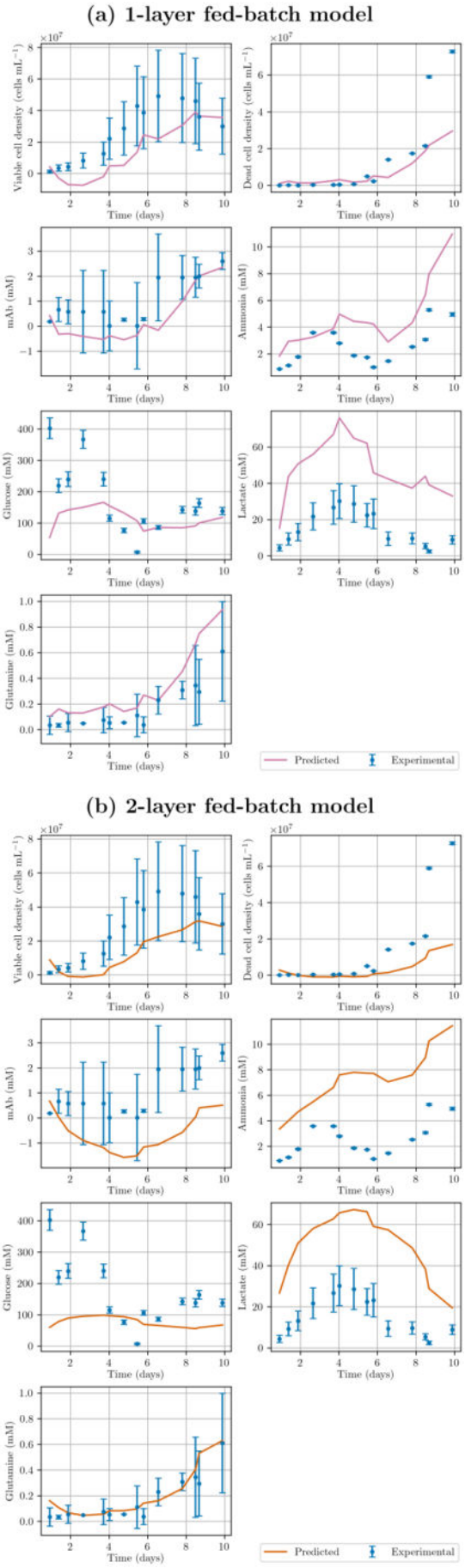


Fig. 4. Predictions of fed-batch data. (a) 1-layer fed-batch model. (b) 2-layer fed-batch model.

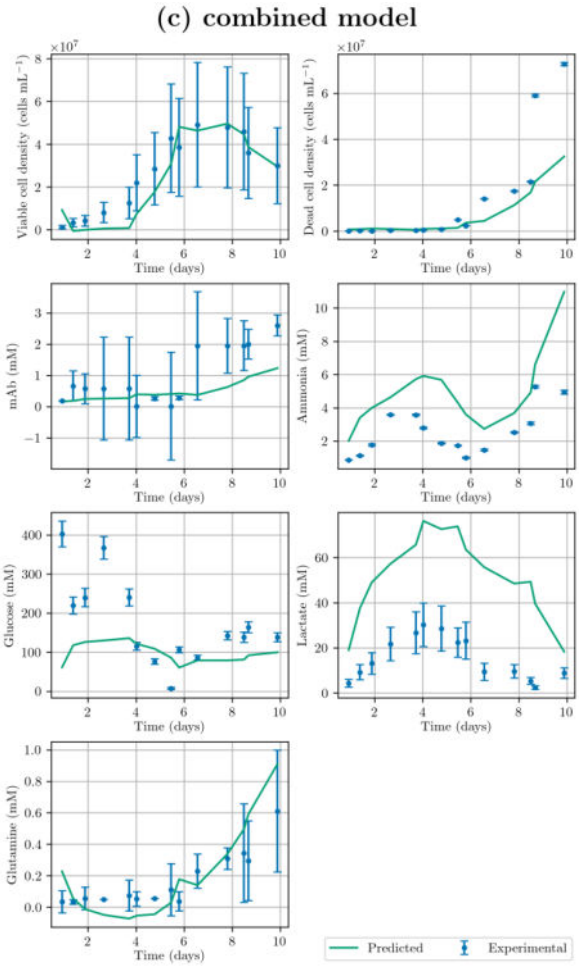


Fig. 4. (*Continued.*) Predictions of fed-batch data. (c) Combined model.

A notable trend across all models was their ability to most accurately predict process variables with higher measurement uncertainty, such as viable cell density and glutamine concentration. It would have been preferable for the accurate prediction of variables with low uncertainty to be prioritized, which highlights an area where the models could be further improved.

For quantitative comparison between models, a table of statistical metrics, Table 3, and bar charts of metrics for individual process variables, Fig. 5, are presented. These confirm that the 1-layer fed-batch model and the combined model performed comparably, while the 2-layer fed-batch model performed the worst. They also show that none of the models performed well overall. Negative R^2 values indicate that the models' predictions were worse than simply predicting the mean value for each process variable, and the MAPE values are high. It should be noted, however, that both metrics were skewed by poor performance on a handful of variables. In the case of R^2, these were ammonia and lactate due to the scaling issues mentioned earlier. For MAPE, this was the monoclonal antibody concentration since the

Table 3. Overview of model metrics for fed-batch data predictions.

Model	nRMSE (-)	R^2 (-)	MAPE (%)
1-layer fed-batch model	1.48	-2.19	475
2-layer fed-batch model	1.76	-4.00	1240
Combined model	1.62	-2.51	440

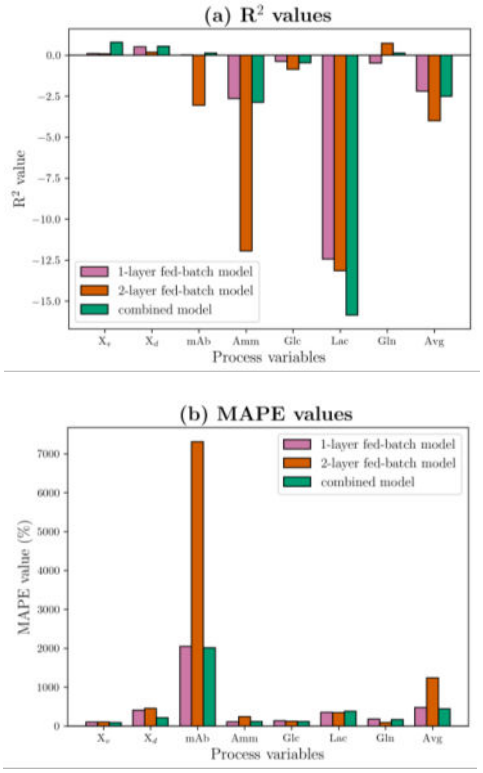


Fig. 5. Bar charts of metrics for fed-batch data predictions, broken down by process variable. (a) R^2 values. (b) MAPE values.

metric is highly sensitive when the true values of a variable are small.

Despite these challenges, the fed-batch model results were deemed sufficient to justify progressing to transfer learning, with opportunities for further optimization identified.

4.2. Performance on Unseen Continuous Data

This sub-section will begin with the two transfer learning approaches, transition to the combined model, and conclude with a summary supported by quantitative metrics.

4.2.1. Transfer Learning Using 1-Layer Base

To provide a further baseline, the unseen continuous sequence was provided as an input to the 1-layer fed-batch model. The model once again captured some of the general trends, but it struggled to accurately predict dead cell density. This is understandable given that in fed-batch processes, dead cell density tends to increase over time, while in continuous processes it remains roughly constant.

As is clear from the results displayed in Fig. 6, the transfer learning approach using this model as a base was unsuccessful. The validation loss started to increase as soon as training commenced and plateaued at a higher value than the initial baseline. Comparison of the prediction plots before and after transfer learning also confirms model degradation.

This failure likely stemmed from the single continuous sequence being insufficient to effectively train an LSTM layer from scratch. A more diverse or larger dataset would be necessary for successful transfer learning in this context. It should be noted, however, that this is difficult to procure.

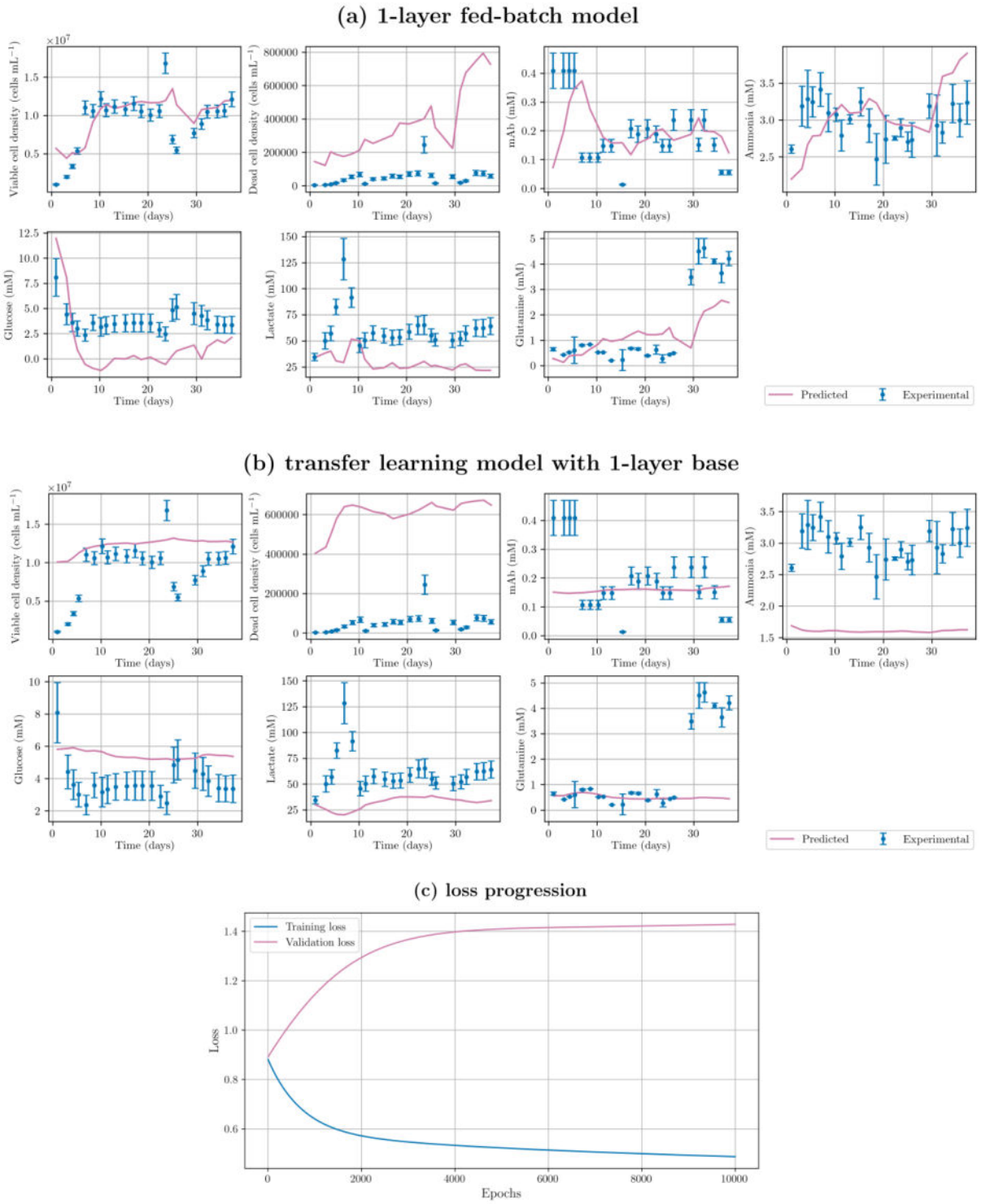


Fig. 6. Transfer learning using 1-layer model as base. (a) 1-layer fed-batch model predictions of continuous data. (b) Transfer learning model with 1-layer base predictions of continuous data. (c) Loss progression during transfer learning.

4.2.2. Transfer Learning Using 2-Layer Base

The baseline predictions for continuous data from the 2-layer fed-batch model were worse than those from the 1-layer model, as illustrated in Fig. 7. This is consistent with the 2-layer model's poorer performance on unseen fed-batch data noted in Section 4.1.

In contrast to the 1-layer transfer learning approach, transfer learning using the 2-layer fed-batch model as the base resulted in a slight decrease in validation loss during training. This is likely because it is computationally easier to adjust pre-trained parameters rather than train them from scratch. A layer initialized with pre-trained parameters is less likely to immediately overfit when faced with a single training sequence.

However, overall, this approach also cannot be considered a success, as the reduction in loss was minimal, and the prediction plots remained visually unchanged.

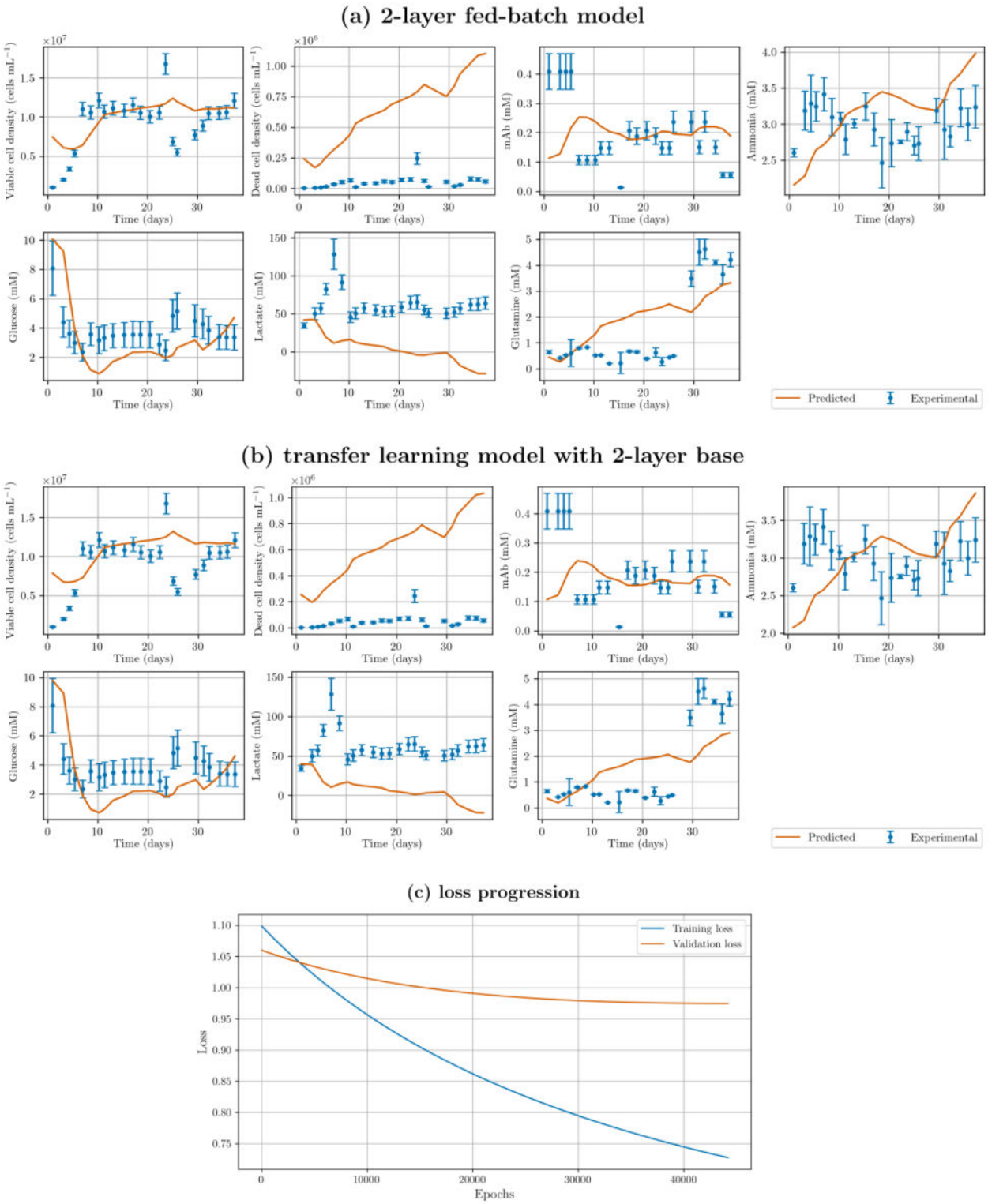


Fig. 7. Transfer learning approach using 2-layer model as base. (a) 2-layer fed-batch model predictions of continuous data. (b) Transfer learning model with 2-layer base predictions of continuous data. (c) Loss progression during transfer learning.

4.2.3. Combined Model

The combined model showed poor performance, as is evident from Fig. 8. While combining fed-batch and continuous training data did not degrade trend predictions for fed-batch data, the model failed to generalize effectively to unseen continuous data. This highlights the model's limited robustness and its strong dependence on the characteristics of the training and test datasets. Notably, no hyperparameter optimization was performed for this model, so there may be room for improvement in future iterations.

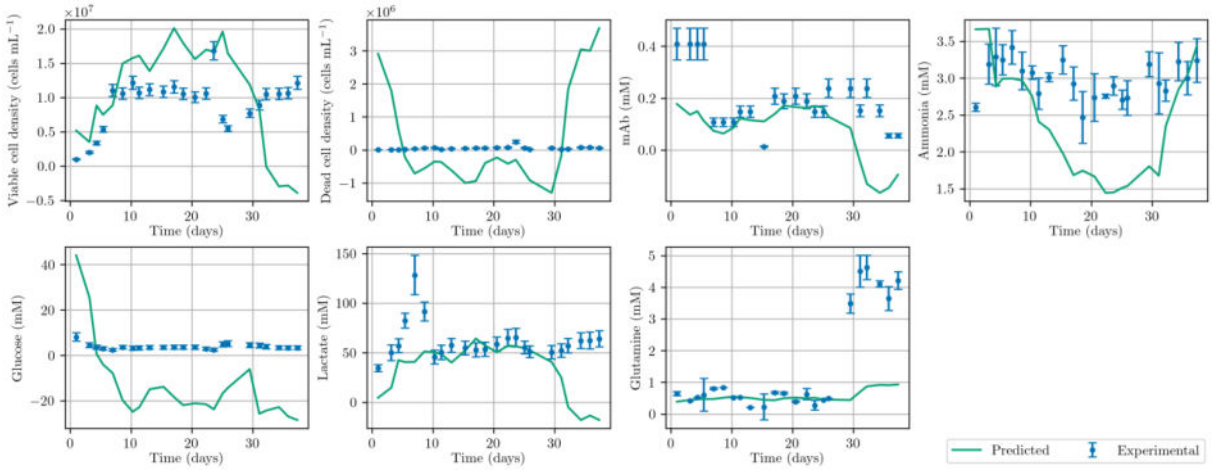


Fig. 8. Combined model predictions of continuous data.

4.2.4. Performance Overview

This section summarizes the performance of the models in predicting unseen continuous data. Statistical metrics, presented in Table 4, and bar charts for individual process variables, shown in Fig. 9, aid comparison between models. The transfer learning model using a 1-layer base was excluded due to poor performance in initial evaluations, as discussed in Section 4.2.1.

Among all the models, the 1-layer fed-batch model exhibited the best performance across all statistical metrics. However, its performance cannot be considered satisfactory, especially given its significantly negative R^2 value.

The transfer learning model with the 2-layer base outperformed the standalone 2-layer fed-batch model in all three metrics, although only marginally. It should be noted, however, that the transfer learning model's hyperparameters and early stopping were tuned on the continuous validation sequence, artificially improving performance. It is likely that predictions for a truly unseen continuous sequence would be worse.

The combined model displayed the poorest performance overall. Its R^2 value is a particularly striking indication of this, reflecting the dramatic deviation of its predictions from actual values.

Across all models, the variable that most significantly impacted the R^2 and MAPE values was the dead cell density. This suggests that the differing trends of dead cell density in fed-batch and continuous processes represent the primary challenge in applying transfer learning to this context.

Table 4. Overview of model metrics for continuous data predictions.

Model	nRMSE (-)	R^2 (-)	MAPE (%)
1-layer fed-batch model	0.81	-10	233
2-layer fed-batch model	1.03	-28	377
Transfer learning model with 2-layer base	0.99	-25	369
Combined model	2.51	-210	1230

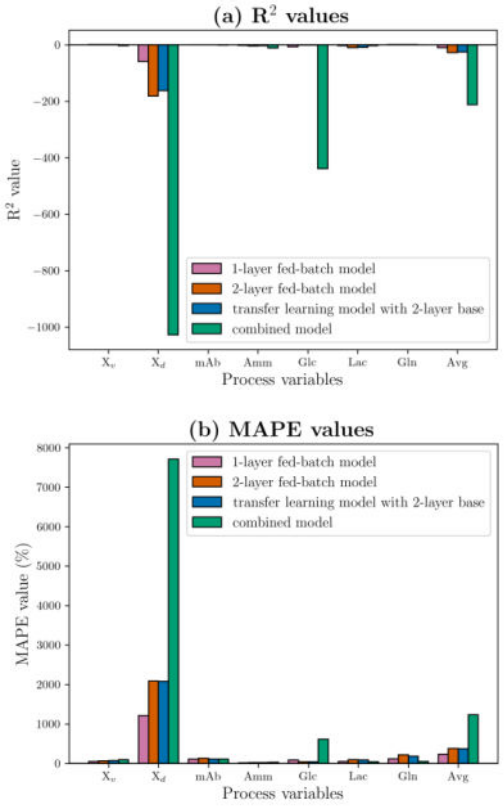


Fig. 9. Bar charts of metrics for continuous data predictions, broken down by process variable. (a) R^2 values. (b) MAPE values.

5. Conclusion

This study aimed to develop an RNN trained on fed-batch data to accurately predict process variables and apply transfer learning to enable reliable predictions for continuous data, within the context of monoclonal antibody production. Among the tested models, the 1-layer fed-batch model showed the best performance on both unseen fed-batch and continuous data, but its accuracy was still unsatisfactory. While transfer learning was implemented, it failed to significantly improve performance on continuous data, largely due to a lack of data and data diversity. Nevertheless, these findings are not sufficient to rule out transfer learning as

a promising approach for transitioning from fed-batch to continuous operations.

Data quantity and normalization emerged as critical factors affecting model performance. Input data normalization significantly enhanced model predictions, suggesting that a more comprehensive evaluation of normalization techniques could be a valuable direction for future work. Additionally, the models generally performed better on variables with higher experimental error, highlighting the need for methods that prioritize accuracy for low-uncertainty variables. Incorporating uncertainty quantification, such as through Bayesian networks, could help address this issue.

Finally, integrating domain knowledge through physics-informed RNNs that capture steady-state continuous process profiles may improve model performance by aligning predictions with known process behaviors. These approaches, combined with increased data availability and refinement of transfer learning methods, offer promising avenues for future work.

References

[1] UK Government, "Biopharmaceuticals," 2021. [Online]. Available: https://www.great.gov.uk/international/investment/sectors/biopharmaceuticals/ . [Accessed 6 December 2024].

[2] G. Sudhakaran, S. R. Babu, H. Mahendra and J. Arockiaraj, "Updated experimental cellular models to study polycystic ovarian syndrome.," *Life Sciences,* vol. 322, p. 121672, 2023. doi: https://doi.org/10.1016/j.lfs.2023.121672.

[3] Y. Fan, D. Ley and M. R. Andersen, "Fed-Batch CHO Cell Culture for Lab-Scale Antibody Production," *Methods in Molecular Biology (Clifton, N.J.),* vol. 1674, pp. 147-161, 2018. doi: https://doi.org/10.1007/978-1-4939-7312-5_12.

[4] P. Bhendale, "Bioprocess Online," 28 March 2024. [Online]. Available: https://www.bioprocessonline.com/doc/how-continuous-bioprocessing-is-shaping-modern-biopharma-manufacturing-0001. [Accessed 9 December 2024].

[5] J. Schellenberg *et al.*, "Stress-induced increase of monoclonal antibody production in CHO cells," *Engineering in Life Sciences,* vol. 22, no. 5, pp. 427-436, 2022. doi: https://doi.org/10.1002/elsc.202100062.

[6] R. Kunert and D. Reinhart, "Advances in recombinant antibody manufacturing," *Applied Microbiology and Biotechnology,* vol. 100, no. 8, pp. 3452-3461, 2016. doi: https://doi.org/10.1007/s00253-016-7388-9

[7] M. S. Croughan, K. B. Konstantinov and C. Cooney, "The future of industrial bioprocessing: Batch or continuous?," *Biotechnology and Bioengineering,* vol. 112, no. 4, pp. 648-651, 2015. doi: https://doi.org/10.1002/bit.25529.

[8] M. Monteiro and C. Kontoravdi, "Bioprocess Control: A Shift in Methodology Towards Reinforcement Learning," *Computer-aided chemical engineering,* pp. 2851-2856, 2024. doi: https://doi.org/10.1016/b978-0-443-28824-1.50476-2.

[9] A. Antonakoudis, B. Strain, R. Barbosa, I. Jimenez del Val and C. Kontoravdi, "Synergising stoichiometric modelling with artificial neural networks to predict antibody glycosylation patterns in Chinese hamster ovary cells," *Computers & Chemical Engineering,* vol. 154, p. 107471, 2021. doi: https://doi.org/10.1016/j.compchemeng.2021.107471.

[10] P. Kotidis and C. Kontoravdi, "Harnessing the potential of artificial neural networks for predicting protein glycosylation," *Metabolic Engineering Communications,* vol. 10, p. e00131, 2020. doi: https://doi.org/10.1016/j.mec.2020.e00131.

[11] D. Fagbuyiro, "Guide To Transfer Learning in Deep Learning," Medium, 20 April 2024. [Online]. Available: https://medium.com/@davidfagb/guide-to-transfer-learning-in-deep-learning-1f685db1fc94. [Accessed 6 December 2024].

[12] Y. González-Hernández and P. Perre, "Building blocks needed for mechanistic modelling of bioprocesses: A critical review based on protein production by CHO cells," *Metabolic engineering communications,* p. e00232, 2024. doi: https://doi.org/10.1016/j.mec.2024.e00232.

[13] W. Wong, E. Chee, J. Li and X. Wang, "Recurrent Neural Network-Based Model Predictive Control for Continuous Pharmaceutical Manufacturing," *Mathematics,* vol. 6, no. 11, p. 242, 2018. doi: https://doi.org/10.3390/math6110242

[14] J. Smiatek *et al.,* "Generic and specific recurrent neural network models: Applications for large and small scale biopharmaceutical upstream processes," *Biotechnology Reports,* vol. 31, p. ee00640, 2021. doi: https://doi.org/10.1016/j.btre.2021.e00640.

[15] X. Zhang, C. Zhong, J. Zhang, T. Wang and W. W. Y. Ng, "Robust recurrent neural networks for time series forecasting," *Neurocomputing,* vol. 526, pp. 143-157, 2023. doi: https://doi.org/10.1016/j.neucom.2023.01.037.

[16] M. Weber, M. Auch, C. Doblander, P. Mandl and H. -A. Jacobsen, "Transfer Learning With Time Series Data: A Systematic Mapping Study," *IEEE Access,* vol. 9, pp. 165409-165432, 2021. doi: https://doi.org/10.1109/access.2021.3134628

[17] "3.1. Cross-validation: evaluating estimator performance - scikit-learn 0.21.3 documentation," Scikit-learn.org, 2009. [Online]. Available: https://scikit-learn.org/stable/modules/cross_validation.html. [Accessed 6 December 2024].

[18] Q. -Q. He, P. C.-I. Pang and Y. -W. Si, "Transfer Learning for Financial Time Series Forecasting," *PRICAI 2019: Trends in Artificial Intelligence,* pp. 24-36, 2019. doi: https://doi.org/10.1007/978-3-030-29911-8_3.

[19] S. M. Al-Selwi, "RNN-LSTM: From applications to modeling techniques and beyond - Systematic review," *Journal of King Saud University - Computer and Information Sciences,* vol. 36, no. 5, p. 102068, 2024. doi: https://doi.org/10.1016/j.jksuci.2024.102068.

[20] L. Ciampiconi, A. Elwood, M. Leonardi, A. Mohamed and A. Rozza, "A survey and taxonomy of loss functions in machine learning," 2023. Available: https://arxiv.org/pdf/2301.05579

[21] T. Akiba, S. Sano, T. Yanase, T. Ohta and M. Koyama, "Optuna: A Next-generation Hyperparameter Optimization Framework," 2019. Available: https://arxiv.org/pdf/1907.10902

Investigation into the Impact of Impurities on the Regeneration of Paracetamol Crystals

Reuben Cambridge-Lau and Charmaine Hong
Department of Chemical Engineering, Imperial College London, U.K.

Abstract Crystal regeneration is the self-healing of a cleaved crystal, maintaining its crystal habit and reforming back to its pre-cleaved state. This phenomenon is commonly observed in the animal kingdom, amphibians such as salamanders demonstrate regeneration through the regrowth of lost limbs. This paper investigates the effects of solvent and solute (additive) impurities on the regeneration process, quantified through mass deposition and growth rates, contrasted against the normal growth of crystals. Preliminary experiments were conducted using a yellow food colouring as the impurity. The effects of the yellow food colouring were observed as an incorporation of curcumin into the crystal, polymorphic changes of the paracetamol crystal and reduced rates of regeneration. The methodology was refined to isolate polymorphic effects, through acetone solvent impurity, and rate inhibiting or impurity incorporations effects through curcumin solute impurity. Curcumin impurities resulted in a decrease of up to 54% in the regeneration rate, and presented defects around the cleavage plane, whereas the acetone increased the aspect ratio. It was proposed that these effects were due to the incorporation of impurity molecules into the lattice through the step pinning mechanism, the changes in surface attachment energy on the crystal facets and the effects of increased solution viscosity on the paracetamol's diffusion rates.

1. Introduction

Crystallisation has widespread separation applications in the pharmaceutical and fine chemical industry, in particular, it is the most common solid isolation method and plays a prominent part of the API manufacturing process. In API manufacturing, where batch processes are predominant, crystallisation is the preferred due to its high reproducibility and ability to manufacture separation products at a high purity. Fine tuning the crystallisation processes ensures control over the desired polymorphic form and size distribution of crystals which, in turn, affects the pharmacokinetic performance of the resultant drug.

Crystal breakage through mechanical means is a common occurrence in the crystallisation process. Secondary nuclei (small broken crystals) are produced via attrition and shear through crystal-crystal or crystal-impeller collision. Surface breeding can also form secondary nuclei, but this study focuses on the secondary nuclei formed through the fragmentation of crystal. Common modelling methods on the generation and growth of these secondary nuclei involves population balance equations in which the attrition contribution is grouped into a breakage term and the growth rate term of nuclei is modelled to be size-dependent (Bosetti & Mazzotti, 2019). However, these population balance equations and models are inadequate in describing the growth of secondary nucleation and the understanding of the underlying mechanisms involved are yet to be satisfactory.

Recently, a novel phenomenon was observed in which crystals broken along their cleavage plane were found to 'regenerate' back into their pre-broken state, after which the crystal will grow isotropically (Bade et al., 2023). This curious phenomenon was coined as crystal regeneration. Growth corresponding to this process will be referred to 'regeneration' throughout this study. It was measured that the rate of regeneration of the broken crystal was 3 times faster than that of the normal (unbroken) growth. Crystal growth is affected by many factors, including temperature, pH and anti-solvent. Crystal regeneration is also controlled by these same factors. The investigation and modelling of the underlying mechanism could aid the accurate assessment of crystal size distribution and efficiencies in seeded crystallisers.

In addition, the presence of impurities impacts the growth of crystals in a crystallisation process (Sangwal, 1993). Even trace amounts of impurity can have a large effect on the quality and stability of the crystallisation product. An array of impurity incorporation mechanisms has been proposed, which can be grouped into impurity incorporation mechanisms on the either the crystal interior or exterior. However, there is significant difficulty in the diagnosis and differentiation of these mechanisms.

This study aims to investigate the effects of impurities on the growth rate and habit of regenerated crystals, through the addition of solute and solvent compounds into the pure mother liquor.

2. Theory

2.1. Crystallisation

Crystallisation is the process of solid particle formation by which molecules in a homogenous liquid phase arrange to form a highly ordered 3D lattice. Crystal growth occurs through various mechanisms such as spiral growth, step-terrace-kink and birth and spread (Boistelle & Astier, 1988). The shape, size and morphology of the crystals are highly dependent on the conditions of the crystallisation process. There are existing thermodynamic and kinetic theories that can accurately predict crystal characteristics (Bosetti & Mazzotti, 2019), which take into account the intermolecular interaction and growth environment of the crystals. The effects of impurity addition to the mother liquor on crystal growth can be described through the step pinning mechanism (Ranganathan & Weeks, 2013), a diffusion-limiting model by (Ó'Ciardhá et al., 2012) and the Cabrera-Vermilyea model of dead zone formations (van Enckevort et al., 1999). The step pinning mechanism serves as a widely accepted explanation to the inhibition of crystal growth due to the adsorption of impurity molecule onto the crystal surface. Consider the edges of the growing crystal layers as advancing steps. When an impurity molecule is adsorbed onto the crystal surface, the step advancement (growth) of the adsorbed position is impeded (pinned), whilst other parts of the step continue to move forward (Figure 1). A kink is therefore created immediately in front of the step. Once the step moves past this position, the impurity molecule is assumed to be desorbed or incorporated into the lattice and would no longer have any effect on the velocity of the steps beyond it. The regular arrangement of the crystal lattice is disturbed through this mechanism, which contributes to the habit modification and defects of the crystal.

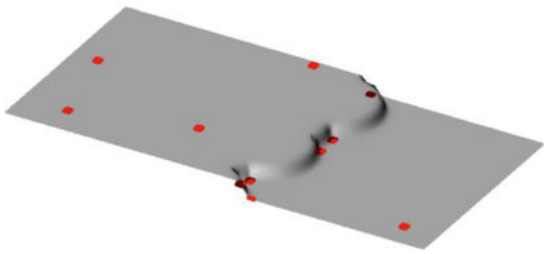


Figure 1 - 3D representation of step pinning , taken from (Ranganathan & Weeks, 2013)

2.2. Crystal Regeneration

The crystal regeneration phenomenon is similar to the autonomous self-healing of minor fissures and defects in crystals (Commins et al., 2020), but at a larger scale. When broken parallel to its cleavage plane, rather than retaining the same shape and growing isotropically, the crystal first grows back to its pre-cleaved state to regain its original morphology and crystal habit, and then undergo overall growth.

There are three distinct phases to the regeneration process: rapid regenerative growth, sluggish regeneration and slow overall growth. The first phase is particularly perplexing due to the paradoxical growth behaviour of the facet (010), which contradicts widely recognised crystal habit and growth theory. The BFDH theory (Prywer, 2005), states that the growth rate of a facet is inversely proportional to it surface attachment energy. However, the growth rate of the facet (010) in the regeneration process contradicts this theory. Research into the surface energy of crystals shows that the facet (010) has the lowest surface energy, according to the BFDH theory this facet should exhibit the slowest growth rate. But when exposed through cleavage in a regeneration experiment, the facet (010) measured the highest growth rate as seen in the rapid regeneration phase. The author of (Bade et al., 2023) attributed this to the fact that facet (010) is not a normally exposed facet in the morphologic form of a paracetamol crystal, leading to an instability which significantly increases the surface energy of the facet, hence an enhanced growth rate. The regeneration scheme exhibits directional dependence, regeneration only occurs when the crystal is cleaved along the facet (010).

The regeneration process can be exploited in milled seeds crystallisers, in which seeds undergo attrition and separated by surface area to ensure a narrow crystal size distribution. Growth rate can be enhanced by intentional attrition that targets the breakage of crystals to expose the facet (010).

2.3. Material Selection

The crystallisation system of interest was chosen to be paracetamol dissolved in ethanol, and the

impurities to be studied were yellow food colouring, acetone and curcumin.

The paracetamol-ethanol system was chosen as the primary focus of the regenerative crystallisation process due to its reproducible nature and well-documented thermodynamic and kinetic behavior. Paracetamol is a common drug of high relevance in the pharmaceutical industry, it serves as the model drug with its multipurpose analgesic and antipyretic effects. Furthermore, the large cleavage plane of a Form I paracetamol crystal from ethanol (Heng et al., 2006) makes the experiment more tractable. The monoclinic polymorph of the paracetamol crystal also provides a more stable form for the ease of handling throughout the experiments (Kolesov et al., 2011).

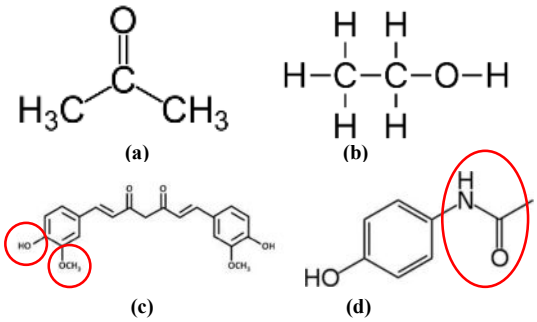


Figure 2 – Chemical structure of (a) acetone, (b) ethanol, (c) curcumin and (d) paracetamol with relevant functional groups highlighted

Curcumin was selected as the solute impurity due to various factors. Given that a curcumin molecule is twice the size of a paracetamol molecule, there might be an effect of steric hindrance on the paracetamol crystal lattice. Furthermore, the hydroxyl and methoxy groups of a curcumin molecule have the potential to interact with the hydroxyl and amide groups of a paracetamol molecule through hydrogen bonding Due to its moderate solubility of 9mg/ml (Carvalho et al., 2015) in ethanol, curcumin allows for a wide range of concentrations to be studied. The main feature

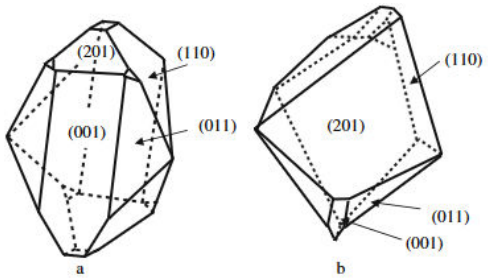


Figure 3 – (a) Diagram of paracetamol grown in ethanol, (b) Diagram of paracetamol grown in acetone. Taken from (Heng et al., 2006).

curcumin possesses is its fluorescent yellow colour, allowing the incorporation of its particles into the paracetamol crystal to be clearly observed.

Acetone was chosen as the solvent impurity. It is a widely used solvent in crystallisation processes (Kumar Bandaru et al., 2021), including the crystallisation of paracetamol. The aspect ratio of the facets on a paracetamol crystal grown from an acetone mother liquor differs from that of a crystal grown in a paracetamol-ethanol system (Heng et al., 2006), which would be of interest as the regeneration may favour one morphology over the other.

3. Methodology

3.1. Materials

Initial experiments used Sainsbury's yellow food colouring (Curcumin, Water, Polysorbate 80 and Propylene glycol). The remaining materials employed throughout the experiments were all of analytical standard and were used without further purification. The following materials were used:

Paracetamol (Acetaminophen) (98%, Sigma-Aldrich), Ethanol absolute (100%, VWR chemicals), Acetone (100%, VWR chemicals), Curcumin (Sigma-Aldrich),

3.2. Dobot photography rig

Static frame, high resolution images of the crystals were acquired to accurately record crystal growth. To record images of the crystals in the Petri dishes and vials suspended in water baths, a USB microscope (RS Pro USB digital microscope) was

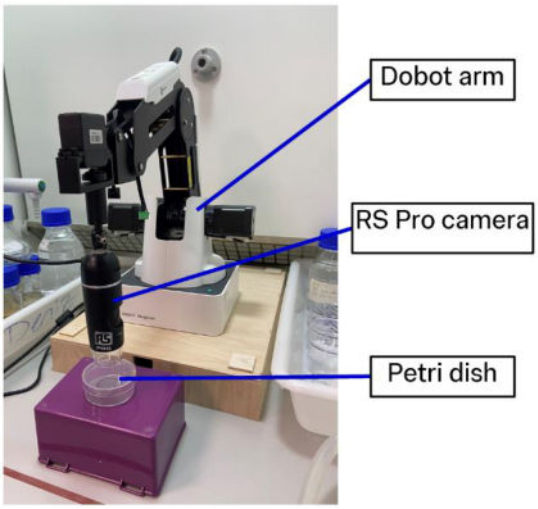


Figure 4 – Dobot rig setup for evaporative regeneration of ethanol-impurity systems.

attached to a Dobot arm (Figure 4). A python script automatically maneuvered the arm to periodically capture images of the growing crystals.

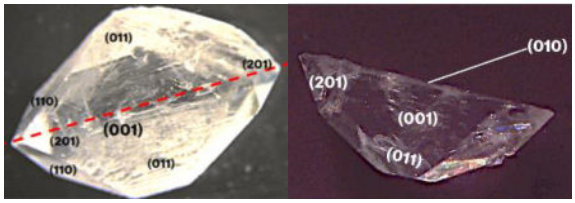
3.3. Evaporative crystal regeneration with curcumin

3.3.1. Crystal preparation

A saturated solution of 500mg/ml of paracetamol in ethanol was stirred overnight, ensuring a solid suspension was observed, then left to settle for 2 hours. 15ml of the supernatant was then filtered through a 45µm PES syringe filter into a vial, covered with parafilm containing a few pierced holes and left in a fume hood at room temperature (22°C) for 48 hours. The seeds formed were then extracted and transferred individually into vials containing 15ml of filtered saturated solution. The vials were covered with parafilm containing a few pierced holes to allow for growth via slow evaporative seeded crystallisation, the setup was placed back into a fume hood at room temperature until the size of the crystals reached 7±1mm.

3.3.2. Cleavage scheme

The cleavage plane in the crystal lattice can be found by identifying the plane with the highest interplanar spacing and the lowest attachment energy. For the paracetamol crystal lattice, the facet (010) is the cleavage plane. It is the plane on which the bonding between paracetamol molecules are weak, mainly due to the apolarity of the methyl groups. When subjected to point pressure, such as through a scalpel, the crystal is susceptible to breaking along the facet (001), exposing the cleavage plane facet (010).



(a) **(b)**

Figure 5 – Facet Miller indices and cleavage scheme for (a) an ethanol-grown paracetamol crystal and (b) a cleaved crystal.

Paracetamol crystals were cleaved into clean halves by following the methods as described in (Bade et al., 2023). A 7±1mm paracetamol crystal was cleaved with a scalpel by applying point pressure on the facet (001), a propagating fissure created by the scalpel fractures the crystal into two clean fragments with the facet (010) exposed.

3.3.3. Evaporative regeneration setup – yellow dye

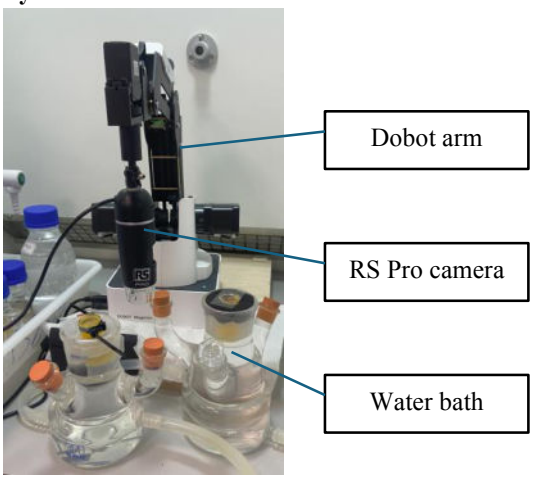


Figure 6 – Cooling regenerative set-up.

A solution of yellow dye and ethanol at a 1:10 (dye: ethanol) volumetric ratio was made, repeated for ratios 0.5-5:10, excess paracetamol was added, ensured a solid suspension was observed and stirred overnight. After being left to settle for 2 hours, 15ml of the supernatant was filtered through a 45µm PES syringe filter into a petri dish. A cleaved paracetamol crystal (section 3.3.2) was allowed to grow and recorded using the dobot photography rig, repeated for the remaining ratios of yellow dye to ethanol.

3.3.4. Evaporative regeneration set-up - curcumin

A 1wt% curcumin in saturated solution of paracetamol in ethanol was made by thoroughly dissolving 16.3mg of curcumin in 10ml of filtered saturated solution in a petri dish, this process was repeated for various wt% between 0.2-4wt%, calculated as describe below.

$$x\ wt\% = \frac{mass\ of\ curcumin\ in\ solution}{total\ mass\ of\ solute\ in\ solution}$$

The cleaved crystal (Section 3.3.2) was then placed in a petri dish and covered with a lid to allow for slow solvent evaporation at room temperature (22°C). The crystal growth was recorded using the robotic-arm-powered photography rig.

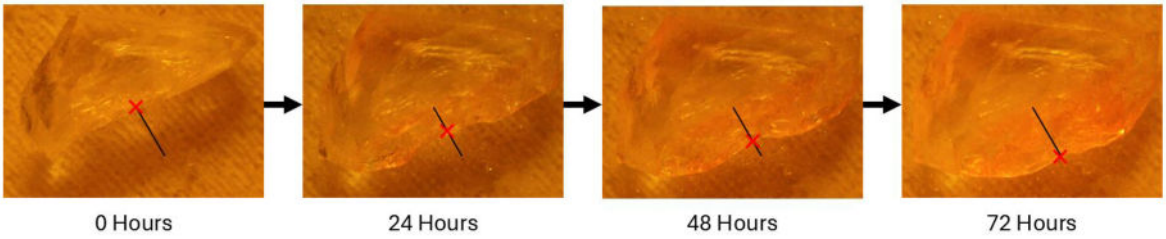


Figure 7 – Point selection for determining crystal growth

3.4. Cooling crystal regeneration in acetone-ethanol solution

3.4.1. Crystal preparation

A saturated solution of 500mg/ml of paracetamol in a solution of acetone-ethanol (1:1 vol.) was stirred overnight, ensuring solid suspension was observed, and left to settle for 2 hours at room temperature. 10ml of the supernatant was filtered and transferred into a glass vial, which was then sealed with parafilm and covered with a lid to prevent evaporation. The solution was cooled down to a set temperature in a water bath to allow for the formation of crystal seeds. The seeds were extracted, placed individually into vials containing the saturated solution at room temperature and allowed to grow via cooling crystallisation in an water bath until the size of the crystal reached 65±5µm.

3.4.2 Regeneration in acetone-ethanol solution

A 10ml vial was filled with the filtered saturated solution at room temperature. It was placed in water bath at a set temperature to achieve supersaturation. A cleaved crystal was placed at the bottom of the vial. To prevent evaporation, the vial was sealed by gluing down a glass slide onto the rim of the vial. The temperature of the water bath was maintained at a set temperature and the growth of the crystal was recorded using the robotic-arm-powered photography rig (Figure 6). The temperature of the mother liquor was cooled according to the temperature profile shown in (Figure 8)

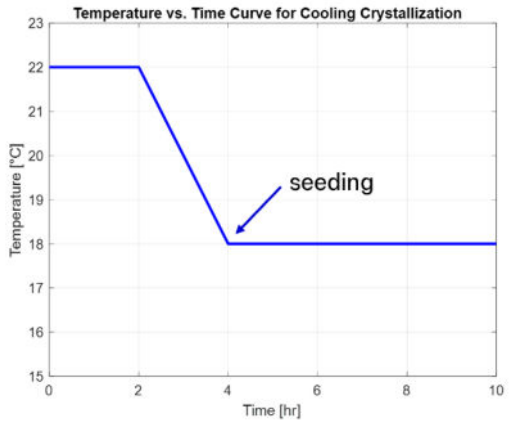


Figure 8 – Cooling profile of mother liquor in cooling regeneration

3.5. Image analysis – determination of growth rate (MATLAB)

An image analysis tool was developed on MATLAB and utilized to measure the growth rate of the crystals. This tool allows for the measurement of the growth rate of the crystals in a given direction. The set of images from the dobot rig were first stabilize and processed, through changing the brightness and sharpness of images, to make the crystal edges clearer. Next, the spatial resolution (mm per pixel) was calibrated by measuring the actual size (mm) of the crystal via an electronic microscope and compared to the same length in pixels. A guideline was selected (black line in figure 7), and the sequence of images were iterated through to record the edge of the crystals in the direction of growth, measuring the distance extended and time elapsed. This same method was employed for the growth rate in the facet (011) direction. The growth rate can be calculated by the gradient of the length-time plots, through the following equation:

$$G = \frac{\Delta L}{\Delta t}$$

G is the growth rate [mm hr^{-1}], L is the distance growth [mm] in specified direction during time [hr]

4. Results and discussion

4.1. Pure Evaporative Regeneration

The regeneration of paracetamol in pure ethanol, evaporating at room temperature, was measured as a control. The crystal rapidly regenerated compared to those with impurity and returned to its original size within 36 hours (Figure 9), growing almost double the length of the crystals with impurities present, and showing the three stages of fast regeneration, sluggish regeneration and slow overall growth.

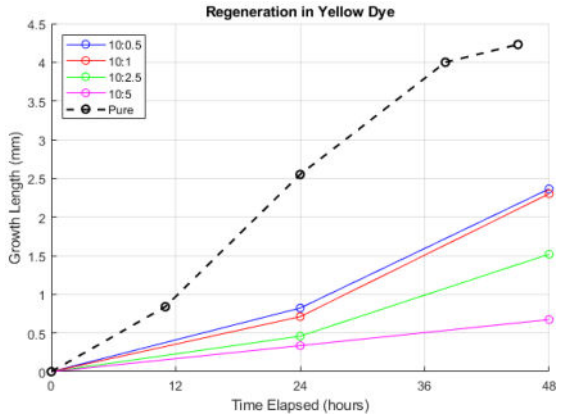


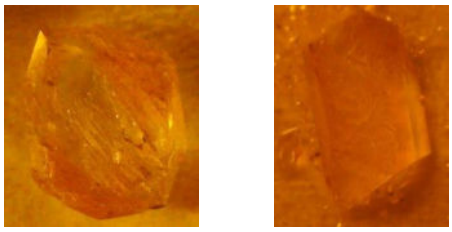
Figure 9 – Length grown against time plots for the regeneration of paracetamol crystals in volumetric ratios (ethanol to yellow food colouring) ranging from 10: 0.5 to 5

4.2 Yellow Food Colouring Impurity

Yellow food colouring was initially used to represent impurity in the system. At low ratios of ethanol to yellow colouring, regeneration is still observed. However, at higher ratios, the regeneration was seen to dramatically decrease such that at 10:5 the regeneration effectively ceased. It can be observed that the increase in volumetric ratio of yellow food colouring from 10:0.5 to 10:5 led to a 72% decrease in regeneration rate, resulting in it having a similar value to the normal growth rate from the vertex of facet (011). Past 10:1 ethanol to yellow colouring, the crystals did not finish regenerating back to their original shapes. Furthermore, the regeneration was seen to accelerate in the later 24 hours. In all cases, except the 10:5

Time period	Ratio of Ethanol:Yellow Colouring			
	10:0.5	10:1	10:2.5	10:5
0 - 24	0.034	0.030	0.019	0.014
24 - 48	0.064	0.066	0.044	0.014

Table 1: Growth rates [mm/hr] for varying ratios of ethanol:yellow dye



(a) (b)

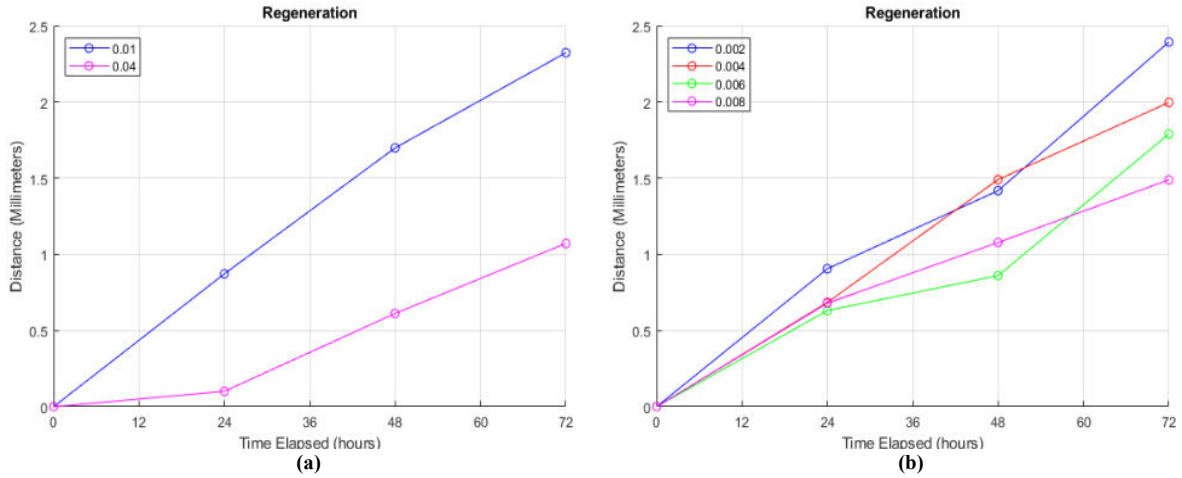
Figure 11 – Change in aspect ratio of yellow dye-ethanol crystal regeneration for (a) 10:1 and (b)10:5 (dye:eth.)

experiment, the regeneration rate during the 24–48 hour period was double that of the 0-24 hour period. The 10:5 consistently stayed at the same regeneration rate through the 48 hours (Table 1)

The resulting retardation of regeneration leads to an altered crystal habit (Figure 11), as the coplanar directions parallel to the cleavage plane grow with equal magnitudes to the regeneration, resulting in prismatic crystals rather than cubic ones, with larger aspect ratios.

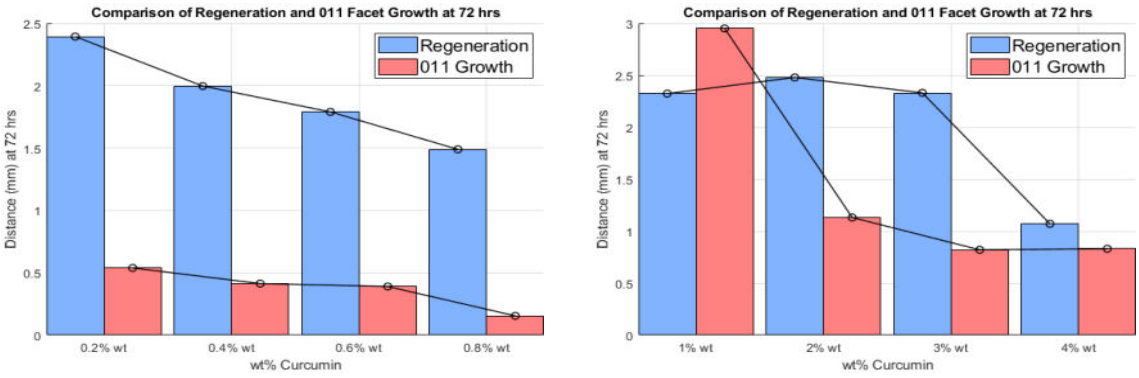
The regeneration rate inhibition could be attributed to the step pinning mechanism and the decreased evaporation rates of the system, given water and polyethylene glycol have lower vapor pressures than ethanol, leading to a delayed onset of supersaturation and a longer time for growth to initiate.

The confounding effects from the solute and solvent impurities present in the colouring mixture resulted in the refined methodology of isolating the solute and solvent effects through separate experiments with curcumin and acetone.



(a) (b)

Figure 10 – Length grown against time plots for the regeneration of paracetamol crystals with varying curcumin concentration (a) 1% and 4%, (b) 0.2-0.8%



(a) **(b)**

Figure 12 – Bar charts comparing the regeneration rate and growth rate in 011 direction of paracetamol crystals with varying curcumin concentration (a) 1wt% and 4wt%, (b) 0.2-0.8wt%

4.3. Curcumin Impurity

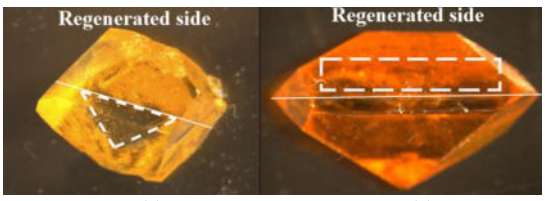
4.3.1. Impact on physical and mechanical

The addition of solute impurity was observed to have a large impact on the physical properties of the regenerated paracetamol crystals. All crystals regenerated were 'dyed' yellow throughout to various degrees, depending on the curcumin concentration in the mother liquor. Crystals regenerated at a high curcumin concentration were observed to attain a more uniform and intense colour. This difference in colour can be linked to the influence of concentration on the adsorption, described by the Langmuir adsorption model (Islam et al., 2021):

$$q_e = \frac{q_0 K_L C_e}{1 + K_L C_e}$$

q_e is the equilibrium adsorption density ($kg_{adsorbate}/kg_{adsorbent}$), q_0 is the adsorption capacity ($kg_{adsorbate}/kg_{adsorbent}$), C_e is the equilibrium adsorbate concentration [kg/m^3], and K_L is the Langmuir equilibrium constant [m^3/kg].

An increase in concentration led to an uptake in amount of solute adsorbed, due to a higher equilibrium curcumin concentration. Additionally, at 0.2wt%, the color in areas that belonged to the initial cleaved crystals was less intense, almost translucent. On the regenerated side, crystal growth occurs through advancing layers that adsorb and incorporate curcumin molecules onto the surface of these layers Section 2.1, resulting in curcumin molecules being dispersed throughout the crystal lattice. However, on the region belong to the original crystal, adsorption only occurs on the surface, which led to a lower color intensity. This phenomenon is exaggerated at low concentrations (Figure 13) where adsorption amounts are low.



(a) **(b)**

Figure 13 – Incorporation (a) and defect lines (b) in regenerated crystals

Kink and defect lines were observed in regions where growth occurs, but not in regions that belonged to the original crystal seed. The density of the defect lines is observed to be high at regions closer to the cleavage plane. The observation is postulated to be caused by the step pinning hypothesis and the rapid regeneration rate. In the step pinning mechanism, curcumin molecules are adsorbed onto the surface of the growth layers of crystals, which created new kinks in the structure. These defects reduced in frequency as regeneration continued.

In addition, the high density of defect lines also led to a change in cleavage behavior. In attempts made to cleave the regenerated crystals via the same method as described in Section 2.1, the facet (010) exposed almost always on the regenerated side of the crystal. The high-density defect lines contributed to the ease of stress fracture when pressure is applied, as these kinks act as weak spots for the crack to propagate.

To assess whether the regeneration occurs with surface adsorption of curcumin or incorporation into the crystal, all crystals were cleaved in the same method described in section 3.3.2. This confirmed the curcumin retention is internal, and that only surface adsorption occurred at lower concentrations

such that the curcumin would not fully diffuse into the uncleaved crystal.

4.3.2. Regeneration Rates and Normal Growth

Experimental results demonstrated that regardless of curcumin concentration, regeneration was observed. The regenerated lengths from the cleavage plane for the 0.2-4wt% curcumin were measured at 24-hour periods (Figure 12), for 72 hours. There is a clear trend of decreasing regenerative growth with increasing wt% of curcumin in the system. Comparing 0.2 to 0.8wt%, there is a decrease in final regenerated length of 38% and for 1 to 4wt%, a decrease of 54%. All experiments were conducted without any primary nucleation occurring; however, the supersaturation was not measured, although it is expected to increase as the ethanol evaporated faster than the curcumin and paracetamol deposited onto the crystal, observed through the solutions becoming a stronger shade of orange. All crystals regenerated the paracetamol crystal whilst maintaining the crystal habit and aspect ratio, however in the allotted time, did not regenerate in totality.

The experiments displayed the effect of the curcumin impurity inhibiting the regeneration process of the paracetamol crystal, through the decreased final growth length. The inhibitory effect reduced the regeneration and normal growth rates to different degrees, suggesting that regeneration on occurs through a different mechanism than normal growth. The normal growth rate of the paracetamol from the facet (011) is observed to be a lot slower and less sensitive to the impurity as it decreased by a small amount with increasing concentration of curcumin. The facet (011) growth has values decreasing from 0.5 to 0.2 mm for curcumin 0.2-0.8wt%, whereas the regeneration decreases from 2.4 to 1.5mm, (Figure 12). The two experiments occurred in different conditions, due to the fluctuations in room temperature, leading to different growth rates. The fluctuation explains why the 1-4wt% curcumin has a larger regeneration rate than the lower concentrations.

At the low impurity concentrations, the growth rate from the facet (011) vertex was on average 20% of the regeneration rate. At high concentrations (1-4wt%), the growth rate from the facet (011) vertex was 66% of the regeneration length, reinforcing the sensitivity of regeneration to curcumin impurities. The 1wt% curcumin presented greater facet (011) growth rates compared to the regeneration rates, an unexpected result that did not fit the trend.

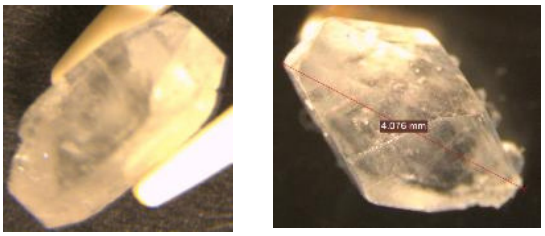
The disparity in growth rates can be explained through the different mechanisms in which regeneration and normal growth occurs. From a solution kinetics perspective, increasing the curcumin concentration will reduce the diffusion rate of paracetamol to the crystal-solution interface, inhibiting the regeneration and growth process. However, the regeneration of the cleaved facet experiences a greater degree of inhibition with increasing curcumin, such that it can be postulated that the faster regeneration process is hindered by the decreasing diffusion rates of paracetamol, leading to a greater reduction in growth rate. The solution of curcumin with higher concentrations was visually more viscous, as described by (Ó'Ciardhá et al., 2012), the molecular mobility of paracetamol is hindered by viscosity. Alternatively, as described in previous sections, the inhibition of regeneration may be caused by the step pinning mechanism, where curcumin is adsorbed into the growing step edges of the crystals, creating kinks that hinders the further advancement of the growth layers.

Aforementioned literature posits that regeneration occurs through three steps of rapid self-healing, sluggish self-healing and slow overall growth. These regions of growth were not explicitly observed, and the crystal regeneration rates did not show a trend within the 24-hour intervals, owing to interactions between the impurity and adsorption.

4.4. Isothermal Regeneration with Acetone

4.4.1. Impact on physical properties

The addition of acetone to the mother liquor altered the paracetamol crystal habit. While the morphology remains the same, paracetamol crystals grown from acetone-ethanol solution have a more prismatic



(a) **(b)**

Figure 14 – Images of (a) regenerated crystal and (b) crystal grown without regeneration, both in the 1:1 acetone:ethanol solution



(a) **(b)**

Figure 15 – Crystal grown for acetone-ethanol solution (a) pre-breakage, (b) post-breakage.

shape when compared to a crystal grown in a paracetamol-ethanol system. This phenomenon can be attributed to the change in surface attachment energy of facets due to the addition of acetone as the solvent impurity. Furthermore, the regenerated crystal from acetone-ethanol mixture was observed to have a larger aspect ratio, appearing to be longer when compared to an uncleaved crystal (Figure 14)

Crystal regeneration in a mother liquor of acetone-ethanol saturated solution was found to produce crystals of higher friability. The regenerated crystals were prone to fragmentation and would crumble at low abrasion and mechanical stress (Figure 15).

4.4.2. Regeneration rates and growth

From (Table 2), the regeneration rate for the cooling crystallisation of acetone is a fraction of the pure regeneration. In 36 hours, the growth of the acetone-ethanol system was 7.5% of the length the pure system. The reason for this could be postulated that at lower operating temperatures (18°C), reduces the crystallisation rate of the process, and that supersaturation induced by evaporation supersedes that by cooling, leading to the diminished rate. Nevertheless, comparing the mass growth rate of the cleaved to the uncleaved crystal, demonstrates that even with the acetone impurity, regeneration leads to a greater total mass than with uncleaved growth. An uncleaved crystal of mass 48mg and cleaved crystal of mass 53mg were subjected to isothermal crystallisation at 18°C, leading to the uncleaved crystal increasing in mass by only 60%, and the cleaved doubling in mass (Figure 16).

Time elapsed [hr]	23	50	71
Distance grown [mm]	0.20	0.32	0.67
Growth rate [mm/hr]	0.0087	0.0064	0.0094

Table 2 – Growth of pure paracetamol crystal in ethanol and acetone

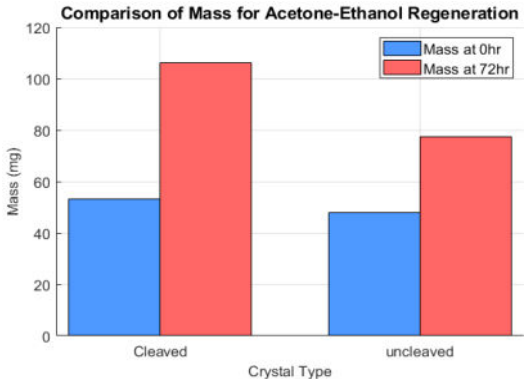


Figure 16 – Bar chart representing the change in mass of crystals grown at 17°C in acetone-ethanol solution for a cleaved an uncleaved crystal.

5. Conclusion

In summary, this paper demonstrated the effects impurities have on the regeneration process. The first impurity investigated was yellow food colouring, it was found there was a change in the crystal habit and a retardation of the regeneration rate. To isolate the effects of solute and solvent impurity on the regeneration process, experiments were conducted using curcumin and acetone as the solute and solvent impurity separately.

Both setups found that the presence of impurity caused a decrease in the rate of regenerative growth, however, the regeneration rate was still higher than that of normal growth. The presence of acetone in the mother liquor was found to cause a change in the crystal habit of the regenerated crystal, which was not observed with the curcumin. The rate of regeneration was found to decrease as the curcumin concentration increases, which was postulated as an effect of the step pinning mechanism, the Langmuir adsorption model and reduction in paracetamol diffusion. The implications of this inhibitory effect are detrimental to industrial scale crystallization processes that employ agitation systems, as the attrition and breakage of the crystal can lead to decelerated overall growth rates, the regeneration of the crystal is hindered by the presence of impurities. Furthermore, the mechanical defects observed by the regeneration of the crystal can lead to a cascade of further breakage due to the stability of the crystal diminishing, resulting in greater friability and crystal size distributions. The addition of solvent impurities into the system may not pose as big a threat to the rate of regeneration, however, does lead to altered crystal habits.

6. Outlook

Further research into the topic can investigate the effect of process conditions (e.g. temperature, evaporation rates and supersaturation level) have on the regeneration process, looking into a wider range of solute and solvent impurities, such as acids, water and metal ions. Analytical methods can be used to quantify the mechanical and physical properties observations, for example crystal habit, friability and hardness. UV-Vis spectroscopy can be utilized to analyze the intensity of fluorescence from the curcumin, to assess its incorporation. SEM and XRD imaging can be used to measure the amount of curcumin and acetone molecules integrated into the crystal lattice. Quantifying qualitative observations can also aid in the development of a robust computational model for the kinetics of crystal regeneration, which, in turn, can improve accuracy of outdated population balance equation in crystallisation processes.

REFERENCES

Bade, I., Verma, V., Rosbottom, I., & Heng, J. Y. Y. (2023). Crystal regeneration – a unique growth phenomenon observed in organic crystals post breakage. Materials Horizons, 10(4), 1425–1430. https://doi.org/10.1039/d2mh01180h

Boistelle, R., & Astier, J. P. (1988). Crystallization mechanisms in solution. *Journal of Crystal Growth*, *90*(1), 14–30. https://doi.org/10.1016/0022-0248(88)90294-1

Bosetti, L., & Mazzotti, M. (2019). Population Balance Modeling of Growth and Secondary Nucleation by Attrition and Ripening. Crystal Growth & Design, 20(1), 307–319. https://doi.org/10.1021/acs.cgd.9b01240

Carvalho, D. de M., Takeuchi, K. P., Geraldine, R. M., Moura, C. J. de, & Torres, M. C. L. (2015). Production, solubility and antioxidant activity of curcumin nanosuspension. Food Science and Technology (Campinas), 35(1), 115–119. https://doi.org/10.1590/1678-457x.6515

Commins, P., Al-Handawi, M. B., Karothu, D. P., Raj, G., & Naumov, P. (2020). Efficiently self-healing boronic ester crystals. Chemical Science, 11(10), 2606–2613. https://doi.org/10.1039/c9sc05640h

Heng, J. Y. Y., Thielmann, F., & Williams, D. R. (2006). The Effects of Milling on the Surface Properties of Form I Paracetamol Crystals. Pharmaceutical Research, 23(8), 1918–1927. https://doi.org/10.1007/s11095-006-9042-1

Islam, M. A., Chowdhury, M. A., Mozumder, Md. S. I., & Uddin, Md. T. (2021). Langmuir Adsorption Kinetics in Liquid Media: Interface Reaction Model. *ACS Omega*, *6*(22), 14481–14492. https://doi.org/10.1021/acsomega.1c01449

Kolesov, B. A., Mikhailenko, M. A., & Boldyreva, E. V. (2011). Dynamics of the intermolecular hydrogen bonds in the polymorphs of paracetamol in relation to crystal packing and conformational transitions: a variable-temperature polarized Raman spectroscopy study. *Physical Chemistry Chemical Physics: PCCP*, *13*(31), 14243–14253. https://doi.org/10.1039/c1cp20139e

Kumar Bandaru, R., Rout, S. R., Kenguva, G., Gorain, B., Alhakamy, N. A., Kesharwani, P., & Dandela, R. (2021). Recent Advances in Pharmaceutical Cocrystals: From Bench to Market. *Frontiers in Pharmacology*, *12*. https://doi.org/10.3389/fphar.2021.780582

Lutsko, J. F., Nélido González-Segredo, Durán-Olivencia, M. A., Maes, D., & Sleutel, M. (2014). Crystal Growth Cessation Revisited - the physical basis of step pinning. *Crystal Growth & Design*, *14*, 6129. https://www.researchgate.net/publication/268771698_Crystal_Growth_Cessation_Revisited_-_the_physical_basis_of_step_pinning.

Ó'Ciardhá, C. T., Mitchell, N. A., Hutton, K. W., & Frawley, P. J. (2012). Determination of the Crystal Growth Rate of Paracetamol As a Function of Solvent Composition. *Industrial & Engineering Chemistry Research*, *51*(12), 4731–4740. https://doi.org/10.1021/ie2020262

Prywer, J. (2005). Kinetic and geometric determination of the growth morphology of bulk crystals: Recent developments. Progress in Crystal Growth and Characterization of Materials, 50(1-3), 1–38. https://doi.org/10.1016/j.pcrysgrow.2005.08.001

Ranganathan, M., & Weeks, J. D. (2013). Theory of Impurity Induced Step Pinning and Recovery in Crystal Growth from Solutions. Physical Review Letters, 110(5). https://doi.org/10.1103/physrevlett.110.055503

Sangwal, K. (1993). Effect of impurities on the processes of crystal growth. *Journal of Crystal Growth*, *128*(1-4), 1236–1244. https://doi.org/10.1016/s0022-0248(07)80129-1

van Enckevort, W. J. P., van der Berg, A. C. J. F., Kreuwel, K. B. G., Derksen, A. J., & Couto, M. S. (1999). Impurity blocking of growth steps: experiments and theory. *Journal of Crystal Growth*, *166*(1-4), 156–161. https://doi.org/10.1016/0022-0248(96)00136-4

Comparison of Levasil CS50-28 and CS40-125 Colloidal Silicas to LUDOX TM-50 for the Purpose of Forming Shake-Gels

Peter Mee and Emma Bell

Department of Chemical Engineering, Imperial College London, U.K.

Abstract

The term 'shake-gel' describes a fluid-like structure that when introduced to shear force via shaking undergoes shear thickening behaviour and produces a viscoelastic solid or gel which, after a period of time, will relax back to the liquid state. Previously in literature, the silica suspension LUDOX TM-50 has been combined with polyethylene oxide (PEO) and water to form the gels. In this paper, Levasil CS50-28 and CS40-125 are compared as alternative silica suspensions for this purpose. The points for comparison are their molecular properties and how they differ in shake gels over various parameters such as relaxation time, viscoelastic character and temperature. Whilst being similar molecularly to LUDOX TM-50, CS40-125 was unable to form a shake-gel due to a large difference in pH. CS50-28 formed a shake gel but showed a different viscoelastic response to LUDOX TM-50 that may suggest an alternative mechanism for relaxation than has previously been quantified. It was concluded that the characteristics of LUDOX TM-50 gels indicate that it remains the best silica suspension for the purpose of making shake-gels.

1. Introduction

Silica shake-gels have undergone extensive research due to their unique properties. When no stress is applied, they remain as a liquid however upon shaking or an applied shear stress they form a gel. They are formed from a mixture of Poly-ethylene oxide, PEO, deionised water, and colloidal silica and can be tuned to change the strength of the gel formed or have additives added so that they can be applied in a variety of fields such as fracturing fluid in drills. Current understanding of why gels form is that when shaken the polymers stretch breaking some PEO silica bonds, leaving more active sites for the adsorption of other PEO chains onto the surface of silica particles, forming a gel structure. After time, the stretched polymer begins to reform into a random coil which causes the gel to go back to liquid. The gel structure forms from polymer bridging where multiple silica particles are bonding to a polymer chain forming a 3D network [1].

Though they have undergone a large variety of research many of the published papers focus on using LUDOX TM-50 [2] as the colloidal silica with little deviation from this. In this paper two alternative silica suspensions, Levasil CS50-28 and Levasil CS40-125 [3], which have been sold as similar to LUDOX TM-50, are compared to the LUDOX TM-50 for the purpose of making shake-gels. From this a better understanding of the key parameters required for a shake gel to form can be explored. Reasons for why LUDOX TM-50 is mainly used and further understanding of the mechanisms that form the gel are secondary aims.

To have a complete comparison of the silica suspensions their properties have been recorded before being implemented into gels and, in some cases, have been changed to match those of the LUDOX TM-50. From there, a selection of results reported for silica shake-gels in literature have been tested against the alternative silica suspensions. Key defining properties of the gels are the relaxation of the gel, viscoelasticity measurements via the storage and loss moduli, and the effects of differing PEO concentrations and temperatures [4].

Despite the volume of research, silica shake-gels currently have no definite applications. Through further understanding into the main ways of modifying their rheological properties and their underlying mechanisms, applications for them could start to be developed.

2. Background

Research into shake-gels began in the late 20th century, starting with the exploration into the addition of polymers such as PEO to silica particles, causing the reversible gelation effect. Through further research it was discovered the shake-gels are non-Newtonian liquids that undergo shear thickening behaviour [5], meaning their viscosity

will increase as an increased shear is applied. For a shake-gel, this effect continues until a rapid rise in viscosity, indicating that a gel has formed. Once in this gel state the rheology becomes unpredictable, which is why shake gels have garnered attention.

There are many focuses of research into shake-gels. Some have focused on the type of polymer that is used in shake-gel production, whilst others have focused on the silica suspensions. Work by Ramos-Tejada et al., indicates that the PEO molecular weight is an important parameter in the type of shake gel produced. The authors present that if the molecular weight is too low, the mixture stays as a liquid and will not gel, whereas a molecular weight that is too high will result in a permanent gel. They found the optimum molecular weight to be 900,000 $g.mole^{-1}$ [6].

The PEO concentration has also been investigated regarding the effects it has on shake-gels by Banerjee et al., giving insight into the shear rate requirements for gels to form. As would be expected, increasing the PEO concentration leads to the gel forming at a lower shear rate, however it also has other interesting results. The first test had samples with a concentration between 0.3 and 0.4 wt.% PEO and showed that all samples gel at constant shear rates, only differing on how long it takes to gel and the resultant gel's viscosity. This showed that PEO does help dictate the strength of gels produced and that the time taken to gel can be used to compare this. This study also investigated the effects of temperature on shake gels and found that gelation time increased from 15-20°C but decreased beyond this. This was justified by polymer coil shrinkage at higher temperatures [4].

Research has also proven that the pH of the silica determines whether the mixture will form a gel. Both the gelation and shear thickening effect only occur within the range of 8.0-9.9 pH, with longer relaxation times occurring at lower pH values. The solution would not gel at the extremes of these values if there was a high ionic strength. Therefore, it was concluded that around the midpoint of the range gave the most re-creatable results [7].

From these reports the optimal properties for shake-gel production were determined so that these could first be measured for each silica suspension and then used to help explain patterns in the results.

3. Methodology

The colloidal silica came from two suppliers. The LUDOX TM-50 was purchased from Sigma-Aldrich and the CS40-125 and CS50-28 were obtained from Nouryon. Each of these were tested for variables thought to effect shake gels, including pH, silica concentration, zeta potential and particle size. The pH was measured using a Jenway pH meter. A light scattering analyser was used for the measurement of particle size and zeta potential. For the shake gels, polyethylene oxide (PEO), also purchased from Sigma-Aldrich, was used with a molecular weight of 900,000. The solutions were completed with de-ionised water from the Imperial chemical engineering department. All materials, except for the CS40-125, underwent no processing or modification via the addition of other chemicals.

The shake gels were made with a fixed 25 wt% concentration of silica and a range of PEO concentrations between 0.3 and 0.45 wt%. Once made they were shaken by hand to form the gel and ensure thorough mixing, left to rest for a day, and then tested on a Thermo Scientific HAAKE MARS rheometer.

The first conducted test was a so-called 'ramp' test on the shear rate to ensure that gels formed and to get a basis for the constant shear experiments. To do this 17.2ml of shake gel was added to the rheometer and the temperature was set to 25°C, with an increasing shear applied from 1-1500 s^{-1}. The rheometer calculated the viscosity of the fluid by first measuring the Torque. From this the shear stress can be calculated:

$$\tau = T \times \frac{r}{J} \qquad (1)$$

Where τ is the shear stress, T is the torque, r is the radius of the container and J is the moment of inertia. From this the viscosity can be calculated using:

$$\eta = \frac{\tau}{\gamma} \qquad (2)$$

With η as the viscosity and γ the shear rate.

After this constant shear rate tests were run at 800 s^{-1} for a range of PEO concentrations, at a set temperature of 25°C. The first step in every experiment was to shear at 1 s^{-1} to gauge the initial viscosity of the solution in its liquid state. This was followed then by an hour of higher constant applied shear rate at 800 s^{-1}, before returning to the lower shear rate of 1 s^{-1} for another hour to allow the gel that had formed during the higher shear to relax.

The constant shear tests were repeated at different set temperatures, at 5-degree intervals from 20 to 45°C. The focus here was on relaxation time.

The oscillatory shear experiments were similarly fixed at a temperature of 25°C. Initially, as seen before, an increasing shear step from 1-1500 s^{-1} is required to induce the formation of the gels before the oscillating shear step is commenced with a shear strain of 0.1 and a frequency of 1 Hz. During this last step, the rheometer induces a sinusoidal shear deformation and measures the stress response, from which the viscoelastic properties of the gels can be measured. This allows, for example, for the storage and loss moduli to be plotted, which can then be used to find the phase angle, using:

$$\tan(Phase\ angle) = \frac{G'}{G''} \quad (3)$$

For CS40-125 multiple modifications were attempted with the aim of inducing a shake-gel behaviour so that it may be compared to the other silica suspensions. First was calculating its concentration. This was done by measuring the samples mass, evaporating the water, and then remeasuring. From this shake-gel calculations were re-calibrated. Then the pH was decreased to match the LUDOX TM-50, first by using 6 molar hydrochloric acid (HCl) and after using the buffer solution HEPES, also known as 4-(2-Hydroxyethyl)piperazine-1-ethane-sulfonic acid.

4. Results

4.1 Chemical and molecular properties of silica suspensions

The data gathered on the molecular properties of each silica suspension is presented in Table 4.1 below.

Table 4.1 – pH, zeta potential and average particle size for all three silica suspensions in testing.

	LUDOX TM-50	CS50-28	CS40-125
pH	9.08	9.45	10.45
Zeta potential (mV)	-37.92	-60.07	-34.57
Particle size (nm)	~ 37	~ 70	~ 41

It is noted that value of the LUDOX's particle diameter presented here is larger than that given by the suppliers who suggest a 22nm average particle size [8]. It is possible that natural aggregation within the solution gave the appearance of a larger particle size. However, the most important feature to note is that CS50-28 has a larger average particle diameter than LUDOX TM-50 and CS40-125.

The difference in zeta potential between CS50-28 and LUDOX TM-50 may be down to the more alkaline environment of CS50-28, which can contribute to a more negative zeta potential. However, it is difficult to predict the extent of this impact [9].

4.2 pH modification of CS40-125

Two different chemicals were added to the CS40-125 solution in the attempt to bring the pH of the solution from its original value of ~10.45 to a closer match to the LUDOX TM-50 at ~9.08, as described in Section 3.

Following the addition of 6M hydrochloric acid, the CS40-125 was able to form a shake-gel. These gels were of notably higher viscosity than either the LUDOX TM-50 or CS50-28 gels, with so-called 'permanent gels' forming at a PEO concentration between 0.35 wt% and 0.4 wt%. About two weeks after the HCl had been added, the CS40-125 solution then seemed to independently aggregate until it had a high enough viscosity and elasticity to support its own weight against gravity, without the addition of PEO (pictured in 4.14).



Figure 4.14 – Solution of CS40-125 and hydrochloric acid after aggregation.

After the addition of HEPES, a gel could also be formed. However, these gels failed to relax within a window that would have allowed experiments to be carried out. Hence, the primary focus for the following results will be on the LUDOX TM-50 and CS50-28 gels only.

4.3 Gel formation under constant shear rate

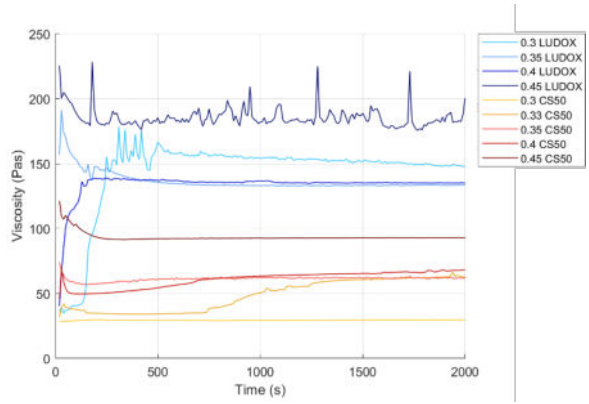


Figure 4.1 – The viscosity against time for a selection of shake gels with different PEO concentrations under 800 s^{-1} constant shear rate. LUDOX TM-50 in blues, CS50-28 in reds/oranges.

In Figure 4.1 it is shown that almost all of the tested solutions for both LUDOX TM-50 and CS50-28 formed gels instantaneously when a constant shear rate of 800 s^{-1} was applied. The exceptions to this are the 0.3 wt% PEO and 0.4 wt% PEO LUDOX TM-50 solutions which gel at the point of a sharp increase in viscosity, and the 0.3 wt% PEO CS50-28 that fails to form a gel under the shear applied here.

In the case of each silica suspension, the gel with the highest weight concentration of PEO formed the most viscous gel, excluding only the 0.3 wt% PEO LUDOX TM-50. The LUDOX TM-50 also consistently forms a more viscous gel than CS50-28 solutions, with a ~100 Pas difference between the 0.45 wt% PEO gels.

When repeating the experiment with a lower shear rate of 700 s^{-1}, the CS50-28 solutions did not transition fully to a gel.

4.4 Relaxation at low shear rate

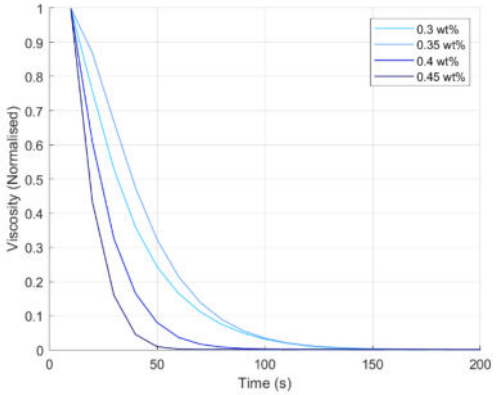


Figure 4.2 – Normalised viscosity for LUDOX TM-50 gels at different PEO concentrations during relaxation under a shear rate of 1 s^{-1}.

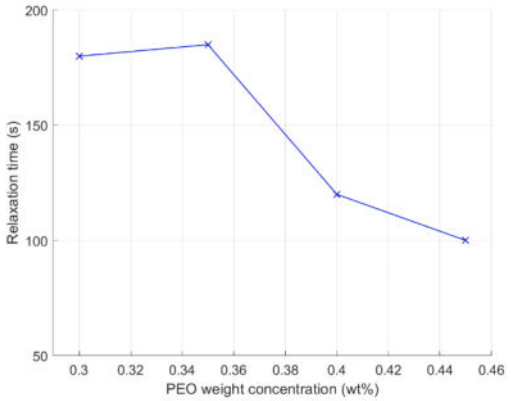


Figure 4.3 – Relaxation time (taken as time for viscosity gradient to reach zero) for LUDOX TM-50 gels compared with PEO concentration.

After forming under a 800 s^{-1} shear rate, the LUDOX TM-50 gels follow a smooth relaxation when exposed to a lower shear rate of 1 s^{-1}, characterised by a steady decrease in viscosity over time (as in Figure 4.2).

The relaxation times are quite fast - all within the window of 100-200 seconds. A noticeable trend is the decrease in relaxation time with increasing PEO concentration (Figure 4.3).

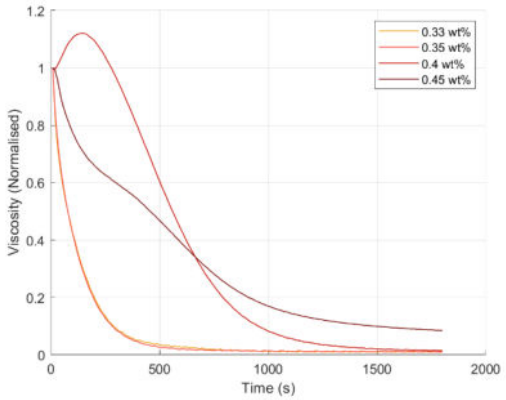


Figure 4.4 – Normalised viscosity against time for CS50-28 gels of different PEO concentrations during relaxation.

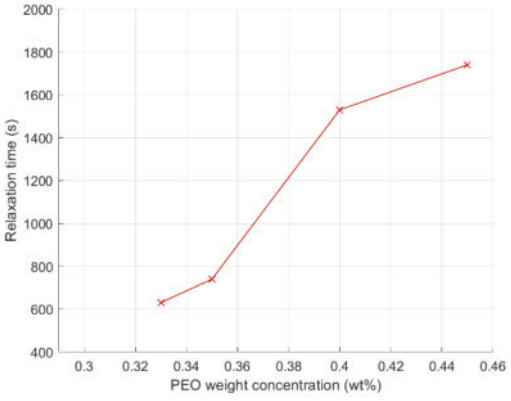


Figure 4.5 - Relaxation time (taken as time for viscosity gradient to reach zero) for CS50-28 gels compared with PEO concentration.

Similarly, Figure 4.4 shows the relaxation of the CS50-128 gels, also with normalised viscosity plotted against time.

In comparison, the CS50-128 gels exhibit much slower relaxation under the same shear conditions. The pattern of faster relaxation of gels with higher PEO concentration is notably not repeated here. In fact, Figure 4.5 clearly shows a sharp increase in relaxation time for the CS50-28 gels when the PEO concentration is increased.

Interestingly, the viscosity of the 0.4 wt% PEO CS50-28 gel increases slightly before decreasing. However, this behaviour is not repeated at any other concentration.

4.5 Oscillatory rheology measurements during gel relaxation

Following a high shear phase to form the gels, an oscillatory shear experiment at 1 Hz and 0.1 shear strain was performed. This allowed for storage and loss moduli, as well as the phase angle, to be monitored whilst the gels relaxed.

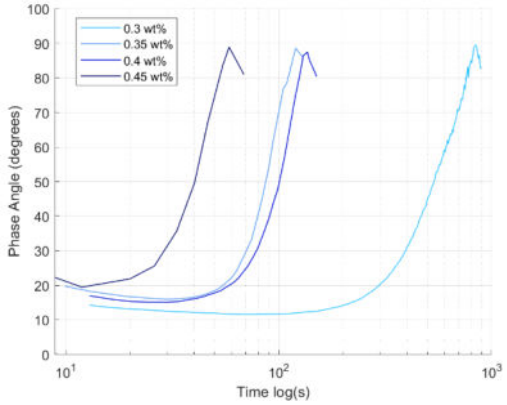


Figure 4.6 – Phase angle in degrees against a log time axis over the relaxation period for LUDOX TM-50 gels at different PEO concentrations.

Figure 4.6 shows the phase angle measured during the relaxation of LUDOX TM-50 gels. The steady increase from a lower value to a peak higher value corresponds with the relaxation of the gel from a viscoelastic solid to a liquid.

The pattern of higher concentration leading to a faster relaxation time is also reflected here, with the peak phase angle value for lower concentrations appearing later.

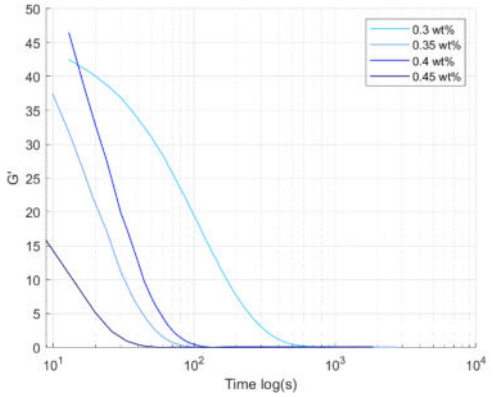


Figure 4.7 - Storage modulus against a log time axis over the relaxation period for LUDOX TM-50 gels at different PEO concentrations.

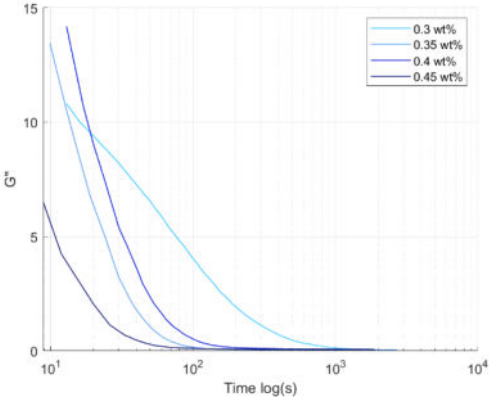


Figure 4.8 – Loss modulus against a log time axis over the relaxation period for LUDOX TM-50 gels at different PEO concentrations.

Figures 4.7 and 4.8 similarly show a decrease in the storage and loss moduli, respectively, whilst the gel relaxes.

The pattern of slower relaxation with increased PEO concentration is also repeated below.

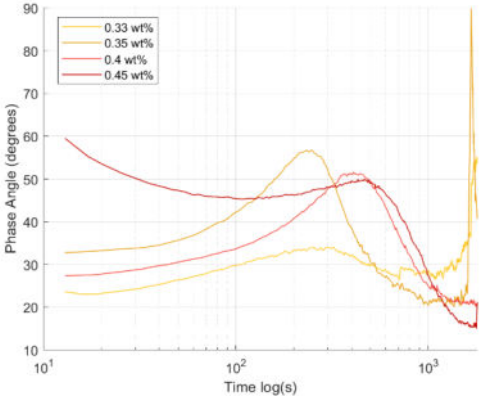


Figure 4.9 - Phase angle in degrees against a log time axis over the relaxation period for CS50-28 gels at different PEO concentrations.

The plot of the phase angle in Figure 4.9 for the relaxation of CS50-28 gels is quite different from that previously seen. With the exception of the 0.45

wt% PEO gel, the phase angle increases with time as seen with LUDOX TM-50. However, there is then a significant drop in value. The 0.33 wt% and 0.35 wt% PEO gels reach a local minimum before spiking sharply in value again. It is likely that the higher concentration gels would also reach this spike if given more time than our experiment allowed for.

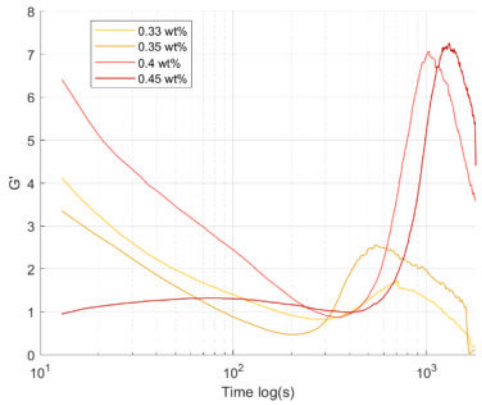


Figure 4.10 – Storage modulus against a log time axis over the relaxation period for CS50-28 gels at different PEO concentrations.

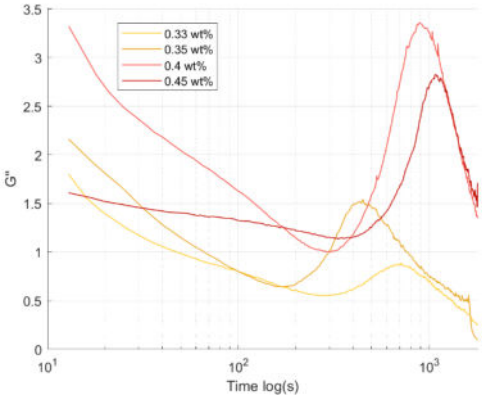


Figure 4.11 - Storage modulus against a log time axis over the relaxation period for CS50-28 gels at different PEO concentrations.

The change in storage and loss moduli of the CS50-28 gels during relaxation also differs noticeably from that of the LUDOX TM-50. In Figures 4.10 and 4.11, both the values of the storage and loss moduli initially decrease before peaking at a later point in the relaxation process. The values then proceed to drop again. In the case of the 0.4 wt% PEO gel, the moduli values reached in the peak equal their initial values at the beginning of relaxation, when the solutions were in gel form. In the 0.45 wt% PEO gel they are even exceeded.

It is worth noting that the storage and loss moduli, as well as the phase angle, for CS50-28 gels follows the same pattern as seen before in which a higher PEO concentration results in a longer relaxation. In these results this can be seen in the delay of the apperance of peaks and troughs.

4.6 Constant shear rate relaxation at different temperatures

The results in Figures 4.12 and 4.13 show a concordant response in relaxation time to decreasing temperatures between the two gel types examined. The viscosity has been normalised in both graphs so that relaxation time at each concentration is easier to compare. The PEO concentration is fixed at 0.35 wt%.

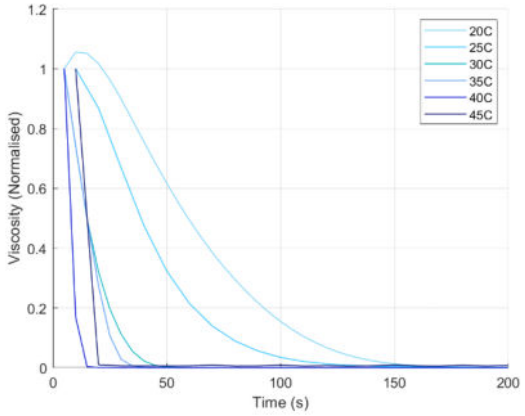


Figure 4.12 – Normalised viscosity against time of LUDOX TM-50 gel fixed at 0.35 wt% PEO concentration during relaxation at temperatures between 20-45°C.

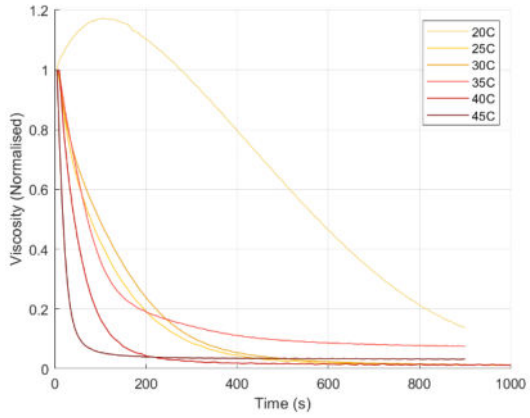


Figure 4.13 - Normalised viscosity against time of CS50-28 gel fixed at 0.35 wt% PEO concentration during relaxation at temperatures between 20-45°C.

Both the CS50-28 and LUDOX TM-50 gels show an increase in relaxation time with decreasing temperature. As before, the CS50-28 relaxes more slowly than the LUDOX TM-50 at each temperature.

5. Discussion

5.1 Characteristics of CS40-125 shake-gels

The aggregation observed in the CS40-125 after the addition of 6M hydrochloric acid can most likely be

attributed to the introduction of a large concentration of chloride ions to the solution. This behaviour has been previously documented by Ghaffari et al. [10].

The high viscosity of the shake-gels that were formed after the addition of each of these materials can likely be attributed to an over-shooting of the acidity (see discussion of pH in Section 5.2), even though in both cases the pH either matched that of the LUDOX TM-50 or was left slightly higher still.

Additionally, it may be that the HEPES buffer also interacts undesirably with the silica or PEO components of the gels as the HCl did. However, the specifics of this are harder to pin down. Work reported by Rubio and Kitchener suggest that the character of functional groups in between isolated silanol hydroxide groups can affect the adsorption of PEO molecules. Specifically, if the regions between silanol groups are hydrated or significantly ionised then PEO adsorption becomes less favourable, whereas hydrophobic regions promote PEO adsorption [11]. Therefore, there is a possibility that the characteristics of the gel are being affected by unobserved chemical interactions between the buffer and non-silanol regions on the silica particle surface, as well as by pH.

5.2 Gelation time

The generally instantaneous formation of gels comprised of both CS50-28 and LUDOX TM-50 is not reflective of what can be found in literature. Other authors using LUDOX TM-50 present a steady, low value of viscosity that suddenly peaks to a much higher value after a given time under constant shear rate. Consistently, time to gelation is shown to increase with higher polymer concentrations and decrease with higher applied shear rates [12] [4] [13]. Faster shear rates than the 800 s^{-1} in this paper are generally used, suggesting that there is another factor contributing to the results presented here. We will now go on to explain where this difference may arise from.

The reported pH range of LUDOX TM-50 is ~9-10. With the solution used here at ~9.08, it's at the lower end of this range as well as lower than that used in other papers [4] [12] [13].

The effect of pH on the interaction between silica and PEO has been widely reported. Lewis bases, in this case the ester groups of PEO, have the ability to strongly adsorb via hydrogen bonds to Bronsted acid sites. In a shake-gel these sites are the silanol groups on the surface of silica particles that tend to lose a proton in aqueous solutions [14].

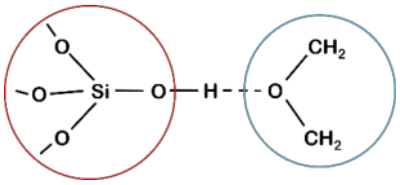


Figure 5.1 - Hydrogen bonding interaction between an isolated silanol group (circled red) and PEO oxygen (circled blue)

As has widely been agreed, the mechanism underpinning the formation of shake gels involves the elongation and adsorption of PEO onto the surface of silica particles under applied shear, resulting in a solution of large silica aggregates connected by a network of PEO 'bridges' [4] [15] [12]. As the pH decreases, the density of deprotonated silanol groups increases and thus the strength of the PEO-silica interaction increases [16] [15]. We suggest, therefore, that the lower pH of the LUDOX TM-50 used in this paper is likely responsible for its seemingly heightened ability to form a gel at lower applied shear rates.

5.3 Characteristics of a LUDOX TM-50 shake-gel

The results presented here suggest that a lower PEO weight concentration results in a longer relaxation time for a LUDOX TM-50 shake gel. This seems to be a widely accepted trend in literature [4] [17] [18]. A generally adopted explanation for this is that as PEO concentration increases, so does the degree of surface coverage on the silica particles. Therefore, fewer bonds are formed when shear is applied, leaving fewer sites for PEO to then desorb from during relaxation [5]. This allows the shorter relaxation times that have been demonstrated here.

The insight given by the oscillatory shear experiments supports trends found by other authors [4] and provides evidence to support the commonly accepted mechanism of LUDOX shake-gel relaxation. If we remember that the extremes of phase angle value at 0° and 90° represent perfectly elastic deformation behaviour (i.e. viscoelastic solid) and ideally viscous flow (i.e. fluid) behaviour respectively, then we can see the steady transition from a gel to a liquid reflected in the phase angle data here [19]. In turn, the storge modulus represents the deformation energy stored within the sample and the loss modulus is indicative of the energy dissipated through viscosity [19]. In the case of a shake-gel, the elastic deformation energy is stored in the elongation of PEO chains from their relaxed state in a randomly coiled form to forming a bridging network, whilst the viscosity is linked to the inter-silica particle distance [17]. Therefore, we can see from the steady decrease in both parameters the underlying relaxation mechanism of a LUDOX

shake-gel. During this process, the removal of the applied shear gives way for thermal energy to overcome the hydrogen bonds between PEO ester sites and deprotonated silanol groups. This allows for the extended PEO chains to return to their randomly coiled orientation. The result is the break-down of the silica-PEO clustered aggregates, which in turn increases the silica inter-particle distance.

5.4 Characteristics of a CS50-28 shake-gel

The first comparison that can be drawn between the LUDOX TM-50 and CS50-28 gels is that the CS50-28 gels generally have a lower viscosity. A possible explanation for this may be that the more negative zeta potential of the CS50-28 silica at -60.07 mV, compared to -37.92 mV in LUDOX TM-50, results in greater particle-particle repulsion. This in turn would lower viscosity as interparticle friction is reduced.

The even more notable trend that arises from the constant shear experiments is the reversal of the relationship between PEO concentration and relaxation time that has been widely observed for LUDOX TM-50 gels. In contrast to what was discussed in Section 5.2, CS50-28 gels tend to increase in relaxation time with a higher PEO concentration.

The oscillatory shear experiments seem to indicate a noticeable difference in relaxation behaviour from the LUDOX TM-50 gels which may be the reason behind this change in pattern. Where the results for the LUDOX TM-50 gels represent a smooth transition from a viscoelastic solid to a more liquid state, the peaking and troughing of the oscillatory shear variables for CS50-28 seem to indicate that the gels begin a process of relaxation but then return to a more solid state. After this, they then decrease in viscoelastic character again. Unfortunately, without performing further experiments, it is challenging to determine what is occurring on the molecular level.

However, since the relaxation graphs in Section 4.3 seem to indicate a smooth decrease in viscosity throughout relaxation, and the storage modulus is of significantly larger magnitude than the loss modulus, it is possible to say that the difference in behaviour concerns elasticity more than viscosity. That is to say, the deformation behaviour of the PEO molecules may be the focal point of difference between the relaxation of the two gel types, as opposed to the movement of the silica particles.

We will now present two theories that could explain this difference in rheological behaviour.

Firstly, the larger average silica particle size of the CS50-28 solution may allow for the PEO to adsorb on multiple active silanol sites on the same silica particle that have the potential to be far away from each other. In this process, we make the safer assumption that the fundamental mechanism behind the relaxation of LUDOX TM-50 gels is still abided by. The only difference suggested is that, after the aggregated silica clusters have dissipated, the PEO molecules remain elongated between multiple active sites on the same particle. Therefore, only elasticity is affected.

Table 5.1 – Number of silica particles per PEO molecule at a range of PEO concentrations 0.3-0.4 wt%.

	No. particles/polymer molecule	
PEO concentration (wt %)	**LUDOX TM-50**	**CS50-28**
0.3	11	0.35
0.35	10	0.3
0.4	8	0.26

Using the previously determined particle sizes of 22nm and 70nm for LUDOX TM-50 and CS50-28, respectively, it is possible to calculate the ratio of silica particles to PEO molecules in each gel (Table 5.1). Compared to the LUDOX TM-50 gels where there is a larger number of silica particles than PEO molecules at each concentration, there is a scarcity of particles in the CS50-28 gels. Having fewer than one silica particle per PEO molecule reduces the ability of a single PEO molecule to adsorb on multiple particles, not only due to their scarcity but also because of physical hinderance from other PEO chains. This increases the likelihood of adsorption onto multiple active sites on the same particle, providing support for this first theory.

Alternatively, the peaking behaviour of the oscillatory shear variables invokes the idea of a two-stage relaxation process. The moduli data suggests that a second viscoelastic state is formed in an intermediary step between the beginning and end stages of the relaxation mechanism that we have seen demonstrated for a LUDOX TM-50 gel. It is possible that, following their desorption, the PEO molecules adsorb again onto other silica surfaces during their recoiling, before breaking this new set of bonds and finally returning to their randomly coiled state in a lower viscosity solution. However, it is unclear where the energy for this second stage of hydrogen bonding would come from.

If the mechanism for the relaxation of CS50-28 gels differs from what has been previously recorded for LUDOX TM-50 gels, then it may also help to explain the difference in effect caused by increasing PEO concentration. This could feasibly

also extend into why the CS50-28 gels stop forming at a PEO concentration of 0.33 wt%, whereas LUDOX TM-50 gels can reportedly reach 0.25 wt% [12].

5.5 Effect of temperature on shake gel relaxation

In a rare moment of concordance between the two silica suspensions, both CS50-28 and LUDOX TM-50 gels relax faster under higher temperatures. This affirms what has been observed for the LUDOX TM-50 in literature [4].

The explanation behind this effect is quite simple. Thermal energy is required to break the bonds that are formed, under shear, between the PEO ester groups and deprotonated silanol hydroxides. Therefore, the more thermal energy that is supplied to the gel, the faster this desorption can occur.

This also confirms that PEO desorption is still crucial for the relaxation of CS50-28 gels in the case that their relaxation mechanism is different from LUDOX TM-50, as suggested.

6. Conclusions

The results presented in this paper lead us to conclude that, for the purpose of making shake-gels, the LUDOX TM-50 is the best silica suspension out of the three tested here. Whilst both the LUDOX TM-50 and CS50-28 are advantageous in that no modification, chemical or otherwise, is required to induce the shake-gel behaviour, the LUDOX TM-50 is able to form a gel more easily over a wider range of PEO concentrations. Depending on the applications of the gel, it may also be advantageous that LUDOX TM-50 gels tend to be more viscous and relax back to their liquid state faster.

The similarity between average particle size and zeta potential in the LUDOX TM-50 and CS40-125 suspensions suggest that it would be possible for them to form similar gels, as advertised by Nouryan, if the difference in pH can be rectified. However, the matter of determining the exact value of pH that the CS40-125 should be reduced to, and deciding which solution should be used to do it, makes the process significantly more complex. As has been displayed here, overshooting the acidity or using an additive that has interactions with the silica beyond the protonation/deprotonation of hydroxide groups can quite easily result in permanent silica aggregation. Further work could be completed to find an appropriate pH modifier, after which the properties of each shake-gel could be compared as has been done here for LUDOX TM-50 and CS50-28.

Building on this, the importance of pH in the formation of shake gels has been demonstrated across each silica solution. The acidity of the LUDOX TM-50 suspension used is the most likely reason for its instantaneous gelation under a relatively low applied shear. Furthermore, whilst there are certainly other factors behind it, it is possible that the higher pH of the CS50-28 plays a role in its slower relaxation and differing viscoelastic characteristics.

Whilst the CS50-28 silica suspension can form a shake-gel when combined with PEO, its increase in relaxation time with increased PEO concentration and its variable elasticity during relaxation make it notably different in rheological character from the LUDOX TM-50 gels. This may open-up the suggestion that shake-gel relaxation has multiple possible mechanisms, depending on the molecular properties of the silica suspension in use. Further work that observes the relaxation of CS50-28 shake-gels on a molecular-level, perhaps through FTIR-ATR spectroscopic imaging or x-ray techniques, might reveal what lies behind the variable elasticity during relaxation.

Finally, it has been affirmed that a lower temperature leads to a slower relaxation for LUDOX TM-50 shake-gels and newly documented that CS50-28 shake-gels exhibit the same trend.

7. References

[1] Y. Huang, S. Sato and M. Kobayashi, “Conditions for Shake-Gel Formation: The Relationship between the Size of Poly(Ethylene Oxide) and the Distance between Silica Particles,” *Molecules,* vol. 27, no. 22, 2022.

[2] Merck, “LUDOX® TM-50 colloidal silica,” [Online]. Available: https://www.sigmaaldrich.com/GB/en/product/aldrich/420778?srsltid=AfmBOopvOPJhpZS0YZxXpx0SmSmQfD0D4AhNu30IA3Y2WfnAaoqhegL7. [Accessed 25 October 2024].

[3] Nouryon, “Levasil® colloidal silica,” [Online]. Available: https://www.nouryon.com/products/colloidal-silica/. [Accessed 25 October 2024].

[4] D. Banerjee, J. H. Lew and P. Luckham, “On the rheological properties of silica – Polyethylene oxide dispersions: Shake gels I the effect of polymer concentration and

temperature," *Colloids and Surfaces A: Physicochemical and Engineering Aspects,* vol. 703, 2024.

[5] J. Zebrowski, V. Prasad, W. Zhang, L. M. Walker and D. Weitz, "Shake-gels: shear-induced gelation of laponite–PEO mixtures," *Colloids and Surfaces A: Physicochemical and Engineering Aspects,* vol. 213, no. 2-3, pp. 189-197, 2003.

[6] M. Ramos-Tejada and P. Luckham, "Shaken but not stirred: The formation of reversible particle – polymer gels under shear," *Colloids and Surfaces A: Physicochemical and Engineering Aspects,* vol. 471, pp. 164-169, 2015.

[7] S. Kawasaki and M. Kobayashi, "Affirmation of the effect of pH on shake-gel and shear thickening of a mixed suspension of polyethylene oxide and silica nanoparticles," *Colloids and Surfaces A: Physicochemical and Engineering Aspects,* vol. 537, pp. 236-242, 2018.

[8] Chempoint, "LUDOX® TM-50," [Online]. Available: https://www.chempoint.com/products/grace/ludox-monodispersed-colloidal-silica/ludox-colloidal-silica/ludox-tm-50. [Accessed 2 November 2024].

[9] Malvern, "Zeta potential - An introduction in 30 minutes," 2015. [Online]. Available: https://www.research.colostate.edu/wp-content/uploads/2018/11/ZetaPotential-Introduction-in-30min-Malvern.pdf. [Accessed 2 November 2024].

[10] Z. Ghaffari, H. Rezvani, A. Khalilnezhad, F. Cortes and M. Riazi, "Experimental characterization of colloidal silica gel for water conformance control in oil reservoirs," *Scientific Reports,* vol. 12, 2022.

[11] J. Rubio and J. Kitchener, "The mechanism of adsorption of poly(ethylene oxide) flocculant on silica," *Journal of Colloid and Interface Science,* vol. 57, no. 1, pp. 132-142, 1976.

[12] H. Collini, M. Mohr, P. Luckham, J. Shan and A. Russell, "The effects of polymer concentration, shear rate and temperature on the gelation time of aqueous Silica-Poly(ethylene-oxide) "Shake-gels"," *Journal of Colloid and Interface Science,* vol. 517, pp. 1-8, 2018.

[13] S. Sato and M. Kobayashi, "The Relationship between Gelation Behavior and the Amount of Polymer Dose per Silica Surface Area of "Shake-Gels" Consisting of Silica Nanoparticles and Poly(Ethylene Oxide)," *Molecules,* vol. 28, no. 8, 2023.

[14] S. Mathur and M. Moudgil, "Mechanisms of nonionic polymer adsorption on oxide surfaces," *Mining, metallurgy & exploration,* vol. 15, pp. 24-28, 1998.

[15] Y. Huang and M. Kobayashi, "Direct Observation of Relaxation of Aqueous Shake-Gel Consisting of Silica Nanoparticles and Polyethylene Oxide," *Polymers,* vol. 12, no. 5, 2020.

[16] R. Gage, E. Currie and M. Cohen Stuart, "Adsorption of Nanocolloidal SiO2 Particles on PEO Brushes," *Macromolecules,* vol. 54, no. 15, pp. 5078-5080, 2001.

[17] Q. Tian, J. Sun, M. Henderson, X. Huang, N. Li, J. Courtois and L. Almasy, "Quantitative analysis of the structural relaxation of silica-PEO shake gel by X-ray and light scattering," *Polymer Testing,* vol. 104, 2021.

[18] Y. Saito, Y. Hirose and Y. Otsubo, "Shear-induced reversible gelation of nanoparticle suspensions flocculated by poly(ethylene oxide)," *Colloids and Surfaces A: Physicochemical and Engineering Aspects,* vol. 284, no. 1-3, pp. 40-46, 2011.

[19] Anton Paar, "Basics of rheology," [Online]. Available: https://wiki.anton-paar.com/uk-en/basics-of-rheology/#viscoelastic-behavior. [Accessed 15 November 2024].

Hydrodynamic modelling to optimise hydrogen production in a photoelectrochemical reactor

David Okorogheye, Miles Lewis

Department of Chemical Engineering, Imperial College London

Abstract

Photoelectrochemical (PEC) reactors are an emerging technology that can be utilised to facilitate green hydrogen production from solar energy via water electrolysis. However, this technology is not currently suitable for commercial applications primarily due to its low achievable solar-to-hydrogen (STH) efficiency, partly attributable to bubble formation on electrode surfaces. Hence, this paper seeks to address this issue by using a COMSOL Multiphysics model to evaluate the impact of increasing the electrolyte flowrate on facilitating the removal of adhering bubbles from the electrode surfaces and consequently, the hydrogen production rates and STH efficiency achievable within the reactor. Upon model implementation, it was found that increasing the electrolyte flowrate up to about 600 ml/min led to an increase in hydrogen production achieved within the PEC reactor system as expected due to enhanced bubble removal. However, further increasing the flowrate led to a decrease in the hydrogen production rates observed as the negative impacts of the corresponding increase in other influencing factors exceed the positive effect of enhanced bubble removal. Hence, future investigations should consider initially improving the optical efficiency, and thus the photocurrent density, of the system before investigating electrolyte flowrate to optimise the overall system efficiency.

1 Introduction

Due to increasing population, industrialization and wealth generation, especially in developing countries, the global energy demand is projected to keep rising until at least 2045 [1]. Currently, fossil fuels serve as the major global energy source, accounting for about 60% of global energy generation [2]. However, a global transition away from fossil fuels is crucial for a number of reasons. First and most importantly, a transition is required to mitigate the emissions associated with fossil fuel use that increase atmospheric carbon dioxide levels and consequently, the rate of global warming. Global warming leads to a myriad of problems, ranging from increasing average global temperatures, rising sea levels, more extreme weather events, loss of biodiversity and ocean acidification [3]. Additionally, since fossil fuels are non-renewable and finite, increasing energy demands have led to a gradual depletion of fossil fuel reserves. At the current rate of fossil fuel consumption, it is projected that by 2100, the vast majority of conventional fossil fuel reserves will be depleted [4].

Thus, it is imperative to mitigate the global dependence on fossil fuels and facilitate the transition to cleaner, more renewable energy sources. Some cleaner energy sources that are currently being explored include solar, wind, hydroelectric and geothermal energy. Recently, hydrogen fuel has also emerged as a promising clean energy alternative to fossil fuels. Hydrogen is considered an environmentally friendly substance that produces no carbon emissions upon utilisation in a fuel cell and is highly versatile and efficient as an energy source, producing the highest energy per unit mass, when compared to any other fuel, at 120 MJ/kg [5]. Moreover, hydrogen provides an avenue to address the issue of intermittent energy supply

associated with renewable energy sources like wind and solar, since hydrogen produced can be stored for future use.

Based on the production process employed for hydrogen generation, there are three main types of hydrogen namely Grey, Blue and Green hydrogen [6]. Grey and Blue hydrogen are both obtained from natural gas via steam methane reforming (SMR) but the latter involves the implementation of carbon capture and storage (CCS) to mitigate carbon emissions. Green hydrogen, however, is considered a particularly interesting prospect for the future as it involves hydrogen fuel generation via electrolysis of water using energy obtained from a renewable energy source like wind or solar. Solar hydrogen is generally preferable to wind hydrogen as it is considered more economically viable in several developing regions with abundant sunlight like most African and Middle Eastern countries [7]. Nonetheless, hydrogen production from water electrolysis using solar energy (photoelectrolysis) poses several challenges that must be addressed to facilitate application on a commercial scale.

The initial approach to achieving photoelectrolysis involved the combination of photovoltaic (PV) panels and electrolysers to conduct the required energy conversions from solar to electrical and subsequently chemical energy respectively. However, this approach was not economically viable for commercial application primarily due to the high capital cost of the PV and electrolyser units, and the cost and energy losses associated with energy transport between the separate units. To address these issues, a photoelectrochemical (PEC) device has been proposed, which integrates the PV and electrolyser units into a single system capable of performing both processes required for the production of hydrogen fuel from solar energy – light capture and water electrolysis [8]. The potential benefits of PECs over PV-electrolysis systems include lower manufacturing costs due to unit integration and the mitigation of financial and energy losses associated with energy transport between system components. Furthermore, the integrated nature of the PEC provides an avenue to utilise thermal management to enhance the efficiency achieved within the system as the aqueous electrolyte extracts heat from the illuminated photoelectrode, maintaining a high photovoltage, while the solution is being heated, lowering the overpotential [9]. All the key reactions occurring are summarised below, with the last two depicting the anodic and cathodic reactions:

$$Fe_2O_3 + hv \rightarrow e^- + p^+ \tag{1}$$

$$2p^+ + H_2O \rightarrow \frac{1}{2}O_2 + 2H^+ \tag{2}$$

$$2e^- + 2H^+ \rightarrow H_2 \tag{3}$$

However, the main limitation associated with PEC devices is poor durability, which stems from the gradual corrosion of the photoelectrode semiconductor coating as photoelectrolysis occurs [8]. The choice of the semiconductor material to be implemented is further restricted by the fact that the semiconductor utilised must be capable of light transmission, not easily corroded, and catalyze one of the reactions above required to facilitate water splitting occurring at the electrolyte-electrode interfaces. Thus, PEC devices currently have a comparatively shorter lifetime of reliable performance than PV-electrolysis systems. Additionally, PEC devices currently have a significantly lower technological readiness level (TRL) and record solar-to-hydrogen (STH) efficiency achieved (16.2%) when compared to PV-electrolysis systems (30%) [10]. Recent research has centred on efforts to increase the economic viability and scalability of PEC systems by enhancing the STH efficiency achievable and durability of proposed PEC systems.

Several approaches have been taken in a bid to improve the STH efficiency achievable and rate of hydrogen production by altering a wide range of potential factors that influence the performance of the PEC system. Many studies have investigated the impact of enhancing the semiconductors, photocatalysts, and electrolytes utilised on the rate of hydrogen production and STH efficiency. Investigations have also been conducted to determine the effect of other influencing factors like reactor geometry and configuration, quantum efficiency, and concentration of solar irradiance on the STH efficiency and rate of hydrogen production observed in the designed PEC system. However, although it has been identified that bubble formation on the photoelectrode surface negatively impacts the rate of hydrogen production and consequently, the STH efficiency of the PEC system, there is still currently insufficient research into precisely the best means to address this issue.

Bubbles could adversely affect the rate of hydrogen production in the PEC system in a

number of ways including scattering the incident light, increasing the ionic and charge transfer resistances, and decreasing key parameters that impact hydrogen generation like ionic conductivity, current density and photoelectrode active area over time [11]. Prior investigations have determined that several factors could influence bubble coverage in the PEC reactor including electrolyte temperature, system operating pressure, and photoelectrode surface properties. However, it has been established that the most effective parameter that can be altered to facilitate the removal of adhering bubbles is the electrolyte velocity with preliminary studies recommending a high electrolyte velocity to mitigate bubble formation on the photoelectrode surface while also facilitating mass transport of species within the system. Hence, the primary objective of this investigation is to comprehensively evaluate the impact of altering electrolyte velocity on the STH efficiency and hydrogen production rate achieved within a PEC reactor system by model simulation using the COMSOL software.

2 Methodology

2.1 Simulation overview

To simulate the potential effect of electrolyte velocity on the PEC performance, the implementation of a numerical model that takes into account all the relevant kinetics, physical phenomena and boundary conditions is required. COMSOL Multiphysics is a suitable modelling software that can be utilised as it provides an avenue to optimise prediction accuracy when simulating systems, like PEC reactors, that involve several key processes ranging from hydrodynamics to ion transfer and convection [12]. COMSOL Multiphysics utilises equations via accessible interfaces that facilitate the modelling of the required processes. Each interface provides the opportunity to merge the in-built interfaces with user-specific parameters and equations, alongside the ability to change default numbers where appropriate. Electrochemical processes are usually covered by the Primary, Secondary, or Tertiary Current Distribution interfaces.

The Primary and Secondary Current Distribution interfaces both presume that the current in the electrolyte, which is the ionically conducting aqueous phase, conforms to Ohm's law and help to determine electrolyte potential at a specified conductivity value. Similarly, at both interfaces, it is assumed that the composition of the electrolyte is fairly unchanged by electrolysis and that potential losses due to the electrochemical processes occurring can be neglected. However, the key difference between both interfaces is that, unlike the Primary interface which fails to consider electrode kinetics, the Secondary interface permits the integration of an electrode kinetic law that correlates the current density associated with a specific reaction to its electrode overpotential. The Tertiary Current Distribution interface directly incorporates the mass transport of species into the model simulation, but this introduces nonlinear equations into the model, thus increasing its computational complexity and is not needed when electrolyte composition remains fairly constant. The model interfaces utilised during the scope of this investigation include the Secondary Current Distribution (SIEC) coupled with the Transport of Diluted Species (TDS) physics and the Euler-Euler model, Laminar-Flow interface for flow simulation.

2.2 Secondary current distribution model interface

The Secondary Current Distribution (SIEC) model interface was ultimately selected as the most appropriate interface for implementation to model the electrochemical processes occurring in the PEC reactor system. This deduction was made because this model interface provided the best avenue to balance model simulation accuracy and computational efficiency. Due to the stable ionic strength of the electrolyte solution, it can be presumed that the current in the electrolyte conforms to Ohm's law, provided conductivity is kept constant. As such, the Secondary Current Distribution model interface is sufficient for implementation to determine electrolyte potential. This model interface is usually applicable when the electrochemical processes occurring are solely dependent on charge transfer and the assumption can be made that the concentration gradient existing is insignificant. The Secondary Current Distribution model interface is best suited for model simulation involving cells with negligible concentration overpotential, resulting from either highly efficient mixing or a concentrated electrolyte solution, and losses emanating from electrode kinetics that are significant in comparison to the ohmic losses incurred.

However, in order to incorporate the impact

of concentration on electrode kinetics, we can couple this model interface to a chemical species transport model which accounts for all the relevant factors influencing the migration of chemical species. For this specific application, the physics interface selected was the Transport of Diluted Species (TDS) interface, which is suitable for modelling the transport of a dilute solute in a solvent [13]. Driving forces for the transport of dilute chemical species include diffusion governed by Fick's law, convection upon coupling to a flow field, and migration, upon coupling to an electric field (electrokinetic flow). Thus, since the model being considered also involves coupling to a flow field, diffusion and convection are the main driving forces for species transport.

2.3 Euler-Euler model, Laminar flow interface

Multiphase flow, as the name indicates, involves the flow of a mixture consisting of species existing in more than one distinct phase, like gas-liquid and liquid-solid mixtures, or the same phase but immiscible, like with a mixture of two immiscible liquids [14]. The modelling of multiphase flow can be performed at different distinct scales, depending on the required application. At smaller scales, the phase boundary properties can be described in detail via separated multiphase flow models in COMSOL [15]. The techniques utilised to simulate these models are known as surface tracking methods, of which there are three available on COMSOL namely Level set, Phase field and Moving mesh. However, at larger scales, it becomes impractical to implement the equations required to model the phase boundary properties in detail and as such, dispersed multiphase flow models are utilised instead. There are four distinct models available, namely Bubbly flow, Mixture, Euler-Euler (continuum), and the Euler-Lagrange (discrete) model.

The bubbly flow model is suitable for modelling flow involving gas bubbles dispersed in a liquid. However, this model fails to integrate the momentum contribution of the dispersed gas phase and thus, is only applicable when the gas density is negligible in comparison to the liquid density [16]. Additionally, this model assumes that both phases share a common pressure field and that the motion of the dispersed phase relative to the continuous phase is determined by the existing pressure forces and viscous drag. The mixture model involves some similar assumptions to the bubbly flow model and is usually preferable when a liquid serves as the continuous phase and the dispersed phase is comprised of either liquid droplets or solid particles [17]. However, unlike the bubbly flow model, it incorporates the momentum contribution of the dispersed phase. Since the Euler-Lagrange (discrete) model is only applicable for multiphase flow simulation when there is a relatively low number of small particles dispersed in the continuous phase, the Euler-Euler model was ultimately selected for modelling multiphase flow. This model was coupled to the other physics interfaces to introduce convection into the system simulation since a flow field is being incorporated. Moreover, the Euler-Euler (continuum) model is considered the best model for simulating dispersed multiphase flow in terms of prediction accuracy and versatility. Though its limitations are that it is only suitable for two-phase flow and needs adequate initial conditions to achieve the required convergence upon solving.

Furthermore, laminar flow was assumed within the system since the required model simulation involves flow at a low to moderate range of Reynolds numbers (Re less than 2000) based on the relevant model parameters. Hence, we coupled the Euler-Euler model, laminar flow interface to the other physics interfaces utilised for the simulation of multiphase flow involving two completely interpenetrating phases [18]. This physics interface is usually suitable for modelling the migration of solid particles dispersed in a gas or liquid or liquid droplets and bubbles dispersed in a liquid. It incorporates two distinct sets of Navier-Stokes equations for each phase to determine their respective velocity fields. Additionally, the transfer of momentum between the two phases is simulated via a drag model. The pressure and the bubble volume fraction are both determined using a continuity equation and a transport equation respectively. Thus, the existing dependent variables are the velocity field of the continuous phase (uc) and the dispersed phase(ud), the pressure (p) and the volume fraction of the bubbles of the dispersed phase (ϕ_d). The Schiller-Naumann drag model was utilised as it was more applicable for a wider range of dispersed phase Reynolds numbers in comparison to Hadamard-Rybcynski which is only suitable for modelling dilute flows with a dispersed phase Reynolds number less than one [19].

Nonetheless, in order to facilitate the im-

plementation of this physics interface in COMSOL, a few assumptions were made. First, both phases were considered incompressible to reduce the computational complexity of the model. Similarly, the bubbles are not modelled individually but instead as a distribution of the appropriate properties as in practice, bubble formation occurs randomly across the electrode surface and it is extremely difficult to design a model that can accurately resolve this. Also, multiphase flow simulation using this physics interface involved the consideration of only one species of gas bubbles in the dispersed phase at a time and the assumption of a fixed bubble size to facilitate model resolution. Furthermore, we assumed no-slip boundary conditions at the walls, that bubble generation only occurred at the relevant electrodes, and that the electrolyte fed into the system at the inlet had no dispersed phase (bubbles) present.

All the relevant equations, including those signifying key assumptions made, incorporated in the model implemented can be found as annotated below. The first seven equations represent the default equations implemented in the Euler-Euler model utilised. Equation (11) represents the no-slip boundary condtions imposed on the stationary walls while Equation (12) reflects the assumptions made on the inlet conditions of the electrolyte utilised. Similarly, Equation (13) outlines the outlet conditions of the electrolyte while Equation (14) represents those at the electrode in the implemented model. As shown by the relevant equations below, only steady-state conditions were taken into account as they provide the best model prediction accuracy and the introduction of time-dependent parameters would unnecessarily increase the model computational complexity.

$$\rho_c(u_c.\nabla)u_c = \nabla.[-pI + \tau_c]$$
$$+\rho_c g + (F_{m,c}/\phi_c) + F_c + (m_{dc}/\phi_c).(u_{int} - u_c) \quad (4)$$

$$\rho_d(u_d.\nabla)u_d = \nabla.[-pI + \tau_d]$$
$$+\rho_d g + (F_{m,d}/\phi_d) + F_d + (m_{dc}/\phi_d).(u_{int} - u_d) \quad (5)$$

$$\nabla.(\phi_c.u_c + \phi_d.u_d) = m_{dc}((1/\rho_c) - (1/\rho_d)) \quad (6)$$

$$\nabla.(\phi_d.u_d) = -(m_{dc}/\rho_d) \quad (7)$$

$$\phi_c = 1 - \phi_d \quad (8)$$

$$\tau_c = \mu_c^m(\nabla.u_c + (\nabla.u_c)^\tau - (2/3).(\nabla.u_c).I) \quad (9)$$

$$\tau_d = \mu_d^m(\nabla.u_d + (\nabla.u_d)^\tau - (2/3).(\nabla.u_d).I) \quad (10)$$

$$u_c = 0;\ u_d = 0 \quad (11)$$

$$u_c = u_{c0};\ u_d = 0;\ \phi_d u_d.n = 0 \quad (12)$$

$$p = p_0;\ u_c.t = 0;\ u_d.t = 0 \quad (13)$$

$$u_c = 0;\ u_d = u_{d0};\ \phi_d u_d.n = -m_d/\rho_d \quad (14)$$

2.4 Model specifications

To enhance computational efficiency, a 2D model, shown above in Figure 1, was selected as the most appropriate for implementation. Model geometry consists of a 7 cm by 10.5 cm rectangle which represents the bulk of the system. This is filled with 1 M of KOH in an aqueous solution. Inside the PEC, there are two electrodes illustrated as two rectangles with heights of 7 cm and widths of 5 mm each. The photoanode, which utilises a tin(IV)-doped hematite semiconductor, is the electrode on the right whereas the cathode is on the left. These electrodes are 1 cm apart to allow for the flow of ions between them and the production of bubbles on the inner surfaces. The electrolyte inlet is at the bottom of the model which has a pipe diameter of 1 cm. The outlet is at the top of the model and has a diameter of 2 cm to sufficiently allow for both liquid and gas to leave the system.

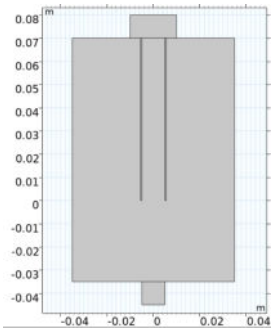


Figure 1: *Geometry of the 2D COMSOL model implemented*

3 Results and Discussion

3.1 Modelling flowrate

3.1.1 Decoupled system

Initially the Euler-Euler physics was established into the system without the photoelectric element in order to model how the electrolyte flows

through the system. Figure 2 shows the velocity profile of this decoupled system, illustrating the presence of a significant amount of unused space within this geometry where the electrolyte does not flow towards the outlet. Here, an inlet flowrate of 800 ml/min was used as it was the maximum flowrate explored within the scope of the research.

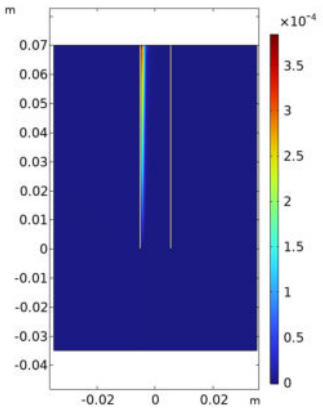


Figure 2: *Velocity profile of the decoupled system with an inlet flowrate of 800 ml/min*

3.1.2 Introduction of Gas Production

The photoelectric element was coupled with the Euler-Euler physics through modelling the convection of ions and the gas production on the surface of the electrode. Figures 3 and 4 show the presence of the gas films when modelling hydrogen and oxygen gases respectively. From this, we can see that the volume fraction of hydrogen gas generated on the cathode is higher than that of oxygen on the anode. This is expected as water splitting produces twice as many moles of hydrogen gas than oxygen. Consequently, this causes the hydrogen bubbles to have a greater impact in reducing the surface area of the electrode than of the oxygen. Due to the limitations of the Euler-Euler interface only allowing the simulation of one specie at a time in the dispersed phase, only the hydrogen gas film was taken into account in determining the impact on gas production. Although it is also expected that the oxygen film will have a similar, but less significant impact on the process.

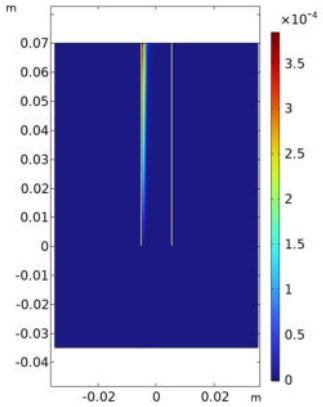


Figure 3: *Hydrogen volume fraction in PEC with inlet flowrate of 500 ml/min*

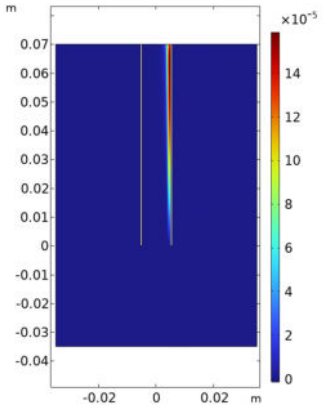


Figure 4: *Oxygen volume fraction in PEC with inlet flowrate of 500 ml/min*

3.1.3 Calculation of film thickness

As the film thickness is not calculated within any of the physics implemented, an estimation for the film thickness was added to the model. This estimation was calculated by setting all the points in the mesh to either 1 or 0 depending on if the volume fraction was greater than 10^{-4}. Then this was integrated across the entire surface and divided by the height of the electrode.

3.1.4 Comparison of film thickness to flowrate

A parametric sweep was run on the inlet flowrate starting at 200 ml/min and increasing it in steps of 100 ml/min up to the maximum flowrate of 800 ml/min. These results were then plotted against flowrate to compare how it changes the film thickness. Figure 5 shows that

as the flowrate was increased, the gas film thickness on the cathode decreased. This can be explained by the momentum transfer between the electrolyte and the gas film. As the flowrate increases, more momentum is transferred onto the film causing the gas to rise to the surface faster. Due to this, the film layer is reduced as flowrates increase.

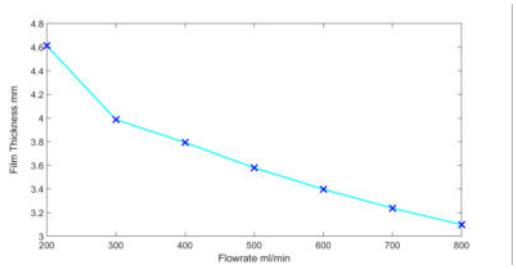


Figure 5: *Hydrogen gas film thickness against inlet flowrate of electrolyte*

3.2 Bubble Impact

3.2.1 Estimation of bubble surface area

Like with the film thickness, the area taken up by bubbles on the surface of the electrode is not directly included in the physics interfaces utilised for the system. Hence, an estimation was introduced to enable the model to predict this.

The area that bubbles take up on the surface is estimated by integrating the volume fraction across the entire surface to calculate the volume of gas in the system. This volume is then divided by the volume of a hemispherical bubble to estimate the number of bubbles. Subsequently, we multiply by the average bubble diameter to estimate the area taken up by all the bubbles formed on the electrode surface. As this model simulation is conducted using a 2D model, the area taken up by bubbles is determined in the standard unit of metres (m).

3.2.2 Effective reduced height determination

To model the impact of the area of bubbles on the rate of production of hydrogen, another variable was introduced known as *erh*. This is the effective reduced height, which is the fraction of the surface of the electrode that is not taken up by bubbles unlike the volume occupied by bubbles (VOB).

This variable was introduced into the equations for local current densities at the cathode by multiplication of the original equation.

$$erh = (height - VOB)/height \qquad (15)$$

3.2.3 Impact on hydrogen production rates

Figure 6 shows the reduction in hydrogen production when the effective reduced height was introduced. Although there is a decrease in the production of hydrogen, it is only a decrease of about 0.2%, so bubbles are unlikely to have much impact on this small-scale model.

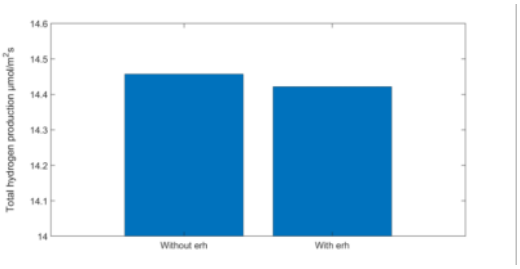


Figure 6: *Comparing hydrogen production with and without including bubble impact on gas production (inlet = 500 ml/min)*

3.2.4 Flowrate impact on hydrogen production

The parametric sweep between 200 and 800 ml/min was reintroduced to evaluate the impact on hydrogen bubble production. Figure 7 shows that the bubble production rate increases with flowrate as expected. However, the production rate decreases at around 600 ml/min. It is suspected that this is caused by the effect of bubbles becoming small enough that convection has a more significant impact on the production rate. An increase in flowrate leads to an increase in the rate of convection. However, as the velocity field is in a different direction from the flow of species, this causes a slight reduction in the overall rate of electrolysis achieved. Additionally, the increase in flowrate causes more hydrogen to remain dissolved in the electrolyte solution leaving the system, though an adequate recycle loop could help to reduce this.

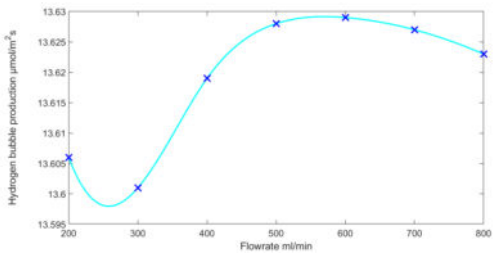


Figure 7: *Hydrogen bubble production against flowrate*

3.3 Other parameters

3.3.1 Bubble diameter

The bubble diameter selected for the model was 0.05 mm, which is quite small in comparison to other applications. This relatively small diameter was specifically chosen for simulation as the model implemented was expected to have a very small hydrogen production rate.

As a result, it was ultimately decided to explore other bubble diameters to evaluate the potential impact on the performance of the system. Figure 8 shows a comparison between the hydrogen bubble production rate at 0.1 mm compared to 0.05 mm. Based on how the bubble area is estimated, it was predicted that a higher bubble diameter would lead to a smaller area being occupied by bubbles, hence the higher production rate. This shows that the flowrate should be chosen based on the size of the bubbles on the surface of the electrodes in order to optimise the production of hydrogen bubbles.

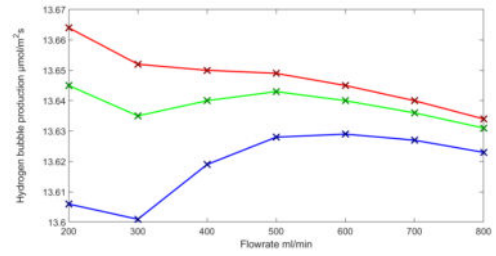


Figure 8: *Hydrogen bubble production against flowrate for different bubble diameters*

3.3.2 Photocurrent density

As mentioned earlier, the effect of electrolyte flowrate on the production of hydrogen bubbles under the current operating conditions was found to be fairly insignificant, especially at high velocities. As a result, different photocurrent densities were tested to evaluate whether flowrate has a higher impact on bubble production when optical efficiency is increased. To simulate increased optical efficiency, the original current density was multiplied by factors of 5 and 10 to give the results shown in Figure 9 below. As expected, increased hydrogen bubble production was observed at higher photocurrent densities. Consequently, the impact of electrolyte flowrate on hydrogen production became more significant, indicating that bubble removal is an effective approach to optimise reactor efficiency, especially in systems with high optical efficiencies.

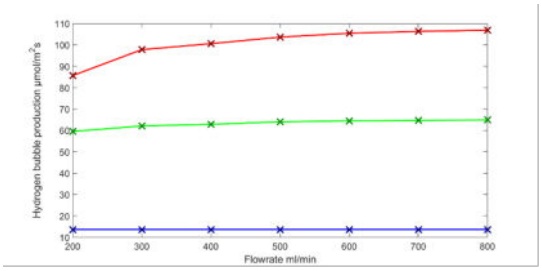


Figure 9: *Hydrogen bubble production as a function of flowrate at different photocurrent densities. Blue: initial, Green: 5 times initial, Red: 10 times initial*

4 Conclusion

The overall aim of this investigation was to determine the impact of facilitating bubble removal from the electrode surface by increasing electrolyte velocity on the hydrogen production, and thus the STH efficiency, achieved within a PEC reactor system. As illustrated in Figure 4 above, increasing the electrolyte velocity from 200 to 800 ml/min led to a general decrease in the hydrogen gas film thickness size observed on the corresponding electrode surface. This can be attributed to the fact that an increasing electrolyte velocity leads to an increase in the momentum transferred unto the gas film, subsequently causing more of the adhering hydrogen gas bubbles forming on the electrode surface to detach and rise in the electrolyte solution.

Furthermore, as illustrated in Figure 6, increasing the electrolyte velocity led to an initial increase in the hydrogen bubble production until a maximum value is achieved at about 600 ml/min. This is expected as the enhanced bubble removal stemming from the increased

flowrate leads to a lower fraction of the electrode surface being covered by bubbles, consequently increasing the effective surface area available for hydrogen production. However, increasing the electrolyte velocity beyond this point led to a slight decrease in the hydrogen production achieved as the negative impact of the corresponding increase in convection and hydrogen dissolution begins to outweigh the positive effect of enhanced bubble removal. This is further reinforced by the fact that bubble production does not have a significant impact on the rate of hydrogen production in the PEC system considered, as shown in Figure 5.

Finally, upon considering the impact of bubble diameter on hydrogen production as shown in Figure 7, it can be seen that at higher bubble diameters, increasing electrolyte velocity actually leads to a general decrease in the hydrogen production achieved. This can be attributed to the fact that less bubbles are formed on the electrode surface and thus, bubble removal has a significantly lower impact on the hydrogen production rates achieved while the corresponding increase in convection and hydrogen dissolution have a more pronounced effect, leading to the trend observed. Additionally, as seen in Figure 8, increasing the photocurrent density within the system, and consequently the overall system efficiency, led to the expected overall trends of increasing hydrogen production with increasing electrolyte velocity. Moreover, the higher the photocurrent density, the more significant the impact of electrolyte velocity on the hydrogen production rates achieved within the system. This is due to the fact that at higher photocurrent densities, the rate of hydrogen bubble formation on the electrode increases, causing bubble removal facilitated by higher flowrates to have a greater impact on the hydrogen production rates achieved within the system.

5 Outlook

During the scope of this investigation, we sought to enhance the hydrogen production rate and thus, the STH efficiency achieved within a PEC reactor by altering the electrolyte velocity. However, there are several other factors that could potentially influence the STH efficiency of a PEC reactor that can be further explored in future investigations to optimise reactor operation. Some prior studies have suggested that PEC reactor operation under concentrated irradiation could help to enhance the overall STH efficiency of the system. This can be achieved by utilising the appropriate solar concentrators, like Fresnel lenses, to enhance the optical efficiency of the system by focusing light onto the photoelectrode. This increases system efficiency by causing an increase in the photocurrent density while also enhancing light utilisation and reaction kinetics due to the associated temperature increase.

However, to optimise the positive effects of PEC operation under concentrated irradiation on the achievable STH efficiency, the impact of the corresponding temperature increase on reactor performance requires mitigation [20]. This can be achieved by utilising photoelectrode materials with high optical and thermal stability to prevent material degradation during reactor operation which subsequently enhances system durability. Hence, further research could evaluate the implementation of other photoelectrode materials that could potentially be less susceptible to material degradation emanating from either an increase in the operating temperature or even the electrolyte velocity. Also, the reactor could be designed to facilitate sufficient cooling by the electrolyte to mitigate the impact of temperature rise on the electrodes while simultaneously slightly increasing electrolyte temperature which improves the rate of water dissociation.

Additionally, the reactor geometry and configuration have both been identified to have a significant impact on the STH efficiency achieved within the PEC reactor as they could potentially affect several other aspects influencing reactor performance like photoelectrode illumination and bubble and thermal management. Hence, subsequent investigations in this field could also seek to optimise these to improve the overall efficiency of the PEC reactor system. Furthermore, future investigations to optimise the achievable STH efficiency of a PEC reactor could also evaluate the impact of other influencing factors like the type and concentration of electrolyte utilised, the incorporation of appropriate catalysts, and membrane implementation to mitigate gas crossover from both electrodes and unintended reactions that could negatively affect the reactor performance.

Acknowledgements

We would like to thank Dr Anna Hankin for her invaluable support and guidance throughout the project. We would also like to thank the COMSOL team for providing the software utilised for

model simulations.

References

[1] Statista. World Energy Consumption by Energy Source 2040 — Statista. Statista. https://www.statista.com/statistics/222066/projected-global-energy-consumption-by-source/.

[2] Energy Institute. Statistical Review of World Energy. Statistical review of world energy. https://www.energyinst.org/statistical-review

[3] United Nations. Climate Action Fast Facts. United Nations. https://www.un.org/en/climatechange/science/key-findings#temperature-rise

[4] Höök, M.; Tang, X. Depletion of Fossil Fuels and Anthropogenic Climate Change—a Review. Energy Policy 2013, 52 (1), 797–809. https://doi.org/10.1016/j.enpol.2012.10.046

[5] University of Michigan. Hydrogen Factsheet. Center for Sustainable Systems. https://css.umich.edu/publications/factsheets/energy/hydrogen-factsheet

[6] Hague, O. 3 Main Types of Hydrogen - Blue, Grey and Green. Brunel. https://www.brunel.net/en/blog/renewable-energy/3-main-types-of-hydrogen

[7] Gado, M. G.; Nasser, M.; Hassan, H. Potential of Solar and Wind-Based Green Hydrogen Production Frameworks in African Countries. International journal of hydrogen energy 2024, 68, 520–536. https://doi.org/10.1016/j.ijhydene.2024.04.272

[8] Kelly, N.; Gibson, T. Design and Characterization of a Robust Photoelectrochemical Device to Generate Hydrogen Using Solar Water Splitting. International Journal of Hydrogen Energy 2006, 31 (12), 1658–1673. https://doi.org/10.1016/j.ijhydene.2005.12.014

[9] Yang, W.; Prabhakar, R. R.; Tan, J.; Tilley, S. D.; Moon, J. Strategies for Enhancing the Photocurrent, Photovoltage, and Stability of Photoelectrodes for Photoelectrochemical Water Splitting. Chemical Society Reviews 2019, 48 (19), 4979–5015. https://doi.org/10.1039/C8CS00997J

[10] Jong, W.A; PEC versus PV-E a Future Potential Comparison; 2018.

[11] Qureshy, A. M. M. I.; Dincer, I. A Study on Energy and Fluid Flows for a Novel Photoelectrochemical Reactor Developed for Hydrogen Production. International Journal of Heat and Mass Transfer 2020, 163, 120523. https://doi.org/10.1016/j.ijheatmasstransfer.2020.120523

[12] Dickinson, E. J. F.; Ekström, H.; Fontes, E. COMSOL Multiphysics: Finite Element Software for Electrochemical Analysis. A Mini-Review. Electrochemistry Communications 2014, 40, 71–74. https://doi.org/10.1016/j.elecom.2013.12.020

[13] Chemical Reaction Engineering Module User's Guide. https://doc.comsol.com/5.4/doc/com.comsol.help.chem/ChemicalReactionEngineeringModuleUsersGuide.pdf

[14] Azzopardi, B. MULTIPHASE FLOW. http://www.eolss.net/sample-chapters/c06/e6-34-01-05.pdf

[15] Modeling and Simulation of Multiphase Flow in COMSOL: Part 1. COMSOL. https://www.comsol.com/blogs/modeling-and-simulation-of-multiphase-flow-in-comsol-part-1

[16] The Bubbly Flow Equations. Comsol.com. https://doc.comsol.com/6.0/doc/com.comsol.help.cfd/cfd_ug_fluidflow_multi.09.142.html.

[17] The Mixture Model Equations. Comsol.com. https://doc.comsol.com/6.0/doc/com.comsol.help.cfd/cfd_ug_fluidflow_multi.09.146.html#2330963.

[18] CFD Module User's Guide. https://doc.comsol.com/5.3/doc/com.comsol.help.cfd/CFDModuleUsersGuide.pdf

[19] The Euler–Euler Model Equations. Comsol.com. https://doc.comsol.com/5.6/doc/com.comsol.help.cfd/cfd_ug_fluidflow_multi.09.142.html

[20] Bicer, Y.; Sprotte, A. F. V.; Dincer, I. Concentrated Solar Light Splitting Using Cold Mirrors for Photovoltaics and Photonic Hydrogen Production Applications. Applied Energy 2017, 197, 169–182. https://doi.org/10.1016/j.apenergy.2017.04.009

Predicting the Competition of SO_2 with CO_2 in an MEA Capture Solvent using the SAFT-γ Mie Equation of State

Angus Kitching-Williams, Frederick Martin

Department of Chemical Engineering, Imperial College London, U.K.

Abstract: This report explores the competitive effect of sulphur dioxide impurities, found in flue gas streams, with carbon dioxide in an aqueous monoethanolamine solution commonly used in carbon capture and storage. We propose a methodology to estimate the distribution of ionic species in solution, providing insights into the degree of competition. The proposed framework utilises the SAFT-γ Mie group-contribution approach. This allows for the modelling of chemical reactions through physical association, enabling an implicit representation of the chemical speciation typical in carbon capture systems. We validate the models used with experimental data where applicable, as well as introduce novel estimated group-contribution parameters for use in the SAFT-γ Mie approach. The results presented in this work confirm sulphur dioxide competition with carbon dioxide, although to a lesser extent than expected. Nevertheless, this study presents the first research to account for the influence of sulphur dioxide on carbon capture using the SAFT-γ Mie equation of state.

1. Introduction

Carbon capture and storage (CSS) stands as an important technology in the net removal of carbon dioxide (CO_2) from the atmosphere [1]. It is anticipated to play a major role in meeting global warming targets [2] and mitigating anthropogenic climate change [3]. The most commonly-used solvents for CO_2 capture are amine-based – specifically alkanolamines – having been used for this purpose since 1930 [4]. Alkanolamines, such as monoethanolamine (MEA), offer a high reactivity with CO_2 [5] leading to the formation of carbamates. Complexity arises when CO_2 is dissolved in an aqueous amine solution, as this leads to speciation [6, 7]. Speciation refers to the different forms a chemical species can take within a given system and their distribution [8]. Different forms of a species can arise through various processes, with dissociation reactions being of particular note for carbon capture [9]. Hence, an accurate theoretical approach to describe the complex fluid-phase behaviour in an aqueous MEA solution presents a significant challenge.

Considerable work has been carried out to model the behaviour of CO_2 in aqueous solutions of amines. Most notably, the use of activity coefficient models, which can be used to calculate the solubility of CO_2 and the distribution of both ionic and molecular species within the liquid phase [10, 11]. A drawback of this approach is the requirement of experimental data and the introduction of empirical parameters [12], making them less versatile. Cubic engineering equations of state (EoS), such as that of Peng-Robinson [13] and Soave-Redlich-Kwong [14], have also been used to model the fluid-phase behaviour of CO_2 in aqueous amine solutions. Critically, these approaches are limited in their ability to account for association effects, placing them at a distinct disadvantage.

The Statistical Associating Fluid Theory (SAFT) EoS [15, 16], and its many variants, is a powerful method developed using Wertheim's thermodynamic perturbation theory (TPT) for associating fluids. A key advantage of a SAFT-type approach is the implicit treatment of the speciation reactions that occur in CO_2 capture. As such, recent work has been undertaken using SAFT group-contribution (GC) approaches to model aqueous solutions of amines with CO_2 [12]. GC methods decompose a species into groups, with each group containing information that contributes to the thermodynamic properties of the species.

In this work, we aim to include the negative impact of flue gas sulphur dioxide (SO_2) on the carbon capture process, using the SAFT-γ Mie EoS. Flue gas, a by-product of industrial combustion processes, typically contains a mixture of gases including CO_2 and impurities such as SO_2. Specifically, we explore the competition expected between SO_2 and CO_2 in reacting with MEA, reducing the effectiveness of the solvent.

The remainder of this report is set out as follows: in Section 2 we provide an overview of the theoretical background, introducing the SAFT-γ Mie EoS (Section 2.1) and the speciation reactions present in CO_2 capture (Section 2.3). We also detail our general workflow (Section 2.4) in predicting thermodynamic properties. We then discuss and analyse our results in Section 3, summarising them in Section 4. Lastly, in Section 5, we outline the space for future work.

2. Methodology

2.1. SAFT-γ Mie

In this research we consider the SAFT-γ Mie GC approach, an adaptation of SAFT developed by Papaioannou et al. [17] that builds on the work of Lafitte et al. [18], by considering a variable-range potential. This approach represents molecules as distinct functional groups of fused spherical segments, which interact through Mie potentials of variable attractive and repulsive ranges. For two given groups, k and l, the Mie potential is given by (Equation 1),

$$\phi_{kl}^{\mathrm{Mie}} = C_{kl}\,\varepsilon_{kl}\left[\left(\frac{\sigma_{kl}}{r_{kl}}\right)^{\lambda_{kl}^{\mathrm{r}}} - \left(\frac{\sigma_{kl}}{r_{kl}}\right)^{\lambda_{kl}^{\mathrm{a}}}\right], \quad (1)$$

$$C_{kl} = \frac{\lambda_{kl}^{\mathrm{r}}}{\lambda_{kl}^{\mathrm{r}} - \lambda_{kl}^{\mathrm{a}}}\left(\frac{\lambda_{kl}^{\mathrm{r}}}{\lambda_{kl}^{\mathrm{a}}}\right)^{\frac{\lambda_{kl}^{\mathrm{a}}}{\lambda_{kl}^{\mathrm{r}} - \lambda_{kl}^{\mathrm{a}}}}, \quad (2)$$

where ε_{kl}, λ_{kl}^{r}, λ_{kl}^{a} ,σ_{kl} and r_{kl}, are the potential depth, the repulsive exponent, the attractive exponent, the average segment diameter and the centre-centre distance. Setting the value of λ_{kl}^{r} to 12 and λ_{kl}^{a} to 6 leads to the familiar Lennard-Jones potential. For a given mixture of molecules, the total Helmholtz free energy A is written as the sum of four contributions,

$$\frac{A}{Nk_BT} = \frac{A_{\mathrm{ideal}}}{Nk_BT} + \frac{A_{\mathrm{mono.}}}{Nk_BT} + \frac{A_{\mathrm{chain}}}{Nk_BT} + \frac{A_{\mathrm{assoc.}}}{Nk_BT}, \quad (3)$$

where N is the number of particles, T is the temperature and k_B the Boltzmann constant. "ideal" denotes the ideal contribution to the free energy of the mixture, "mono." refers to the term accounting for interactions between monomer Mie segments, "chain" denotes the contribution to the free energy due to the formation of molecules from segments and "assoc." refers to the association interactions. For a more detailed explanation on the above theory, we direct the reader to the original publications regarding SAFT-γ Mie [17, 19].

In Figure 1 we illustrate the SAFT-γ Mie representation of $\mathrm{MEA} + H_2O$. Of particular note is the inclusion of e and H sites used to describe the association between a given group k and a given group l.

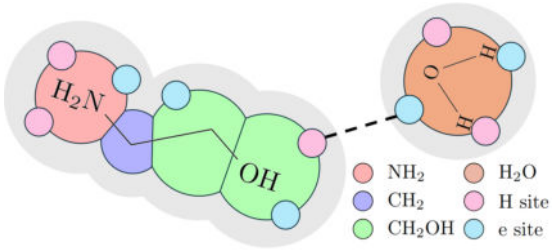


Figure 1: $\mathrm{MEA} + H_2O$, described for a SAFT-γ Mie GC approach. The dotted line shows CH_2OH H site associating with H_2O e site, with grey shading marking a chain of segments.

2.2. Groups Considered

Given SAFT-γ Mie is a GC approach, an aspect of our methodology is the gathering of relevant parameters used to define the groups that make up the molecules included in this study. In addition to the aforementioned parameters ε_{kl}, λ_{kl}^{r}, λ_{kl}^{a} and σ_{kl} we define ν_k^* and S_k referring to the number of segments and shape factor of group k. In Table 1, we summarise the like group parameters.

2.3. Speciation

2.3.1. Disassociation Reactions of Absorbates

When CO_2 dissolves in an aqueous MEA solution, bicarbonate ions (HCO_3^-) and carbamate ions (RNHCOO^-) are formed. Reaction with H_2O also leads to the formation of carbonic acid, H_2CO_3 [25]:

$$CO_2 + 2\mathrm{RNH}_2 \rightleftharpoons [\mathrm{RNH}_3^+ + \mathrm{RNHCOO}^-], \quad (4)$$

$$\mathrm{RNHCOO}^- + H_3O^+ \rightleftharpoons [\mathrm{RNH}_3^+ + HCO_3^-], \quad (5)$$

$$H_2O + CO_2 \rightleftharpoons H_2CO_3 \rightleftharpoons [H^+ + HCO_3^-]. \quad (6)$$

The equations above are written for a general primary amine, RNH_2. For MEA, the R group is equivalent to a CH_2CH_2OH.

In Equation 4 we show the formation of a zwitterionic form of RNHCOO^-, which can then participate in a proton exchange reaction to form HCO_3^- in Equation 5. Formation of HCO_3^- ions is also achieved through de-protonation of H_2CO_3 as illustrated in Equation 6.

Two key disassociation reactions are identified for the case of SO_2 dissolving in an aqueous MEA solvent [26]:

$$SO_2 + H_2O \rightleftharpoons [HSO_3^- + H^+], \quad (7)$$

$$SO_2 + \mathrm{RNH}_2 + H_2O \rightleftharpoons [\mathrm{RNH}_3^+ + HSO_3^-]. \quad (8)$$

Both reactions relate to the formation of bisulfite ions, HSO3$^-$. Notably, both Equations 7 and 8 involve reaction with H_2O, while only Equation 8 involves MEA.

Table 1: Like group parameters, for the SAFT-γ Mie GC approach. The number of association types for group k is given by $N_{ST,k}$, with $n_{k,H}$, $n_{k,e}$, n_{k,α_1}, and n_{k,α_2} representing the number of sites for each specific association type.

k	Group k	ν_k^*	S_k	$\lambda_{kk}^{\mathrm{r}}$	$\lambda_{kk}^{\mathrm{a}}$	σ_{kk} /Å	$(\frac{\varepsilon_{kk}}{k_B})$ / K	$N_{ST,k}$	$n_{k,H}$	$n_{k,e}$	n_{k,α_1}	n_{k,α_2}	Ref.
1	NH_2	1	0.79675	10.254	6.0000	3.2477	284.78	2	2	1	–	–	[20]
2	CH_2	1	0.22932	19.871	6.0000	4.8801	473.39	–	–	–	–	–	[21]
3	CH_2OH	2	0.58538	22.699	6.0000	3.4054	407.22	2	1	2	–	–	[22]
4	H_2O	1	1.0000	17.020	6.0000	3.0063	266.68	2	2	2	–	–	[21]
5	CO_2	2	0.84680	26.408	5.0550	3.0500	207.89	2	–	–	1	1	[23]
6	SO_2	3	0.62547	15.213	6.0000	3.1605	237.54	2	1	1	–	–	[24]

2.3.2. Calculating the Degree of Speciation

To calculate the degree of speciation, we require knowledge of the associations occurring. To this end, we can obtain the fraction of unassociated sites of type a on group k of component i, $X_{i,k,a}$. This is obtained through the solution of the mass-action equation in SAFT theory as: [15, 27, 28]:

$$X_{i,k,a} = (1 + \sum_{j=1}^{N_C} \sum_{l=1}^{N_G} \sum_{b=1}^{N_{ST,l}} \rho x_j \nu_{l,j} n_{l,b} X_{j,l,b} \Delta_{ij,kl,ab})^{-1}. \quad (9)$$

Where N_C and N_G are the number of components and groups respectively. ρ is the number density, x_j the mole fraction of component j, $\nu_{l,j}$ the number of occurrences of group l in component j, $n_{l,b}$ the number of sites of type b on group l and $\Delta_{ij,kl,ab}$ the association strength between sites a and b of groups k and l on components i and j.

Equations 4 - 8 are classified by their reversibility and formation of products that preserve much of the chemical identity of the reactants. Hence, it is possible to consider an implicit reaction model [12]. Such treatment involves considering the charged species produced in the square brackets on the right hand side of Equations 4 - 8 as a single neutral species with no net charge. The mole fraction of each species is deduced by considering the fraction of molecules associated to the other groups appearing in the ionic complex; made available through Equation 9.

2.3.3. Calculating CO_2 Speciation

As shown in Table 1, CO_2 is treated as having two association sites referred to as α_1 and α_2. These sites are capable of interacting with the NH_2 group of MEA and with H_2O; relevant parameter values are described in Table 3. CO_2 as described through the SAFT-γ Mie approach is depicted in Figure 2.

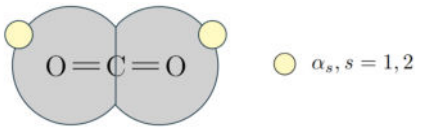


Figure 2: CO_2, described for a SAFT-γ Mie group-contribution approach.

In previous work, the mole fraction of $RNHCOO^-$ ($x_{\mathrm{RNHCOO^-}}$) and HCO_3^- ($x_{\mathrm{HCO_3^-}}$) formed through reaction with an amine group, has been deduced using Equations 12 and 13. Calculations assume no interaction between CO_2 and H_2O [12]. To simplify the equations we introduce β_1 and β_2:

$$\beta_1 = X_{\mathrm{CO_2,CO_2},\alpha_1}, \quad (10)$$

$$\beta_2 = X_{\mathrm{CO_2,CO_2},\alpha_2}, \quad (11)$$

$$x_{\mathrm{RNHCOO^-}} = x_{\mathrm{CO_2}}(1-\beta_1)(1-\beta_2), \quad (12)$$

$$x_{\mathrm{HCO_3^-}} = x_{\mathrm{CO_2}}(\beta_1 + \beta_2 - 2(\beta_1\beta_2)), \quad (13)$$

with $x_{\mathrm{RNHCOO^-}}$ corresponding to the fraction of CO_2 associated at both α sites. Similarly, $x_{\mathrm{HCO_3^-}}$ can be calculated as the fraction of CO_2 associated at only a single α site. This is the result of not considering CO_2 - H_2O association. Accordingly, the product on the right hand side of Equation 5 comprises a single association between CO_2 and MEA. An issue with this treatment is that association with H_2O accounts for H_2CO_3 formation.

Including the H_2O with CO_2 interaction will enable calculation of $x_{\mathrm{HCO_3^-}}$ formed through dissociation of H_2CO_3 (Equation 6). We can keep the form of Equations 4 and 5 the same for comparability, but consider this time association between H_2O and CO_2. Now, the product of Equation 4 is still considered as a double association between CO_2 and MEA. HCO_3^- from Equation 5 includes two associations of CO_2 with MEA and H_2O. The HCO_3^- from H_2CO_3 in Equation 6 is a single association between CO_2 and H_2O. Thus, Equation 12 now represents the $RNHCOO^-$ and

HCO_3^- formed through reaction with MEA (both involving two associations of CO_2). Equation 13 represents HCO_3^- formed through disassociation of H_2CO_3 (single association of CO_2).

2.3.4. Calculating SO_2 Speciation

Just as for CO_2, by considering the reaction schemes in Equations 7 and 8 implicitly, HSO_3^- from Equation 7 can be interpreted as a single association between SO_2 and H_2O. HSO_3^- from Equation 8 will correspond to an association between SO_2 and the NH_2 group of MEA, as well as an association between SO_2 and H_2O. Accordingly, the concentrations of HSO_3^- from each reaction can be calculated as:

$$x_{\mathrm{HSO_3^-}} = x_{\mathrm{SO_2}}(\mathcal{F}_{\mathrm{SO_2,H\text{ - }NH_2,e}})(\mathcal{F}_{\mathrm{SO_2,H\text{ - }H_2O,e}}) \tag{14}$$

$$x_{\mathrm{HSO_3^-}} = x_{\mathrm{SO_2}}(\mathcal{F}_{\mathrm{SO_2,H\text{ - }H_2O,e}}) \tag{15}$$

Where $\mathcal{F}_{k,a\text{ - }l,b}$ is the fraction of associated sites a and b between groups k and l. This is obtained, for a general case, using:

$$\mathcal{F}_{k,a\text{ - }l,b} = \rho_l n_{l,b} \Delta_{ij,kl,ab} X_{i,k,a} X_{j,l,b} \tag{16}$$

Here, we assume all associations of SO_2 with NH_2 and H_2O correspond to the reactions in Equations 7 and 8. We have also not accounted for association with the CH_2OH group of MEA in these equations.

2.4. Property Prediction

In this section, we detail the general workflow of our approach for property prediction. We start with the development of a pure component model, adding complexity to form binary, ternary and finally quaternary mixtures within the SAFT-γ Mie framework. To solve for the desired properties we employ `Clapeyron.jl` [29], and utilise `Metaheuristics.jl` [30] for the estimation of parameters.

Starting with the pure component model, we use the `crit_pure` and `saturation_pressure` functions in `Clapeyron.jl`. This allows for the estimation of the critical point and the saturation properties respectively. Equations 17 and 18 define the critical point whereas Equations 19 and 20 are used to define the saturation envelope. P refers to the pressure, μ is the chemical potential and $V_{\mathrm{Liq.}}$ and $V_{\mathrm{Vap.}}$ refer to the liquid and vapour volumes respectively.

$$\left(\frac{\partial^2 A}{\partial V^2}\right)_T = 0, \tag{17}$$

$$\left(\frac{\partial^3 A}{\partial V^3}\right)_T = 0, \tag{18}$$

$$P(V_{\mathrm{Liq.}}, T) = P(V_{\mathrm{Vap.}}, T), \tag{19}$$

$$\mu(V_{\mathrm{Liq.}}, T) = \mu(V_{\mathrm{Vap.}}, T), \tag{20}$$

In the case of binary mixtures, `dew_pressure` and `bubble_pressure` functions provided by `Clapeyron.jl` [29] are used. The most notable binary mixture we develop is that of MEA + H_2O, this mixture describes the aqueous phase used in carbon capture. This serves as the foundation of our work, where we explore the addition of further components to develop ternary and quaternary mixtures.

Of interest in this study is the variation of the partial pressure of CO_2 as a function of the CO_2 loading ($\theta_{\mathrm{CO_2}}$), where loading of a component i refers to the amount of i absorbed by a solvent (Equation 21).

$$\theta_{\mathrm{i}} = \frac{n_{\mathrm{i}}}{n_{\mathrm{MEA}}} \tag{21}$$

In order to calculate this, we assume a given w_{MEA} %, hence the initial composition of the liquid phase (before any CO_2 loading) is fully defined. On addition of CO_2, the liquid phase compositions are recalculated ensuring a fixed ratio of $x_{\mathrm{H_2O}}$ to x_{MEA}. Hence – for a given amount of CO_2 loading – the vapour composition can be calculated using the `bubble_pressure` function, as before. We highlight this procedure in Algorithm 1.

Algorithm 1 Calculate CO_2 partial pressure

Require: x_{MEA}, $x_{\mathrm{H_2O}}$, SAFT-γ Mie parameters, range of $\theta_{\mathrm{CO_2}}$

1: $solvent_ratio \leftarrow \frac{x_{\mathrm{H_2O}}}{x_{\mathrm{MEA}}}$
2: $\Gamma \leftarrow$ SAFT-γ Mie parameters
3: $\boldsymbol{y}_{\mathrm{CO_2}} \leftarrow [\ ]$
4: $\boldsymbol{P}_{\mathrm{CO_2}} \leftarrow [\ ]$

Recalculate x_{MEA} **and** $x_{\mathrm{H_2O}}$ **and calculate** $\boldsymbol{P}_{\mathrm{CO_2}}$**:**

5: **for** $\theta_{\mathrm{CO_2}} = \theta_{\mathrm{CO_2,min}}, \ldots, \theta_{\mathrm{CO_2,max}}$ **do**
6: $\quad x_{\mathrm{MEA}} \leftarrow (1 + \theta_{\mathrm{CO_2}} + solvent_ratio)^{-1}$
7: $\quad x_{\mathrm{CO_2}} \leftarrow x_{\mathrm{MEA}} * \theta_{\mathrm{CO_2}}$
8: $\quad x_{\mathrm{H_2O}} \leftarrow 1 - x_{\mathrm{MEA}} - x_{\mathrm{CO_2}}$
9: $\quad \mathbf{x} \leftarrow [x_{\mathrm{MEA}}, x_{\mathrm{H_2O}}, x_{\mathrm{CO_2}}]$
10: $\quad \boldsymbol{y}_{\mathrm{CO_2}} \leftarrow$ `bubble_pressure`$(\Gamma, T, \mathbf{x})$
11: $\quad \boldsymbol{P}_{\mathrm{CO_2}} \leftarrow \boldsymbol{y}_{\mathrm{CO_2}} * P$
12: **end for**
13: **return** $\boldsymbol{P}_{\mathrm{CO_2}}$

Algorithm 1 can be easily adapted to incorporate constant SO_2 loadings for calculations relating to the quaternary mixture (Section 3.5), by including a parameter, $\theta_{\mathrm{SO_2}}$, of constant SO_2 loading. $\theta_{\mathrm{SO_2}}$ is then used in the calculation of x_{MEA} by simply adding it into the reciprocal in line 6 of Algorithm 1. The mole fraction of SO_2 ($x_{\mathrm{SO_2}}$) is calculated with $\theta_{\mathrm{SO_2}}$ in an analogous way to $x_{\mathrm{CO_2}}$ in Algorithm 1, line 7.

Table 2: Unlike group dispersion energies ε_{kl} and repulsive exponent λ^r_{kl} between groups k and l. CR denotes the value obtained via a relevant combining rule [12].

k	l	Group k	Group l	$(\frac{\varepsilon_{kl}}{k_B})$ / K	$\lambda^{\mathrm{r}}_{kl}$	Ref.	k	l	Group k	Group l	$(\frac{\varepsilon_{kl}}{k_B})$ / K	$\lambda^{\mathrm{r}}_{kl}$	Ref.
1	2	NH_2	CH_2	348.39	CR	[20]	2	6	CH_2	SO_2	301.96	CR	[24]
1	3	NH_2	CH_2OH	528.21	52.305	[12]	3	4	CH_2OH	H_2O	358.02†	CR†	[31]
1	4	NH_2	H_2O	339.89†	CR†	[20]	3	5	CH_2OH	CO_2	312.30	CR	[12]
1	5	NH_2	CO_2	134.58†	50.060†	[12]	3	6	CH_2OH	SO_2	302.89	CR	[24]
1	6	NH_2	SO_2	453.85‡	CR	‡	4	5	H_2O	CO_2	226.38	CR	[23]
2	3	CH_2	CH_2OH	423.17	CR	[22]	4	6	H_2O	SO_2	239.53	CR	[24]
2	4	CH_2	H_2O	423.63	100.00	[22]	5	6	CO_2	SO_2	233.15	CR	[24]
2	5	CH_2	CO_2	276.45	CR	[23]							

† indicates a second-order interaction between groups k and l due to the molecular environment of group k.
‡ indicates the value is obtained as part of this report.

Table 3: Unlike group association energies $\varepsilon_{assoc.,kl}$ and bonding-volumes $K_{kl,\,ab}$ between site a on group k and site b on group l. CR denotes the value obtained via a relevant combining rule [12].

k	l	Group k	Group l	site a of group k	site b of group l	$(\frac{\varepsilon_{\mathrm{assoc.},kl}}{k_B})$ / K	$K_{kl,\,ab}$ / Å^3	Ref.
1	1	NH_2	NH_2	H	e	1070.8	95.225	[20]
1	3	NH_2	CH_2OH	H	e	629.88	346.08	[12]
1	4	NH_2	H_2O	H	e	1364.4†	22.450†	[12]
1	6	NH_2	SO_2	H	e	663.02‡	CR	‡
3	1	CH_2OH	NH_2	H	e	2403.8	26.192	[12]
3	3	CH_2OH	CH_2OH	H	e	2097.9	62.309	[22]
3	4	CH_2OH	H_2O	H	e	600.00†	490.19†	[31]
3	6	CH_2OH	SO_2	H	e	CR	CR	[24]
4	1	H_2O	NH_2	H	e	1877.5†	459.18†	[12]
4	3	H_2O	CH_2OH	H	e	2000.6†	130.02†	[31]
4	4	H_2O	H_2O	H	e	1985.4	101.69	[21]
4	6	H_2O	SO_2	H	e	1343.9	CR	[24]
5	4	CO_2	H_2O	α_1	e	1398.1	91.419	[23]
5	1	CO_2	NH_2	α_1	e	3313.0†	3280.3†	[12]
5	1	CO_2	NH_2	α_2	e	4943.6†	142.64†	[12]
6	1	SO_2	NH_2	H	e	663.02‡	CR	‡
6	3	SO_2	CH_2OH	H	e	CR	CR	[24]
6	4	SO_2	H_2O	H	e	1343.9	CR	[24]

† indicates a second-order interaction between groups k and l due to the molecular environment of group k.
‡ indicates the value is obtained as part of this report.

3. Results and Discussion

3.1. Pure MEA

In this section we present the result of the pure MEA model formulated with the parameters given in Table 1, 2 and 3. As shown in Figure 3 (a), the accurate prediction of vapour-liquid equilibrium (VLE) indicates that the model effectively captures the pure phase properties of MEA. Due to the lack of available experimental data, a NIST correlation [37] was used to validate the prediction at high temperatures. In Figure 3 (b), we show the model provides a satisfactory prediction for the specific isobaric heat capacity (c_p), a second-derivative property of the Helmholtz free energy. As before, a lack of experimental data meant only the saturated liquid curve could be compared to experimental data. We report a percentage absolute average deviation AAD(%) in c_p / J mol^{-1}K^{-1} of 4.51. Throughout this work, the AAD(%) between experimental values and calculated values, $Z^{\mathrm{exp.}}$ and $Z^{\mathrm{calc.}}$ respectively, for a given property Z of N data points is given as:

$$\mathrm{AAD}(\%) = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{Z_i^{\mathrm{exp.}} - Z_i^{\mathrm{pred.}}}{Z_i^{\mathrm{exp.}}}\right|. \qquad (22)$$

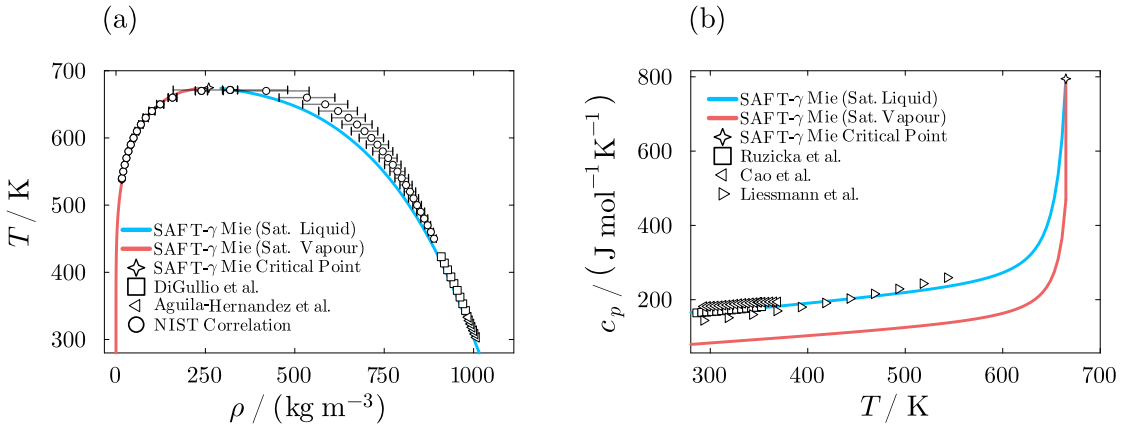


Figure 3: SAFT-γ Mie predictions for MEA: (a) temperature-density VLE envelope and (b) isobaric heat capacity as a function of temperature. The SAFT-γ Mie critical point is represented by a star marker in both cases. Experimental data [32–36] is shown using unfilled markers, including data from a NIST correlation used for validation at high temperatures in (a).

3.2. MEA + H_2O Mixture

To represent the behaviour of the aqueous phase involved in CO_2 capture, we present isothermal pressure-composition slices for the binary mixture of MEA + H_2O, as shown in Figure 4. We report the following AAD(%) in P / bar as: 6.10 at $T = 363.15$ K, 4.65 at $T = 343.15$ K and 4.85 at $T = 323.15$ K, illustrating a strong agreement. While most MEA capture solvents are 30 w_{MEA}% [38], it is encouraging to see accurate agreement between the model prediction and experiential data for a range of x_{H_2O}. It should be noted that in the development of the binary mixture model, second-order interactions were included to account for the proximity effects induced by MEA. [12, 31].

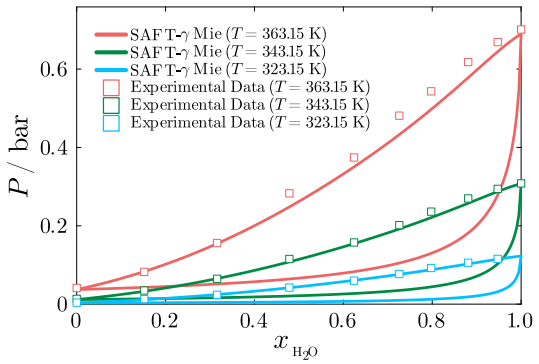


Figure 4: Isothermal pressure-composition slices of the vapour-liquid equilibria for a binary mixture of MEA + H_2O. Experimental data [39] is shown using unfilled square markers.

3.3. MEA + H_2O + CO_2 Mixture

Being able to accurately predict phase equilibrium behaviour for a "pure" flue gas (one containing only CO_2) and capture solvent system is necessary. This enables confident analysis of, and comparison with, results from our approach when the added complexity of SO_2 impurity is introduced.

Figure 5 is a reproduction of a published result [12], created using the approach outlined in Section 2. We verify the validity of our methodology, particularly group parameters and property prediction approaches, for modelling carbon capture systems.

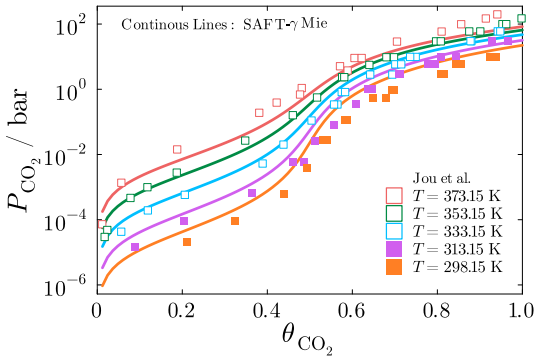


Figure 5: CO_2 loading in 30 w_{MEA}% aqueous MEA [12]. Experimental data is depicted as square markers [40], with filled squares indicating the data was used in parameter estimation.

In Figure 6 we show the predicted mole fractions, x_{RNHCOO^-} and $x_{HCO_3^-}$. As detailed in Section 2.3.3, the filled black line can be interpreted as the mole fractions of both $RNHCOO^-$ and HCO_3^- arising through reaction with MEA. The filled blue line represents $x_{HCO_3^-}$ deriving from disassociation of H_2CO_3.

Physically, the fixed quantity of MEA in the

solution imposes limitations on its availability for reaction. As θ_{CO_2} increases, more CO_2 reacts with MEA to form ionic complexes. Consequently, the amount of free MEA available for further reaction diminishes, reducing the absorption capacity of the solvent. This behaviour is consistent with the plateauing of the filled black line above a θ_{CO_2} of approximately 0.5. Achieving greater loadings of CO_2 past this point will therefore rely heavily on reaction with H_2O to form H_2CO_3. This behaviour is evidenced by the steady increase in $x_{HCO_3^-}$ produced through reaction with H_2O.

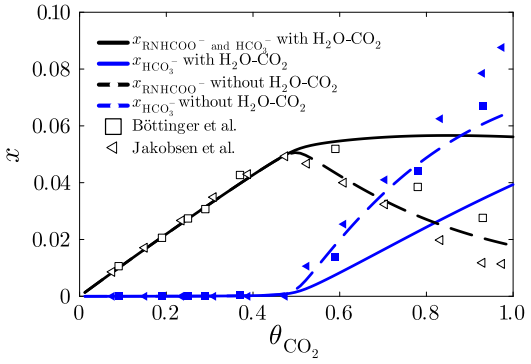


Figure 6: Predicted mole fractions of $NHCOO^-$ and HCO_3^- in 30 $w_{MEA}\%$ aqueous MEA at $T = 293.15$ K. Continuous lines indicate CO_2-H_2O interaction, dashed lines indicate no CO_2-H_2O interaction. Experimental data is depicted as square and triangular markers [41, 42], with filled markers indicating the data was used in parameter estimation.

3.4. MEA + H_2O + SO_2 Mixture

To investigate the role of SO_2, we estimate two previously undeveloped parameters on a range on θ_{SO_2} experimental data for varying $w_{MEA}\%$, illustrated in Figure 7. The estimated parameters are ε_{NH_2,SO_2} and $\varepsilon_{assoc.,NH_2,SO_2}$, with the latter assuming that symmetric association is considered. That is, the association energy between an H site on SO_2 and an e site on NH_2 is equivalent to that between an e site on SO_2 and an H site on NH_2. We report the value of $\varepsilon_{NH_2,SO_2}/k_B$ as 453.85 K, and $\varepsilon_{assoc,NH_2,SO_2}/k_B$ as 663.02 K (Tables 2 and 3).

As presented in Figure 7, the fit to experimental data worsens at low $w_{MEA}\%$. As the parameter estimation was conducted considering only the interactions between SO_2 and NH_2, this result is expected, due to reduced $w_{MEA}\%$ in the solution. We report an AAD(%) of 36.13 in P_{SO_2} / bar. This value somewhat misrepresents the goodness of fit, due to the division of very small values of P_{SO_2} / bar (Equation 22) that occur at low values of θ_{SO_2}. Hence, when considering the AAD(%) in $\ln(P_{SO_2}$ / bar) we report a value of 4.24, indicating a satisfactory fit.

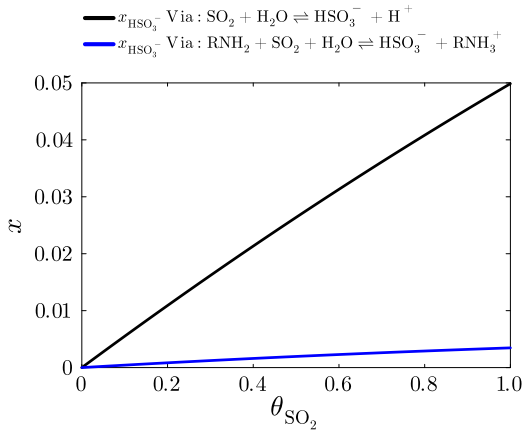


Figure 8: Predicted mole fractions of HSO_3^- in 30 $w_{MEA}\%$ aqueous MEA at $T = 298.15$ K, accounting for the two different speciation routes.

To examine the affinity of SO_2 for MEA, the speciation of SO_2 using the newly estimated parameters is investigated, with the results illustrated in Figure 8.

While SO_2 can be seen to react with MEA, it appears the main mechanism for dissolution of SO_2 into the solvent arises through reaction with H_2O. Based on the discussion in Section 2.3.4, particularly around Equation 15, this result is to be expected. Our implicit interpretation of the reaction scheme, requires SO_2 to be associated with both NH_2 and H_2O to facilitate the formation of HSO_3^- through reaction with MEA. As a result, the molar composition formed through Equation 8 is a subfraction of that formed through Equation 7.

While small, in Figure 8 we show that SO_2 does react with MEA. $x_{HSO_3^-}$ formed through Reaction 8 is depicted by the blue line. This reaction is the SO_2 equivalent for HCO_3^- and $NHCOO^-$ formation through Reactions 4 and 5 in the case of CO_2 speciation with MEA. This gives a preliminary indication that competition between SO_2 and CO_2 for reaction with MEA in the quaternary mixture is likely.

3.5. MEA + H_2O + CO_2 + SO_2 Mixture

In Figure 9 we present the effect of doping in SO_2 at fixed θ_{SO_2} to a MEA + H_2O + CO_2 mixture.

Upon first examination, Figure 9 (b) appears identical to the result for the MEA + H_2O + CO_2 mixture at 298.15 K presented in Figure 5. However, as can be seen in Figure 9 (a), a reduction in the partial pressure of CO_2 is observed at a given θ_{CO_2} as the set θ_{SO_2} is increased.

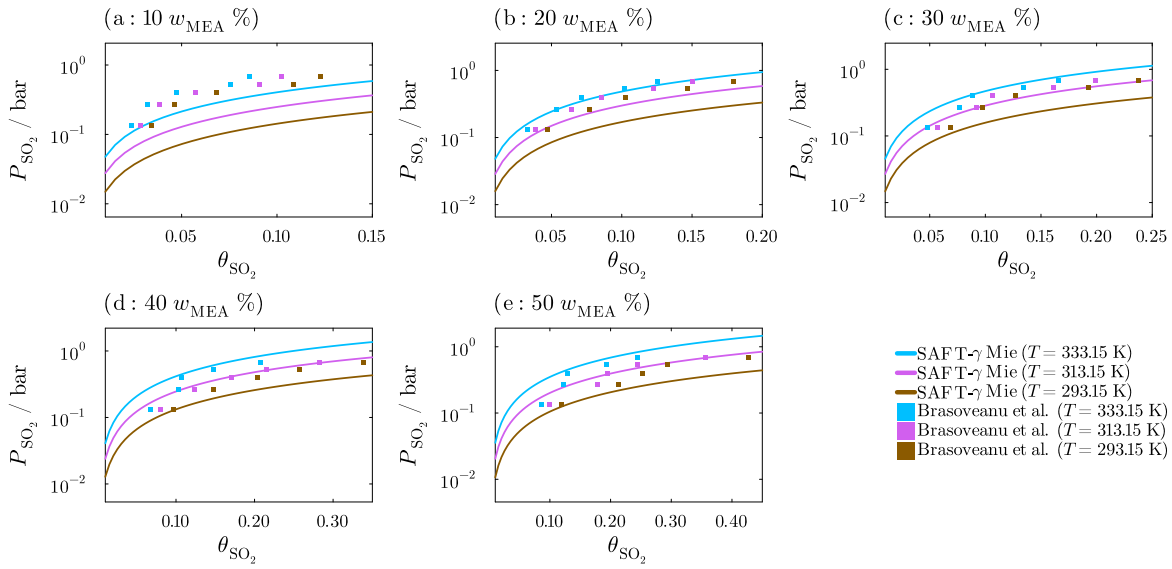


Figure 7: SO_2 loading in 10-50 $w_{\mathrm{MEA}}\%$ in H_2O . Experimental data is depicted as square markers [43], with filled squares indicating the data was used in parameter estimation.

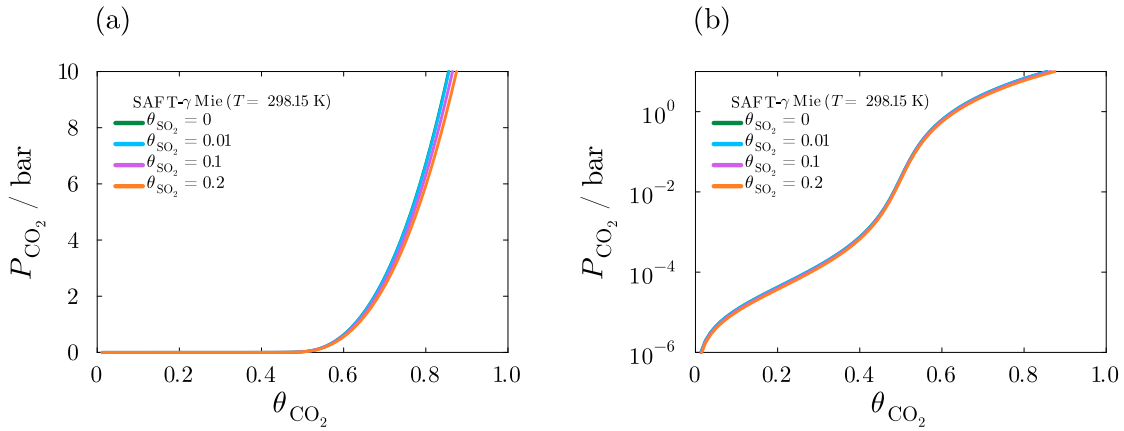


Figure 9: Predicted partial pressure of CO_2 at variable CO_2 loading in 30 $w_{\mathrm{MEA}}\%$ solution, with lines of fixed $\theta_{\mathrm{SO_2}}$. Linear plot (a), logarithmic plot (b).

Reduction in CO_2 partial pressure ($P_{\mathrm{CO_2}}$) as $\theta_{\mathrm{SO_2}}$ is increased is an unexpected result. Intuitively, it's expected that competition for association with MEA and H_2O, arising from introduction of SO_2 into the mixture, will lead to a greater CO_2 partial pressure. This behaviour is believed to be due to a reduction in the vapour pressure of the mixture, due to the presence of less volatile SO_2, in combination with reduced vapour fractions of CO_2.

The reduction in $P_{\mathrm{CO_2}}$ observed is small and mainly occurs at high $\theta_{\mathrm{CO_2}}$ and pressures non-typical of flue gas carbon capture systems. This result does, however, prompt examination of CO_2 speciation in this system. Speciation is used to verify there are no unexpected competing effects observed due to introduction of SO_2.

Prediction of the combined $\mathrm{RNHCOO^-}$ and $\mathrm{HCO_3^-}$ mole fractions ($x_{\mathrm{RNHCOO^-\ and\ HCO_3^-}}$), at fixed $\theta_{\mathrm{SO_2}}$ for the quaternary mixture of MEA + H_2O + CO_2 + SO_2 are presented in Figure 10 (a) and (b). Composition of $\mathrm{HCO_3^-}$ arising from H_2CO_3 disassociation are omitted from Figure 10, since reaction with MEA is deemed more significant for evaluating competition between CO_2 and SO_2.

As expected, $x_{\mathrm{RNHCOO^-\ and\ HCO_3^-}}$ for a given $\theta_{\mathrm{CO_2}}$ take on smaller values as $\theta_{\mathrm{SO_2}}$ is increased. This reduction is most prominent at higher $\theta_{\mathrm{SO_2}}$, consistent with the increase in $x_{\mathrm{HSO_3^-}}$ through reaction with MEA observed at higher $\theta_{\mathrm{SO_2}}$ in Figure 8. Confirming that SO_2 is competing with CO_2 for association with MEA.

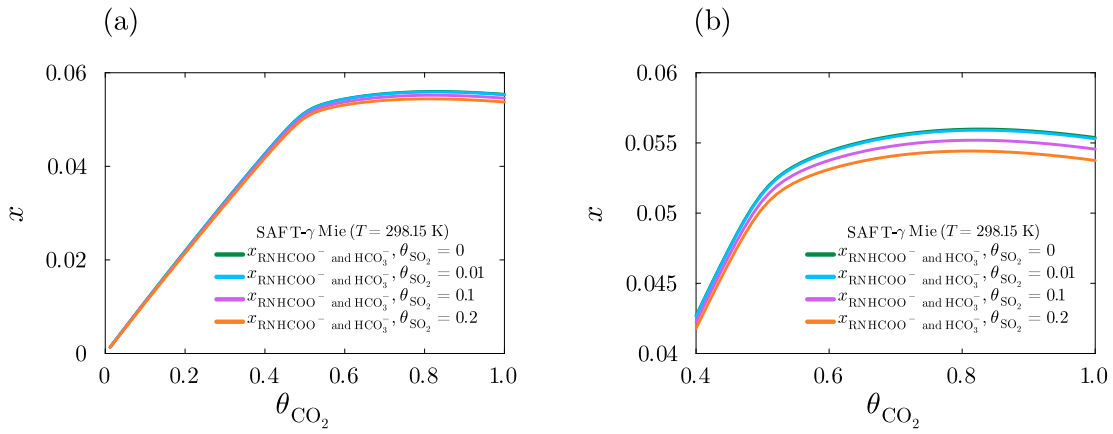


Figure 10: Predicted mole fractions of $RNHCOO^-$ at variable CO_2 loading in 30 w_{MEA}% solution, with lines of fixed SO_2 loading. Unmagnified plot (a), magnified plot (b)

4. Conclusion

In this study we have demonstrated the impact of SO_2 flue gas on CO_2 capture using the SAFT-γ Mie EoS. We developed a methodology to predict the distribution of chemical species resulting from speciation reactions in an aqueous MEA solution, highlighting the competitive interactions between CO_2 and SO_2. This approach incorporates the dispersion and association energy parameters, for use in the SAFT-γ Mie EoS, specifically between NH_2 and SO_2, for which we have proposed novel values.

Critically, we demonstrate that while there is competition, the degree of which is minimal. This is attributed to the greater affinity of CO_2 for MEA, suggested by the relatively smaller formation of ionic species from the reaction of SO_2 with MEA compared to the reaction of CO_2 with MEA.

5. Future Work

Having demonstrated the impact of SO_2 in flue gas, a logical progression of this work would involve examining the effects of other common flue gas components. These species include nitrogen oxide (NO), nitrogen dioxide (NO_2), and nitrogen (N_2) [24]. It is also worth noting that this report considers an MEA solution only, where more advanced solvents are also being used for CO_2 capture, hence modelling the competition in these solvents would be of interest. As with introduction of SO_2 in our work, the addition of more components would also call for the estimation of the parameters that describe them.

Considering the influence of oxygen (O_2) could also present an exciting opportunity for more detailed investigation. Inclusion of O_2 would lead to the formation of sulphur trioxide (SO_3), and hence sulphuric acid (H_2SO_4) in solution. H_2SO_4 reacts with MEA, reducing the ability for CO_2 capture. While work has been done in this area, Wang et al. [44] investigated the absorption kinetics of CO_2 in MEA, it would be of interest to examine in further detail the competition that arises.

6. Acknowledgements

We thank Pierre Walker for his continued advice and insights throughout this project. Additionally, we thank Chris Weston for his help.

References

(1) Bui, M. et al. *Energy Environmental Science* **2018**, *11*, 1062–1176.
(2) UNFCCC *Treaties and international agreements registered or filed and recorded with the Secretariat of the United Nations* **2016**.
(3) Kazlou, T.; Cherp, A.; Jewell, J. *Nature Climate Change* **2024**, *14*, 1047–1055.
(4) Rochelle, G. T. *Science* **2009**, *325*, 1652–1654.
(5) Huang, K.-H.; Wei, Z.; Cooks, R. G. *Chemical Science* **2021**, *12*, 2242–2250.
(6) Laddha, S.; Danckwerts, P. *Chemical Engineering Science* **1981**, *36*, 479–482.
(7) Vaidya, P. D.; Mahajani, V. V. *Industrial Engineering Chemistry Research* **2005**, *44*, 1868–1873.
(8) Crea, F.; Pettignano, A. *Molecules* **2020**, *25*, 826.
(9) Da Silva, E. F.; Svendsen, H. F. *International Journal of Greenhouse Gas Control* **2007**, *1*, 151–157.
(10) Sadegh, N.; Stenby, E. H.; Thomsen, K. *Fuel* **2015**, *144*, 295–306.
(11) Hessen, E. T.; Haug-Warberg, T.; Svendsen, H. F. *Chemical Engineering Science* **2010**, *65*, 3638–3648.
(12) Perdomo, F. A.; Khalit, S. H.; Graham, E. J.; Tzirakis, F.; Papadopoulos, A. I.; Tsiv-

intzelis, I.; Seferlis, P.; Adjiman, C. S.; Jackson, G.; Galindo, A. *Fluid Phase Equilibria* **2023**, *566*, 113635.
(13) Peng, D.-Y.; Robinson, D. B. *Industrial Engineering Chemistry Fundamentals* **1976**, *15*, 59–64.
(14) Soave, G. *Chemical Engineering Science* **1972**, *27*, 1197–1203.
(15) Chapman, W.; Gubbins, K.; Jackson, G.; Radosz, M. *Fluid Phase Equilibria* **1989**, *52*, 31–38.
(16) Chapman, W. G.; Gubbins, K. E.; Jackson, G.; Radosz, M. *Industrial Engineering Chemistry Research* **1990**, *29*, 1709–1721.
(17) Papaioannou, V.; Lafitte, T.; Avendaño, C.; Adjiman, C. S.; Jackson, G.; Müller, E. A.; Galindo, A. *The Journal of Chemical Physics* **2014**, *140*, DOI: 10.1063/1.4851455.
(18) Lafitte, T.; Apostolakou, A.; Avendaño, C.; Galindo, A.; Adjiman, C. S.; Müller, E. A.; Jackson, G. *The Journal of Chemical Physics* **2013**, *139*, DOI: 10.1063/1.4819786.
(19) Dufal, S.; Lafitte, T.; Haslam, A. J.; Galindo, A.; Clark, G. N.; Vega, C.; Jackson, G. *Molecular Physics* **2015**, *113*, 948–984.
(20) Perdomo, F. A.; Khalit, S. H.; Adjiman, C. S.; Galindo, A.; Jackson, G. *AIChE Journal* **2021**, *67*, DOI: 10.1002/aic.17194.
(21) Dufal, S.; Papaioannou, V.; Sadeqzadeh, M.; Pogiatzis, T.; Chremos, A.; Adjiman, C. S.; Jackson, G.; Galindo, A. *Journal of Chemical Engineering Data* **2014**, *59*, 3272–3288.
(22) Hutacharoen, P.; Dufal, S.; Papaioannou, V.; Shanker, R. M.; Adjiman, C. S.; Jackson, G.; Galindo, A. *Industrial Engineering Chemistry Research* **2017**, *56*, 10856–10876.
(23) Papaioannou, V.; Calado, F.; Lafitte, T.; Dufal, S.; Sadeqzadeh, M.; Jackson, G.; Adjiman, C. S.; Galindo, A. *Fluid Phase Equilibria* **2016**, *416*, 104–119.
(24) Haslam, A.; Jackson, G.; Galindo, A.; Adjiman, C. Unpublished data from the HiRECORD project, HiRECORD project.
(25) McCann, N.; Phan, D.; Wang, X.; Conway, W.; Burns, R.; Attalla, M.; Puxty, G.; Maeder, M. *The Journal of Physical Chemistry A* **2009**, *113*, PMID: 19338322, 5022–5029.
(26) Chang, L.; Zhongyang, Z.; Lingyu, S.; Linhang, Z.; Feng, X.; Xuanze, J.; Chenghang, Z.; Xiang, G. *Chemical Engineering Journal* **2022**, *448*, DOI: 10.1016/j.cej.2022.137751.
(27) Wertheim, M. S. *Journal of Statistical Physics* **1986**, *42*, 477–492.
(28) Lymperiadis, A.; Adjiman, C. S.; Galindo, A.; Jackson, G. *The Journal of Chemical Physics* **2007**, *127*, DOI: 10.1063/1.2813894.
(29) Pierre J. Walker, H.-W. Y.; Riedemann, A. *Ind. Eng. Chem. Res.* **2022**, *61*, 7130–7153.
(30) Mejía-de-Dios, J.-A.; Mezura-Montes, E. *Journal of Open Source Software* **2022**, *7*, 4723.
(31) Haslam, A. J.; González-Pérez, A.; Lecce, S. D.; Khalit, S. H.; Perdomo, F. A.; Kournopoulos, S.; Kohns, M.; Lindeboom, T.; Wehbe, M.; Febra, S.; Jackson, G.; Adjiman, C. S.; Galindo, A. *Journal of Chemical Engineering Data* **2020**, *65*, 5862–5890.
(32) DiGuilio, R. M.; Lee, R. J.; Schaeffer, S. T.; Brasher, L. L.; Teja, A. S. *Journal of Chemical Engineering Data* **1992**, *37*, 239–242.
(33) Águila-Hernández, J.; Trejo, A.; García-Flores, B. E.; Molnar, R. *Fluid Phase Equilibria* **2008**, *267*, 172–180.
(34) Růžička, K.; Fulem, M.; Mahnel, T.; Červinka, C. *Fluid Phase Equilibria* **2015**, *406*, 34–46.
(35) Cao, L.; Gao, J.; Zeng, S.; Dong, H.; Gao, H.; Zhang, X.; Huang, J. *International Journal of Greenhouse Gas Control* **2017**, *66*, 120–128.
(36) Liessmann, G.; Schmidt, W.; Reiffarth, S. *Data compilation of the Saechsische Olefinwerke Boehlen Germany* **1995**, *1*.
(37) Kroenlein, K.; Muzny, C.; Kazakov, A.; Diky, V.; Chirico, R.; Magee, J.; Abdulagatov, I.; Frenkel, M. NIST/TRC Web Thermo Tables (WTT).
(38) Notz, R.; Mangalapally, H. P.; Hasse, H. *International Journal of Greenhouse Gas Control* **2012**, *6*, 84–112.
(39) Belabbaci, A.; Razzouk, A.; Mokbel, I.; Jose, J.; Negadi, L. *Journal of Chemical Engineering Data* **2009**, *54*, 2312–2316.
(40) Jou, F.-Y.; Mather, A. E.; Otto, F. D. *The Canadian Journal of Chemical Engineering* **1995**, *73*, 140–147.
(41) Böttinger, W.; Maiwald, M.; Hasse, H. *Industrial Engineering Chemistry Research* **2008**, *47*, 7917–7926.
(42) Jakobsen, J.; Svendsen, H. *Industrial engineering chemistry research* **2005**.
(43) Brasoveanu, D.; Mihai, M.; Belcu, M. *Revista de Chimie -Bucharest- Original Edition-* **2000**, *51*, 687–696.
(44) Wang, M. H.; Ledoux, A.; Estel, L. *Procedia Engineering* **2012**, *42*, 397–407.

Tuning Fluid Models to Predict Sustainable Gasoline Substitutes Properties

Xinyu Cheng and Jing Xie
Department of Chemical Engineering, Imperial College London, U.K.

Abstract: In this paper, we report the sound speed in pure toluene, cyclopentane and binary liquid mixtures of toluene-ethanol and toluene-cyclopentane at temperatures between (273 to 373)K with pressures up to 60MPa. The measurements were carried out using a dual path pulse echo apparatus which was calibrated in pure toluene. The relative expanded combined uncertainty (k=2) varied from 0.049% to 0.15% for toluene, 0.043% to 0.11% for cyclopentane, 0.052% to 0.22% for the toluene ethanol mixture, and from 0.043% to 0.23% for the toluene-cyclopentane mixture. The reducing parameters and the scaling parameter of the KW0 model in the REFPROP[10] software was tuned in order to improve the sound speed prediction accuracy.

1. Introduction and Background

In response to the Paris Agreement[1], technologies in aid of decarbonization has been explored extensively. Many traditional fuel-burning applications have been electrified. However, some socially critical areas, such as the aviation and shipping industries, appear to be inadaptable to electric systems due to the state of the art battery energy density and capacity. Carbon prices have significantly increased in less than a decade, in the European Union, the 2022 total carbon revenue increased by the 600% in comparison to 2017[2]. This rapidly increasing operation cost incentivizes the exploration of sustainable gasoline alternatives to support the decarbonization of liquid fuel-dependent industries. A four-component system composed of toluene, ethanol, cyclopentane, and isooctane has been proposed as a surrogate for gasoline.

All four components are used as fuel additive to gasoline because of their high octane rating.[3][4][5][6] This property is related to engine operation, a fuel blend with a high octane number mitigates the risk of engine knocking or pinging. Additionally, blending ethanol into gasoline can effectively reduce air pollutant released during combustion.[2] Cyclopentane is also enhances heat efficiency and reduces carbon emission. [6]

Blends of these components can exhibit complex behaviours as a result of molecular interactions. Therefore, the thermodynamic and transport properties of these novel must be studied to predict the combustion performance and the compatibility with existing fuel engines. This calls for the design of a property calculation model that is capable of characterising mixtures based on composition alone. The multi-fluid Helmholtz equation of state (EoS) serves as a robust thermodynamic framework upon which the model could be built. The development of such a model relies on abundant experimental measurements. The speed of sound is amongst the various properties which are considered essential for model tuning in this work. This thermodynamic property is correlated to density and compressibility, making it crucial indicator of fuel response to changes in pressure and temperature.

A literature review was conducted on the published sound speed investigations (p, T, w) of binary mixture systems composed of the components introduced above as listed in Table 1. While there is readily available experimental data on pure component systems, few research was conducted on measuring the speed of sound in binary systems under high temperature high pressure conditions. This study presents speed of sound data for ethanol-toluene and cyclopentane-toluene mixtures of different compositions under various (p,T) conditions. In addition to establishing an accurate EoS model for the two binary mixtures under investigation, this study also aims to broaden the scope and fill the gap in speed of sound investigations regarding binary mixture systems. We have good faith that these experimental data would make valuable contributions in future studies of novel fuel mixtures in petrochemical industry along with improvements in establishing accurate EoS for binary systems.

2. Methods

2.1. Apparatus Description

The dual path pulse-echo method was employed to measure the sound speed in this research. The ultrasonic cell used in this study is fabricated from Invar 36. This material is an iron-nickel alloy (consisting of 64 mass % Fe and 36 mass % Ni) whose thermal expansion coefficient is approximately a tenth of that of stainless

Table 1. Review of sound speed measurements in binary systems

Author	Mixture	Year	Points	Temperature (K)	Pressure (MPa)	Phase
Moravkova et al.[18]	Toluene-Ethanol	2011	72	298.15-328.15	0.101325	liquid
Nikam et al. [19]	Toluene-Ethanol	2000	22	303.15-308.15	NA	liquid
Thiyagarajan R. and Palaniappan L.[20]	Toluene-Ethanol	2008	11	303	NA	liquid
Gonzalez-Olmos et al.[16]	Toluene-Ethanol	2010	195	288.18-323.15	0.101325	liquid
Calvar et al.[14]	Toluene-Cyclopentane	2009	13	313.15	0.101325	liquid
Gonzalez et al.[15]	Cyclopentane-Ethanol	2007	39	293.15-303.15	0.1	liquid

steel. [7] A gold-plated ceramic piezoelectric transducer is immersed inside the sample volume. The generator excites the transducer with a five-cycle tone burst at a carrier frequency of 5 Hz comprised of several sine waves. The transducer would in turn generate two ultrasonic pulses that travel through the liquid under investigation. The reflectors located on either ends of the cell will reflect the two pulses back to the transducer, which will generate an electrical response when a pulse is sensed. A digital oscilloscope then captures the signals from the transducer. Two responses will be recorded since the pulses the distance between the transducer and the two reflectors are different in Figure 1 (a). The echo arrival time difference Δt_{echo} between the two responses could be obtained from the two signals. Superimposing the two returning echoes (as demonstrated in Figure 1 (b).) provides insights to the datum quality. The speed of sound in the sample liquid could be calculated using the equation (1). This equation also incorporates a diffraction correction τ in terms of phase advance and angular frequency, whose contribution is less than a relative magnitude of 0.023%.

$$c = \frac{2(L_2 - L_1)}{\Delta t_{echo} + \tau} \quad (1)$$

The path length difference $(L_2 - L_1)$ was calculated from the equation (2) where ΔL_0 represents the path length difference at a reference state of (T_0, P_0); α stands for the average isobaric expansivity of the cell between the temperatures T_0 and T at constant pressure P_0; β is the mean thermal compressibility and is the average change in volume per unit volume for a unit change in pressure. The expansivity and compressibility were expressed as equations (3) and (4). The coefficients involved in these equations were established in the Scholz et al[8] work where the same ultrasonic cell was employed, their values are listed in Table 2.

Table 2. Coefficients for Invar expansivity and compressibility[8]

	$i = 0$	$i = 1$	$i = 2$	$i = 3$
a_i	1.65E+00	8.55E-04	2.94E-05	4.4E-08
b_i	1.06E+02	3.16E-02	3.27E-04	

$$(L_2 - L_1) = \Delta L(T, p) = \Delta L_0 \left[1 + \alpha(T - T_0) - \frac{\beta}{3}(P - P_0)\right] \quad (2)$$

$$10^6 \alpha / K^{-1} = \sum_0^3 a_i (T - 298.15)^i \quad (3)$$

$$K / GPa = \sum_0^2 b_i \left(T / K - 298.15\right)^i \quad \beta = \frac{1}{K} \quad (4)$$

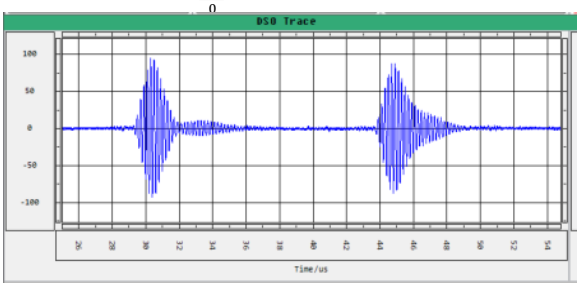


Figure 1 (a). Capture of the returning signals to the pulse-echo sensor

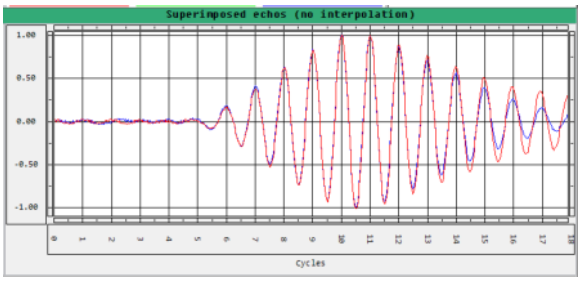


Figure 1 (b). Superposition of the returning signals, showing good overlap at the major peaks

A reference path length difference was established by calibrating the apparatus with pure toluene. The reference conditions are T=298K and p=1MPa. Measurements were carried out along five isotherms of T= (273.15, 298.15, 323.15, 348.15, and 373.15)K respectively over a pressure range of up to 60 MPa. The sound speed was then calculated with a nominal reference path length difference. The sound speeds obtained were compared with the measurements done by Meier et al[9]., and the reference path length difference was adjusted so that our measured sound speed match Meier's value[9] at the reference point. Discrepancies between the sound speed measurements in this work and Meier's work[9] were plotted in Figure 2. The deviations lay within ±0.03%, demonstrating a good agreement between measured sound speeds and previous literature data. Incidentally, the deviations were more pronounced under high temperature and low-pressure conditions and reached their minimum values near ambient temperatures.

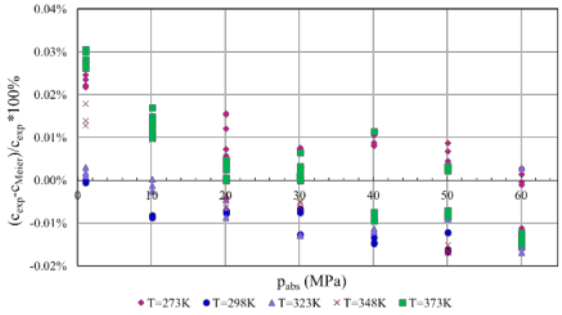


Figure 2. Deviation of toluene calibration experimental results from Meier's[9] data against absolute pressure

2.2. Experimental Procedures

In the calibration measurement using pure toluene and all subsequent toluene-ethanol system measurements, the liquid sample was degassed under vacuum before introducing into the apparatus. However, degassing was omitted when conducting measurements in pure cyclopentane and in the toluene-cyclopentane binary mixture to account for the highly volatile cyclopentane.

Before conducting any measurements, a vacuum must first be established between V6 and V5 for a substantial amount of time. This not only removes any impurities but also enables sample suction. A syringe pump, which was controlled through the software Pumpworks, varied the system pressure. When filling the apparatus, the syringe would retract until it was filled with fresh sample and would then extend until reaching a safe pressure limit of 5MPa. This step eliminates the possibility of vapour spaces remaining in the ultrasonic cell. To avoid sample vaporization and to ensure mixture

homogeneity, pump retraction-extension cycles were then performed.

The sound speed was measured along isotherms of 273.15, 323.15, and 373.15 K over a pressure range of 1MPa to 60MPa with the aid of the custom software HTBasic that acquires and analyses data. Each time when pressure was increased to the next desired testing value, the system temperature would experience a slight increase as a result of adiabatic pressurization. Reestablishing the thermo-equilibrium took around 20 minutes between measurements. Each binary pair was investigated at three different compositions. Check measurements were consistently conducted for all samples after recoding the last data set to ensure reliability.

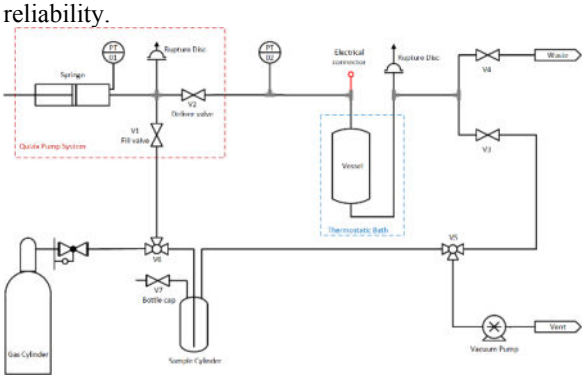


Figure 3. Process flow diagram of the apparatus involved

2.3. Uncertainty Analysis

To determine the measurement uncertainty, sound speed must first be correlated to temperature and pressure. equations (1) and (2) were combined, and since the uncertainty of ΔL_0 was difficult to obtain, this term was eliminated and replaced by Δt_0. This would result in the following equation:

$$c(T,p) = c_0 \frac{\Delta t_0 + \tau_0}{\Delta t + \tau}\left[1 + \alpha(T - T_0) - \frac{\beta}{3}(p - p_0)\right] \quad (5)$$

The numerically insignificant diffraction terms were neglected in the uncertainty analysis. Temperature and pressure's contribution to the variance of the sound speed was not accounted for in the squared uncertainty, where the uncertainty of the measurable and calculated variables is considered.

$$u^2(c) = \left[\left(\frac{\partial c}{\partial c_0}\right)u(c_0)\right]^2 + \left[\left(\frac{\partial c}{\partial \Delta t_0}\right)u(\Delta t_0)\right]^2 + \left[\left(\frac{\partial c}{\partial \Delta t}\right)u(\Delta t)\right]^2 + \left[\left(\frac{\partial c}{\partial \alpha}\right)u(\alpha)\right]^2 + \left[\left(\frac{\partial c}{\partial \beta}\right)u(\beta)\right]^2$$
$$= \left[\left(\frac{c}{c_0}\right)u(c_0)\right]^2 + \left[\left(\frac{c}{\Delta t_0}\right)u(\Delta t_0)\right]^2 + \left[\left(\frac{c}{\Delta t}\right)u(\Delta t)\right]^2 + \left[\left(\frac{c_0 \Delta t_0 (T - T_0)}{\Delta t}\right)u(\alpha)\right]^2 + \left[\left(\frac{-c_0 \Delta t_0 (p - p_0)}{3\Delta t}\right)u(\beta)\right]^2 \quad (6)$$

And the overall expanded combined uncertainty $U_c(c)$ is given by:

$$U_c(c) = k\{u^2(c) + [(\partial c/\partial p)_T u(p)]^2 + [(\partial c/\partial T)_p u(T)]^2 + u^2(c(w))\}^{\frac{1}{2}} \quad (7)$$

The term $u(c(w))$ accounts for the uncertainty of the sound speed which arises from the sample impurities. The sample purity was obtained from their respective lot analysis. The ethanol purity; however, was determined via a moisture content analysis due to its hydroscopic nature. Potential impurities were proposed for each pure component, and their effect on the sound speed was assessed using REFPROP[10]. The impurity with the largest effect on a pure component sound speed was carried forth to the binary mixture sensitivity analysis. The uncertainty which arises from the impurities is estimated using Equation 6 proposed by Scholz et al.[8] The chemical sample details and the result of the sensitivity analysis are summarised in Table SI (1).

$$u^2(c(w)) = \frac{(\Delta c)^2}{\sqrt{3}} \quad (8)$$

The uncertainty budget of the speed of sound for the toluene-ethanol mixture with the composition of 54 mol % ethanol at T=323K and P=30 MPa is exemplified in Table SI (2). The coverage factor, k, takes a value of 2 over the entire range of investigation.[11]

2.4. EoS Tuning

The Kunz-Wagner (KW0) model[12] was employed to analyze the experimental data and was later tuned to improve the model's predictive abilities. This model is an EoS based on the expression of Helmholtz free energy as a function of temperature, density, and composition.

$$\alpha(\delta, \tau, \bar{x}) = \alpha^o(\rho, T, \bar{x}) + \alpha^r(\delta, \tau, \bar{x}) \quad (9)$$

Where α^o describes ideal gas mixture behavior at the given conditions, and α^r accounts for the residual Helmholtz free energy. α^r could be further broken down to the contribution of pure component i (α^r_{oi}) and the departure function ($\Delta\alpha^r$). The departure function captures the non-ideal behaviors of the mixture and is essential for accurate predictions. The departure function for a binary mixture of i and j is expressed as:

$$\Delta\alpha^r_{i,j} = x_i x_j F_{ij} \alpha^r_{ij}(\delta, \tau) \quad (10)$$

The scaling factor, F_{ij}, adjusts magnitude of the generalized reducing function so that the EoS accurately calculates thermodynamic properties for a specific binary mixture. Departure functions are defined with respect to δ is the reduced mixture density and τ is the inversed reduced mixture temperature. These two parameters are defined as:

$$\delta = \frac{\rho}{\rho_r(\bar{x})} \; and \; \tau = \frac{T_r(\bar{x})}{T} \quad (11) \; and \; (12)$$

In the KW0 model[12], the reducing functions, $\rho_r(\bar{x})$ and $T_r(\bar{x})$, are developed from the quadratic mixing rule:

$$\frac{1}{\rho_r(\bar{x})} = \sum_{i=1}^{N} x_i^2 \frac{1}{\rho_{c,i}} + \sum_{i=1}^{N-1}\sum_{j=i+1}^{N} 2x_i x_j \beta_{v,ij}\gamma_{v,ij} \frac{x_i + x_j}{\beta_{v,ij}^2 x_i + x_j} \frac{1}{8}\left(\frac{1}{\rho_{c,i}^{1/3}} + \frac{1}{\rho_{c,j}^{1/3}}\right)^3 \quad (13)$$

$$T_r(\bar{x}) = \sum_{i=1}^{N} x_i^2 T_{c,i} + \sum_{i=1}^{N}\sum_{j=1}^{N} 2x_i x_j \beta_{T,ij}\gamma_{T,ij} \frac{x_i + x_j}{\beta_{T,ij}^2 x_i + x_j}\left(T_{c,i} \cdot T_{c,j}\right)^{0.5} \quad (14)$$

Two reducing parameters (β and γ), which are fitted to experimental data, are responsible for adjusting the reducing functions' shape, hence allowing the EoS to capture non-ideal mixture behaviors.

The composed database contained both our experimental results and literature data. In addition to deviation, defined as the difference between the experimentally determined data and the model prediction, a reduced deviation term was employed when tuning the EoS parameters. This term is obtained by dividing the deviation by the uncertainty associated with the experimental data. By calculating this, weightings could be assigned to individual experimental data according to their robustness.

Least-squares fitting is carried out to refine the parameters, and the objective function is set as follows:

$$objective\ function = \sum \frac{\left(x_{i,EoS} - x_{\mathrm{i,exp}}\right)^2}{\mathrm{U(x)}} \quad (15)$$

Where x_i is a thermodynamic property. This definition of the objective function considers the measurement quality with the rigorous data (characterized by their low uncertainty) receiving more weighting in the objective function. The objective function is minimized by tuning the EoS parameters. Parameters that yield the minimized objective function are recorded as the optimum parameters.

3. Results and Discussion

3.1. Pure Cyclopentane

Two illustrations of the general trend of speed of sound in relation with pressure and temperature in pure cyclopentane were plotted. The sound speed would increase with pressure and decreases temperature. Additionally, the speed of sound is more sensitive to temperature changes in the measurement range.

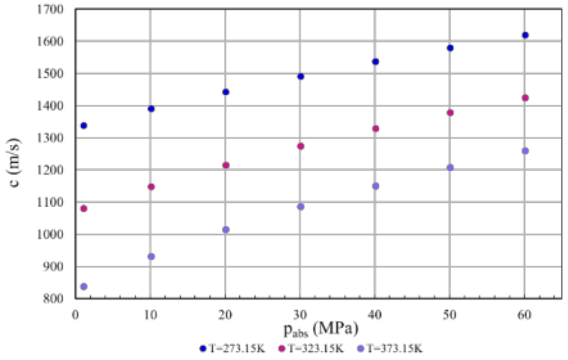


Figure 4. Sound speed against absolute pressure in pure cyclopentane

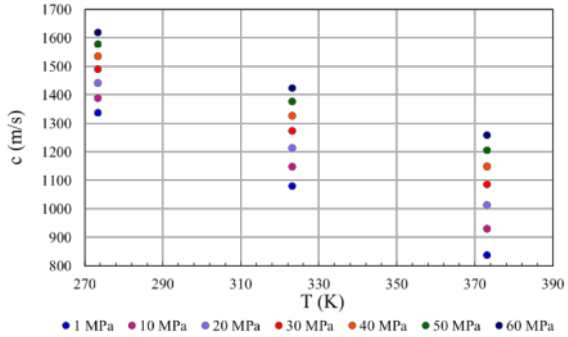


Figure 5. Sound speed against temperature in pure cyclopentane

3.2. Toluene-Ethanol Mixtures

Figure 6 displays the speed of sound as a function of composition at T=273K for all measured isobars. Pure ethanol sound speed was obtained from Wegge et. al. Nonlinearity could be observed from the set of graphs. This could be attributed to the non-ideal mixing behavior of toluene and ethanol, which is a mixture with a minimum boiling azeotrope.

The EoS sound speed calculations are in substantial disagreement with the experimentally measured values. The greatest deviation, which arose from the calculation at a temperature of 373K and 60MPa at a toluene mole fraction of 0.46, approaches 11%. Despite the intolerable inaccuracy, the performance of the EoS is sensitive to temperature, pressure, and composition. Figure 7 demonstrates the calculated value's deviation from the experimentally determined speed of sound. Amongst pressure, temperature and composition, the EoS performance is most sensitive to pressure changes. The model accuracy significantly improves with pressure. The deviation at 60 MPa is less than 60% of the deviation at 1MPa across all isotherms of this binary mixture. The predictions at low temperatures are also relatively more accurate, evidenced by the measurements at a temperature of 273K consistently having the lowest set of deviations across the three compositions. Additionally, the model performed better when calculating the speed of sound for the 25% mole toluene mixture and the 75% mole toluene mixture. This indicates that non-ideal mixture behaviors are not well represented in the EoS; consequently, the parameters of the departure function need to be adjusted accordingly.

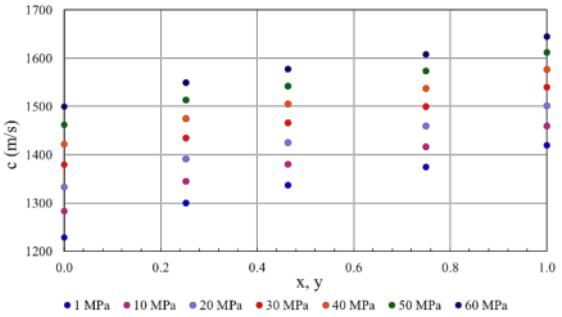


Figure 6. Sound speed in toluene-ethanol mixture against toluene composition at constant temperature of 273K

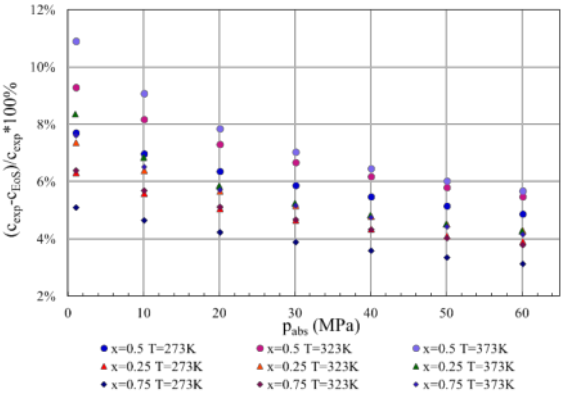


Figure 7. Deviation of measured sound speed in toluene-ethanol mixture from the EoS predictions against absolute pressure, x stands for the toluene composition

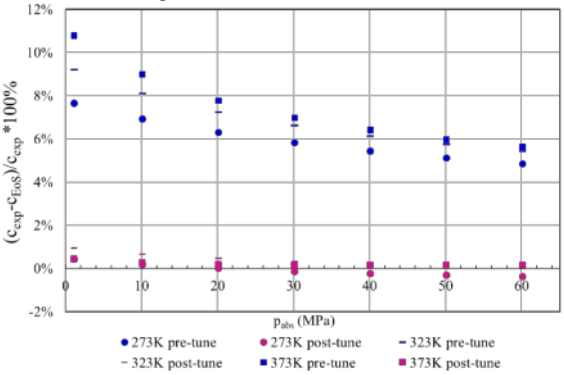


Figure 8. Deviation of measured sound speed in toluene-ethanol mixture from the EoS predictions against absolute pressure before and after parameters tuning

The sound speed prediction ability of the EoS increased drastically after the parameters were tuned. Figure 8 compares the calculated EoS sound speed's deviation from experimental data before and after parameter tuning. The calculated values' deviation from

their corresponding experimentally determined counterparts decreased to less than 1/100 of the original value, with the largest deviation being a 0.95% overestimation. However, this improvement comes with the tradeoff of sacrificing accurate predictions of VLE data, as shown in Figure 9 (a) and 9 (b). The departure function accounts for the non-ideal behaviors of a mixture. Although the scaling factor, Fij, determines the magnitude of the non-ideal adjustment, it is the reducing parameters βv,ij, γv,ij, βT,ij, and γv,ij that depict the non-ideal behaviors. Therefore, adjusting these reducing parameters would greatly affect the VLE predictions, such adverse effects are further magnified for this binary mixture due to the presence of an azeotrope. To compensate for this drawback, the weighting on the VLE experimental data is increased. These attempts to accurately predict both thermodynamic properties; however, led to increases in the objective function value at the optimal solution. This increase indicates that the overall model accuracy is compromised and that the KW0 model cannot account for both sound speed and VLE at the same time for the toluene-ethanol mixtures. Future studies should focus on exploring alternative departure functions for this binary mixture.

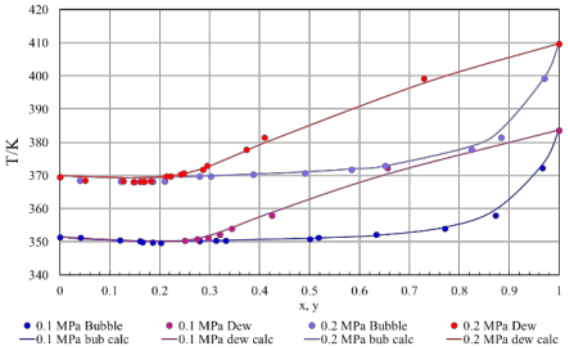


Figure 9 (a). Phase diagram of the toluene-ethanol mixture before parameters tuning

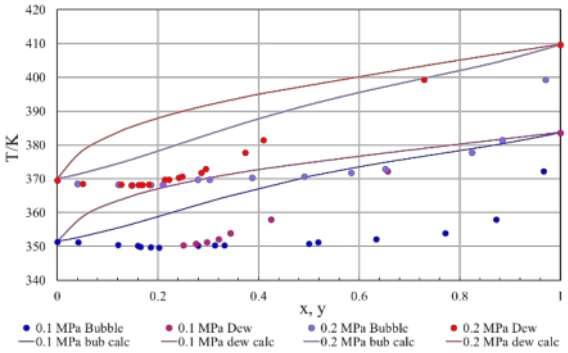


Figure 9 (b). Phase diagram of the toluene-ethanol mixture before parameters tuning

3.3. Toluene Cyclopentane Mixtures

Figure 10 displays the speed of sound as a function of composition at the three measured temperatures for all measured isobars. Nonlinearity that arises from non-ideal mixing behavior could be observed from the set of graphs.

The disagreement between the experimentally determined speed of sound and the EoS predicted sound speed for the toluene cyclopentane mixture is less significant than that of the toluene-ethanol mixture with the greatest deviation from the EoS prediction amounting to -1.33%. The EoS performance's relation to pressure and composition is more complex when compared to the model for the toluene-ethanol mixture. Figure 11 displays the calculated value's deviation from the experimentally determined speed of sound. The EoS predictions are more accurate at low temperatures, demonstrated by the isotherms at a temperature of 273K having the lowest deviation magnitude at a certain composition. The model performance could not be related to pressure or the mixture composition since the model's response to these two parameters is related to temperature. The inconsistencies in the model performance enhancement trends could be attributed to the negligence of non-ideal mixing behaviors.

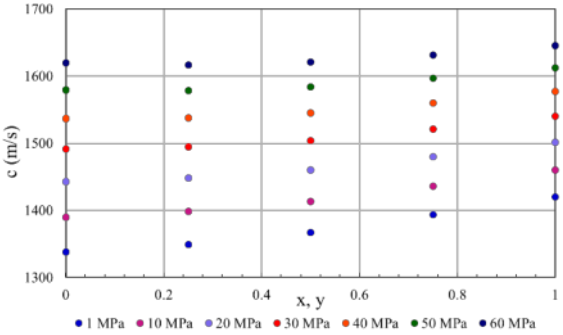


Figure 10. Sound speed in toluene-cyclopentane mixture against toluene composition at constant temperature of 273K

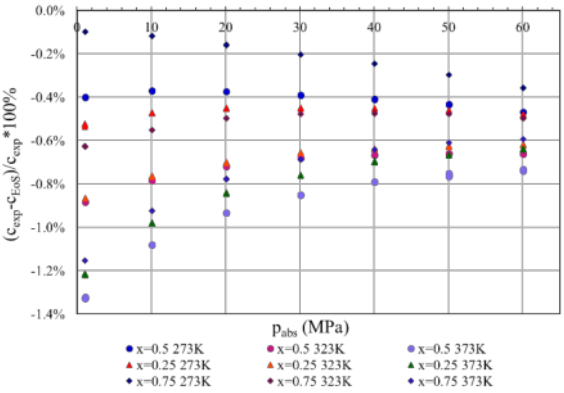


Figure 11. Deviation of measured sound speed in toluene-cyclopentane mixture from the EoS predictions against absolute pressure, x stands for the toluene composition

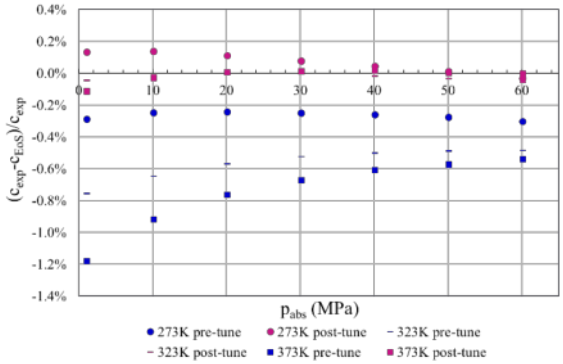


Figure 12 (a). Deviation of measured sound speed in toluene-cyclopentane mixture from the EoS predictions against absolute pressure before and after parameters tuning

The greatest challenge of tuning the EoS for the toluene-cyclopentane mixture arises from the lack of available data. Previous studies on this mixture often employed empirical correlations instead of fitting the experimental data to an EoS. Consequently, this binary mixture was assumed to behave ideally when

establishing the initial guesses for the reducing parameters and the scaling parameter values. Aside from the experimentally determined speed of sound measurements, data sets on density, heat capacity, and surface tension were downloaded from the literature in order to consider as many thermodynamic properties as possible when tuning the EoS parameters. Notably, no data on this binary mixture's VLE were available from the literature. VLE conveys important mixture behavior information; hence, it is generally considered quintessential for mixture EoS parameter fitting[12]. Absence of this information could result in inaccurate parameter adjustments. Nevertheless, the performance of the equation of state is significantly enhanced after tuning the parameters. As demonstrated in Figure 12 (a) and 12 (b), the discrepancy between the modelled values and the experimentally obtained values decreased for both sound speed and density. The weighting on the density data was adjusted to determine if the value of the objective function could be further reduced. However, the model performance worsened as more weight was assigned to the density data. Future studies should focus on determining the VLE for this binary mixture since this property could be an indicator of the tuned EoS's validity.

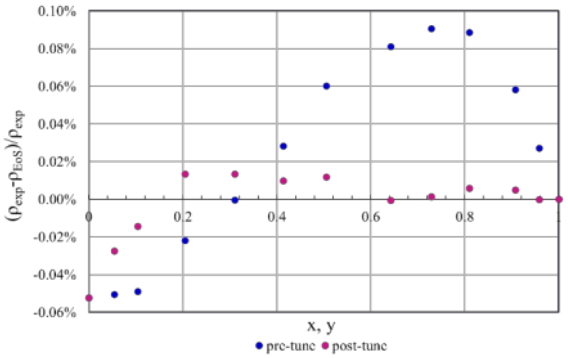


Figure 12 (b). Deviation of literature density data in toluene-cyclopentane mixture from the EoS predictions against toluene composition before and after parameters tuning

4. Speed of Sound Data

4.1. Toluene-Cyclopentane Mixtures

	T (K)	p (MPa)	c (ms^{-1})	U(c) (ms^{-1})
x=0.24	273.289	1.089	1349.39	1.18
	273.29	10.095	1398.46	1.10
	273.291	20.104	1448.49	1.04
	273.292	30.093	1494.63	0.98
	273.29	40.087	1537.90	0.94
	273.291	50.084	1578.62	0.90
	273.292	60.092	1617.12	0.87
	323.144	1.098	1104.19	1.51
	323.141	10.102	1167.40	1.35
	323.146	20.099	1229.61	1.22
	323.149	30.092	1285.77	1.12
	323.149	40.087	1337.04	1.05
	323.149	50.085	1384.59	0.99
	323.151	60.090	1429.09	0.94
	373.061	1.092	874.58	2.05
	373.063	10.103	958.72	1.70
	373.064	20.104	1036.75	1.46
	373.064	30.098	1104.21	1.30
	373.064	40.091	1164.41	1.18
	373.065	50.090	1218.93	1.09
	373.064	60.097	1269.23	1.02

	T (K)	p (MPa)	c (ms^{-1})	U(c) (ms^{-1})
x=0.48	273.294	1.097	1367.07	0.789
	273.293	10.104	1413.25	0.772
	273.296	20.104	1460.47	0.755
	273.295	30.100	1504.33	0.735
	273.295	40.091	1545.56	0.722
	273.293	50.086	1584.46	0.714
	273.293	60.089	1621.26	0.707
	323.152	1.096	1132.49	0.816
	323.149	10.114	1191.22	0.786
	323.148	20.108	1249.61	0.761
	323.148	30.098	1302.54	0.738
	323.149	40.094	1351.23	0.720
	323.149	50.091	1396.53	0.706
	323.148	60.093	1439.01	0.694
	373.061	1.115	914.14	0.953
	373.063	10.108	990.61	0.897
	373.064	20.108	1063.03	0.848
	373.064	30.098	1126.35	0.809
	373.063	40.091	1183.29	0.776
	373.065	50.090	1235.27	0.750
	373.063	60.093	1283.26	0.728

	T (K)	p (MPa)	c (ms^{-1})	U(c) (ms^{-1})
x=0.76	273.29	1.10	1393.39	1.13
	273.29	10.11	1436.14	1.06
	273.29	20.11	1480.20	1.00
	273.29	30.10	1521.39	0.95
	273.29	40.09	1560.29	0.91
	273.29	50.09	1597.02	0.88
	273.29	60.09	1631.90	0.85

323.15	1.09	1169.03	1.41
323.15	10.11	1222.90	1.27
323.15	20.10	1277.15	1.16
323.15	30.09	1326.64	1.08
323.15	40.09	1372.45	1.01
323.15	50.09	1415.27	0.96
323.15	60.09	1455.45	0.92
373.05	1.09	961.06	1.86
373.06	10.10	1030.15	1.58
373.07	20.10	1096.76	1.38
373.06	30.10	1155.81	1.24
373.06	40.09	1209.13	1.13
373.06	50.09	1258.22	1.06
373.07	60.09	1303.74	0.99

4.2. Toluene-Ethanol Mixtures

	T (K)	p (MPa)	c (ms^{-1})	U(c) (ms^{-1})
x=0.25	273.31	1.09	1300.89	1.19
	273.29	10.11	1345.72	1.09
	273.29	20.11	1392.25	1.01
	273.29	30.10	1435.33	0.94
	273.29	40.09	1475.60	0.89
	273.29	50.09	1513.65	0.85
	273.29	60.09	1549.80	0.82
	323.15	1.10	1109.91	1.44
	323.15	10.11	1166.06	1.27
	323.15	20.11	1221.92	1.14
	323.15	30.10	1272.85	1.05
	323.15	40.09	1319.69	0.98
	323.15	50.09	1363.26	0.92
	323.15	60.09	1404.19	0.88
	373.06	1.10	922.33	1.82
	373.06	10.10	994.78	1.52
	373.06	20.10	1063.81	1.32
	373.06	30.10	1124.48	1.18
	373.06	40.09	1179.20	1.08
	373.06	50.09	1229.08	1.00
	373.06	60.09	1275.28	0.94

	T (K)	p (MPa)	c (ms^{-1})	U(c) (ms^{-1})
x=0.46	273.30	1.09	1337.48	0.77
	273.30	10.10	1380.69	0.75
	273.29	20.10	1425.32	0.74
	273.29	30.10	1466.79	0.72
	273.29	40.09	1505.77	0.71
	273.29	50.09	1542.73	0.70
	273.29	60.09	1577.83	0.69
	323.15	1.09	1136.36	0.80
	323.15	10.10	1190.39	0.78
	323.15	20.10	1244.60	0.75
	323.15	30.09	1294.09	0.73
	323.15	40.08	1339.73	0.71
	323.15	50.09	1382.53	0.70
	323.15	60.09	1422.53	0.69
	373.06	1.09	942.46	0.94
	373.06	10.10	1012.32	0.89
	373.06	20.10	1079.40	0.84
	373.06	30.09	1138.64	0.80
	373.06	40.09	1192.18	0.77
	373.06	50.09	1241.23	0.74
	373.06	60.09	1286.75	0.72

	T (K)	p (MPa)	c (ms^{-1})	U(c) (ms^{-1})
x=0.75	273.29	1.10	1375.19	1.08
	273.29	10.11	1417.06	1.01
	273.29	20.10	1459.91	0.95
	273.29	30.10	1500.16	0.91
	273.29	40.09	1538.08	0.87
	273.29	50.09	1574.01	0.84
	273.29	60.09	1608.17	0.82

323.15	1.09	1167.02	1.30
323.15	10.11	1218.93	1.18
323.15	20.11	1271.48	1.08
323.15	30.10	1319.46	1.01
323.15	50.09	1405.67	0.90
323.15	60.10	1444.89	0.87
373.06	1.09	969.94	1.64
373.06	10.10	1036.87	1.42
373.06	20.10	1101.48	1.26
373.06	30.10	1159.04	1.14
373.06	40.10	1211.10	1.05
373.06	50.09	1259.08	0.98
373.06	60.09	1303.61	0.93

4.3. Pure Toluene and Pure Cyclopentane

Pure toluene

T (K)	p (MPa)	c (ms^{-1})	U(c) (ms^{-1})
273.29	1.09	1420.13	1.01
273.29	10.10	1460.11	0.96
273.29	20.10	1501.55	0.92
273.29	30.10	1540.55	0.88
273.29	40.09	1577.38	0.85
273.29	50.09	1612.25	0.83
273.29	60.10	1645.58	0.81
298.15	1.10	1309.88	1.10
298.15	10.12	1354.58	1.03
298.15	20.11	1400.56	0.97
298.15	30.11	1443.30	0.93
298.15	40.10	1483.20	0.89
298.15	50.09	1521.00	0.86
298.15	60.09	1556.74	0.83
323.15	1.09	1203.58	1.20
323.15	10.10	1253.77	1.11
323.15	20.10	1304.50	1.04
323.16	30.10	1351.18	0.98
323.15	40.09	1394.68	0.93
323.15	50.09	1435.36	0.89
323.15	60.09	1473.62	0.85
348.10	1.10	1101.60	1.33
348.10	10.11	1157.94	1.21
348.11	20.10	1213.95	1.11
348.11	30.10	1264.96	1.03
348.11	40.09	1311.90	0.97
348.10	50.09	1355.44	0.92
348.11	60.09	1396.39	0.88
373.06	1.10	1002.78	1.50
373.06	10.11	1066.29	1.33
373.06	20.10	1128.29	1.20
373.06	30.10	1183.88	1.10

Pure cyclopentane

T (K)	p (MPa)	c (ms^{-1})	U(c) (ms^{-1})
273.29	1.09	1338.01	0.78
273.29	10.10	1390.11	0.76
273.29	20.10	1443.00	0.75
273.29	30.09	1491.61	0.73
273.29	40.09	1537.04	0.72
273.29	50.09	1579.55	0.71
273.29	60.09	1619.69	0.70
323.15	1.09	1080.48	0.81
323.16	10.10	1148.57	0.78
323.16	20.10	1214.92	0.76
323.15	30.10	1274.39	0.74
323.15	40.09	1328.43	0.72
323.16	50.09	1378.23	0.70
323.16	60.09	1424.77	0.69
373.06	1.09	838.05	0.95
373.06	10.10	930.48	0.89
373.06	20.10	1014.43	0.84
373.07	30.09	1086.32	0.80
373.06	40.09	1149.90	0.77
373.06	50.08	1207.21	0.75
373.06	60.09	1259.90	0.73

373.06	40.09	1234.37	1.02
373.06	50.09	1281.27	0.96
373.06	60.10	1324.53	0.92

5. Conclusion

The sound speed was measured in binary mixtures of toluene-ethanol and toluene-cyclopentane at temperatures ranging from 273K to 373K and pressures ranging from 1MPa to 60MPa via the dual path pulse echo method. The system was calibrated using pure toluene against the empirical relationship established by Meier et al. The sound speed measurements in the binary mixtures were compared to the Kunz-Wagner (KW0) model calculations in REFPROP[10]. The model predictions significantly deviated from the experimentally determined speed of sound values, with the largest deviation approaching 11% for the toluene-ethanol mixtures and -1.33% for the toluene-cyclopentane mixture. The reducing parameters and scaling parameter was tuned to minimize the EoS prediction's deviation from experimental data. The deviation of the sound speed predictions of the tuned EoS reduced to 0.95% for the toluene-ethanol mixtures and 0.16% for the toluene-cyclopentane mixture. However, fitting the sound speed data to the EoS destroyed the accurate VLE prediction for the toluene-ethanol mixture, and this could not be compensated by adjusting the weighting on the VLE data. This suggests that future studies focusing on this binary mixture should explore employing other departure functions which could be more suitable. VLE data, which are essential for EoS tuning, was not available for the toluene-cyclopentane mixture. Future works on this mixture should first measure this property and then proceed to assess the validity of our model

6. References

[1] IPCC, 2023: Sections. In: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland, pp. 35-115, doi: 10.59327/IPCC/AR6-9789291691647

[2] European Environment Agency (2023). *Use of Auctioning Revenues Generated under the EU Emissions Trading System*. [online] www.eea.europa.eu. Available at: https://www.eea.europa.eu/en/analysis/indicators/use-of-auctioning-revenues-generated.

[3] Jiao, L., Liu, H., Qu, L., Xue, Z., Wang, Y., Wang, Y., Lei, B., Zang, Y., Xu, R., Zhang, Z., Li, H. and Abdulaziz, O. (2020). QSPR Studies on the Octane Number of Toluene Primary Reference Fuel Based on the Electrotopological State Index. *ACS omega*, 5(8), pp.38783888.doi:https://doi.org/10.1021/acsomega.9b03139.

[4] www.iea-amf.org. (n.d.). *Ethanol properties*. [online] Available at: https://www.iea-amf.org/content/fuel_information/fuel_info_home/ethanol/e10/ethanol_properties.

[5] American Chemical Society. (2024). *Isooctane*. [online] Available at: https://www.acs.org/molecule-of-the-week/archive/i/isooctane.html [Accessed 12 Dec. 2024].

[6] *Chemical & Engineering Data*, 65(7), pp.3679–3689. doi:https://doi.org/10.1021/acs.jced.0c00299.

Song, H., Dauphin, R. and Vanhove, G. (2020). A kinetic investigation on the synergistic low-temperature reactivity, antagonistic RON blending of high-octane fuels: Diisobutylene and cyclopentane. *Combustion and Flame*, 220, pp.23–33. doi:https://doi.org/10.1016/j.combustflame.2020.06.030.

[7]Al Ghafri, S.Z.S., Matabishi, E.A., Trusler, J.P.M., May, E.F. and Stanwix, P.L. (2019). Speed of sound and derived thermodynamic properties of para-xylene at temperatures between (306 and 448) K and at pressures up to 66 MPa. *The Journal of Chemical Thermodynamics*, [online] 135, pp.369–381. doi:https://doi.org/10.1016/j.jct.2019.03.022.

[8]Scholz, C.W., Sanchez-Vicente, Y., Tananilgul, T., Thol, M., Trusler, J.P.M. and Richter, M. (2020). Speeds of Sound in *n*-Pentane at Temperatures from 233.50 to 473.15 K at Pressures up to 390 MPa. *Journal of Chemical & Engineering Data, 65(7), pp.3679–3689. doi:https://doi.org/10.1021/acs.jced.0c00299.*

[9]Meier, K. and S. Kabelac (2013). Measurements of the Speed of Sound in Liquid Toluene. *Journal of Chemical & Engineering Data*, 58(5), pp.1398–1406. doi:https://doi.org/10.1021/je4001697.

[10] Lemmon, E.W., Huber, M.L. and McLinden, M.O. (2013). NIST Standard Reference Database 23: Reference Fluid Thermodynamic and Transport Properties-REFPROP, Version 9.1. NIST. [online] Available at: https://www.nist.gov/publications/nist-standard-reference-database-23-reference-fluid-thermodynamic-and-transport.

[11]Wagner, W. and Pruß, A. (2002). The IAPWS Formulation 1995 for the Thermodynamic Properties of Ordinary Water Substance for General and Scientific Use. *Journal of Physical and Chemical Reference Data*, 31(2), pp.387–535. doi:https://doi.org/10.1063/1.1461829.

[12]Kunz, O. and Wagner, W. (2012). The GERG-2008 Wide-Range Equation of State for Natural Gases and Other Mixtures: An Expansion of GERG-2004. *Journal of Chemical & Engineering Data*, 57(11), pp.3032–3091. doi:https://doi.org/10.1021/je300655b.

[13]Wegge, R., Richter, M. and Span, R. (2015). Speed of Sound Measurements in Ethanol and Benzene over the Temperature Range from (253.2 to 353.2) K at Pressures up to 30 MPa. *Journal of Chemical & Engineering Data*, 60(5), pp.1345–1353. doi:https://doi.org/10.1021/je501065g.

[14]Calvar, N., González, B., Gómez, E. and Canosa, J. (2009). Density, Speed of Sound, and Refractive Index for Binary Mixtures Containing Cycloalkanes and Aromatic Compounds at T = 313.15 K. *Journal of Chemical & Engineering Data*, 54(4), pp.1334–1339. doi:https://doi.org/10.1021/je8009336.

[15]González, B., Calvar, N., Domínguez, Á. and Tojo, J. (2007). Dynamic viscosities of binary mixtures of cycloalkanes with primary alcohols at T=(293.15, 298.15, and 303.15)K: New UNIFAC-VISCO interaction parameters. *The Journal of Chemical Thermodynamics*, 39(2), pp.322–334. doi:https://doi.org/10.1016/j.jct.2006.06.008.
[16]Gonzalez-Olmos, R., Iglesias, M. and Mattedi, S. (2010). Influence of temperature on thermodynamics of ethanol + hydrocarbon gasoline additives. *Physics and Chemistry of Liquids*, 48(3), pp.337–384. doi:https://doi.org/10.1080/00319100902894223.
[17]Lam, V.T. and Benson, G.C. (1970). Surface tensions of binary liquid systems. I. Mixtures of nonelectrolytes. *Canadian Journal of Chemistry*, 48(24), pp.3773–3781. doi:https://doi.org/10.1139/v70-637.
[18]Morávková, L., Wagner, Z., Sedláková, Z. and Linek, J. (2011). Volumetric behaviour of binary and ternary liquid systems composed of ethanol, isooctane, and toluene at temperatures from (298.15 to 328.15) K. Experimental data and correlation. *The Journal of Chemical Thermodynamics*, [online] 43(12), pp.1906–1916. doi:https://doi.org/10.1016/j.jct.2011.06.021.
[19]Nikam, P.S., Jagdale, B.S., Sawant, A.B. and Hasan, M. (2000). Densities and Viscosities of Binary Mixtures of Toluene with Methanol, Ethanol, Propan-1-ol, Butan-1-ol, Pentan-1-ol, and 2-Methylpropan-2-ol at (303.15, 308.15, 313.15) K. *Journal of Chemical & Engineering Data*, 45(4), pp.559–563. doi:https://doi.org/10.1021/je990317i.
[20]R. Thiyagarajan and Palaniappan, L. (2008). Effect of toluene in the dipolar interaction of some aliphatic alcohols. *Physics and Chemistry of Liquids*, 46(4), pp.366–371. doi:https://doi.org/10.1080/00319100701312807.
[21]Vega-Maza, D., Segovia, J.J., Carmen Martín, M., Villamañán, R.M. and Villamañán, M.A. (2011). Thermodynamic properties of biofuels: Heat capacities of binary mixtures containing ethanol and hydrocarbons up to 20MPa and the pure compounds using a new flow calorimeter. *The Journal of Chemical Thermodynamics*, [online] 43(12), pp.1893–1896. doi:https://doi.org/10.1016/j.jct.2011.06.019.
[22]Wright, W.A. (1933). A Method for a More Complete Examination of Binary Liquid Mixtures. *Journal of physical chemistry*, 37(2), pp.233–243. doi:https://doi.org/10.1021/j150344a009.
[23]Zéberg-Mikkelsen, C.K., Lugo, L., García, J. and Fernández, J. (2005). Volumetric properties under pressure for the binary system ethanol+toluene. *Fluid Phase Equilibria*, 235(2), pp.139–151. doi:https://doi.org/10.1016/j.fluid.2005.05.023.

The Assessment of Readiness for Small-scale Carbon Capture & Storage Systems Applied on Diesel Fuel Power Generators in Remote Areas.

Yihao Zhang and Yuhan Wang
Department of Chemical Engineering, Imperial College London, U.K.

Abstract:

The implementation of carbon capture and storage (CCS) on a considerably wide scale is vital for achieving global climate goals, yet its application in small-scale systems remains underdeveloped. This study explores the feasibility of deploying CCS on diesel-powered generators in remote areas, using the Meadowbank open-pit gold mine in Canada as a case study. A post-combustion CO2 capture approach utilizing monoethanolamine (MEA) absorption is analyzed, along with compression and transportation systems modelled in Aspen Plus & Aspen HYSYS. Results reveal significant challenges due to the small scale and remote location, including high energy and transportation costs. Alternatives such as membrane-based capture technologies and direct air capture (DAC) are discussed as potential solutions. While technically viable, small-scale CCS faces economic and logistical barriers, highlighting the need for technological innovation and policy support to enhance its practicality in decarbonization efforts.

1. Introduction

The carbon capture and storage (CCS) system is widely regarded as a crucial technology that aligns with the objectives of the Paris Agreement. While global CCS capacity experienced substantial growth in the 2020s, with a 48% increase noted in 2023[1], experts argue that its deployment must be expedited to keep pace with the rapid expansion of wind turbines during the 2000s [2]. To expand its deployment range, CCS for small-scale plants presents an intriguing area for research, given that only 3% of the current planning and constructing CCS facilities were designed for small-scale (below 0.1 Mt annually) [3].

Diesel fuel-powered generators are crucial in remote and underdeveloped areas that lack access to the power grid, and they also provide reliable energy backup for industrial and residential buildings during emergencies. In addition to providing electricity, diesel fuel combustion engines can generate heat from waste gases when necessary, which cannot be offered directly by some alternatives, like wind turbines. This non-intermittent energy source is crucial, particularly in the mining and construction sectors. Therefore, the market for diesel fuel generators is estimated to grow by 8% due to rapid industrialization, urbanization, and the development of new mines in developing countries [4].

Various cutting-edge technologies have been proposed to reduce CO2 emissions from diesel fuel power generators. Other portable energy sources, including small-scale wind turbines and micro-nuclear reactors, have unique advantages [5] [6]. However, many of these methods are underdeveloped or face significant geographic limitations.

Direct air capture (DAC) is another alternative for small-scale CCS. It is an essential approach for addressing historical emissions and mitigating emissions from discrete sources, such as small-scale diesel power generators [7].

In contrast, implementing small-scale CCS to collect CO2 from ‘point emission sources’ is still an appealing choice for achieving ‘zero emissions.’

2. Background

2.1 Scenario specification

To illustrate the challenges of implementing CCS on small-scale diesel power generators in remote areas, a Canadian gold mine in the subarctic region has been chosen for a case study.

The Meadowbank open-pit gold mine in the Kivalliq District of Nunavut, Canada, operated by Agnico Eagle, was the company's first sub-arctic gold mine, beginning commercial production on March 1, 2010 [8]. In 2019, the ore in the main gold mines had been depleted. Still, the processing complex continued to be used for the newly discovered Amaruq open-pit gold mine, located 50 kilometres away.

Both Meadowbank and Amaruq install sets of low-sulphur diesel generators with 29 MW and 900 kW capacities, respectively [9]. These generators provide power and heat for the ore processing facilities and support the living infrastructure for 500 people. As a result, the total diesel consumption amounts to 33.5 million litres, leading to annual CO2 emissions of 90,000 tons [8].

In 2023, the site produced 431,666 ounces (12 tons) of gold and 125,000 ounces (3.5 tons) of silver. However, due to its geographical remoteness, the cost of production reached $1,214 per ounce of gold [8]. Raw materials, equipment, and supplies are typically transported via large cargo ships over the ocean to Hudson Bay; from there, shallow-draft vessels transport these materials through the Chesterfield Inlet to the floating port located in Baker Lake [9].

Meadowbank Mine serves as an extreme example of decarbonization. A relatively small amount of CO2, harsh environmental conditions, and low accessibility all contribute to the challenges of deploying CCS, as the following discussion will demonstrate.

2.2 Method specification

In this study, CO2 capture, separation and transportation techniques will be considered as thoroughly as possible.

As the first step, an appropriate CO2 capture method should be selected from three basic types: pre-combustion, oxyl-fuel or post-combustion. The first two methods involve the gasification of fuel to produce syngas (CO and H2) before burning. High purity of CO2 could be reached as a result, which makes the separation easier and less energy-intensive [33]. However, these two methods require a complete overhaul of the entire system, which is economically unfeasible for replacing small-scale diesel power generators, considering that this miniaturization is still under development.

Since the post-combustion CO2 capture method is the only viable option, the next step is to decide on the separation technique. Several methods have been developed for extracting CO2 from exhaust gas streams, including absorption, adsorption, membrane separation, and distillation [33]. Monoethanolamine (MEA) absorption is a well-studied and modelled benchmark for CO2 capture, making it the primary separation technique. Other alternatives will be discussed in section 4.5.

Given that the major planning and operating permanent CO2 storage sites in Canada are primarily situated in Alberta and Saskatchewan [10]—over 1600 kilometres away from the Meadowbank Mine—shipping emerges as the only viable method for transporting CO2 to other storage sites. Consequently, it is necessary to construct onshore infrastructure near Barker Lake, which will include temporary CO2 storage tanks, pumps, and corresponding dock facilities to transport CO2 to the containers on boats. This setup will be linked to the main carbon capture facility through a 100-kilometre pipeline. In the summer months, ships should be equipped with the appropriate technology to transport CO2 via Hudson Bay.

3. Methodology

The CO2 capture and compression facilities designed to mitigate the carbon footprint of the diesel power generators at the Meadowbank open-pit gold mine have been modelled separately using Aspen Plus. A series of sensitivity analyses have been conducted to optimize the equipment's operating parameters. Additionally, heat integration across two flowsheets has been assessed to ensure the balance between energy consumption and capital costs. As a result, capital and operating costs have been effectively reduced.

3.1 Flowsheet explanations

Temperature	**45℃**
Pressure	**1.1bar**
Mass flowrate	**200 tonne/hour**
CO2 mass fraction	**0.05**
N2 mass fraction	**0.77**
O2 mass fraction	**0.11**
H2O mass fraction	**0.07**

table 1 Exhaust gas composition

As shown in Figure 1, the whole plant could be divided into three sectors: CO2 capture, CO2 compression, and CO2 transportation.

The exhaust gases from the diesel power generators are directed to the bottom of the absorber after electricity and heat are supplied to the ore processing facility. Before entering the CO2 capture sector, its temperature was reduced to 45ºC.

The gas stream encounters a 30% MEA solution in a counter-current flow within the absorber, effectively reducing the concentration of CO2 in the exiting gas to below 1% on a weight basis. The resulting dirty MEA solution is then pressurized to 2 bars and pumped to the top of the desorber, where it is heated to regenerate the MEA. After condensation, a gas stream containing 20% water and 80% CO2 exits the top of the desorber. To achieve a higher concentration of CO2 for the compression cycle, an additional flash vessel is needed to separate water from that stream. The purified MEA passes through a heat exchanger to preheat the dirty stream before returning to the absorber's top.

During the compression cycle, nearly pure CO2 is pressurized and cooled to a sub-cooled state at 130 bars and 15°C. After travelling 100 km through the pipeline,

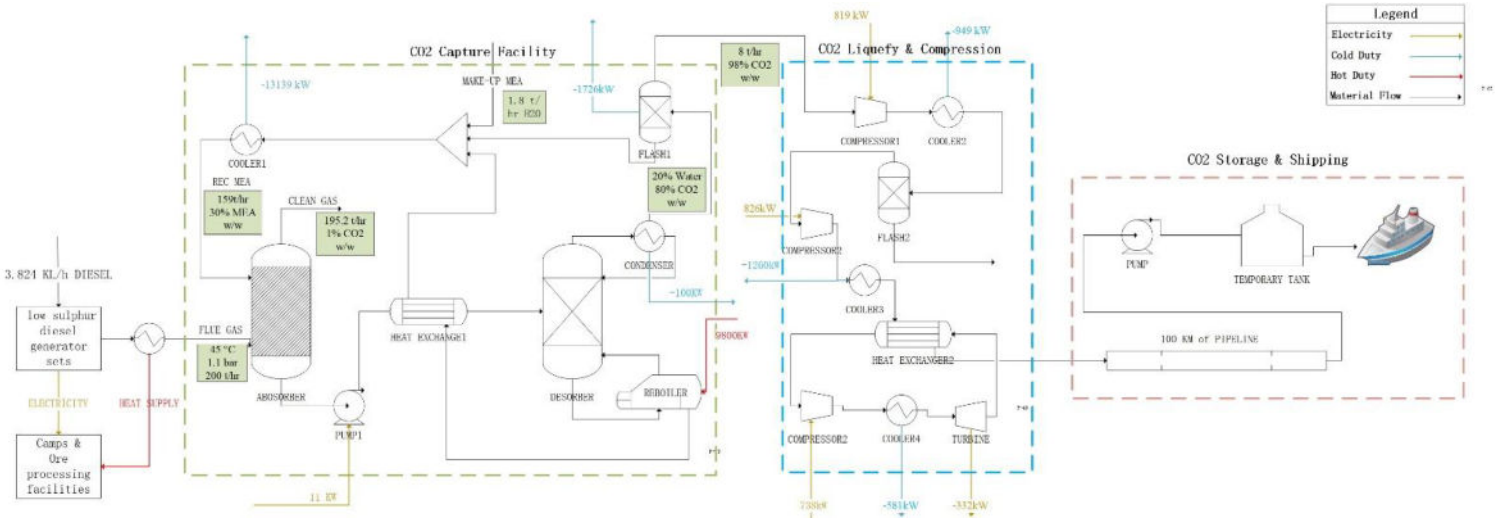


Figure 1: The layout of the CO2 capture, compression and transportation complex

the CO2 is reliquefied and temporarily stored in tanks at the shore before being transported by ships to a

permanent storage site.

3.2 Reaction Mechanism

The NRTL-RK electrolyte property method predicts thermodynamic properties for both liquid and vapour phases. This model combines the Electrolyte Non-Random Two-Liquid (NRTL) model, which predicts the behaviour of non-ideal liquid-phase electrolyte systems, with the Redlich-Kwong equation of state. As a result, it accurately describes the characteristics of liquid-liquid equilibrium (LLE) involving electrolytes and salts, as well as the non-ideal gas state and the corresponding vapour-liquid equilibrium (VLE) [11].

Aspen Electrolyte Wizard has provided the following equations for the reaction between MEA and CO2:

$$2H_2O \leftrightarrow OH^- + H_3O^+ \quad (1)$$

$$H_2O + HCO^{3-} \leftrightarrow CO_3^{2-} + H_3O^+ \quad (2)$$

$$2H_2O + CO_2 \leftrightarrow HCO^{3-} + H_3O^+ \quad (3)$$

$$H_2O + MEACOO^- \leftrightarrow MEA + HCO^{3-} \quad (4)$$

$$H_2O + MEA^+ \leftrightarrow MEA + H_3O^+ \quad (5)$$

Equation 1 describes the intrinsic ionization of H_2O. Equations 2 and 3 illustrate how CO_2 dissolves in water and forms bicarbonate. Meanwhile, Equations 4 and 5 explain how MEA significantly enhances the solution's capacity to absorb CO_2 by forming MEA^+ and $MEACOO^-$ ions.

The rate constant is estimated by the following equation:

$$\ln(K_{eq}) = A + \frac{B}{T} + Cln(T) + DT \quad (6)$$

	A	B	C	D
(1)	132.90	-13446	-22.477	0
(2)	216.05	-12432	-35.482	0
(3)	231.47	-12092	-36.782	0
(4)	-0.5214	-2545.5	0	0
(5)	-3.0383	-7008.4	0	-0.0031

Table 2 Reaction kinetics parameter

3.3 Design for absorber and desorber

In this study, the Aspen Plus rate-based column model (RadFrac) is employed to simulate both the absorber and desorber. This model incorporates film resistance calculations based on the two-film theory [12]. Each stage within the columns is represented as a liquid-gas interface, considering mass, heat transfer, and film resistance phenomena. In contrast to the equilibrium-stage model, which assumes that each stage achieves thermodynamic equilibrium, the rate-based model provides a more accurate and realistic representation of the highly non-ideal mixture in this case.

Parameters	Absorber	Desorber
Module	Rate-based column	Rate-based column
Internal type	packed	Packed
Packing material	Sulzer 250Y	Sulzer 250Y
Stage number	10	12
Diameter (m)	4.5	4.5
Packing height(m)	10	8
Liquid insert stage	1	1
Liquid film resistance	Discretize film	Discretize film
Vapor film resistance	Ignore film	Ignore film
Reaction condition factor	0.9	0.9
Discrete point in a liquid film	5	5
Film Discretization Ratio	10	10

Table 3 Summary of Design Parameters for CO2 Capture

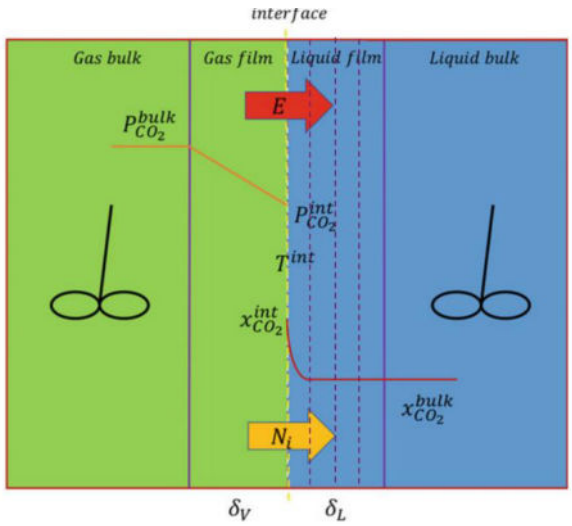


Figure 2 Concentration profile gas-liquid interface.[13]

In the absorber, no reaction occurs in the gas phase, which leads to a linear material profile in the gas zone. That leads to reasonable disregarding of the gas film in the model. However, when CO_2 is absorbed into the MEA solution, it reacts quickly, creating a strongly non-linear material profile. In this case, it is necessary to analyze the liquid film by discretizing it and activating the Discrxn option to describe the gradient of the profile accurately. A similar film model can also be applied to the reverse reaction of CO_2 and MEA in the desorber, as

described in [13]. The profile of CO2 concentration is shown in Figure 2.

Table 3 provides the relevant parameters used in the model, referencing [14] and [15].

3.4 Design for the compression cycle

Stream	H2O molar flow (kmol/h)	CO2 molar flow (kmol/h)	Temp (℃)	Pressure （bar）
Input	**11.8**	**214**	**35**	**1.1**
COMP1 out	**11.8**	**214**	**334**	**20**
COOLER 2 out	**11.8**	**214**	**35**	**10**
FLASH1 TOP	**1.1**	**214**	**35**	**10**
FLASH2 bottom	**10.7**	**0**	**35**	**10**
COMP2 out	**1.1**	**214**	**315**	**130**
COOLER 3 out	**1.1**	**214**	**35**	**130**
EX2 out	**1.1**	**214**	**10**	**130**

Table 4 stream table of the compression process

Device	Process	Electricity power required (kW)	Heat required (kW)
COMP1	Isentropic compressing	1024	-
COOLER2	Isobaric cooling	-	-949
FLASH2	Separation	-	-
COMP2	Isentropic compressing	826	-
COOLER3	Isobaric cooling	-	-1260
COMP3	Isentropic compressing	738	-
COOLER4	Isobaric cooling	-	-580
TURBINE	-	-332	-

Table 5 Summary of Design Parameters for Compression

To meet the pipeline transportation requirements, CO_2 must be compressed into either a subcooled or supersaturated fluid state. Given the relatively low ambient temperatures during transportation, a subcooled liquid state at 15°C and 130 bars has been selected. The reasoning behind this choice will be explained further in section 3.6. Furthermore, the accuracy of the thermodynamic property diminishes significantly as the impurity level increases [16]. For the purpose of accuracy of the modelling, as well as the concern of preventing corrosion in the pipeline, water content must be reduced to less than 50 ppm [17].

A two-stage compression system with inter-stage cooling is applied in the compression cycle. Compared with single-stage compression, multi-stage compression consumes less energy. Since the compression process is assumed to be isentropic and adiabatic, the temperature of the compressor outlet stream will increase. By applying multi-stage compression with inter-stage cooling, the process approaches isothermal compression [18]. An inter-stage separator can also be installed between two compression cycles to further enhance the CO_2 purity in the transportation stream. The process flow is shown in figure 1.

The h-p diagram and the compression route is shown in figure 3. We use Aspen Plus to stimulate this process, the stream result is shown in table 4 and the duty of devices is given in table 5.

3.5 heat integration

Category	Inlet T(℃)	Outlet T(℃)	Heat duty （kW）
Desorber (condenser)	**110**	**106**	**-100**
Desorber (reboiler)	**110**	**136**	**9800**
Flash1	**106**	**25**	**-1780**
Cooler1	**105**	**10**	**-13000**
Desorber inlet	**47.9**	**110**	**9900**
Desorber bottom outlet	**136**	**69**	**-9900**
Cooler2	**334**	**35**	**-949**
Cooler3	**315**	**35**	**-1260**
Cooler4	**83.7**	**35**	**-581**

Table 6 Heat integration table for all devices

Heat integration is an effective approach to minimize energy consumption and enhance economic potential. The devices and streams involved in heat transfer are listed in Table 6. Obviously, the inlet stream to the desorber and the outlet stream from its bottom have complementary heat duties. Both inlet and outlet streams have temperature differences exceeding 10°C, so Heat Exchanger 1 is installed to meet their specific temperature requirements.

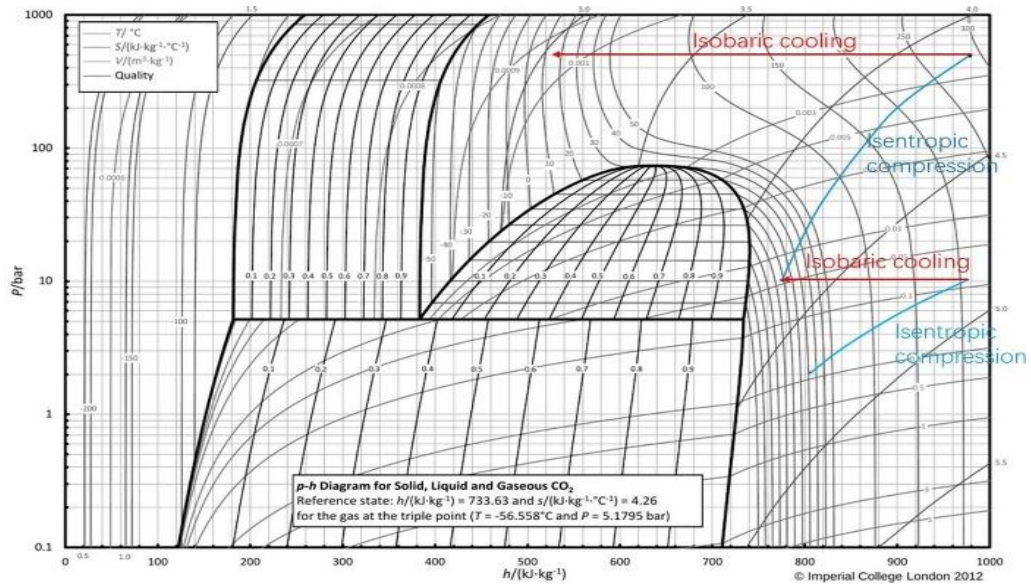


Figure 3 CO2 h-p plot and compression pathway

Only the desorber's reboiler requires heating among the remaining devices. Considering the initial temperature, only Cooler 2 and Cooler 3 are suitable for heat exchange with the reboiler. By installing two additional heat exchangers, energy savings of 657 kW and 878 kW can be achieved. However, considering the capital costs, this option has limited economic feasibility. Furthermore, since the plant is expected to operate in a cold climate, the excess heat could be redirected to support the facility's heating system, providing additional benefits.

The pump station also has a cooler, but it cannot be used for heat integration due to geographical limitations. An air cooler is applied to avoid unnecessary pipelines.

3.6 Design for Transportation Strategy

As mentioned in section 2.2, CO2 is captured and compressed near the gold mine before being transported via a 100 km pipeline to a storage facility near Baker Lake. When Baker Lake is accessible, ships with suitable tanks can transport the CO2 to permanent storage sites. Therefore, the transportation design is divided into two components: pipelines and shipping.

The fluid must have a high density for efficient transportation to minimise pipe size. Therefore, CO2 should be transported as a dense liquid or supercritical state. As the fluid moves through the pipe, the friction between the stream and the walls decreases energy, resulting in an inevitable pressure drop. As the pressure decreases, the density of the dense phases notably declines, which increases the transportation speed and exacerbates the pressure loss. Eventually, this pressure drop can lead to a significant decrease in pressure, known as a "choking point," at which gas formation occurs after a certain distance [19].

Aspen HYSYS was used to evaluate the pressure drops under various inlet and ambient conditions. As illustrated in Figure 4, both streams initially have a pressure of 140 bar; however, the temperatures have been adjusted to ensure they are in different states. The ambient temperature is set to 0°C.

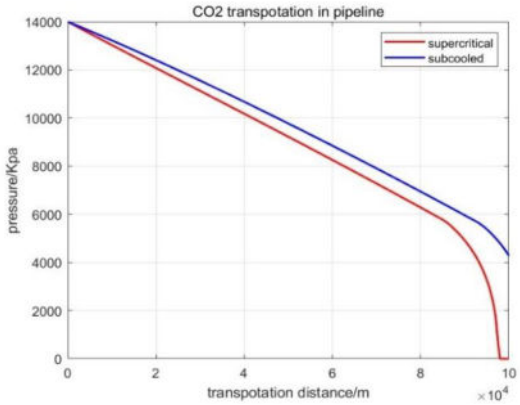


Figure 4 CO2 transportation via pipeline (Supercritical & sub-cooled)

The ‘choking point’ for both streams is observed, but the supercritical is inferior to the subcooled state. This is mainly due to:

1. Subcooled liquids usually have higher density, which leads to a slower volumetric flow rate and fewer friction losses.
2. Subcooled liquid has a smaller compressibility factor, which implies less change in the volumetric flow rate as pressure decreases. [19].
3. More heat is transferred to the surroundings for the supercritical stream since it has a much more significant temperature difference.

The main drawback of the subcooled state is its low operating temperature, which significantly limits transportation distances in warmer regions due to rapid gasification as a considerable amount of heat is absorbed from the surroundings. However, the annual temperature in the Baker Lake region ranges from -30 °C to 17 °C [20], making it an ideal location for transport in a

subcooled state.

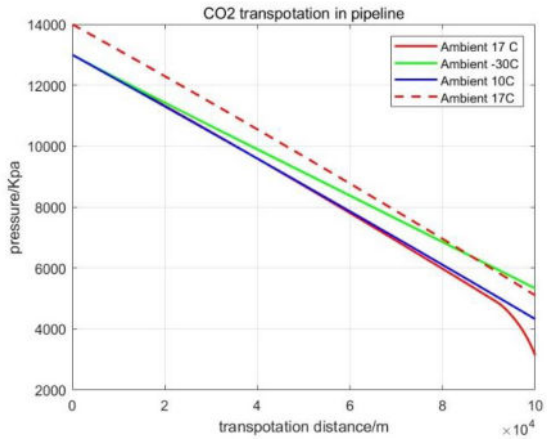


Figure 5 CO2 transportation simulation for different ambient conditions

Figure 5 illustrates the validation of the selected operating condition (130 bar, 10 °C). A noticeable 'choking point' occurs at maximum ambient temperature. To ensure safe transportation during the summer, a higher pressure of 140 bar must be used.

Implementing an insulation layer could potentially benefit transportation by reducing the heat transfer from the surroundings. However, this approach may add complexity to the calculation of capital costs. It could also limit the advantages of the colder surroundings for most of the year, which helps maintain the fluid in a dense state [19].

Temporary storage is the last step in our research before CO2 is transported to a permanent storage site, such as an oil field or saline formation. The typical operating conditions for CO2 tanks on ships are -52 °C and 5~7 bar[19]. By applying this condition and simulation in Aspen under the thermodynamic model of Predictive Soave−Redlich−Kwong (PSRK), the monthly CO2 captured is 4800 m^3. Baker Lake is accessible for only 2.5 months yearly, requiring CO2 to be stored in temporary tanks for the remaining 9.5 months. [9]. As a result, the maximum volume of CO2 within the tanks (V_{year}) is assumed to be 45624 m^3

$$V_{ship} = \frac{V_{year}}{N_{v,year}} \times \frac{1}{0.98} \quad (7)$$

Assume Only one voyage could be completed during the summer months $N_{v,\,year}$=1. The multiplier 1/0.98 reflects the international Code, which states that a tank should only be filled to 98% of its capacity to prevent the explosion from liquid expansion [21].

As a result, the total volume of tanks on the cargo ships should be 46555 m^3.

The Chesterfield Inlet is a long, narrow, relatively shallow channel connecting Baker Lake to Hudson Bay. Because of its depth and width, large cargo ships cannot enter Baker Lake, even in the summer months. Instead, smaller vessels, approximately 100 meters long, 20 meters wide, and with a draft not exceeding 6 meters, should be used. However, navigating this shipping lane remains highly risky due to complex environmental conditions and potential human factors. [22]. This implies the potential instability of the CO2 transportation and hence influences the design of the temporary CO2 storage tank.

The temporary CO2 storage tank's volume, or "buffer capacity," ranges from 150% to 200% of the V_{ship}. Optimising the tank's volume involves balancing capital expenditures against unavailability costs, which are the losses incurred when insufficient storage capacity exists [23]. Due to the high instability mentioned earlier, a capacity of 200% of the V_{ship} is selected for this case.

3.7 compression station

Due to the continuous pressure drop of carbon dioxide during pipeline transportation, a compression station needs to be installed every 100 kilometres to transport the CO_2 to more distant storage sites. At an ambient temperature of 10°C, carbon dioxide begins transportation at a pressure of 130 bar. Simulations reveal that after travelling 100 kilometres, the pressure decreases to 59.07 bar. A compressor and a subsequent air cooler are applied in the station. The compressor will restore the pressure to 130 bar. Since this process is isentropic, an air cooler must be installed to prevent the hot stream from damaging the pipe walls. Such a method is necessary for transporting longer distances via pipeline.

4. Results and discussion

4.1 Energy analysis

Category	Energy consumption(kW)	Energy percentage
Capture	**9811**	**81.5%**
Compression	**2051**	**17.0%**
Transportation	**-**	**-**
Pump station	**162**	**1.5%**
Total	**12033**	**1**

Table 7 energy consumption and percentage of each unit.

The efficiency of the CCS facility is evaluated by energy consumption per ton of CO2 capture.

The heat duty of the system is 9,800 kW in desorber and the total electricity requirement for the pumps and compressors is 2224 kW. Assuming a 75% efficiency for small-scale combined heat and power generators [24], this results in an additional annual diesel consumption of 8.5 ML and 22500 tons of CO2 emissions, which is a 25% increase compared to the origin.

As a result, the average energy cost per ton of CO2 captured is 3.7 GJ, which lies between the typical value of MEA absorption, 3.6 ~4 GJ/ton of CO2 [4.5.6].

This example did not take into account the extra energy needed to capture the CO2 emitted from the CCS process. The iteration method is necessary for the entire flowsheet to accurately determine the stable reboiler duty

and power requirements. Nonetheless, the energy cost per ton of CO2 captured for this process should remain unchanged.

4.2 Cost analysis

All costs in this report are estimated using Aspen simulations and presented in USD. In this example, the transport distance is assumed to be 200 km, including one compression station.

The simulation results for costs are summarised in Table 8. Based on the literature [25], the project lifespan (n) is set to be 10 years, with a discount rate (r) of 5%. The annualised capital cost (ACC) is calculated using the following equations.

$$ACC = CAP \times CRF \tag{8}$$

$$CRF = \frac{r \cdot (1+r)^n}{(1+r)^n - 1} \tag{9}$$

CRF is used to determine the annual payments required to amortise the capital over specific r and n [27]. Based on ACC, the levelized cost can be specified with a given annual CO2 absorption weight (ACM):

$$LC = \frac{ACC}{ACM} \tag{10}$$

Category	Operating cost ($/year)	Capital cost ($)	ACC($)	LC ($/tCO2)
Capture	**2370000**	**6350000**	**823000**	**30.2**
Compression	**1970000**	**5010000**	**649000**	**24.6**
Pipeline	**-**	**25200000**	**3260000**	**46.8**
Compression station	**1490000**	**3360000**	**43000**	**17.8**
Storage tank	**-**	**23682360**	**3066865**	**44.1**
Total	**5830000**	**63582360**	**8226865**	**163.6**

Table 8 cost analysis for different section

If we consider only the carbon capture and compression processes, the levelized cost (LC) amounts to $54.85 per ton of CO_2. In comparison, for large-scale CCS, the capture cost is $22.46 per ton of CO_2 [25], and the cost of two-stage compression is $21.80 per ton of CO_2 [26]. The higher LC for smaller-scale CCS systems can be attributed to size constraints, which result in lower energy efficiency and equipment utilisation compared to large-scale CCS operations. However, the transportation process, including pump stations, accounts for 54.08% of the total levelized cost. Additionally, this cost increases with distance, making transportation a critical factor in determining the economic feasibility of the project. This relationship is further elaborated in Section 4.3.

Consider the uncertainty during installation and operation. Engineering and contingency costs are considered by referencing the ratio recommended in [27]. The detailed cost is shown in Table 9.

Category	Cost ($)	Percentage
ACC	5160000	25%
Operating cost	5830000	29%
Engineering	2240000	11%
contingency	7110000	35%

Table 9 Summary of costs

4.3 Alternative carbon capture method

Membrane-based technology (MBT) is widely acknowledged as a promising alternative for carbon capture systems. Membranes can selectively separate gas mixtures due to the differing permeabilities of various gas molecules, which arise from their physical or chemical properties. Novel research is focused on identifying membrane materials and structures that can achieve optimised separation. For instance, graphene oxide nanosheets exhibit high thermal and mechanical stability and perform effectively in N_2-CO_2 separation due to its 2D structure [28]. Besides that, Carbon-based membranes, such as carbon hollow fibre (CHF), are also suitable for carbon capture due to their high selectivity and permeability [29].

The primary benefit of the membranes is their compactness, allowing all processes to be integrated into a single unit, significantly reducing size and weight [30]. Compared to absorption methods, which rely on a continuous supply of absorbent materials, membrane-based carbon capture systems only need the membranes replaced for ongoing operation. Additionally, the primary energy consumption comes from the electricity required to operate the pump, which is significantly lower than absorption systems. Therefore, membrane technology suits small-scale carbon capture in remote areas.

However, two main factors restrict membrane performance: permeability and selectivity. These two parameters have a trade-off relationship; as the permeability of gas molecules increases, selectivity tends to decrease [31], leading to the requirement for a larger membrane area. The plasticisation of the membrane must be considered, as it causes the pores to swell over time, leading to the degeneration of the membrane. [32]. Finally, membrane technology has not yet reached a commercially viable level, and significant financial investment is needed to drive further advancements and breakthroughs [33].

4.4 Geospatial analysis of small-scale CCS

As highlighted in Section 4.2, transportation, including pipelines and temporary storage tanks—constitutes a significant portion of the overall capital cost. This creates a geospatial limitation for deploying carbon capture and storage (CCS), as these facilities must be located near the emission sources. In contrast, DAC can be established near permanent storage sites, thereby eliminating

transportation costs. While DAC is generally considered less efficient than CCS due to the difference of feed in CO2 concentration, the potential savings on transportation may offset this efficiency difference.

This section examines the relationship between transportation costs and distance. A preliminary analysis is conducted based on levelized costs (LC) to identify suitable locations for small-scale CCS deployment by comparing costs with those of DAC. If deploying small-scale CCS in a region is less cost-effective than DAC, the project may be considered unjustified from a fundamental perspective. The distance at which CCS costs are equal to DAC costs is defined as the "competitive radius", which is centred at permanent storage sites, such as oil fields; CCS is more effective than DAC within this range. DAC cost is based on the literature [34].

Assumptions are shown below:

1. In cold area CO2 is translated under 10℃, 130bar (sub-cooled state) and in warm area it is 30℃, 155bar (supercritical state).
2. In this simulation, we assume that the pipeline is built on flat terrain. However, in reality, pipelines often need to navigate varying elevations to traverse complex landscapes, significantly affecting the pressure gradient within the pipe. Additionally, constructing tunnels or bridges to cross hills or rivers incurs extra capital costs. Therefore, the competitive radius used in the analysis should be considered an overestimation.
3. The average ambient temperature in the warm area is set to be 25℃, and in the cold regions, it is 10℃.
4. The cost for CO2 capture in CCS is assumed to be constant and irrelevant to the humidity of the air.

Figure 6 shows the relationship between transport distance and LC. From the constant cost for CO2 capture facilities, the cost of CCS increases linearly as longer pipes are implemented. The pump stations account for the step changes, as they must be installed every 100 kilometres in cold regions and every 94 kilometres in warm areas. The competitive radius of CCS in cold regions is 222 km for dry conditions and 232 km for wet conditions. In warm areas, the radius is 286 km for dry conditions and 330 km for wet conditions.

While transportation distances are significantly shorter in warm regions, CCS is still more competitive in these areas than in cold ones. This is due to a series of efficiency declines in DAC, including less adsorption capacity, higher energy regeneration, etc. Humidity significantly affects the performance of DAC, particularly in warm regions, which is explained by greater water loading at the adsorbents (air) [35].

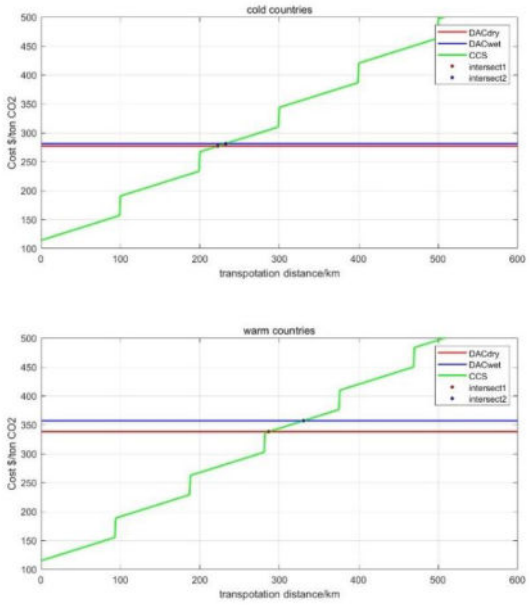


Figure 6 DAC-CCS comparison by transportation distance

Ship-based carbon dioxide transport is also considered to broaden the range of competitive radius. The shipping cost is considered less effective by distance, which is 20 $/ton of CO2. [36] This suggests that the previous analysis remains relevant for the radius centred on the coastline, with an upward adjustment of $20 per ton. Additionally, the distance should consider both the location of the CCS facility relative to the port and the distance from the port to the storage site.

However, in cold areas, seasonal freeze-up may cause the requirement of significant enlargement of CO2 temporary storage tanks. Figure 7 shows the relationship between port frozen time and LC. For the starting point, the pipeline's total length is assumed to be 200km with one pump station. The tank's cost is calculated with the following equation [37]: each can store $1000m^3$ CO2, and the cost is $252000. Each step in the diagram represents the addition of one storage tank.

$$C_p = 475V^{0.51} \quad (11)$$

Two lines are plotted for low instability and high instability scenarios. High instability assumes that the tank's maximum storage capacity is 1.5 times the required amount of carbon dioxide. In contrast, low instability assumes a maximum storage capacity of twice the required amount. Instability reflects the risk tolerance level; for instance, in the case of a port that remains frozen for 11 months each year, an unexpected incident preventing the port from opening would require the preparation of an additional year's storage capacity.

For both cases, the CCS seems unsuitable for the area that may freeze more than half the year.

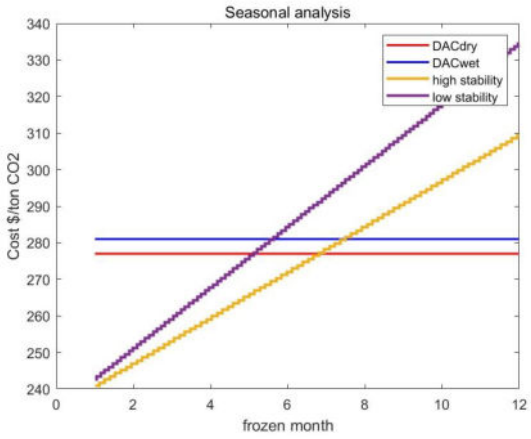


figure 7 DAC-CCS comparison by frozen seasons

5. Conclusion

This study has demonstrated the feasibility and challenges of implementing small-scale CCS systems for diesel power generators in remote areas. Analysing the levelized cost for the Meadowbank Mine highlights the complexities introduced by geographic remoteness, harsh climatic conditions, and the small scale of operation.

There are several constraints that limit the potential for integrating MEA-based CCS with small-scale, off-grid fossil fuel power plants. Some of these constraints are inherent, such as the higher costs associated with lower economies of scale. Others stem from methodological flaws, which could be addressed by adopting advanced technologies. For instance, MEA-based absorption significantly hinders the overall system's potential due to its high heat duty requirements and the unavoidable emission of additional CO2. Membranes could be a promising alternative for MEA-based CCS, offering compact designs and lower energy requirements.

Geospatial analysis shows that DAC is likely more economically favourable than CCS in colder regions. In principle, DAC should be integrated with specific renewable energy sources to exploit its flexibility, which can be constrained by geographical features[7]. The concept of ‘competitive radius’ shows its potential in combining both methods and providing an effective solution for decarbonising remote areas.

While small-scale CCS systems can be technically viable, their broader adoption will necessitate addressing significant economic and logistical barriers to make a real impact on decarbonisation efforts. How to implement policies and subsidies to support the deployment of CCS in remote and small-scale operations is still a point of interest.

References

[1] INSTITUTE, GLOBAL CCS. GLOBAL STATUS OF CCS REPORT 2023. 2024. 23 10 2024. <https://status23.globalccsinstit

[2] KazlouT., CherpA. and JewellJ. “Feasible Deployment of Carbon Capture and Storage and the Requirements of Climate Targets.” Nat. Clim. Chang. (2024): 1047-1055.

[3] CCUS projects database. 3 2024. https://www.iea.org/data-and-statistics/data-product/ccus-projects-database. 30 10 2024.

[4] ICT - Diesel Generator Market, By Power Rating (Low Power Generator, Medium Power Generator, and High Power Generator), By Application (Industrial, Commercial, and Residential), and By Region Forecast to 2032. 3 2023.

[5] Idaho National Laboratory. Microreactors could replace diesel generators, enable renewables, report says. 3 11 2021. https://inl.gov/nuclear-energy/microreactors-could-replace-diesel-generators-enable-renewables-report-says/. 30 10 2024.

[6] uprise energy. comparing Costs and Carbon Dioxide Emissions of Diesel Generators vs Wind Turbines. 29 11 2021. https://upriseenergy.com/blog/2021/11/4/environmental-impact-and-carbon-dioxide-offset. 30 10 2024.

[7] Budinis, S. (2024). Direct Air Capture - Energy System. [online] IEA. Available at: https://www.iea.org/energy-system/carbon-capture-utilisation-and-storage/direct-air-capture.

[8] www.agnicoeagle.com. (n.d.). Agnico Eagle Mines Limited - Home. [online] Available at: https://www.agnicoeagle.com/Home/default.aspx.

[9] Bilodeau, D.P., Badiu, R., McMullen, P. and Leetmaa, K. (2017). T[echnical Report on the Mineral Resources and Mineral Reserves at Meadowbank Gold Complex including the Amaruq Satellite Mine Development, Nunavut, Canada as at December 31, 2017. Agnico Eagle Mines Limited.

[10] Canada, Government (2024). CER – Market Snapshot: Canada is expanding its CO2 pipeline network. [online] Cer-rec.gc.ca. Available at: https://www.cer-rec.gc.ca/en/data-analysis/energy-markets/market-snapshots/2024/market-snapshot-canada-is-expanding-its-co2-pipeline-network.html [Accessed 31 Oct. 2024].

[11]Poling, B.E., Prausnitz, J.M. and John Paul O'connell (2007). The properties of gases and liquids. Boston: Mcgraw-Hill, Cop.

[12] Lewis, W.K. and Whitman, W.G. (1924). Principles of Gas Absorption. Industrial & Engineering Chemistry, 16(12), pp.1215–1220. doi:https://doi.org/10.1021/ie50180a002.

[13] Zhang, Y., Chen, H., Chen, C.-C., Plaza, J.M., Dugas, R. and Rochelle, G.T. (2009). Rate-Based Process Modeling Study of CO2 Capture with Aqueous Monoethanolamine Solution. Industrial & Engineering Chemistry Research, 48(20), pp.9233–9246. doi:https://doi.org/10.1021/ie900068k.

[14] Amirkhosrow, M. et al. (2021) ‘Rigorous rate-based model for CO 2 capture via monoethanolamine-based solutions: effect of kinetic models, mass transfer, and holdup correlations on prediction accuracy’, Separation Science and Technology, 56(9), pp. 1491–1509. Available at: https://doi.org/10.1080/01496395.2020.1784943.

[15] Madeddu, C., Errico, M. and Baratti, R. (2019). CO2 capture by reactive absorption-stripping : modeling, analysis and design. Cham, Switzerland: Springer.

[16] Onyebuchi, V.E., Kolios, A., Hanak, D.P., Biliyok, C. and Manovic, V. (2018). A systematic review of key challenges of CO2 transport via pipelines. Renewable and Sustainable Energy Reviews, 81, pp.2563–2583. doi:https://doi.org/10.1016/j.rser.2017.06.064.

[17] Lee, U., Yang, S., Jeong, Y.S., Lim, Y., Lee, C.S. and Han, C. (2012). Carbon Dioxide Liquefaction Process for Ship Transportation. Industrial & Engineering Chemistry Research, 51(46), pp.15122–15131. doi:https://doi.org/10.1021/ie300431z.

[18] Duan, L., Chen, X. and Yang, Y. (2012). Study on a novel process for CO2compression and liquefaction integrated with the refrigeration process. International Journal of Energy Research, [online] 37(12), pp.1453–1464. doi:https://doi.org/10.1002/er.2951.

[19] Witkowski, A., Majkut, M. and Rulik, S. (2014). Analysis of pipeline transportation systems for carbon dioxide sequestration. Archives of Thermodynamics, 35(1), pp.117–140. doi:https://doi.org/10.2478/aoter-2014-0008.

[20] https://www.weather-atlas.com/en/canada/baker-lake-climate

[21] Anon, (1983). International Code for the Construction and Equipment of Ships Carrying Liquefied Gases in Bulk (IGC Code).

[22] Canada, (2014). Marine Investigation Report M14C0219 - Transportation Safety Board of Canada. [online] Tsb.gc.ca. Available at: https://www.tsb.gc.ca/eng/rapports-reports/marine/2014/m14c0219/m14c0219.html [Accessed 5 Nov. 2024].

[23] Seo, Y., Lee, S., Kim, J., Huh, C. and Chang, D. (2017). Determination of optimal volume of temporary storage tanks in a ship-based carbon capture and storage (CCS) chain using life cycle cost (LCC) including unavailability cost. International Journal of Greenhouse Gas Control, 64, pp.11–22. doi:https://doi.org/10.1016/j.ijggc.2017.06.017.

[24]Rosen, M.A. and Dincer, I. (1999). Exergy analysis of waste emissions. International Journal of Energy Research, 23(13), pp.1153–1163. doi:https://doi.org/10.1002/(sici)1099-114x(19991025)23:13%3C1153::aid-er545%3E3.0.co;2-y.

[25] Beckman, W.A. & Duffie, J.A. (1991) Solar Engineering of Thermal Processes. 2nd ed. National Renewable Energy Laboratory (NREL). Available at: https://www.nrel.gov/docs/legosti/old/5173.pdf

[26] Øi, L.E., Anuththara, M., Kermani, S.H., Mirzapour, M., Shamsiri, S. & Karunarathne, S. (2023) Process Simulation and Cost Optimization of CO_2 Capture Configurations in Aspen HYSYS. Presented at SIMS 64, Västerås, Sweden, 26–27 September. University of South-Eastern Norway.

[27] Øi, L.E., Eldrup, N., Adhikari, U., Bentsen, M.H., Badalge, J.L. and Yang, S. (2016). Simulation and Cost Comparison of CO2 Liquefaction. Energy Procedia, 86, pp.500–510. doi:https://doi.org/10.1016/j.egypro.2016.01.051.

[28] Zhang, J., Xin, Q., Li, X., Yun, M., Xu, R., Wang, S., Li, Y., Lin, L., Ding, X., Ye, H. and Zhang, Y. (2019). Mixed matrix membranes comprising aminosilane-functionalized graphene oxide for enhanced CO2 separation. Journal of Membrane Science, 570-571, pp.343–354. doi:https://doi.org/10.1016/j.memsci.2018.10.075.

[29] Ali, A., Pothu, R., Siyal, S.H., Phulpoto, S., Sajjad, M. and Thebo, K.H. (2019). Graphene-based membranes for CO2 separation. Materials Science for Energy Technologies, 2(1), pp.83–88. doi:https://doi.org/10.1016/j.mset.2018.11.002.

[30] Yun, Y. (2017). Recent advances in carbon capture and storage. Rijeka, Croatia: Intech.

[31] Robeson, L.M. (2008). The upper bound revisited. Journal of Membrane Science, 320(1-2), pp.390–400. doi:https://doi.org/10.1016/j.memsci.2008.04.030.

[32] Luis, P., Van Gerven, T. and Van der Bruggen, B. (2012). Recent developments in membrane-based technologies for CO2 capture. Progress in Energy and Combustion Science, 38(3), pp.419–448. doi:https://doi.org/10.1016/j.pecs.2012.01.004.

[33] Bui, M.; Adjiman, C. S.; Bardow, A.; Anthony, E. J.; Boston, A.; Brown, S.; Fennell, P. S.; Fuss, S.; Galindo, A.; Hackett, L. A.; Hallett, J. P.; Herzog, H. J.; Jackson, G.; Kemper, J.; Krevor, S.; Maitland, G. C.; Matuszewski, M.; Metcalfe, I. S.; Petit, C.; Puxty, G.; Reimer, J.; Reiner, D. M.; Rubin, E. S.; Scott, S. A.; Shah, N.; Smit, B.; Trusler, J. P. M.; Webley, P.; Wilcox, J.; Mac Dowell, N. Carbon Capture and Storage (CCS): The Way Forward. Energy Environ. Sci. 2018, 11 (5), 1062–1176. https://doi.org/10.1039/C7EE02342A.

[34] Sendi, M., Bui, M., Mac Dowell, N. and Fennell, P. (2022). Geospatial analysis of regional climate impacts to accelerate cost-efficient direct air capture deployment. One Earth, 5(10), pp.1153–1164. doi:https://doi.org/10.1016/j.oneear.2022.09.003.

[35] S. (2024). Shipping and offshore pipeline transportation costs of CO2 by distance – Charts – Data & Statistics - IEA. [online] IEA. Available at: https://www.iea.org/data-and-statistics/charts/shipping-and-offshore-pipeline-transportation-costs-of-co2-by-distance?utm_source=chatgpt.com [Accessed 29 Nov. 2024].

[36] Towler, G. and Sinnott, R. (2008). Chemical engineering design : principles, practice and economics of plant and process design. Amsterdam ; Boston: Elsevier/Butterworth-Heinemann.

Enhancing and Optimising TiO_2's Photothermal Properties as a Semiconductor for Clean Water Production

Emaad Khan and Sohaib Elyamani

Department of Chemical Engineering, Imperial College London, U.K.

Abstract (200 words).

The objective of this research is to develop an ideal semiconductor for producing clean water via seawater desalination or sewage treatment by enhancing the currently suboptimal photothermal properties of titanium (IV) dioxide (TiO_2). In its anatase phase, TiO_2 has a relatively wide bandgap of approximately 3.05 eV, limiting its light absorption to the ultraviolet light range. Moreover, the effects of radiative electron-hole recombination weaken its thermal properties, as energy is lost as photons, while the beneficial non-radiative Shockley-Read-Hall and Auger recombination producing thermal and kinetic energy respectively is limited. To maximise the photothermal efficiency of TiO_2, these limitations must be overcome. Ball milling mechanically modifies TiO_2 by reducing its particle size and increasing its surface area for greater light absorption and adds defects in the form of oxygen vacancies to its structure. These defects result in charge carrier separation, reducing radiative electron-hole recombination while enhancing beneficial non-radiative recombination and aiding in dopant dispersion. Doping with copper (II) oxide (CuO) narrows the effective bandgap by creating mid-range impurity levels, which results in a greater spectral absorption range and more beneficial than detrimental electron-hole recombination. By developing a stable and effective accompanying heating system and investigating other factors like glass magnification and nitrogen milling, TiO_2 shows great potential in its enhanced photothermal efficiency.

1. Introduction:

With every passing day, the importance of developing technology in the sustainability sector grows more and more significant across various fields, including environmental protection and reclamation, renewable energy, and waste management. However, in a world where the global population is on the rise and freshwater sources continue to dwindle, one of the most important areas where development is crucial and global demand is high is in clean water production. Titanium (IV) dioxide (TiO_2) is emerging as a promising semiconductor material that could potentially serve in both sewage treatment and seawater purification because of its natural photothermal properties and its environmental friendliness and high chemical stability. Unfortunately, TiO_2 must be improved and optimised before it can be used, on account of its wide band gap and low visible-light absorption.

One factor to improve the mechanical properties of TiO_2 is via ball milling. This mechanical process induces structural modifications in order to reduce the particle size and improve the surface area to volume ratio of the powder. In this way, photothermal activity is improved by increasing the number of available reaction sites, thereby promoting light absorption. Another factor worth exploring is the doping of TiO_2 with transition metals, namely copper (Cu). Copper doping narrows the effective semiconductor bandgap by introducing new energy levels within the gap, discouraging electron-hole recombination and enabling light absorption in a greater spectral range.

The primary aim of this research is to investigate the impacts of the combination of ball milling and copper doping to optimise the photothermal properties of TiO_2. It is important to characterise these induced changes as well for study and analysis, utilising advanced techniques like ultraviolet-visual (UV-vis) spectroscopy, X-ray diffraction (XRD), and scanning electron microscopy (SEM). By analysing and characterising the absorption and morphological properties of the improved semiconductor, optimal synthesis conditions can be identified to maximise TiO_2's performance in the clean water production fields. This research can provide further insights and direction into the optimisation of future semiconductor designs, serving as a guideline for enhancing other photothermal materials, and understanding the implications of these results when considering technology scale-up.

2. Background:

2.1 TiO_2 uses and limitations

TiO_2 is a semiconductor of interest due to its favourable properties for use towards clean water production. Being non-toxic and stable both chemically and biologically, it is a safe and abundant material to use. Additionally, its resistance to corrosion and insolubility in water make it ideal as an environmentally safe material. On top of being a photothermal semiconductor to convert light into heat, its photocatalytic mechanism involves the generation of reactive oxygen species under UV light. This prompts the degradation of organic compounds and pollutants in water, which could be utilised to treat sewage or desalinate seawater.

Despite its promising materials, TiO_2 on its own is not suitable for this task [1]. Its bandgap is relatively wide at approximately 3.05 eV (for its anatase phase), which limits its light absorbing activity to only the UV light range, which just makes up less than 5% of total solar radiation. Moreover, losses to photothermal efficiency due to radiative electron-hole recombination is prevalent in semiconductors. This makes it necessary to adjust the mechanical and photothermal properties of TiO_2 to enable it to absorb light in the far more abundant visible-light range [2], and to prevent these radiative losses, through processes such as ball milling and copper oxide doping.

2.2 Ball Milling

Ball milling is a cost-effective and mass production feasible technique to induce structural modifications within the material including reducing the particle size, increasing the surface area to volume ratio, and introducing defects to the crystal lattice arrangement, giving the powder an amorphous quality [10]. Ball milling involves a transfer of energy from colliding balls within the high-strength ceramic holder to the material causing localised stress and deformation. Through ball milling, defects can be introduced in the TiO_2 structure known as oxygen vacancies essentially leaving behind positive charge sites attracting two free electrons and trapping them so localized states can be created within the bandgap typically closer to the conduction band rather than the valence band. Creating oxygen vacancies through ball milling increases the number of energy levels within the band gap allowing the material to absorb light at longer wavelengths hence a greater portion of the solar spectrum. Furthermore, as we introduce more defects into the material structure the concentration of Ti^{4+}ions differ, essentially having a positive effect on the photothermal properties of titanium dioxide again allowing a greater portion of sunlight to be absorbed and a greater light to thermal energy conversion.

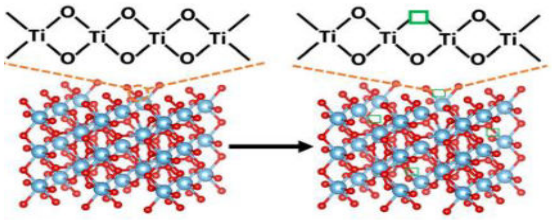


Figure 1: *Oxygen Vacancy in TiO_2 nanostructure*

Because of these defects introduced into TiO_2, ball-milled TiO_2 particles have been found to display higher photocatalytic efficiency due to the defects exposing more reactive sites [3], and in the same way the photothermal efficiency also increases, as heat is more readily released. Moreover, this process can also be beneficial to compliment effective doping, as ball milling promotes a greater dispersion of doping materials within TiO_2, so that the resulting powder is still homogeneous [4].

2.3 TiO_2 phases

Titanium dioxide (TiO_2) can exist in various polymorphic variants including a rutile phase, a phase that is thermodynamically stable under normal conditions, an anatase and brookite phase, a phase that is metastable at low temperatures and pressures, and a high-pressure titanium dioxide phase called srilankite or columbite which can be formed from anatase and brookite under high pressures [12]. TiO_2 naturally exists in anatase, rutile, and brookite form but through high-energy ball milling high temperatures and pressures can be achieved which induces a phase transformation in TiO_2 from the anatase phase to srilankite and rutile phases decreasing the band gap as shown in the figure below:

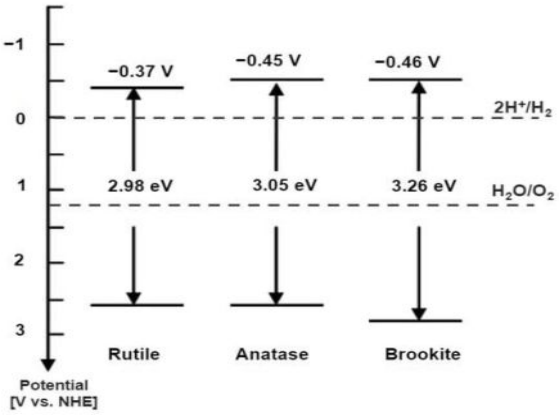


Figure 2*: Bandgaps of TiO_2 phases [5]*

The anatase crystalline structure has a better photothermal activity than the other TiO_2 phases due to its large band gap enhancing TiO_2's oxidization ability and electron transfer. Anatase also has more active sites than rutile due to its significantly larger specific surface area essentially enabling more surface area for photon absorption, converting more light to heat. However, even though the anatase phase has better photothermal activities when compared to the rutile phase, the anatase phase has a low thermodynamic stability so it can only be synthesized under specific conditions; none of which can be controlled in our experiments. Furthermore, the rutile phase absorption of UV light near the visible region makes it an attractive polymorph for TiO_2 photothermal applications. Since rutile is thermodynamically more stable than anatase it can be easily synthesized at low temperatures of 50 °C through hydrolysis of aqueous $TiCl_4$ ethanol solution [13].

2.4 Copper Oxide Doping

The main drawback of pure TiO_2 is its wide bandgap, meaning it can only absorb UV radiation which only accounts for ~5% of all solar radiation. To extend the proportion of sunlight absorbed, a CuO dopant is introduced into the TiO_2 structure to help reduce the band gap by creating defects and d-band states of CuO in TiO_2 [6][14]. Similarly to oxygen vacancy formation in undoped TiO_2, these d-band states create localised mid-ranged energy levels within the band gap of TiO_2 with Cu^{2+} ions creating electron trap centres that reduce radiative electron-hole recombination, a process where excited electrons (in the conduction band) drop down to the empty spaces or "holes" located in the valence band, releasing energy as photons and diminishing the material's photothermal capabilities, as shown on Figure 3. At the same time, this promotes the more beneficial non-radiative electron-hole recombination, including Shockley-Read-Hall and Auger recombination. These phenomena produce thermal energy and kinetic energy (as a phonon) respectively, both of which aid the photothermal efficiency of the semiconductor. Reducing radiative electron-hole pair recombination through electron trap centres allows the material to retain its energy and enhance its ability to convert light to heat. Additionally doping TiO_2 with CuO narrows the band gap and improves the charge carrier separation of photo-generated electron-hole pairs: pairs of electrons and holes formed when TiO_2 absorbs photons with an energy equal to or greater than its bandgap. This allows excited electrons to remain free in the conduction band making it less likely for these electrons to combine with the holes in the valence band reducing the (e−)/(h+) recombination rate in TiO_2 and promoting a greater range of the solar spectrum absorbed.

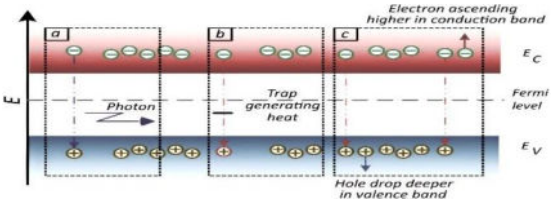


Figure 3: *Radiative electron-hole recombination (Left), Shockley-Read-Hall recombination (Middle), Auger recombination (Right) [7]*

3. Methodology:

3.1 Ball Milling

The ball milling process was performed in a planetary ball mill in air where 500 mg of P25 (from ALDRICH) TiO_2 powder consisting of 75% anatase and 25% rutile was weighed on a mass scale and placed in a 45 ml zirconium oxide ceramic holder containing ZrO_2 balls of 5 mm diameter and milled under a rotation speed of 500rpm for different milling times varying discretely from 1 to 7 hours. The TiO_2 is then collected and stored in labelled small vials to further test its photothermal behaviour. Before each run, the zirconium oxide ceramic holder and balls must be cleaned using sand and deionized water. The ball milling process has three main steps: preparation of the sample, the ball milling process, and collection of the sample. Each step has several parameters that should be evaluated to determine the optimal milling time as shown in the schematic below:

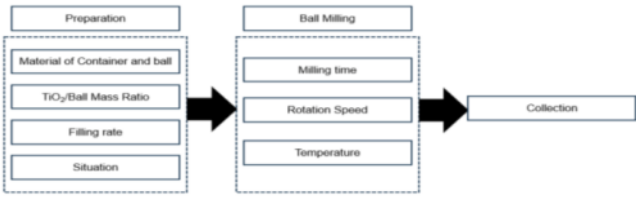


Figure 4: *Ball milling steps and parameters*

For preparation of the sample, four factors must be considered when determining the optimal milling time: the material of container and ball, the TiO_2/ball mass ratio, the filling rate, and the situation.

The material of the container and balls in this experiment are both made from zirconium oxide (ZrO_2) a ceramic material that has a high abrasion resistance, a high resistance to crack propagation, and excellent grinding qualities perfect for grinding coarse TiO_2 powders into smaller particles. Here, the material of the balls and

container are kept the same to avoid any contamination of the particles due to any wear or abrasion leading to unwanted heterogeneity in the final product.

The mass ratio between TiO_2 and the balls must be controlled for each ball mill. The greater the powder-to-ball mass ratio is the slower the kinetics, therefore less force applied to the TiO_2 powder. To optimise the kinetics of the process a mass ratio of 1:100 was considered to add as much force to the TiO_2 as possible and avoid excessive wear and heat generation. A ball-to-powder ratio of 1:100 corresponds to a powder mass of 500 mg for 50 g balls.

The third factor in the preparation step is the filling rate which refers to the volume of powder and balls occupied over the total internal volume of the container. The filling rate is vital in determining optimal milling time as too high a filling rate can reduce the motion of the balls, lowering milling efficiency and too low a filling rate can result in inefficient energy transfer and slower milling progress. Therefore, a standard filling rate of 50% was considered, as this provides enough energy to introduce wear into the material and gives enough volume for greater motion of the balls.

The final factor is the situation or environment, which refers to the atmosphere that the material is milled in. Typically milling in nitrogen is preferred as this creates an inert atmosphere preventing TiO_2 from oxidising further which encourages the formation and stabilisation of oxygen vacancies, useful in improving the electronic properties of TiO_2 including enhancing its photothermal applications, especially under visible light.

For ball milling, three factors must be considered when determining the optimal milling time: the milling time, the rotation speed, and the temperature.

Determining the most efficient milling time is a key objective in this objective. Milling time plays a key role in the size reduction and structural change of TiO_2 due to high energy impacts and shear forces. However long milling times can saturate the thermal efficiency of the sample either by bringing in contaminants due to wear and abrasion of the ZrO_2 balls which may cause degradation to the product or heat generation due to frictional impacts which may cause thermal effects to the temperature-sensitive TiO_2.

The temperature inside the container also plays a key part in determining the optimal milling time. As milling time increases, temperature increases due to frictional forces acting on the material causing TiO_2 particles to agglomerate rather than break apart. Therefore, it's vital that the temperature within the container is not allowed to increase too much to prevent this phenomenon.

The final factor for the milling process is the rotational speed which determines the revolutions per minute (rpm) of the ball. The rotational speed must be sufficiently high so that there's sufficient grinding between the ball and powder but not too high so that the balls are not pressed in the inner walls due to high centrifugal forces. Thus, the optimal rotational speed for a dry mill is 50% of the critical speed which is given by Equation 1 below:

$$\mathrm{Nc} = \frac{1}{2\pi}\sqrt{\frac{g}{R-r}}$$

Equation 1: *Critical ball milling speed [8]*

Where Nc is the critical speed (rpm), g is the gravitational acceleration (m/s^2), R is the inner radius of the container (m) and r is the radius of the balls (m). Here, a constant rotational speed of 500 rpm was used.

3.2 Photothermal behavior of samples

The photothermal behavior of the milled and unmilled samples were determined by conducting evaporation experiments under a 1000W Xenon lamp setup for different devices in which a temperature and mass change was recorded after one hour, allowing for a light-to-heat efficiency to be calculated using the equation below:

$$\eta = \frac{cm_l\Delta T + \Delta m_g H}{pst}$$

Equation 2: *Photothermal Efficiency [9]*

Where η is the thermal efficiency, c is the specific heat capacity of the material (J/g°C), m_l is the mass of liquid (g), ΔT is the temperature change (°C), Δm_g is the vapour mass change (g), H is the latent heat of vaporisation (J/g), p is the power density of light (W/m^2), s is the radiation area (m^2) and t is the radiation time (s).

The temperature increase ΔT is measured by a thermocouple, while the mass of water evaporated from the surface of filter paper m is recorded using a mass balance. The mass and temperature change are recorded over an hour, and then subsequent graphs of temperature change and mass of evaporation can be plotted against time. From these data recordings, the evaporation rate and thermal efficiency can be calculated using Equation 2.

The devices to be tested include a classical system where 50 mg of the sample is poured directly into a 500 ml beaker of tap water, a direct contact system where the powder is deposited onto a 4x8 cm filter paper and then placed directly on top of the beaker of water (held aloft by a filter paper mesh), and an isolated system where a 6x5 cm cut-out polystyrene device was created, allowing for the filter paper to be deposited inside it. This is then examined under light equivalent to 1 sun, ensuring the filter paper end remains in contact with the water. Efficiencies of all three systems were compared for a 1-hour milled sample, and the best system would proceed through to test the other milled (2-7 hours) and copper oxide doped samples where similarly the efficiencies of each run were calculated using Equation 2.

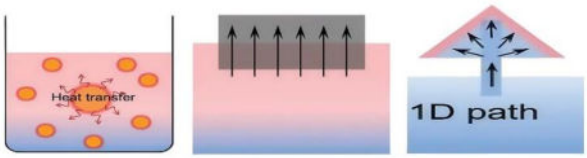


Figure 5: *Classical heating system (Left), Direct contact system (Middle), Isolated confined path system (Right) [10]*

3.3 Nitrogen-milled sample

A milled nitrogen sample was also analysed to compare the efficiency of milling the TiO_2 powder in air and in N_2. To prevent air from entering the high-strength ceramic holder, preparation of the powder in the ball mill was conducted in an inert nitrogen-filled glove box chamber, preventing any air from seeping into the container.

3.4 Filter paper preparation

Correct preparation of the filter paper is critical in supporting the homogeneous TiO_2 film and transporting water to the measuring equipment. To prepare a sample for testing, 50 mg of the material is weighed on a mass scale and then placed in a glass vial where 5 ml of deionized water is added through a calibrated pipette, creating a suspension of TiO_2 and water. This suspension must subsequently be well-mixed in a sonicator for 10 minutes to ensure a homogeneous suspension free of any clumps of material. Once the suspension is sonicated, it is poured onto an 8x4 cm filter paper through a glass beaker placed atop an O-ring of 3.5 cm diameter. This deposits a homogeneous ring of powder on the filter paper which must be dried on a hot plate for 20 minutes at 70 °C, allowing the water to evaporate slowly to prevent any bubble formation. Once the filter paper has dried, testing the sample's photothermal efficiency can be conducted, as mentioned in section 3.2.

3.5 Copper Oxide (CuO) Doping

To maximise the efficiency of light absorption and light-to-heat conversion copper oxide doping was introduced into TiO_2 powder at different weight percentages, starting from 5% to 25%. Similarly to the undoped samples light-to-heat efficiencies were calculated using Equation 2, and its morphology was characterised through different spectroscopic techniques mentioned in section 2.2.

3.6 Characterisation

Characterisation of the milled and doped powders was examined through XRD, SEM, and UV-Vis spectroscopy.

XRD characterisation of the milled, unmilled and CuO doped TiO_2 will show any possible phase transformations between the anatase and rutile and the intensity of the peaks defects are introduced. SEM characterisation will show the particle size reduction for greater milling times and CuO doping, and UV-vis characterisation will show the change in absorbance and reflectance thus determining if more light has been absorbed.

4. Results:

4.1 Comparison of systems

The first aim of this project was to establish which system between the isolated confined path system, direct contact, and classical heating is more efficient. To compare the efficiencies of the different systems the temperature changes of 1 hour milled TiO_2 in air were measured giving the following results:

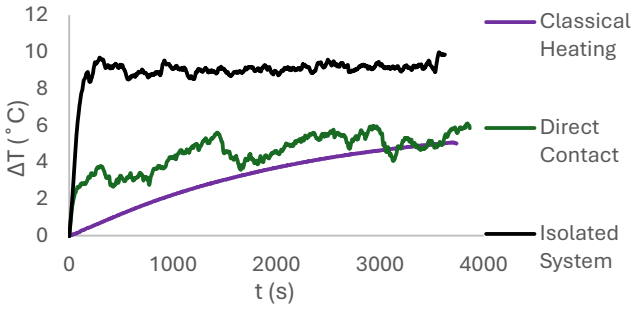


Figure 6: *Temperature change for each heating system*

The temperature change increase for the classical heating system took much longer than the other two systems, indicating a low photothermal efficiency due to significant heat loss to the environment through convection as the

entire bulk water gets heated, rather than just a targeted area.

The direct contact system gave better efficiencies as the heat generated was transferred to a target area of the water on the surface rather than the bulk water so only water on the surface absorbed heat and evaporated. However, preventing convection between the cold water at the bottom and the heated water on the surface was impossible, leading to slight inefficiency.

To prevent this phenomenon, adding an insulator such as polystyrene will trap as much heat thus the generated heat was only used to evaporate the water on the surface leading to a greater temperature increase than the other two models. Since the isolated system gave the greatest efficiency, it was used to optimise the light-to-heat efficiency of ball milling and CuO doping.

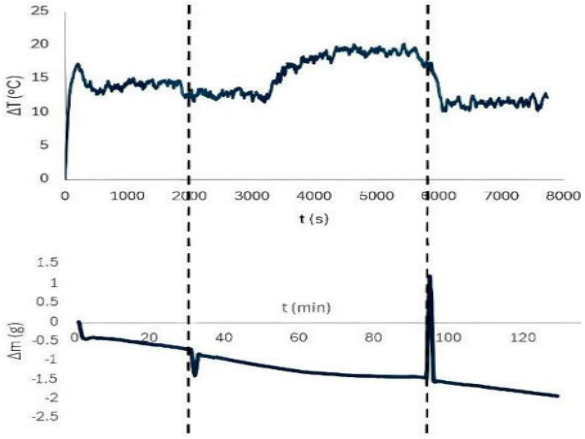


Figure 7: *Temperature change (Above) and Mass change (below)*

Another test corroborating the effectiveness of the confined path system is measuring the temperature and mass changes over time when the submerged end of the filter paper is removed, left to dry, and then placed back into the water. As shown in Figure 7, initially the experiment ran smoothly, and there is a clear peak on the mass graph when the paper was removed at approximately 30 minutes. Then, for just over an hour, the paper was allowed to dry under the lamp, and after 16 more minutes, there is a distinct increase in temperature before it stabilises. This is because all the water on the filter paper dries after 16 minutes of being out, and all subsequent heat energy goes exclusively towards heating the powder. Then, at 100 minutes, the paper end was submerged once again, and over the next 30 minutes, the temperature stabilised once again, and the evaporation rate remained relatively unaffected, demonstrating the stability of both the semiconductor and this heating system.

4.2 Comparison of milling times

Milling times were tested discretely from 1-7 hours, and its efficiencies were calculated to find the optimal milling time. The figures below show the temperature changes, mass of water evaporated from the surface of the filter paper, and the photothermal efficiency for each milling time:

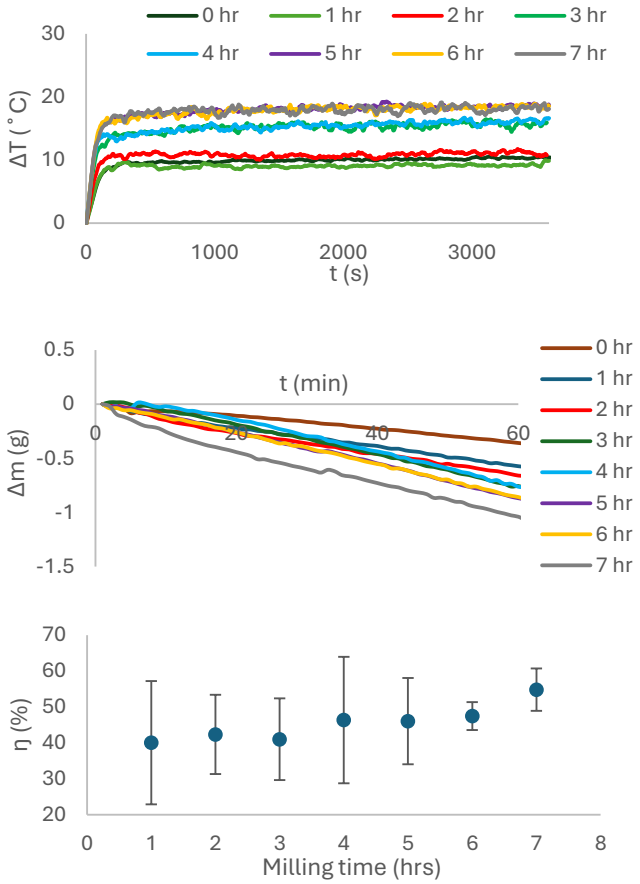


Figure 8: *Temperature change (Top), Mass change (Middle), Photothermal efficiency (Bottom) for each milling time*

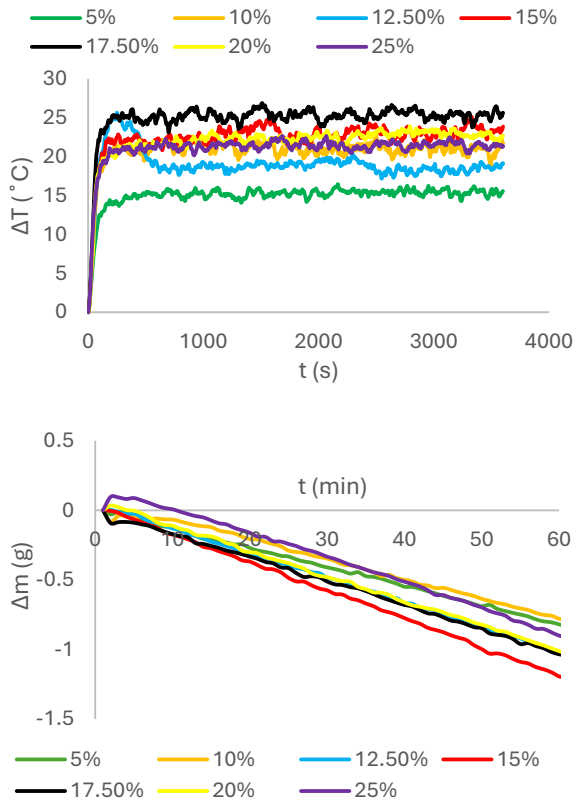
Here, a clear trend is shown where increasing milling times increases temperature increase and mass of water evaporated and thus the photothermal efficiency. Again, this is expected as increased milling times create more defects in the crystalline lattice of TiO_2 thus leading to reduced band gap and greater light absorption and light-to-heat conversion. Therefore, the optimal milling time was chosen to be 7 hours and consequently to be used in finding the optimal CuO doping percentage.

However, one limitation to this experiment was the unreliability of the results shown in the large error bars in Figure 8 due to the rate of water supply. A large water

supply means a greater rate of water flow from the bulk to the water surface to avoid a dry-out condition on the material that could reduce performance. However excessive water supply can increase convection between the colder bulk water and the warmer water surface, leading to energy losses and less evaporation. Also, the time the water is exposed with the material is reduced, leading to lower evaporation rates and lower efficiencies. Conversely, a low water supply rate means a smaller water flow from the bulk to the evaporation surface. This limits heat dissipation to the bulk water through convective heat transfer, so the energy is focused on evaporating the water from the surface, enhancing the evaporating rate and thus the photothermal efficiency. But the water supply rate must never be too low, as it may introduce dry spots on the material, causing overheating and material degradation.

4.3 Comparison of CuO doping %

Once the optimal milling time of 7 hours was established, CuO doping experiments were conducted to establish the optimal dopant wt%. CuO doping percentages increased discretely from 5% to 25% and each sample's evaporation rate, temperature increase, and light-to-heat conversion efficiency is measured and calculated:



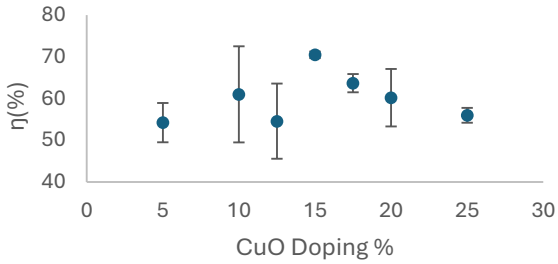


Figure 9: *Temperature change (Top), Mass change (Middle), Photothermal efficiency (Bottom) for each doping percentage*

From Figure 9, an optimal CuO doping of 15% gave photothermal efficiencies of 70% and water evaporation rates greater than 1.2 g/hr.

However excessive doping of CuO in TiO_2 can also have a negative impact on the materials performance due to compositional imbalances which may cause the composite's qualities to degrade. For example, despite CuO having a smaller band gap of 1.3 eV, excessive CuO doping can reduce the proportion of UV light absorbed by TiO_2 leading to lower light absorption. Additionally, TiO_2 with a high CuO content may cause CuO particles to agglomerate into bigger particles, decreasing the surface area available for water contact with the material and light absorption, which could result in uneven heating and thus lower efficiencies.

5. Characterisation

5.1 Ball milling XRD

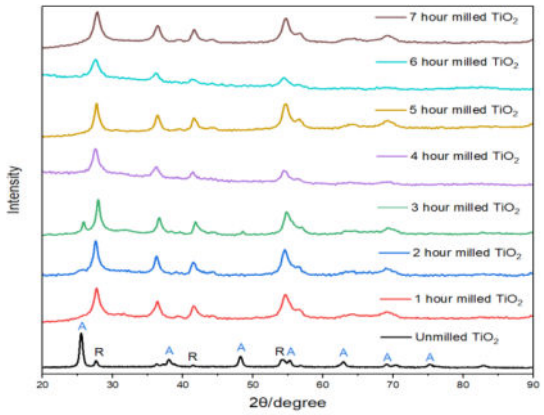


Figure 10: *XRD results of milled TiO_2 samples*

As expected, the XRD of the unmilled TiO_2 (black line) shows there are more anatase TiO_2 (A peaks) than rutile TiO_2 (R peaks). As the duration of ball milling increases the intensity of rutile peaks increases whilst the anatase peaks are not revealed anymore indicating a possible

phase change from the anatase phase to the rutile phase. Therefore, this indicates that the rutile phase is thermodynamically stable during ball milling and hence used in the evaporation measurement. As the milling times increase, the peaks broaden causing a slight decrease in the light intensity of the rutile peak. Another observation is the shift of peaks to higher angles as ball milling time increases from 1 hour to 7 hours corresponding to smaller d-spacing again indicating a potential phase transition from anatase to rutile [15].

5.2 CuO doping XRD

Figure 11 shows the XRD peaks of TiO_2-CuO (5%-25%):

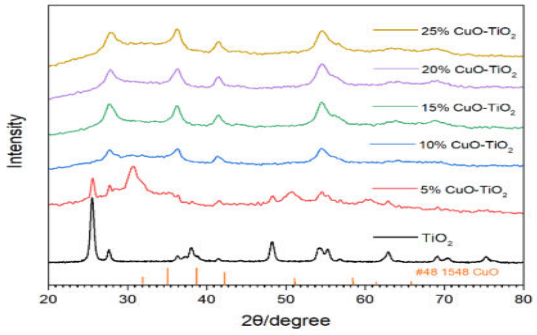


Figure 11: *XRD results of CuO doped TiO_2 samples*

The XRD for the 5% CuO doping (red line) shows the TiO_2-CuO composite was not mixed well due to clear peaks of TiO_2 and CuO. During ball milling, the phases of TiO_2 and CuO should change so a presence of any CuO peaks should not be present as seen in the 10% to 25% TiO_2-CuO composites.

5.3 Ball milling UV-Vis absorbance

Figure 12 illustrates the UV-Vis absorbance spectra of TiO_2 samples measured at different milling times.

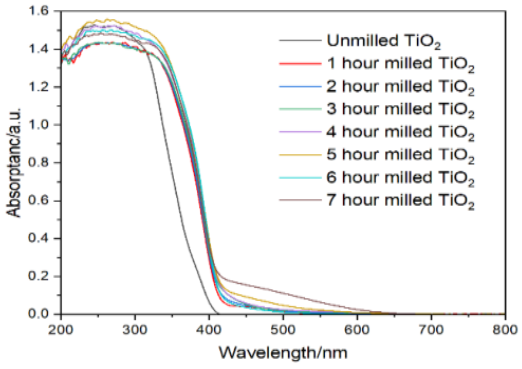


Figure 12: *UV-Vis absorbance results of milled TiO_2 samples*

The milled samples were able to absorb light in the 430-600 nm range corresponding to absorption of visible light indicating a bandgap decrease and formation of new energy levels with this bandgap. New energy levels provide various energy differences for electrons to absorb light so increasing the number of energy levels within a bandgap can cause electrons to absorb different wavelengths and thus more light can be absorbed. Combining this result with the ball milling XRD results, the rutile phase after ball milling absorbs more light compared with the unmilled TiO_2.

5.4 CuO Doping UV-Vis absorbance

Figure 13 shows the UV-Vis absorbance spectra of TiO_2-CuO samples measured at different doping percentages:

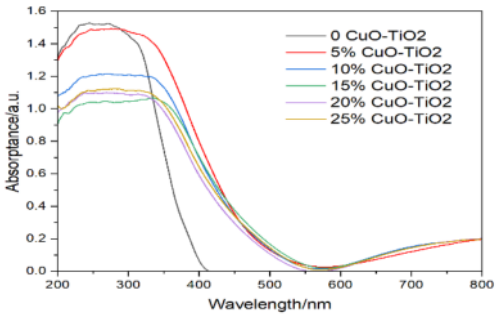
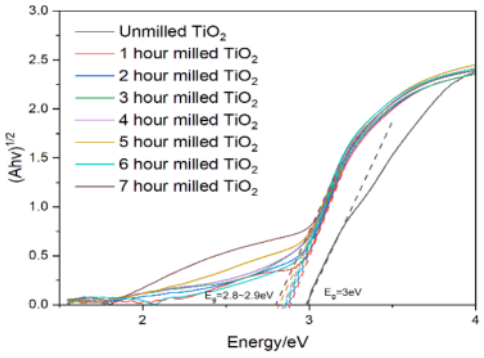


Figure 13: *UV-Vis absorbance results of doped TiO_2 samples*

As expected, the absorbance curve for 5% TiO_2-CuO shows a similar trend to the unmilled TiO_2 between 200-400nm wavelengths due to poor mixing of the composite. The doped CuO should decrease absorbance in this range and increase absorption greater than 400nm because of the CuO bandgap of 1.3 eV reducing the amount of UV radiation (λ~200-400nm) absorbed and increasing the amount of visible light absorbed.

5.5 Tauc plots

The bandgaps are calculated by extrapolating the linear portion of the tauc plot to the x-axis.



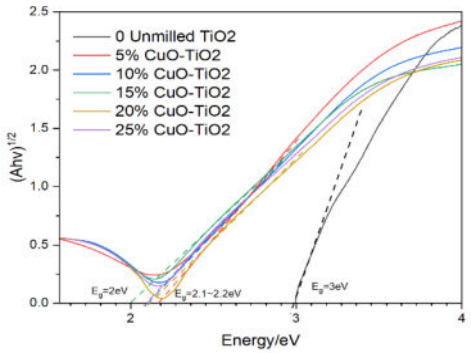


Figure 14: *Tauc plots of milled (Above) and doped (Below) TiO_2 samples*

Using these, the average bandgap of the milled TiO_2 is 2.8-2.9 eV whereas the bandgap for unmilled TiO_2 is 3.0 eV (rutile phase bandgap) indicating the presence of new energy levels within the bandgap. From Figure 14, the bandgap for the optimal 15% doped TiO_2-CuO at 7 hours is 2 eV, which is smaller than the other doped samples approving the claim that the 15% TiO_2-CuO is the optimal sample for absorbing light.

5.6 SEM Images

The SEM images for unmilled TiO_2, 7 hour milled TiO_2 and 15% CuO doped TiO_2 are shown below:

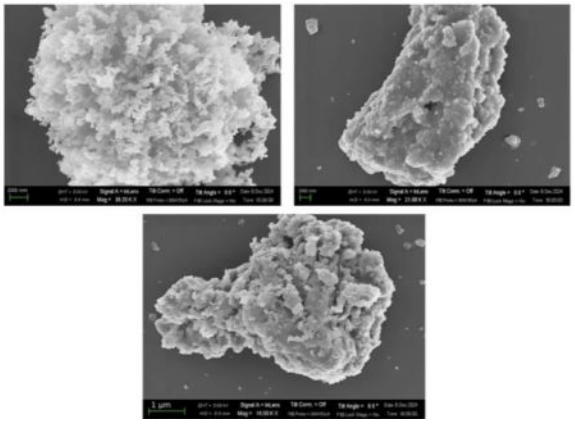


Figure 15: *SEM images of unmilled (Left), 7 hour milled (Right) and 15 wt% doped TiO_2 samples (bottom)*

The unmilled TiO_2 images look like clusters of small nanoparticles while the 7 hour milled TiO_2 and 15% doped TiO_2-CuO show no significant morphology differences between them due to the amorphous structure of TiO_2.

6. Outlook and Improvements:

6.1 Comparison of air and nitrogen

The efficiency of 1 hour milled TiO_2 was measured, calculated, and compared for air and nitrogen.

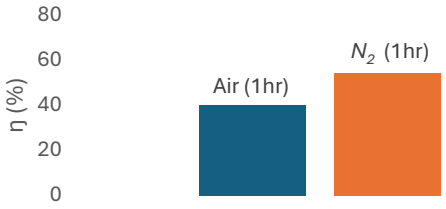


Figure 16: *Photothermal efficiencies of air milled vs N_2 milled TiO_2*

The results show that milling in nitrogen gave better photothermal behavior than when milling in air which is expected as fewer oxygen vacancies are displaced in the structure of TiO_2 creating more energy levels within its band gap thus reducing it. Despite greater photothermal efficiencies for milling in nitrogen, milling in air was chosen for ease and time-saving purposes, and to account for the effect of temperature over longer milling times.

6.2 Adding a glass lens

A modification to the isolated confined path system was tested, in which a circular glass lens was placed atop the polystyrene vessel to create a confined magnification of light onto the 7-hour milled powder. This data was then compared to the normal isolated system results for 7 hours, as shown in Figure 17 below:

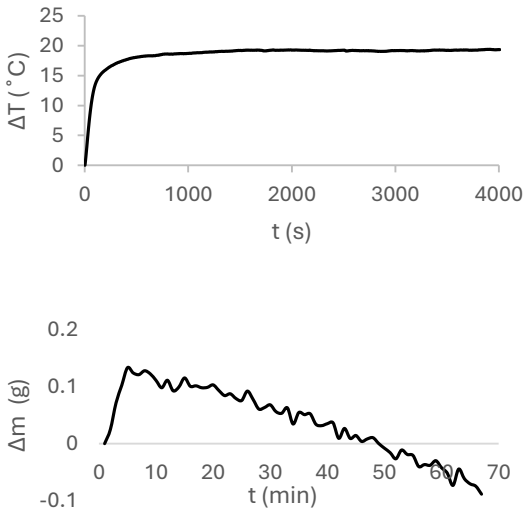


Figure 17: *Temperature change (Above), Mass change (Below)*

While the temperature change no longer fluctuates due to a lack of air flow, water vapour condensation causes a lessened impact on mass change, and the overall efficiency of this modified system is consequently lower.

7. Conclusions:

Synergistically combining these transformative techniques is crucial to augment TiO_2's structure and properties, as employing this double modification has been shown to enhance water mass evaporation and thermal stability than either technique on its own [16]. By narrowing the bandgap to ~2 eV to absorb a larger spectrum of light and minimizing detrimental radiative electron-hole recombination while maximizing beneficial non-radiative electron-hole recombination, TiO_2 shows potential as a lucrative semiconductor for solar-powered water treatment and desalination. Beyond just its photocatalytic abilities, this research demonstrates the viability in exploring the potential of magnifying TiO_2's photothermal conversion efficiency for sustainably producing clean water.

References:

[1]- Pelaez, M., Nolan, N.T., Pillai, S.C., Seery, M.K., Falaras, P., Kontos, A.G., Dunlop, P.S.M., Hamilton, J.W.J., Byrne, J.Anthony., O'Shea, K., Entezari, M.H. and Dionysiou, D.D. (2012). A review on the visible light active titanium dioxide photocatalysts for environmental applications. *Applied Catalysis B: Environmental*, [online] 125, pp.331–349.

[2]- Li, R., Li, T. and Zhou, Q. (2020). Impact of Titanium Dioxide (TiO2) Modification on Its Application to Pollution Treatment—A Review. *Semantic Scholar*

[3]- Kabuyaya Kighuta, Kim, S.-W., Hou, Y.-L., Lee, K.-P. and Kim, W.-J. (2023). Facile and Simple Post Treatment Ball Milling Strategy for the Production of Low-Cost TiO2 Composites with Enhanced Photocatalytic Performance and Applicability to Construction Materials. *Materials*, [online] 16(14), pp.4931–4931

[4]- Ye, M., Pan, J., Guo, Z., Liu, X. and Chen, Y. (2020). Effect of ball milling process on the photocatalytic performance of CdS/TiO2composite. *Nanotechnology Reviews*, 9(1), pp.558–567.

[5]- A. Di Paola, Palmisano, L. and Palmisano, L. (2010). TiO2 photocatalysts prepared by thermohydrolysis of TiCl4 in aqueous solutions. *Elsevier eBooks*, pp.225–228.

[6]- Abbas, M.M. and Rasheed, M. (2021). Solid State Reaction Synthesis and Characterization of Cu doped TiO_2 Nanomaterials. *Journal of Physics: Conference Series*, 1795(1), p.012059.

[7]- Solís-Cisneros, H., Hernández-Gutiérrez, C.A., Anwar, A.-R., Sevilla-Camacho, P.Y., Camas-Anzueto, J.L., Grajales-Coutiño, R., Trejo-Hernández, R., Casallas-Moreno, Y.L. and L´ópez-López, M. (2022). Physical and technological analysis of the AlGaN-based UVC-LED: an extended discussion focused on cubic phase as an alternative for surface disinfection. *Revista Mexicana de Física*, 68(2 Mar-Apr).

[8]- www.pauloabbe.com. (n.d.). *Variables in Ball Mill Operation | Paul O. Abbe®*. [online]

[9]- Chen, L., Liu, J., Fang, X. and Zhang, Z. (2017). Reduced graphene oxide dispersed nanofluids with improved photo-thermal conversion performance for direct absorption solar collectors. *Solar Energy Materials and Solar Cells*, 163, pp.125–133.

[10]- Zhu, L., Gao, M., Connor and Ghim Wei Ho (2018). Solar-driven photothermal nanostructured materials designs and prerequisites for evaporation and catalysis applications. *Materials horizons*, 5(3), pp.323–343.

[11] - Kołodziejczak-Radzimska, A. and Jesionowski, T. (2014). Zinc Oxide—From Synthesis to Application: A Review. *Materials*, [online] 7(4), pp.2833–2881.

[12] - Valeeva, A.A., Sushnikova, A.A. and Rempel, A.A. (2024). Phase composition tuning by high-energy ball milling of titanium dioxide powders. *Inorganic Chemistry Communications*, 159, p.111727.

[13]- Eddy, D.R., Permana, M.D., Sakti, L.K., Sheha, G.A.N., Solihudin, Hidayat, S., Takei, T., Kumada, N. and Rahayu, I. (2023). Heterophase Polymorph of TiO2 (Anatase, Rutile, Brookite, TiO2 (B)) for Efficient Photocatalyst: Fabrication and Activity. *Nanomaterials*, [online] 13(4), p.704.

[14] - T. Perarasan, Peter, I.J., Kumar, A.M., N. Rajamanickam, Ramachandran, K. and Mohan, C.R. (2019). Copper doped titanium dioxide for enhancing the photovoltaic behavior in solar cell. *Materials Today Proceedings*, 35, pp.66–68.

[15] - Mariño-Gámez, A.E., Guillermo-E Acosta-González, M.I. Pech-Canul, M.B. Hernández, S. García-Villarreal, Zambrano-Robledo, P., Silvana, B. and J.A. Aguilar-Martínez (2021). Influence of high energy ball milling on structural, microstructural and optical properties of TiO2 nanoparticles. *Ceramics International*, 48(3), pp.3362–3367.

[16]- Zhao, Y., Li, Y. and Sun, L. (2021). Recent advances in photocatalytic decomposition of water and pollutants for sustainable application. *Chemosphere*, 276, p.130201.

Applications of the Adversarial Robust Real-Time Optimization and Control Algorithm to Multi-Compressor Problems

Boren Chen and Guanhao Lu
Department of Chemical Engineering, Imperial College London, U.K.

Abstract: Compressors play a crucial role in industrial applications, but their control remains challenging due to disturbances such as input flow fluctuations and mechanical wear. Conventional optimization approaches often prioritize set-point optimality without adequately addressing these disturbances, leading to inefficiencies and increased control effort. This study evaluates the performance of the Adversarial Robust Real-Time Optimization and Control (ARRTOC) algorithm, a robust optimization framework developed to model disturbances as adversarial inputs and generate set-points resilient to worst-case scenarios. Applied to a multi-compressor system, the evaluation focuses on the algorithm's flexibility across varying total flow rates, its robustness through comparisons of nominal and worst-case objective functions and individual compressor efficiency maps, and the trade-off between robustness and power consumption as influenced by the maximum exploration radius Γ. These findings highlight ARRTOC's potential to enhance both robustness and operational efficiency in compressor control systems, with promising applicability to other industrial processes requiring robust optimization.

Key word: Compressor control, Adversarial optimization, Operational robustness, Power consumption minimization

1. Introduction

Compressors are indispensable across a wide range of industries, including oil and gas, chemical manufacturing, power generation, and refrigeration systems. In the oil and gas sector, compressors facilitate the transportation of natural gas through pipelines by maintaining required pressure levels. In chemical manufacturing, they enable precise control of reaction conditions, ensuring product quality and efficiency (Robison, Arvada and Beaty, n.d.).

The versatility and criticality of compressors make their reliability and efficiency essential for minimizing operational costs and ensuring uninterrupted industrial processes. To achieve this, control systems calculate and implement set-points—desired operational targets that maintain optimal performance. These set-points are relayed to lower-level control layers, where the controllers are tasked with three main objectives: maintaining set-point accuracy, mitigating disturbances, and reducing noise (Seborg et al., 2017).

However, in real-world operations, disturbances such as input flow fluctuations, mechanical wear, and sensor inaccuracies often undermine steady performance. These disturbances create significant challenges in maintaining optimal set-points, which are typically calculated under idealized conditions. When disturbances occur, deviations from these set-points force the control layer to exert additional effort to restore balance (Jäschke, Cao and Kariwala, 2017). This can lead to inefficiencies, increased energy consumption, and potential risks to the control system's integrity. Addressing these challenges requires optimization methods that not only identify ideal set-points but also ensure their robustness under varying conditions.

The Adversarial Robust Real-Time Optimization and Control (ARRTOC) algorithm addresses these challenges by integrating adversarial optimization principles with real-time feedback mechanisms. Unlike conventional optimization algorithms, which assume disturbance-free conditions and prioritize optimality under narrow assumptions, ARRTOC explicitly accounts for uncertainties by modelling disturbances as adversarial inputs to the system (Ahmed, Rio-Chanona and Mercangoz, 2023). This approach ensures that the solutions produced are inherently robust against worst-case scenarios, making the algorithm particularly suited for compressor systems, where disturbances are ever-present and stable performance under fluctuating conditions is critical.

Despite the potential advantages offered by the ARRTOC algorithm, it remains a novel approach that has yet to be widely tested in practical applications (Ahmed, Rio-Chanona and Mercangoz, 2023). This paper provides a detailed investigation into the ARRTOC algorithm with a specific focus on its application to compressor control problems. Through this study, the algorithm's performance is comprehensively evaluated, demonstrating its capability to optimize compressor control with an emphasis on both energy efficiency and robustness under real-world conditions.

2. Methodology

2.1 Problem Definition

The nominal objective is to minimize the total power consumption of a process system of 3 parallel interconnected compressors, and the visualization of the process is shown in Fig. 1.

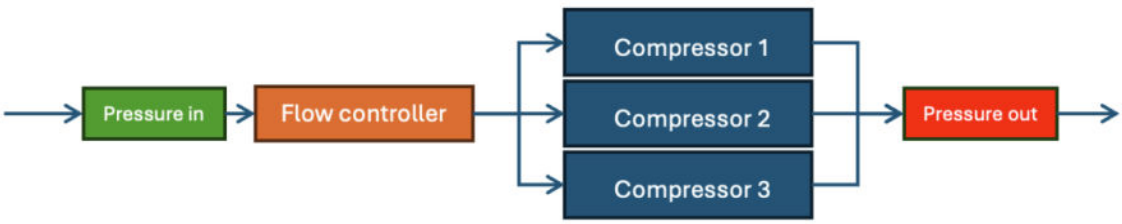


Figure 1. Simplified schematics of the multi-compressor problem

Each compressor operates under specific thermodynamic and efficiency conditions, and those properties are determined by the mass flowrates, polytropic head and efficiency at each compressor stream. Consequently, the total power consumption can be expressed by the following equation:

$$power = \sum_{i=1}^{N} \frac{y_{p,i} m_{comp,i}}{\eta_i} \tag{1}$$

where N is total number of compressors, $y_{p,i}$ and η_i is the polytropic head and efficiency for each compressor respectively, which can be calculated by the equations below:

$$y_{p,i} = \left(0.9 + \frac{8314 \cdot 293}{16.04}\right)\left(\frac{k}{k-1}\right)\left(p_{ratio,i}^{\frac{k-1}{k}} - 1\right), \quad k = 1.27 \tag{2}$$

$$\eta_i = c_{0,i} + c_{1,i} \cdot m_{comp,i} + c_{2,i} \cdot p_{ratio,i} + c_{3,i} \cdot m_{comp,i} \cdot p_{ratio,i} + c_{4,i} \cdot {m_{comp,i}}^2 + c_{5,i} \cdot p_{ratio,i}^2 \tag{3}$$

$$p_{ratio,i} = \begin{pmatrix} m_{comp,i}^2 \\ m_{comp,i} \\ 1 \end{pmatrix}^T \cdot \begin{pmatrix} pp_{1,i} \\ pp_{2,i} \\ pp_{3,i} \end{pmatrix} \tag{4}$$

From the above, all variables are becoming a function of $m_{comp,i}$, hence, the problem can be formulated as:

$$\text{Minimize } f \quad m_{comp,1}, m_{comp,2}, m_{comp,3}$$
$$f = \sum_{i=1}^{3} \frac{y_{p,i} \cdot m_{comp,i}}{\eta_i}$$
$$\text{Subject to:} \quad \sum_{i=1}^{3} m_{comp,i} = T_{flow} \tag{5}$$
$$65 \le m_{comp,1} \le 120$$
$$65 \le m_{comp,2} \le 120$$
$$65 \le m_{comp,3} \le 120$$

As outlined in the problem definition, this is a constrained non-linear programming (NLP) optimization problem. To address it, a constraint-penalty approach integrated with the ARRTOC algorithm is employed. Instead of solving the problem directly using the original objective function (Eq.5), a modified objective function (Eq.6) is used. This modified function incorporates a penalty term $M \cdot (T_{flow} - \sum_{i=1}^{3} m_{comp,i})^2$, which integrates the mass balance constraint into the objective function.

$$\text{Minimize } f \quad m_{comp,1}, m_{comp,2}, m_{comp,3}$$
$$f = \left(\sum_{i=1}^{3} \frac{y_{p,i}\, m_{comp,i}}{\eta_i}\right)^2 + M\left(T_{flow} - \sum_{i=1}^{3} m_{comp,i}\right)^2$$
$$\text{S. t.:} \quad \sum_{i=1}^{3} m_{comp,i} = T_{flow} \tag{6}$$
$$65 \le m_{comp,1} \le 120$$
$$65 \le m_{comp,2} \le 120$$
$$65 \le m_{comp,3} \le 120$$

2.2 Algorithm Overview

The ARRTOC algorithm is a data-driven optimization framework designed to iteratively refine decision variables for improved solutions. The algorithm systematically explores the search space, identifies improving directions through a Second Order Cone Program (SOCP), and calculates optimal step sizes to update the current iterate (Ahmed, Rio-Chanona and Mercangoz, 2023). Key components include neighborhood exploration, constraint handling, and robust optimization, and a high-level schematic of the framework is shown in Fig. 2.

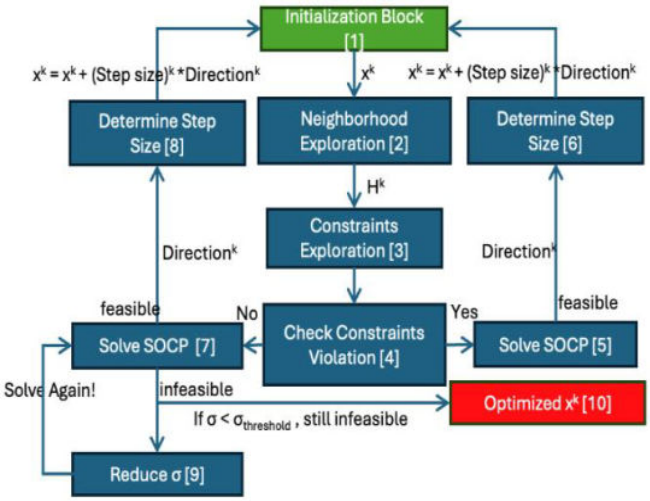


Figure 2. High-level schematic of the ARRTOC Framework

For the multi-compressor problem, since the mass balance constraint has already been incorporated into the objective function via the penalty term $M \cdot (T_{flow} - \sum_{i=1}^{3} m_{comp,i})^2$, the remaining constraints are the individual compressors' operating flow ranges. These are handled using Python's np.clip command to ensure all flow values remain within their specified bounds.

As a result, the constraint-handling blocks, [3], [4], [5], and [6] are not required for this application. The subsequent sections provide detailed explanations of the remaining algorithmic blocks and their roles within the ARRTOC framework.

2.3 Algorithm Blocks Explanations

2.3.1 Neighborhood Exploration [2]

Pseudo code EXPLORE_NEIGHBORHOOD (*f, perturbed_func,* $\mathbf{x^k}$*, Γ, dim,* H^{k-1})

INPUT: *f*: Objective function; *perturbed_func*: Function to be maximized during exploration; $\mathbf{x^k}$: Current point in the search space; *Γ*: Maximum exploration radius; *dim*: Dimensionality of the search space; H^{k-1}: Previous history set

OUTPUT: H^k: History set

Set $H^k \leftarrow H^{k-1}$ **if** $H^{k-1} \neq \emptyset$, **else** $\emptyset$.

For *i = 0* **to** 2 × *dim*:

1. **If** *i = 0*:
   - Set $\mathbf{\Delta x} \leftarrow 0$ (no perturbation).
2. **Else**:
   - Set *val* ← *1* **if** $i \leq dim$, **else** *-1*.
   - Compute $\mathbf{\Delta x} \leftarrow \frac{\Gamma}{2} e_{(i-1) \bmod dim}\, val$, where $e_{(i-1) \bmod dim}$ is the unit vector in the *[(i - 1) mod dim]-th* dimension.
3. Perform diminishing gradient ascent on *perturbed_func* starting from $\mathbf{\Delta x}$, ensuring:
   - A diminishing step size is used.
   - All function evaluations *(x, f(x))* during the ascent are recorded in H^k.

Return H^k

The first step of the algorithm, as outlined in the pseudocode above, is neighborhood exploration. In this phase, the algorithm systematically generates neighboring points around the current iterate $\mathbf{x^k}$, by applying controlled perturbations in each dimension of the search space.

Starting from these perturbations, diminishing gradient ascent is performed and the points along the search path are recorded in the history set H^k. This set serves as a critical input for solving the Second Order Cone Program (SOCP), which is crucial for identifying an improving direction and determine the optimal step size for updating $\mathbf{x^k}$, as discussed in Section 2.3.3 and Section 2.3.4

2.3.2 Visualization of the Exploration Process in 2D

The diagram below illustrates the neighborhood exploration process applied to a generic 2-dimensional objective function.

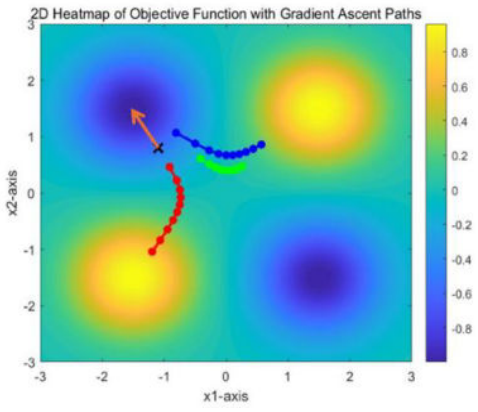


Figure 3. Visual illustration of gradient ascent search paths

Starting from the black cross $\mathbf{x^k}$, the algorithm generates three search paths, represented by red, green, and blue dotted lines. Gradient ascent explores these paths by evaluating points along the way and recording their positions and function values in the history set H^k.

The optimal descent direction from solving the SOCP is depicted by the orange arrow in the diagram.

2.3.3 Solve SOCP [7]

After obtaining the history set H^k, the next step is to construct M^k , a set of high-cost neighbors selected from H^k based on the criteria in Eq. 9.

The construction of M^k requires the calculation of σ, as defined in Eqs. 7 and 8. The points in M^k are then used to solve the Second Order Cone Program (SOCP), represented by the Minimization problem in Eq. 10. In this SOCP, ϵ is a small positive scalar set to 0.01, and $\mathbf{d_{cost}}$ represents the direction vector for the descent direction.

$$\sigma^1 = 0.2\,(\arg\max\,(H^1) - f(\mathbf{x^1})) \quad (7)$$

$$\sigma^k = \sigma^{k-1} \quad (8)$$

$$M^k = \{\mathbf{x} | \mathbf{x} \in H^k,\ f(\mathbf{x}) > \arg\max\,(H^k) - \sigma^k\} \quad (9)$$

If solving the SOCP with H^k is infeasible, σ is iteratively reduced following the update rule $\sigma = \frac{\sigma}{1.05}$, and M^k is reconstructed until the SOCP becomes feasible or σ reaches a predefined threshold. This iterative process ensures that $\mathbf{d_{cost}}$ is directed toward a feasible minimization of $\mathbf{x^k}$

$$\begin{aligned} &\text{Minimize } \beta \\ &\quad \mathbf{d_{cost}}, \beta \\ &\text{S.t.:} \\ &\|\mathbf{d_{cost}}\| \leq 1 \\ &\|\mathbf{d_{cost}}\| \leq 1 \\ &\mathbf{d_{cost}} \cdot \left(\frac{\mathbf{x_i} - \mathbf{x^k}}{\|\mathbf{x_i} - \mathbf{x^k}\|}\right) \leq \beta, \forall \mathbf{x_i} \in M^k \\ &\beta \leq -\epsilon \end{aligned} \quad (10)$$

2.3.4 Determine Step Size [8]

In the case of a feasible SOCP solution, the descent direction $\mathbf{d_{cost}}$ is found, and the next step is to find the optimal step size ρ_{best} by solving the following minimization problem:

$$\begin{aligned} &\text{Minimize } \rho_{best} \\ &\quad \rho \\ &\text{S.t:} \\ &\rho_{best} = \arg\min\,(\rho) \\ &\rho \geq (\mathbf{x_i} - \mathbf{x^k}) \cdot \mathbf{d_{cost}} \\ &\quad + \sqrt{((\mathbf{x_i} - \mathbf{x^k}) \cdot \mathbf{d_{cost}})^2 - \|\mathbf{x_i} - \mathbf{x^k}\|^2 + \Gamma^2} \\ &\forall \mathbf{x_i} \in M^k \end{aligned} \quad (11)$$

Once ρ_{best} is obtained, $\mathbf{x^k}$ is updated using Eq. 12.

$$\mathbf{x^k} = \mathbf{x^k} + \rho_{best} \cdot \mathbf{d_{cost}} \quad (12)$$

2.3.5 Exit Criteria

The algorithm terminates the iteration process in one of two ways. First, it may reach the maximum number of user-defined iterations, which risks premature convergence and potentially meaningless results.

Second, the algorithm exits when two conditions are satisfied: the SOCP is infeasible, indicating no further improving descent direction, and the σ value falls below the pre-set threshold. When these criteria are met, the current $\mathbf{x^k}$ is confirmed as the robust optimum.

2.3.6 Importance of Gamma Selection

Inspired by the work of Bertsimas, Nohadani, and Teo (2010) on constructing worst-case evaluation maps, a similar 3D comparison between the nominal and worst-case objective functions was produced for a generic test function, as shown in Fig. 4. This comparison serves as a crucial tool for analyzing solution robustness under varying Γ values in Section 3.3.

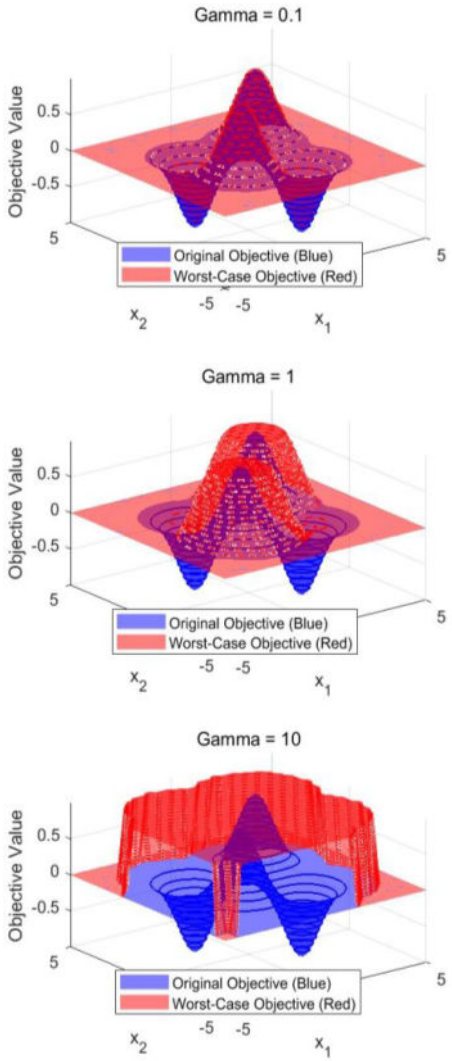


Figure 4. 3D comparison of nominal and worst-case objective functions under varying Γ values for $f(x_1, x_2) = e^{-[(x_1-1.5)^2+(x_2-1.5)^2]} + e^{-[(x_1+1.5)^2+(x_2+1.5)^2]} - e^{-[(x_1-1.5)^2+(x_2+1.5)^2]} - e^{-[(x_1+1.5)^2+(x_2-1.5)^2]}$

To generate the worst-case objective function, a gradient ascent algorithm was applied to every point in the space, tracking the exploration trajectory in a history set. The highest evaluation within the history set was recorded as the new evaluation for each point. Repeating this process across the entire space results in the construction of the worst-case objective function.

The parameter Γ, defined as the maximum exploration radius for each point, plays a critical role in shaping the worst-case objective function. As illustrated in Fig. 4, when Γ is set to a very small value, the worst-case objective function closely resembles the original objective function. This occurs because a small Γ limits exploration to a narrow neighborhood, effectively reducing the search to a local region. This principle underlies the behavior of the ARRTOC algorithm, which, as Γ approaches zero, increasingly mimics a standard optimization algorithm and yields solutions like those obtained from direct optimization.

Conversely, when Γ is too large, the algorithm permits extensive exploration, potentially causing all points to converge to the same local maximum. In this case, the worst-case objective function becomes a horizontal plateau where all points share the same evaluation, as depicted in the third graph in Fig. 4. This scenario eliminates meaningful minima and may lead to premature termination of the ARRTOC algorithm.

Therefore, careful selection of Γ is essential—too small a value limits exploration, while too large a value flattens the worst-case objective function—depending on the desired level of robustness.

3. Results and Discussion:

3.1 Impact of *M* on Mass Balance Violation

For solving the minimization problem of the modified objective function defined in Eq. 6, selecting an appropriate M value is crucial.

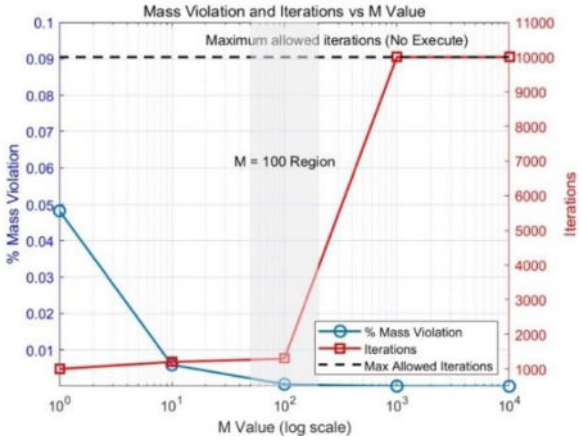


Figure 5. Mass violation and number of iterations versus M value at $\Gamma = 0.6$ (log scale)

As illustrated in Fig. 5, smaller M values lead to faster execution but often result in noticeable violations of the mass balance constraint, referred to as mass

violation (*MV*) in this study. *MV* is calculated using the formula:

$$MV = \frac{\left|\sum_{i=1}^{3} m_{comp,i} - Total\ flow\right|}{Total\ flow} \qquad (13)$$

While increasing M reduces MV, the graph shows that the number of iterations grows sharply beyond $M = 100$. For M values exceeding 1000, the system becomes infeasible due to excessive computational demands. To balance these competing effects, $M = 100$ was identified as the optimal compromise, ensuring minimal MV while maintaining feasible execution times.

3.2 Adherence to Operating Ranges under Varying Total Flow Demand

Fig. 6 illustrates the effectiveness of the np.clip command in enforcing constraints within the ARRTOC algorithm with $\Gamma = 0.6$, ensuring individual compressor flows remain within the operating range of [65, 120] across varying total flow demands.

In the constrained case, represented by orange markers, np.clip successfully enforces flow values within the specified bounds in most cases, except when total flow demands approach the upper limit.

Conversely, in the unconstrained case, represented by blue markers, the absence of constraint-handling capabilities leads to some flow values exceeding the operating range, particularly near both the upper and lower limits.

Notably, within the total flow demand range of [210, 350], both the constrained and unconstrained cases produce similar results. This suggests that the constraints are not active in this region and, therefore, have minimal impact on optimality within this range.

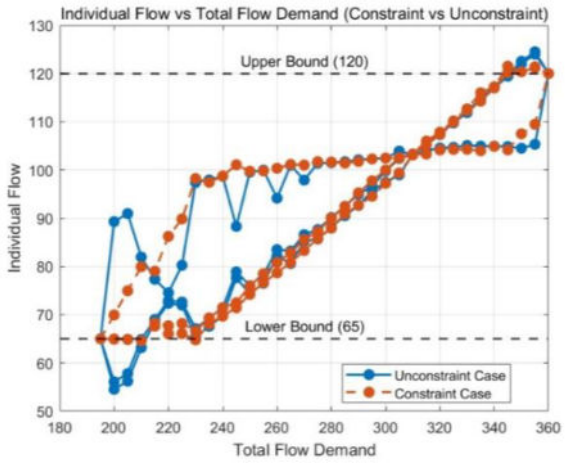


Figure 6. Individual flow versus total flow demand for unconstrained and constrained cases

3.3 Worst-case Objective and ARRTOC Solution at Total Flow Demand = 290

Fig. 7 compares the original objective function with the worst-case objective function in a 3D surface plot at a total flow demand of 290, with increasing Γ values.

Since the problem involves three variables—$m_{comp,1}$, $m_{comp,2}$, and $m_{comp,3}$—dimension reduction is performed by expressing $m_{comp,3}$ as a function of $m_{comp,1}$ and $m_{comp,2}$ using the mass balance equation. This allows for a 3D visualization where the blue surface represents the original objective function, with its nominal minimum marked in green, and the red surface represents the worst-case objective function, with the ARRTOC minimum marked in pink.

Notably, the difference between the two optima increases as Γ increases, and the 3D surface graphs suggest increased robustness at higher Γ values. This robustness is quantified using the relative vertical distance ratio, defined as $\frac{Nominal\ Dist.}{ARRTOC\ Dist.}$. A higher ratio indicates a more resilient ARRTOC optimum to disturbances, and a summary is presented is Table. 1.

Table 1. Relative vertical distance under varying gamma values

Gamma	*Ratio*	*ARRTOC Stability*
5	1.87	less stable
10	8.44	moderate stable
20	41.14	highly stable

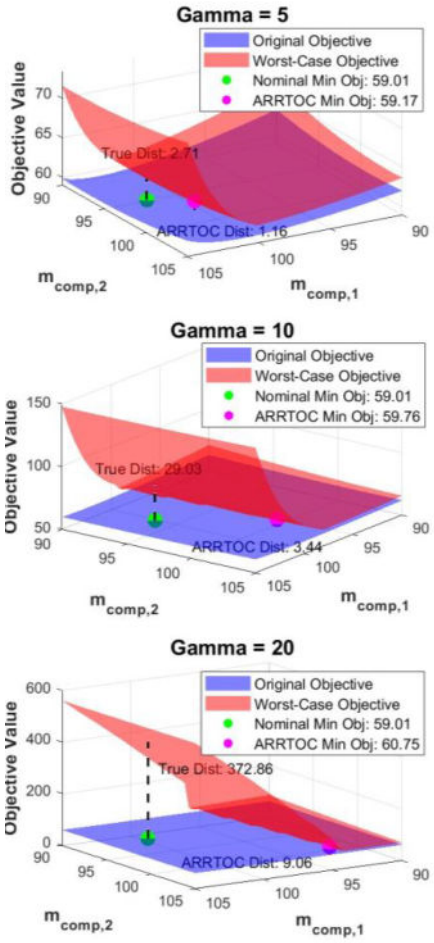


Figure 7. 3D comparison of nominal and worst-case objective functions under varying Γ values for the multi-compressor problem

These results indicate that although the ARRTOC minimum diverges further from the nominal minimum in terms of objective function value, it maintains a closer relative alignment to the original objective surface compared to the nominal minimum's projection on the worst-case evaluation map.

However, increasing Γ also increases the objective function value at the ARRTOC optimum. For example, the objective function value of the ARRTOC minimum, rises from 59.17 at $\Gamma = 5$ to 59.76 at $\Gamma = 10$, and further to 60.75 at $\Gamma = 20$, while the nominal minimum objective value remains constant at 59.01.

This increasing gap highlights the trade-off in the algorithm: as Γ grows, the algorithm prioritizes the robustness of the optimum, effectively guarding the ARRTOC solution against worst-case scenarios. However, this comes at the cost of reduced optimality, as the ARRTOC solution diverges further from the nominal minimum in terms of objective value.

3.4 Verifying Robustness of the ARRTOC Solution Using Compressor Efficiency Curve

Another way to evaluate the robustness of the ARRTOC optimum is by analyzing the corresponding η value at different optima for each compressor.

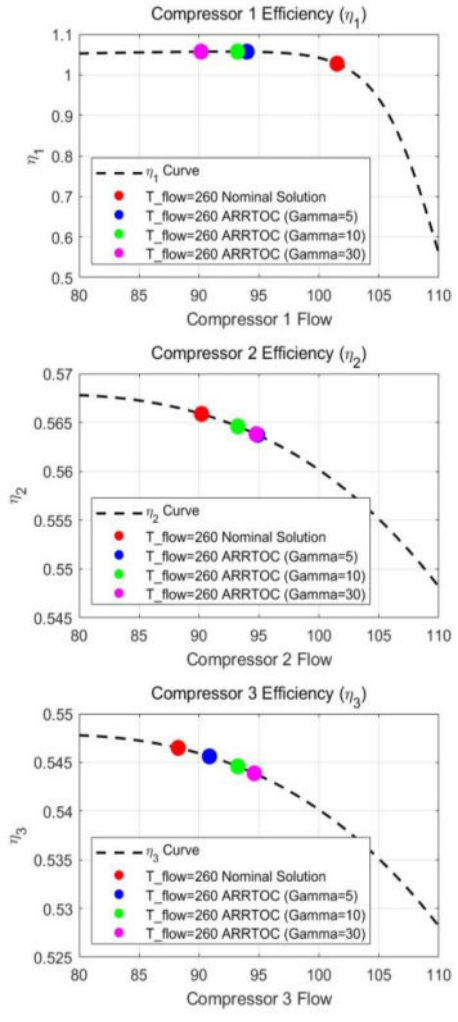


Figure 8. Nominal minimum versus ARRTOC minimum (at varying Γ) shown on the efficiency maps of compressor 1 to 3

In Fig. 8, the red dots represent the nominal optimum obtained from the conventional optimization. Notably, Compressor 1 is far more vulnerable than Compressors 2 and 3, as indicated by the sharp cliff on the right side of the red point in the first graph. This cliff suggests that even small flow disturbances in Compressor 1 can led to significant fluctuations in η_1, making the system highly unstable. As a result, when the ARRTOC algorithm is applied, the solution for Compressor 1 shifts to a smaller flowrate as Gamma increases. This shift moves Compressor 1 further away from the cliff, demonstrating that higher Gamma values improve robustness by enabling the system to tolerate larger disturbances compared to the standard optimization solution.

3.5 Flow Distribution Behavior under Different Gamma

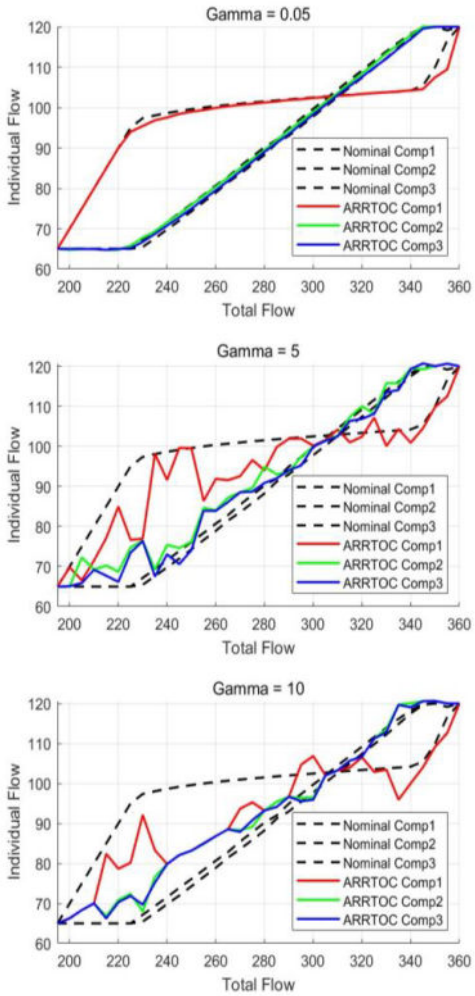


Figure 9. Nominal minimum individual flows versus ARRTOC minimum flows under varying Γ

Fig. 9 illustrates the impact of varying Γ on individual compressor flows across the total flow range of 200 to 360. When Γ is small (e.g., $\Gamma = 0.05$), the ARRTOC solutions align closely with the true solution, as observed earlier in Section 2.3.6.

However, as Γ increases, the flow profiles exhibit fluctuations across most total flow values. These fluctuations arise from the inherent instability of

solutions located near or on the constraint boundaries, where small perturbations can lead to significant deviations.

3.6 Power Requirement under Different Gamma

The relationship between power consumption and the impact of Γ on optimization outcomes is depicted in Fig. 10.

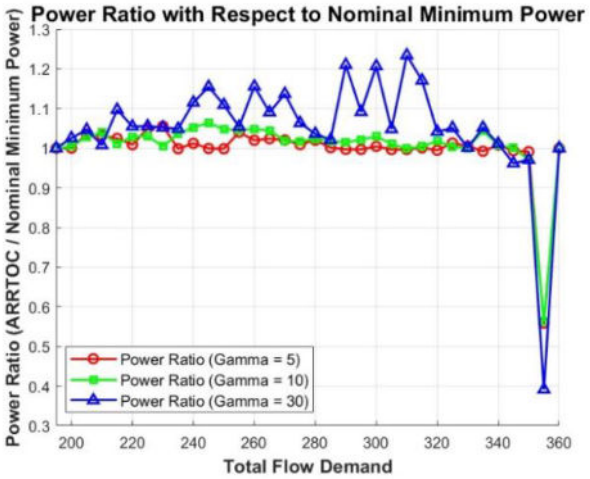


Figure 10. Power ratio relative to nominal minimum power under varying Γ values across different total flow demands

Generally, the power ratio increases as Γ increases across most total flow demands, aligning with the conclusion in Section 3.3 that higher Γ values reduce optimality. However, at a flow demand of 355, the power ratio unexpectedly drops below 1, deviating from the trend.

This anomaly arises from the limitation of using np.clip to enforce the flow rate closer to the upper bounds, leading to a mass balance violation, which is discussed in Section 3.7.

3.7 Mass Balance Violation under Different Gamma

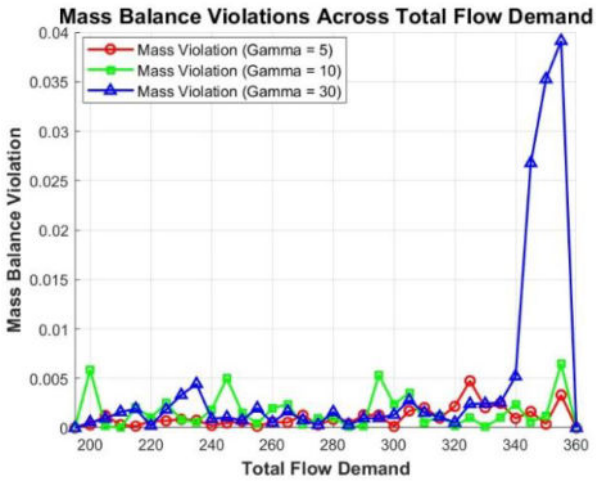


Figure 11. MV under varying Γ values across different total flow demands

In Fig. 11, the average mass balance violation remains within 10% across the entire range of total flow demand. However, a noticeable increase in violations occurs at a flow demand of 355, suggesting that the penalty term in the objective function becomes less effective near the boundary conditions, failing to adequately control the flow rate. This is corroborated by Fig. 10 in Section 3.6, where the power ratio drops below 1 at the same flow demand.

The increased mass balance violation at a total flow demand of 355 arises because the optimal solution lies very close to the constraint boundaries. At this point, the penalty applied to enforce the mass balance equation introduces small errors, causing the ARRTOC algorithm to explore solutions near or beyond the constraint boundaries. Interestingly, some of these infeasible solutions outperform the nominal ones, leading to a power ratio below 1, as seen in Fig. 10. This behavior is amplified at high total flow demands, where the objective function's sensitivity to flow variations exacerbates the errors introduced by the penalty term, occasionally driving the power ratio significantly lower.

4. Conclusion & Outlook

This study explored the theoretical foundation of the ARRTOC algorithm and its application to multi-compressor optimization. By introducing a penalty term into the nominal objective function and applying the np.clip command in Python, the study effectively addressed key constraints such as Mass balance and individual compressor operating flow ranges, and successfully solved the problem using the unconstrained version of the ARRTOC algorithm.

Through an analysis of the percentage of mass violations for varying M values in Section 3.1, an optimal M was identified that balances execution feasibility with mass balance adherence.

In Section 3.2, the study evaluated the use of np.clip in place of the constraint-handling capabilities of the ARRTOC algorithm, demonstrating its effectiveness in maintaining adherence to constraints across varying total flow demands, despite some limitations observed when total flow demands approached upper bounds.

From Section 3.3, an in-depth analysis of the tuning of the key parameter, Γ, was presented to evaluate solution robustness. Through visual 3D representations of the worst-case objective function versus the original objective function of the multi-compressor problem, the relationship between Γ and solution robustness was established. The robustness of the derived optima was further validated through plots of individual compressor efficiency versus flow rates, showing that increasing Γ shifted the robust optima away from unstable regions ("cliffs") in the efficiency map.

In Section 3.6, it was observed that the power ratio generally increased with Γ, but an anomaly occurred at a total flow demand of 355, where the power ratio unexpectedly dropped below 1. This behavior, further explored in Section 3.7, was due to the limitations of the penalty term, which introduced errors near constraint boundaries. These errors caused the algorithm to explore infeasible solutions that occasionally outperformed nominal solutions. The study also highlighted the sensitivity of the objective function to flow variations at

high total flow demands, amplifying the impact of these errors. These findings outline the importance of selecting parameters such as M and Γ based on application-specific requirements and demonstrate the potential of the ARRTOC algorithm for achieving both efficient and robust optimization in multi-compressor systems.

Looking forward, while this study focused on the multi-compressor problem, the ARRTOC algorithm shows significant promise in other domains requiring real-time optimization and uncertainty management. Potential applications include renewable energy systems, such as wind and solar power integration in distribution networks, where fluctuating energy inputs and dynamic supply-demand conditions are critical challenges (Coppitters, De Paepe and Contino, 2020). ARRTOC's ability to provide robust optimization under uncertainty makes it a valuable tool for ensuring stability and efficiency in such complex environments.

Acknowledgement

We would like to express our heartfelt gratitude to everyone in the Autonomous Industrial Systems Lab for their consistent support throughout the duration of this project. A special thanks goes to Akhil Ahmed for his guidance and advice on implementing the ARRTOC algorithm.

References

Ahmed, A., Rio-Chanona, del and Mercangoz, M. (2023). *ARRTOC: Adversarially Robust Real-Time Optimization and Control*. [online] arXiv.org. Available at: https://arxiv.org/abs/2309.04386 [Accessed 4 Oct. 2024].

Coppitters, D., De Paepe, W. and Contino, F. (2020). Robust design optimization and stochastic performance analysis of a grid-connected photovoltaic system with battery storage and hydrogen storage. *Energy*, 213, p.118798. doi:https://doi.org/10.1016/j.energy.2020.118798.

Dimitris Bertsimas, Omid Nohadani and Kwong Meng Teo (2010). Robust Optimization for Unconstrained Simulation-Based Problems. 58(1), pp.161–178. doi:https://doi.org/10.1287/opre.1090.0715.

Jäschke, J., Cao, Y. and Kariwala, V. (2017). Self-optimizing control – A survey. *Annual Reviews in Control*, 43, pp.199–223. doi:https://doi.org/10.1016/j.arcontrol.2017.03.001

Robison, D., Arvada, P. and Beaty (n.d.). *COMPRESSOR TYPES, CLASSIFICATIONS, AND APPLICATIONS Market Specialist Sundstrand Fluid Handling*. [online] Available at: https://core.ac.uk/download/pdf/147259078.pdf?utm_source [Accessed 20 Nov. 2024].

Seborg, D.E., Edgar, T.F., Mellichamp, D.A. and Doyle, F.J. (2017). *Process Dynamics and Control*. 4th ed. Hoboken, Nj: Wiley.

Speed of Sound for Normal Hydrogen at Temperatures of 198 K to 298 K and Pressures of 2 to 50 MPa - 4th Year Research Project Report

Jeremy Mazauric, Shiam Srikumar

Department of Chemical Engineering, Imperial College London, U.K.

Abstract

The speed of sound in normal hydrogen was obtained for temperatures ranging from 198K to 298K, at pressures from 2 MPa to 50 MPa. The results were obtained via operation of a cylindrical acoustic resonator. Calibrations of the path length of vibration, transducer compliance and transducer quality factor were performed at 198K and 298K using helium, with results varying from the EOS within ±0.012% and ±0.001% at 298K and 198K respectively. Excellent repeatability of the resonator was also demonstrated, with respective runs varying within ±0.003% for helium. The speed of sound data in n-hydrogen was able to agree to within ±0.177% of EOS obtained data, with the expanded uncertainty (k=2) varying from 0.004% to 0.021%. The excess half-widths obtained, relative to the perturbed frequency, varied within ±0.078%, highlighting the strong performance of the acoustic model's accuracy.

1 Introduction

Despite a widely accepted consensus being reached in recent years over the threat that increasingly frequent consumption of fossil fuels poses to global sustainability, consistent population growth has lent itself in accelerating the consumption of these fuels. This is indicative in the fact that from 2008 to 2021, the global consumption of oil in millions of tons has risen from 3928 to 5300 [1]. As a result, over the past decade, the energy industry has made frequent efforts in developing sustainable processes with reduced greenhouse emissions to replace the previously dominant fossil-based processes. This has ultimately led to the current period of energy transition in which hydrogen has played a pivotal role in driving the implementation of more sustainable methods of energy production [2].

Hydrogen's main enticing factor lies in the lack of greenhouse and harmful emissions, with the only by-product produced via its combustion being water [3]. Furthermore, there are various sustainable methods of production for H_2, which include using renewable energy processes to drive production of H_2 via electrolysis, as well as incorporating carbon capture and storage (CCS) technologies to reduce the environmental burden of steam-methane reforming processes [4]. Industrially, as of 2023, at least 228 large-scale projects have commenced exploring the implementation of green hydrogen storage processes [5]. However, in order for hydrogen storage processes to be made feasible, highly accurate thermophysical properties of normal hydrogen are essential in improving the performance of the state-of-the-art equations of state (EOS) being used to model H_2 behavior at various temperatures and pressures.

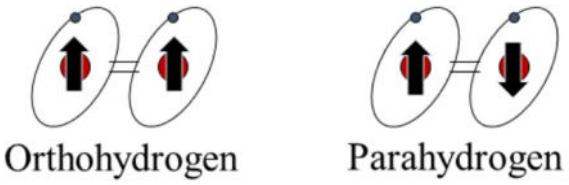


Figure 1: Diagram of the nuclear spin isomers of hydrogen

At any state, pure hydrogen comprises of its two nuclear spin isomers, para-hydrogen and ortho-hydrogen, with opposed and parallel nuclear spins respectively [6]. The composition of hydrogen therefore refers to the ratio of the para to ortho content. At temperatures above 298 K, hydrogen consists of 25% para-hydrogen, which is the composition denoted as n-hydrogen. However, due to its lower energy state, parahydrogen content increases as temperature drops, reaching approximately 50% at 78 K and almost pure para-hydrogen at 20 K [7]. Although, the conversion process is incredibly slow and can take up to many weeks or even months [8], due to the triplet-singlet transformation involved in the mechanics of the interconversion between the two states.

Of the thermophysical properties that are able to be measured, the speed of sound is particularly favorable for its versatility in relating to a number of other properties such as isothermal compressibility, density and specific heat capacity [9]. The highly precise nature of speed of sound measurements also causes the subsequently derived thermodynamic properties or constants to be obtained to very high degrees of accuracy. For example, as is reported in Rowlinson and Tildesley [10], where the gas constant was determined very accurately using speed of sound measurements in Argon at 273.16K. Therefore, highly accurate speed of sound data for hydrogen would enable corresponding equa-

tions of state to extract a greater range of properties and predict these with greater accuracy. With regards to industrial applications of speed of sound measurements, it is of particular interest when concerned with the operation of Venturi nozzles or ultrasonic flow meters [8], which are especially relevant to liquefied hydrogen storage processes.

2 Background

Based on data collected throughout the 1970s and early 1980s, an equation of state predicting the properties of parahydrogen was developed by Younglove et al. [11], and later modified to extrapolate for n-hydrogen, making it applicable for pressures up to 121 MPa and 400 K, which was then incorporated into previous versions of REFPROP. However, this was eventually replaced by the EOS developed by Leachman et al. [12], which was to extend the applicability of the fundamental equation of state to pressures of 2000 MPa and temperatures of 1000 K.

Nevertheless, these correlations were limited because of the experimental data used to perform the necessary regressions, with the data used in Leachman only ranging up to 100 K and 32.5 MPa. Therefore, the EOS that served as a reference for this study was the model developed by Nguyen et al. [9], an EOS that is explicit in the Helmholtz free energy, similar to Leachman. However, unlike Leachman, the experimental data used in Nguyen varied from 273 K to 323 K at pressures of 1 to 100 MPa. Therefore, whilst the data being used may have suggested a more accurately regressed model, it still remained indicative of the lack of speed of sound data for n-hydrogen, specifically at temperatures between 100 K to 273 K at these pressures.

From the data studied and available in the literature, as shown in table 1, this absence of speed of sound data in the temperature-pressure region stated was clear. Therefore, this was the main motivation behind the project's main objective of obtaining speed of sound data to fill this 'gap' in the literature. Of the experimental data cited, the absolute average relative deviation (AARD) from the EOS of Nguyen et al. [9] was calculated using data that shared the ranges of temperature and pressure tested, including those collected in this work. The AARD was defined as shown below in Eq. 1.

$$AARD = \frac{1}{N}\sum_{i=1}^{N} 100\frac{x_{data} - x_{EOS}}{x_{data}} \tag{1}$$

3 Methodology

3.1 Cylindrical Acoustic Resonator

The cylindrical resonator used to perform the sound of speed measurements in helium and hydrogen followed the design outlined in Ruffine and Trusler [19]. The resonator enclosed a cylindrical cavity of 15mm diameter and a length of 60mm. The external dimensions of the resonator shell were 24mm nominal diameter and 85mm nominal length. Piezoelectric disc transducers were used to produce the sound pulses within the cavity, and one was bonded to the inside of each of the two end plates, on either end of the shell. This was bonded via an Epoxy resin with a stainless-steel spring providing the means for an electrical contact to the electrical connection. Each transducer was connected to an AC function generator necessary to provide the electrical signal that produced the mechanical disturbance.

Fig. 2 shows the process flow diagram (PFD) of the apparatus being used for the filling and evacuating of the cylindrical resonator with the respective compressed gases. The configuration is able to be isolated into three different compression stages: low-pressure, intermediate-pressure and high-pressure stages. The initial low-pressure stage, used to compress the system to 10 MPa, is outlined in the PFD as the section from the GSV to V2. The intermediate-pressure stage, situated between V2 and V3, was largely defined by the syringe pump (ISCO, Model 260D). Here, the system was compressed up to 35 MPa. The final high-pressure section, starting with V3 and ending with V5, was defined by the manual pump (Sitec, Model 750.1100-H2) and the cylindrical acoustic resonator. This area of the process was pressurized to 50 MPa.

The acoustic resonator was mounted within a pressure vessel designed to withstand pressures well above the maximum pressure to be tested for, being 50 MPa. This pressure vessel was then suspended in temperature bath E.1 (Fluke, Model 7831) using ethanol as the bath fluid as a means of controlling the temperature of the resonator.

Prior to measurements of the resonance frequency being taken, 5 cycles of purging and evacuation procedures were carried out in order to ensure the purity of the gas sample entering the cylindrical resonator. A computer-controlled operating system was used to test ranges of frequency under a constant voltage of 15V in order to determine the unperturbed resonance frequency of the 2nd order longitudinal mode of vibration within the gas-filled cavity.

3.2 The Acoustic Model

In order to calculate the speed of the sound from the frequency obtained, the following expression was used.

$$w = \frac{2f_{corr}L}{I}[1 + \alpha(T - T_{ref})] \tag{2}$$

Where w is the calculated speed of sound ($m.s^{-1}$), f_{corr} is the perturbed resonance frequency, I is the longitudinal mode of vibration ($I = 2$) and α is the linear thermal expansivity of the resonator. However, before computing the speed of sound, as the frequency measured via the cylindrical resonator was an unperturbed frequency, it was firstly corrected to account for the various contributions to the dissipation occurring

Source	Number of Points	Temperature Range (K)	Pressure Range (MPa)	AARD (Nguyen et al.)
Nguyen et al. (2024)[9]	38	273.139 - 298.154	1.01 - 99.60	0.0215
Carey et al. (1974)[13]	18	295.150 - 299.150	0.15 - 11.02	0.6462
Giacobbe (1993)[14]	3	294.450 - 294.950	0.01	0.8660
Hodge (1937)[15]	11	273.150	0.01 - 10.13	5.1738
Liebenberg et al. (1978)[16]	10	75.040 - 163.90	473.30 – 1871.00	-
Liebenberg et al. (1977)[17]	314	75.000-300.000	200.00 - 2250.00	-
Van Dael et al. (1965)[18]	212	18.210 – 33.070	0.17 – 24.66	-
This Work	29	198.277 - 298.156	1.95 - 50.12	0.0687
Total	635	18.210-300.000	0.01 - 2250.00	0.7412

Table 1: Speed of sound Data for n-hydrogen available in literature

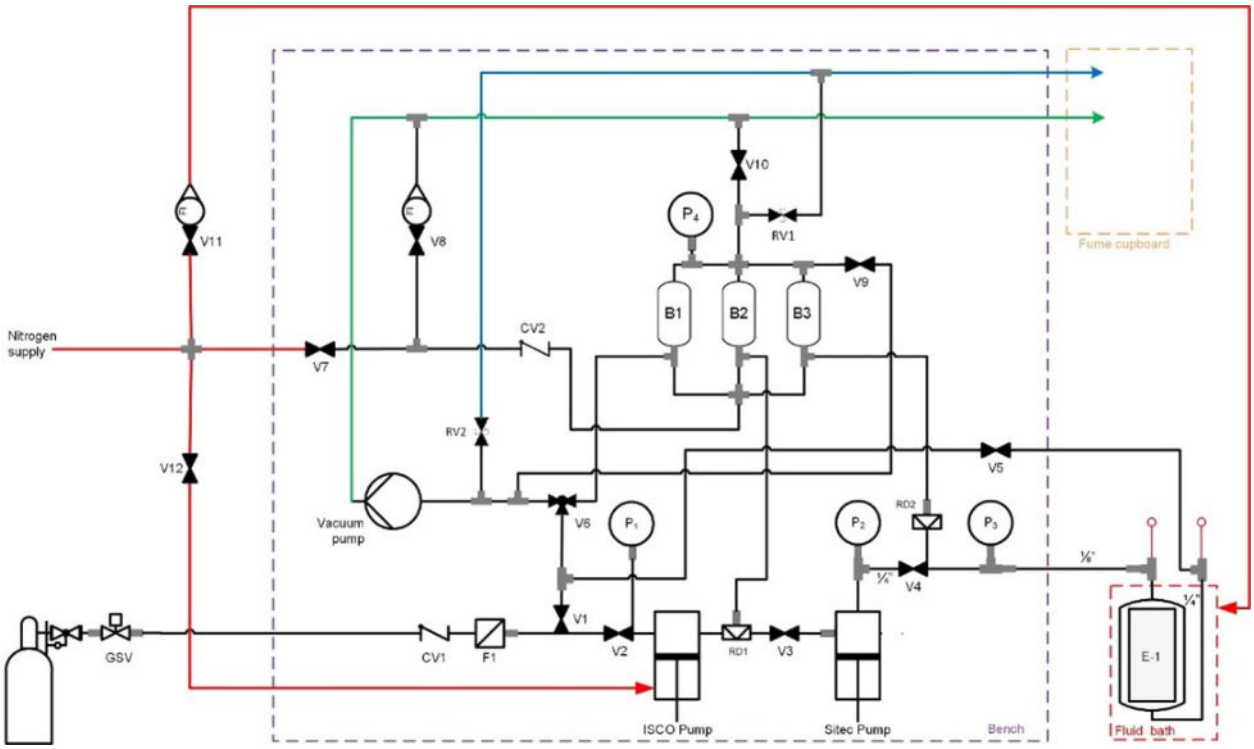


Figure 2: PFD of cylindrical acoustic resonator apparatus

within the resonator. In order to calculate these corrections, the equations described in Trusler, 2020 [20] were used.

The first of these contributions referred to the admittance at the viscous and thermal boundary layers between the cavity and the enclosing shell of the resonator. As the frequency of the propagated sound wave was increased via excitation of the end plate transducers, variations in density and pressure within the fluid generated more considerable thermal and viscous gradients, causing dissipation of sound energy to occur. This, in turn, resulted in a perturbation in the resonance frequency obtained. Within the bulk of the fluid, the relatively long wavelength caused these gradients to be negligible. However, at the boundary between the fluid and the solid surface of the inner vessel, these gradients were much larger. Therefore, the dissipation at the boundary was considerable and needed to be modeled. The correction used to model the effect of viscous damping was expressed as described in Trusler, 2020 [20].

$$\Delta f_s = g_s \left(1 - 2\frac{l_s}{\delta_s}\right) \tag{3}$$

Where l_s is the viscous accommodation length given by:

$$l_s = \frac{\eta}{P}\left(\frac{2-h_s}{h_s}\right)\sqrt{\frac{\pi RT}{2M}} \tag{4}$$

and δ_s is the penetration length of the viscous layer given by:

$$\delta_s = \sqrt{\frac{D_s}{\pi f_0}} \tag{5}$$

where D_s is the viscous diffusivity and f_0 the unperturbed resonance frequency. The term g_s refers to the contribution of the viscous effect to the total half-width of the axially symmetric mode. This was given by:

$$g_s = \frac{f\delta_s}{2b} \tag{6}$$

where b is the radius of the cylindrical cavity. An analogous set of equations was used to compute the contribution of the admittance from the thermal boundary

layer:

$$\Delta f_h = g_h \left(1 - 2\frac{l_h}{\delta_h}\right) \tag{7}$$

Where l_h is the thermal accommodation length given by:

$$l_h = \frac{\chi}{P}\left(\frac{2-h_h}{h_h}\right)\sqrt{\frac{\pi MT}{2R}} \tag{8}$$

and δ_h is the penetration length of the thermal layer given by:

$$\delta_h = \sqrt{\frac{D_h}{\pi f_0}} \tag{9}$$

where D_h is the thermal diffusivity. Similarly to the viscous layer contribution, g_h referred to the contribution of the thermal boundary layer to the total half width and was given by:

$$g_h = \frac{f(\gamma-1)\delta_h}{2b}\left(1 + 2\frac{b}{L}\right) \tag{10}$$

where:

$$\gamma = \frac{c_p}{c_v} \tag{11}$$

Another set of contributions to the perturbance of the resonance frequency that was considered, stemmed from the mechanical admittance due to deformations in the shell. The corrections resulting from the axial and radial stretch of the shell were modeled based on the equations of Zhang et al. [21] and Nguyen et al. [9] as in Eq. 12 and Eq. 13 respectively. Where Δf_{ax} and Δf_{rad} are the perturbations in the frequency due to the axial and radial stretches respectively.

$$\Delta f_{ax} = \frac{\rho w_{EOS}^2 f_0 \left(\frac{C_{ax}}{L}\right)}{1 - \left(\frac{f_0}{f_{ax}}\right)^2} \tag{12}$$

$$\Delta f_{rad} = \frac{\rho w_{EOS}^2 f_0 \left(\frac{C_{ax}}{L}\right)}{1 - \left(\frac{f_0}{f_{rad}}\right)^2} \tag{13}$$

C_{ax} and C_{rad} are the axial and radial compliance of the transducers respectively, f_{ax} and f_{rad} are the axial and radial resonance frequencies respectively, and w_{EOS} is the EOS-estimated speed of sound. As shown in Eq. 12 and Eq. 13, the two terms increased as the unperturbed frequency approached the radial and axial resonance frequencies of the vessel, therefore representing the amplified dissipation that occurred as the resonance frequency of the vessel material was approached.

The next of the perturbations originating from the deformation of the shell was from the effect of the end plates of the cylinder, more specifically the diaphragms upon which the piezoelectric disc transducers were bonded via the epoxy resin. When estimating the contribution of this factor to the overall perturbations, both the effect on the perturbed frequency and the resulting half-width were computed.

$$\Delta f_{ends} = \frac{\rho w_{EOS}^2 f_0 \left(\frac{C_{trans}}{L}\right)\left[1 - \left(\frac{f_0}{f_{trans}}\right)^2\right]}{\left[1 - \left(\frac{f_0}{f_{trans}}\right)^2\right]^2 + \left(\frac{f_0}{f_{trans}Q_{trans}}\right)^2} \tag{14}$$

$$g_{ends} = \frac{\rho w_{EOS}^2 f_0 \left(\frac{C_{trans}}{L}\right)\left(\frac{f_0}{f_{trans}Q_{trans}}\right)}{\left[1 - \left(\frac{f_0}{f_{trans}}\right)^2\right]^2 + \left(\frac{f_0}{f_{trans}Q_{trans}}\right)^2} \tag{15}$$

Where C_{trans} is the compliance of the plate transducer, f_{trans} is the resonance frequency of the plate transducer and Q_{trans} is the quality factor of the plate transducer. While not originally included, the quality factor allowed the model to account for possible damping arising from mechanical resistance of the spring used to provide an electrical connection between the function generator and transducer. As well as these sources, dissipation originating from recoil of the rigid body and the flow of gas via the filling ducts of the resonator, were also computed. However, they were negligible for even-ordered modes of vibration.

Using these perturbation estimations, the corrected frequency could then be computed, as shown below, and then substituted into Eq. 2 to calculate the measured speed of sound:

$$f_{corr} = f_0 + \Delta f_s + \Delta f_h + \Delta f_{ax} + \Delta f_{rad} + 2\Delta f_{ends} \tag{16}$$

As well as the corrected frequency, the acoustic model's projection of the half-widths was also calculated and compared to the experimentally obtained half-widths, relative to the perturbed frequency. This provided insight into the acoustic model's performance in accurately modeling the dissipation within the resonator. As well as the half-width contributions already discussed in Eq. 10, Eq. 6, and Eq. 15, the contribution due to bulk absorption was also expressed through Eq. 17.

$$g_b = \pi\left(\frac{f_0}{w_{EOS}}\right)^2\left(\frac{4D_s}{3} + (\gamma-1)D_h\right) \tag{17}$$

The main mechanics behind the bulk absorption that took place within the vessel, and the subsequent propagation of sound, were the flows of heat and momentum that arose within the fluid during excitation of the transducers [20]. However, once computed, this contribution was found to be negligible, as was expected, for the same reasons explained earlier. As a result, the half widths of resonance predicted by the acoustic model were computed via Eq. 18.

$$g_{corr} = g_h + g_s + g_b + g_{ends} \tag{18}$$

3.3 Uncertainty Analysis

In order to calculate the above correction terms, the thermodynamic properties c_p, c_v, D_h, D_s, ρ and w_{EOS} were obtained via equations of states implemented through REFPROP V10.

For the calibration testing completed using helium, the heat capacities, density and speed of sound were obtained using the EOS by Ortiz-Vega et al. [22]. Thus, the reported uncertainties were used with c_p and c_v having uncertainties of 3%, ρ with an uncertainty of 0.03% for pressures lower than 40 MPa and 0.1% for pressures higher than 40 MPa, and the speed of

sound having an uncertainty of 0.01%. The thermal conductivity data for helium was computed using the model by Hands and Arp [23], with the uncertainty being 10%. For the viscous diffusivity, the viscosity model by Arp and McCarty [24] was used and the uncertainty was given as 5%. The calibration testing ultimately provided values of C_{trans}, Q_{trans} and L. In order for accurate standard uncertainties to be used for the calibration parameters, uncertainties were obtained via a root-sum-square method, due to their involvement in the uncertainty analysis of the n-hydrogen results.

For the testing carried out using n-hydrogen, the EOS by Nguyen et al. [9] for normal hydrogen was implemented via REFPROP to obtain estimates for the isobaric and isochoric heat capacities, density and the speed of sound. As reported in the corresponding literature, the uncertainty for the heat capacities was 0.2%. For the density, at pressures up to 30 MPa, the uncertainty was 0.05%, and 0.3% for pressures above 30 MPa. For the speed of sound, the uncertainty was 0.05% for pressures up to 10 MPa and 0.1% for pressures above 10 MPa. Thermal conductivity data was extracted using the model of Assael et al. [25], with the uncertainty being 4%. The viscosity was determined via the viscosity model of Muzny et al. [26], with the uncertainty being 4%. Therefore, the uncertainties of the speed of sound results were computed via the root-sum-square method using the expression shown below in Eq. 19.

$$u(w) = \Bigg[\left(\frac{\delta w}{\delta f_0}u(f_0)\right)^2 + \left(\frac{\delta w}{\delta \gamma}u(\gamma)\right)^2 + \left(\frac{\delta w}{\delta w_{EOS}}u(w_{EOS})\right)^2 + \left(\frac{\delta w}{\delta \rho}u(\rho)\right)^2 + \left(\frac{\delta w}{\delta f_{trans}}u(f_{trans})\right)^2 + \left(\frac{\delta w}{\delta P}u(P)\right)^2 + \left(\frac{\delta w}{\delta T}u(T)\right)^2 + \left(\frac{\delta w}{\delta D_h}u(D_h)\right)^2 + \left(\frac{\delta w}{\delta D_s}u(D_s)\right)^2 + \left(\frac{\delta w}{\delta C_{trans}}u(C_{trans})\right)^2 + \left(\frac{\delta w}{\delta Q_{trans}}u(Q_{trans})\right)^2 + \left(\frac{\delta w}{\delta L}u(L)\right)^2\Bigg]^{0.5} \quad (19)$$

An example uncertainty budget analysis is given in table 2 for a particular measurement taken for n-hydrogen at 50 MPa and 298.15 K. Inspired by Nguyen et al. [9], a 95% confidence level for the expanded uncertainty was used with k = 2.

4 Results and Discussions

4.1 Model Calibration

The acoustic model was first calibrated using helium as the calibration gas. After conducting three separate sets of helium measurements at a range of pressures from 2 MPa to 50 MPa at 298K, the acoustic model parameters C_{trans}, Q_{trans} and L were calculated by minimizing the square sums of the speed of sound deviations from the EOS of Ortiz-Vega et al. [22] and the excess half-width relative to the perturbed frequency, using a nonlinear solver. As a result, the path length was found to be 0.060029 m, the value for C_{trans} was 0.853 $mTPa^{-1}$, and the Q_{trans} value as 0.72.

However, to test for possible temperature dependence in the 3 calibration parameters, the calibration experiment was repeated at the lowest temperature measured, 198K, from pressures of 2 MPa to 30 MPa. Following the recalibration, new values for the three parameters were obtained, with the path length being 0.060074, the value for C_{trans} being 0.638 and the new value for Q_{trans} was found to be 0.61. As shown in Demirdesen et al. [27], a similar calibration method was used. However, unlike this study, 5 isotherms were tested, yielding a linear correlation as an appropriate approximation for the temperature dependence of C_{trans}. Therefore, the following expression was used to interpolate within the temperature range of results.

$$C_{trans} = 0.2119775 + (2.15 \times 10^{-3})T \quad (20)$$

For the path length, the findings in Demirdesen et al. [27] uncovered an exponential relationship between temperature and L. However, given that only the highest and lowest temperatures were investigated in this study, a linear interpolation was used, as was the case with the quality factor Q_{trans}.

$$L = 0.0601631675 - (4.5 \times 10^{-7})T \quad (21)$$

$$Q_{trans} = 0.392035 + (1.1 \times 10^{-3})T \quad (22)$$

4.2 Helium Results

From the calibration testing carried out as mentioned previously, the speed of sound data obtained provided a means of evaluating the accuracy and performance of the acoustic resonator, as well as the acoustic model's ability to correctly model the various perturbations, prior to n-hydrogen testing. This included 3 sets of speed of sound data collected for a $298.15K$ isotherm at pressures from 2 MPa to 50 MPa. As shown in Fig. 3, the maximum deviation of the experimental data from the EOS was found to be 0.012% at 50 MPa, comparing it to the EOS by Ortiz-Vega et al. [22].

As was expected, the results were found to deviate from the EOS much more for larger pressures, suggesting possible inaccuracies at higher pressures. This was also a concern in Nguyen et al. [9], as a result of the measured frequencies approaching the resonance frequencies of the cylindrical vessel at these higher pressures. Therefore, the additional deviations were attributed to excessive deformations in the vessel that the radial and axial correction terms were failing to account for more accurately at these pressures. Nonetheless, the data for the three runs depicted the strong performance of the acoustic model in accounting for the

Quantity	Value	Standard Uncertainty	Sensitivity Coefficient	Uncertainty Contribution
f_0	28,999.324 Hz	0.011 Hz	0.060 $ms^{-1}Hz^{-1}$	0.0006 ms^{-1}
γ	1.414	0.006	0.378 ms^{-1}	0.0019 ms^{-1}
w_{EOS}	1,748.970 ms^{-1}	1.749 ms^{-1}	0.007	0.0128 ms^{-1}
ρ	30.886 kgm^{-3}	0.093 kgm^{-3}	0.208 $m^4s^{-1}kg^{-1}$	0.0193 ms^{-1}
f_{trans}	70,000 Hz	1 Hz	2.735 x 10^{-5} $ms^{-1}Hz^{-1}$	0.0000 ms^{-1}
P	50.110 MPa	0.03 MPa	0.156 $ms^{-1}MPa^{-1}$	0.0047 ms^{-1}
T	298.154 K	0.015 K	-0.007 $ms^{-1}K^{-1}$	-0.0001 ms^{-1}
D_h	4.94 x 10^{-7} m^2s^{-1}	1.98 x 10^{-8} m^2s^{-1}	199.010 m^{-1}	0.0000 ms^{-1}
D_s	3.33 x 10^{-7} m^2s^{-1}	1.33 x 10^{-8} m^2s^{-1}	384.273 m^{-1}	0.0000 ms^{-1}
C_{trans}	0.853 TPa	0.389 $mTPa_{-1}$	4.471 $s^{-1}TPa$	1.7345 ms^{-1}
Q_{trans}	0.61	0.0193	1.389 ms^{-1}	0.0269 ms^{-1}
L	0.6 m	1.212 x 10^{-5} m	28872.73 s^{-1}	0.3500 ms^{-1}
Expanded Uncertainty, u(w) = 0.3539 ms^{-1} (k = 0.2)				

Table 2: Uncertainty budget for n-hydrogen speed of sound measurement at 50 MPa and 298.15 K

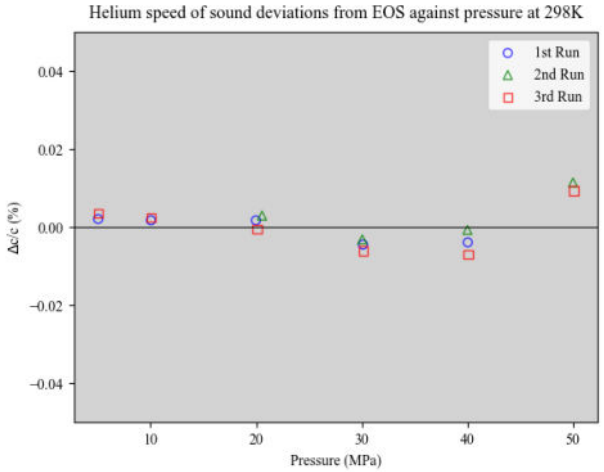


Figure 3: Speed of sound deviations of helium from the EOS

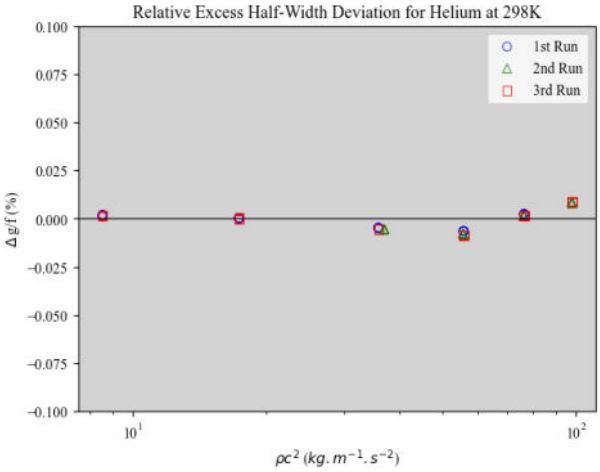


Figure 4: Excess half-widths of helium at 298K

sound energy dissipation occurring within the cylindrical vessel.

Furthermore, Fig. 3 emphasized the high level of repeatability in the results obtained, with the largest deviation between same state measurements being 0.003 %. The comparison of the half-widths projected via the acoustic model as per Eq. 18 to the experimentally obtained half-widths yielded the results shown in Fig. 4. Over all three runs at the 298K isotherm, the excess half-widths, relative to the corrected frequency, was found to be within ± 0.009% across all tested pressures, confirming that the acoustic model was able to predict the perturbations within the real system to a very high level of accuracy for helium.

As explained previously, helium testing was also carried out at 198K with pressures of 2 MPa to 20 MPa tested. Fig. 5 depicts the deviation of the collected data from the same EOS of Ortiz-Vega et al. [22], which further confirmed the acoustic model's accurate performance in accounting for the perturbations. The experimental data at 198K performed incredibly well in agreeing with the EOS, deviating within ± 0.003%, with slightly larger deviations at higher pressures, similar to the isotherm at 298K.

4.3 Hydrogen Results

Following calibration of the necessary parameters and verifying the validity of the acoustic resonator, 5 isotherms were tested using n-hydrogen within the cylindrical vessel. These temperatures being 298K, 273K, 248K, 223K and 198K. The speed of sound measurements obtained can be found in table 3. For the isotherms at 298K and 273K, the full pressure range from 2 MPa to 50 MPa was investigated, however at 223K, the vessel was only pressurized to 30 MPa due to excessive pressure drifts experienced when attempting higher pressure measurements. The vessel was particularly susceptible to this issue at lower temperatures due to the seal's security being compromised, which is discussed in greater detail in section 4.5. Therefore, the issue was further amplified at the coldest isotherm of 198K, where only pressures below 20 MPa were tested.

Following comparison of the obtained data with the equation of state of Nguyen et al. [9] as implemented in REFPROP, the projections in the plane of speed of sound versus pressure showed that the experimental

T (K)	P (MPa)	w (ms-1)	U(w)	% U(w)	T (K)	P (MPa)	w (ms^{-1})	U(w)	% U(w)
298.155	50.12	1747.63	0.354	0.0203	273.220	49.90	1713.85	0.360	0.0210
298.154	40.01	1662.72	0.266	0.0160	273.220	39.98	1625.67	0.269	0.0165
298.154	29.96	1576.00	0.188	0.0119	273.220	30.02	1534.69	0.189	0.0123
298.156	19.94	1487.93	0.122	0.0083	273.220	19.97	1441.29	0.121	0.0084
298.156	9.93	1399.82	0.074	0.0053	273.220	9.97	1348.72	0.072	0.0053
298.155	4.95	1356.48	0.059	0.0043	273.219	4.99	1303.50	0.057	0.0044
298.155	2.05	1330.51	0.054	0.0041	273.220	1.99	1276.74	0.052	0.0041
248.242	40.66	1597.89	0.281	0.0176	223.260	29.83	1452.74	0.192	0.0132
248.242	30.34	1497.31	0.193	0.0129	223.277	20.08	1348.74	0.120	0.0089
248.242	20.10	1395.87	0.121	0.0087	223.269	10.08	1243.69	0.068	0.0055
248.242	10.01	1296.85	0.070	0.0054	223.277	5.06	1193.42	0.053	0.0044
248.244	4.98	1248.87	0.055	0.0044	223.262	1.95	1163.47	0.047	0.0041
248.243	1.99	1221.03	0.050	0.0041					
198.278	19.9	1299.66	0.120	0.0092					
198.277	10.024	1187.23	0.066	0.0055					
198.278	4.788	1131.43	0.050	0.0044					
198.278	2.042	1103.90	0.045	0.0041					

Table 3: Speed of sound data of n-hydrogen at temperatures of 198K-298K and pressures of 2MPa-50MPa

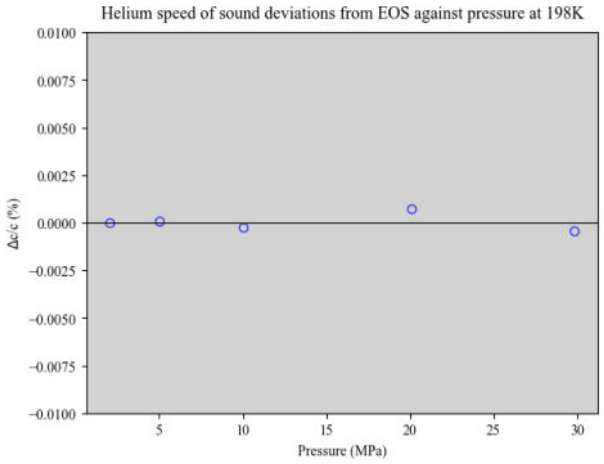


Figure 5: Speed of sound deviations of helium from the EOS by Ortiz-Vega et al.[22] at 198K

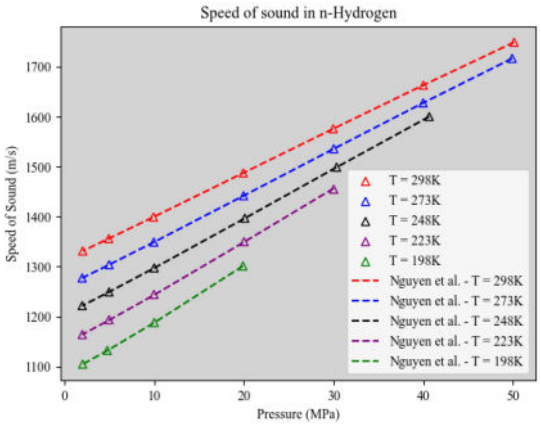


Figure 6: Speed of sound in n-hydrogen against pressure

data agreed with the EOS to a very high degree of accuracy as shown in Fig. 6. As is shown in Fig. 7, at the first isotherm at 298K, the experimental results agreed very well with the EOS, varying within ±0.08% of the REFPROP values, with the maximum expanded uncertainty of the speed of sound obtained at 298K being 0.021%, and decreasing as lower pressures were tested. The highest deviations were found at the 273K isotherm at a pressure of 50 MPa, where the data point deviated from the EOS by -0.177%. However, given that the colder isotherms do not have results for 50 MPa, yet still reached deviations as high as -0.155% at 248K and 40 MPa, and -0.150% at 223K and 30 MPa, this was expected.

The specific deviations of each data point against the EOS obtained data of Nguyen et al. [9] were also plotted as shown in Fig. 7. It was clear to see that, for each of the isotherms tested, a common trend was observed where the absolute deviations were decreasing as the pressure of the fluid decreased, indicating that the EOS was possibly failing to accurately model the speed of sound at higher pressures, relative to lower pressure predictions. On the other hand, the deviation trajectory also corroborated the obtained helium results in suggesting that the acoustic resonator may have been experiencing excessive perturbations due to the resonance frequency of the vessel being approached, for which the acoustic model was failing to account for. However, contrary to the helium data in Fig. 3 and Fig. 5, the deviations from the EOS were becoming increasingly negative with increasing pressure. A possible explanation for the observation lies in the involvement of the quality factor, Q_{trans}, in the end-plate perturbation term in Eq. 14. Whilst the parameter was included to account for possible damping of the end-plates' perturbation due to the spring being used, this spring was replaced prior to calibration test-

ing. Therefore, the inclusion of the quality factor may have resulted in the end-plate correction term overestimating the perturbation caused by dissipation by the transducers' vibrations. Furthermore, deviations were seen to become more severe as the temperature decreased, suggesting failings in the EOS at these lower temperatures, but can also be attributed to the poor approximations of the calibration parameters' temperature dependence. This could be improved by running calibration experiments at several more temperatures within the appropriate temperature range.

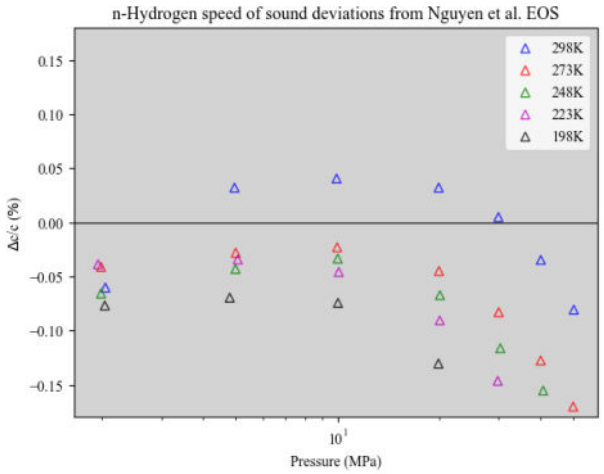


Figure 7: Deviations of n-hydrogen data from Nguyen et al.[9]

As shown in Fig. 7, a common trend was also observed where for lower pressures, there was still considerable deviation from the EOS obtained value, especially at the lower temperatures. Given the strong ability that the EOS by Leachman et al. [12] had in predicting the speed of sound for normal hydrogen at lower pressures (≤ 10 MPa), the experimental data was then compared to this EOS with the deviations plotted against pressure.

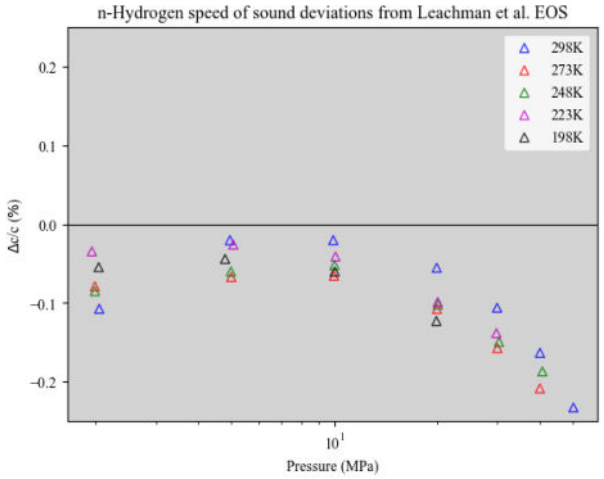


Figure 8: Deviations of n-hydrogen data from Leachman et al.[12]

As seen in Fig. 8, the use of the EOS of Leachman et al. [12] was able to improve the fit of the experimental data to the model at pressures lower than 10 MPa, especially for the lower isotherms, as well as improving the underestimation of Nguyen et al. [9] for the 298K isotherm. However, at higher pressures, the deviations were worsened slightly when using Leachman et al. [12].

4.4 Half-Widths of Resonance

Through comparison of the half-widths projected by the acoustic model via Eq. 18, to the experimentally obtained half-widths, the accuracy of the acoustic model's perturbations were evaluated, as shown in Fig. 9. The half-widths were very consistent across all 5 isotherms. For lower pressures, minimal excess half-widths were observed with values varying within $\pm 0.008\%$ for pressures below 10 MPa across all isotherms, indicating accurate prediction of dissipation at low pressures by the acoustic model. However, the magnitude of the excess in the half-widths became larger as pressures increased, with the largest deviation being 0.078%, suggesting weaker modeling at higher pressures, most likely in the end-plate perturbation. However, relative to the frequency, this was negligible. Therefore, although not as accurate as in helium, the results confirmed the accuracy of the acoustic model.

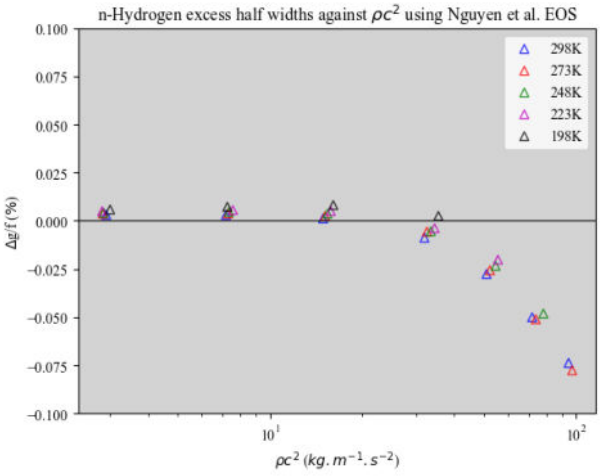


Figure 9: Excess half-widths of n-hydrogen data

4.5 Sources of Error

Although the uncertainties of the experimental data for n-hydrogen varied within $\pm 0.021\%$, there were several aspects and situations within the data recording that may have contributed to errors within the frequencies processed. One such precaution taken involved addressing the thermodynamic equilibrium within the cylindrical vessel itself during operation of the temperature bath. The cylindrical resonator was suspended within the pressure vessel, therefore reducing the rate of heat transfer to the fluid. Whilst this may not have been an issue for the higher temperature isotherms (298K, 273K and 248K), at temperatures as low as

223K, high pressure tests were found to be indicating a leak within the resonator. The most probable cause for this was postulated to originate from the o-ring seal located on the top of the resonator. This seal was made of a PTFE-graphite composite material, which was susceptible to losing elasticity at lower temperatures. This may have been aggravated due to it being the second use of the seal. Therefore, ample time was given before taking measurements to ensure thermodynamic equilibrium within the resonator, with the subsequent deviations from the EOS remaining relatively constant for each specific pressure, indicating that the results were accurate and thermodynamic equilibrium had been reached. Furthermore, higher pressure measurements were not taken for these lower temperatures. Another situation of interest refers to a potentially anomalous data point within the deviations of the experimental n-hydrogen data from the EOS of Nguyen et al. [9]. As shown in Fig. 7, a sharp drop is experienced in a particularly large deviation of -0.06% at 298K and 2 MPa.

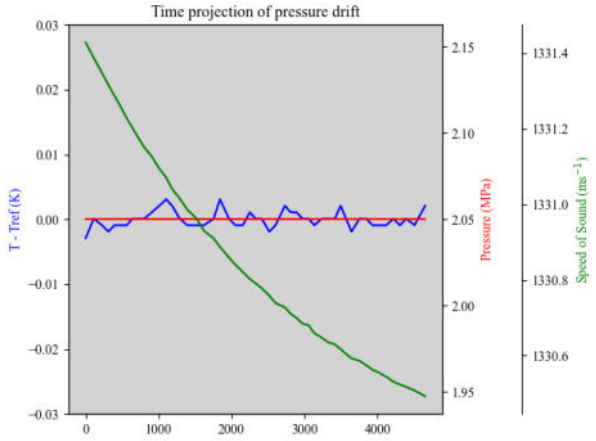


Figure 10: Time series data for n-hydrogen measurement at T = 298K and P = 2 MPa

Fig. 10 shows the time series data for the temperature, pressure and collected speed of sound over the testing conducted at the data point's conditions. The projections clearly outlined the anomalous behavior that may have caused the excessive deviation from EOS, as the speed of sound was found to consistently drop despite the temperature and pressure being constant, suggesting thermodynamic equilibrium within the resonator. However, given that for a time span of 4653 seconds ($\approx$ 1.3 hours), the speed of sound only varied within 0.07% of its original value, the extent of the deviation was considered to be minimal.

5 Conclusion

In this study, the speed of sound of 5 n-hydrogen isotherms between 198K and 298K was measured at pressures ranging from 2 MPa to 50 MPa using a cylindrical acoustic resonator. This allowed for results to fill a gap in the literature at this temperature and pressure range. The acoustic model for the resonator was calibrated by testing with helium at 198K and 298K, which yielded a linear temperature dependence for the model parameters. An uncertainty analysis was then performed on the three calibration parameters. Through this calibration, the deviation of the speed of sound in Helium from the EOS was reduced to $\pm$0.012% at 298K and $\pm$0.001% at 198K, and the excess half-widths, relative to the perturbed frequency, to $\pm$0.01%. Additionally, the repeatability of the apparatus was demonstrated with respective runs in helium varying within $\pm$0.003% at 298K. The speed of sound in n-hydrogen showed larger deviations from the EOS: within $\pm$0.177% of Nguyen et al. [9] with an AARD of 0.069% and within $\pm$0.233% of Leachman et al. [12]. The expanded uncertainty corresponding to 95% confidence ranged from 0.004% to 0.021%. The excess half-widths relative to the perturbed frequency was found to vary within $\pm$0.078% for n-hydrogen and $\pm$0.009% for helium. These results indicated that the acoustic model was performing well in accurately modelling the dissipation of the real system and the subsequent perturbations in the frequency.

As for possible future works, improvements to the calibration could be made by gathering more helium isotherm data, allowing for a more accurate representation of the model parameters with respect to temperature. Another approach could be to model the linear expansivity α as a function of temperature to then use it to calculate L and Q_{trans}. Additionally, it is necessary to determine and remedy the cause of the pressure leak at low temperatures. If it is caused by the existing sealing o-ring, it must be re-torqued, or a replacement seal must be tested. If it still fails, alternative materials like indium must be considered. Lastly, the frequency drift at low n-hydrogen pressures must be investigated.

6 Acknowledgments

The authors would like to extend their gratitude to Yessica Arellano for their continuous support and advice throughout the duration of the project.

References

[1] Behdad Shadidi, Gholamhassan Najafi, and Talal Yusaf. A review of hydrogen as a fuel in internal combustion engines. *Energies*, 14(19):6209, 2021.

[2] Ankica Kovač, Matej Paranos, and Doria Marciuš. Hydrogen in energy transition: A review. *International Journal of Hydrogen Energy*, 46(16):10016–10035, 2021.

[3] Marco Martino, Concetta Ruocco, Eugenio Meloni, Pluton Pullumbi, and Vincenzo Palma. Main hydrogen production processes: An overview. *Catalysts*, 11(5):547, 2021.

[4] Andrea Mio, Elena Barbera, Alessandro Massi Pavan, Alberto Bertucco, and Maurizio Fermeglia.

Sustainability analysis of hydrogen production processes. *International Journal of Hydrogen Energy*, 54:540–553, 2024.

[5] Nan Ma, Weihua Zhao, Wenzhong Wang, Xiangrong Li, and Haiqin Zhou. Large scale of green hydrogen storage: Opportunities and challenges. *International Journal of Hydrogen Energy*, 50:379–396, 2024.

[6] Mitsuru Matsumoto and James H Espenson. Kinetics of the interconversion of parahydrogen and orthohydrogen catalyzed by paramagnetic complex ions. *Journal of the American Chemical Society*, 127(32):11447–11453, 2005.

[7] Roy W Harkness and W Edwards Deming. The equilibrium of para and ortho hydrogen. *Journal of the American Chemical Society*, 54(7):2850–2852, 1932.

[8] Christian Günz. Good practice guide to ensure complete conversion from para to normal hydrogen of vaporized liquified hydrogen. *PTB-OAR*, 2022.

[9] Tan-Trieu-Giang Nguyen, Carsten Wedler, Sven Pohl, Dan Penn, Roland Span, J.P. Martin Trusler, and Monika Thol. Experimental speed-of-sound data and a fundamental equation of state for normal hydrogen optimized for flow measurements. *The Journal of Chemical Thermodynamics*, 198:107341, 2024.

[10] John Shipley Rowlinson and DJ Tildesley. The determination of the gas constant from the speed of sound. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 358(1694):281–286, 1978.

[11] BA Younglove. Speed of sound in fluid parahydrogen. *The Journal of the Acoustical Society of America*, 38(3):433–438, 1965.

[12] Jacob W Leachman, Richard T Jacobsen, SG Penoncello, and Eric W Lemmon. Fundamental equations of state for parahydrogen, normal hydrogen, and orthohydrogen. *Journal of Physical and Chemical Reference Data*, 38(3):721–748, 2009.

[13] C Carey, EH Carnevale, S Uva, and T Marshall. *Experimental Determination of Gas Properties at High Temperatures and/or Pressures*. National Technical Information Service, 1974.

[14] FW Giacobbe. Precision measurement of acoustic velocities in pure gases and gas mixtures. *The Journal of the Acoustical Society of America*, 94(3):1200–1210, 1993.

[15] Alfred H Hodge. An experimental determination of ultrasonic velocity in several gases at pressures between one and one hundred atmospheres. *The Journal of Chemical Physics*, 5(12):974–977, 1937.

[16] DH Liebenberg, RL Mills, and JC Bronson. Measurement of p, v, t, and sound velocity across the melting curve of n-h 2 and n-d 2 to 19 kbar. *Physical Review B*, 18(8):4526, 1978.

[17] DH Liebenberg, RL Mills, and JC Bronson. Thermodynamic properties of fluid nh/sub 2/in the range 75 to 307 k and 2 to 20 kbar. Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 1977.

[18] W Van Dael, A Van Itterbeek, A Cops, and J Thoen. Velocity of sound in liquid hydrogen. *Cryogenics*, 5(4):207–212, 1965.

[19] L Ruffine and JPM Trusler. Sound-speed sensor for gas pipeline applications. *International Journal of Thermophysics*, 30:1106–1117, 2009.

[20] Martin Trusler. *Physical acoustics and metrology of fluids*. CRC Press, 2020.

[21] JT Zhang, Hong Lin, JP Sun, XJ Feng, KA Gillis, and MR Moldover. Cylindrical acoustic resonator for the re-determination of the boltzmann constant. *International Journal of Thermophysics*, 31:1273–1293, 2010.

[22] Diego Ortiz Vega, Kenneth Hall, James Holste, Allan H Harvey, and Eric W Lemmon. An equation of state for the thermodynamic properties of helium. *Technical Report No. NISTIR 8474*, 2023.

[23] BA Hands and VD Arp. A correlation of thermal conductivity data for helium. *Cryogenics*, 21(12):697–703, 1981.

[24] Vincent D Arp and Robert D McCarty. Thermophysical properties of helium-4 from 0.8 to 1500 k with pressures to 2000 mpa. Technical report, 1989.

[25] Marc J Assael, ML Huber, RA Perkins, Y Takata, et al. Correlation of the thermal conductivity of normal and parahydrogen from the triple point to 1000 k and up to 100 mpa. *Journal of Physical and Chemical Reference Data*, 40(3), 2011.

[26] Chris D Muzny, Marcia L Huber, and Andrei F Kazakov. Correlation for the viscosity of normal hydrogen obtained from symbolic regression. *Journal of Chemical & Engineering Data*, 58(4):969–979, 2013.

[27] Defne Demirdesen, Carsten Wedler, Robert E Low, and JP Martin Trusler. Speed of sound in gaseous 1, 1-difluoroethene (r1132a) at temperatures between 193 k and 383 k at pressures up to 3 mpa. *International Journal of Thermophysics*, 44(6):90, 2023.

Analysis of Policies to Facilitate Heat Pump Deployment for Building Decarbonisation

Hannah Corbett, Macey Utting

December 12, 2024

Abstract

This paper examines the changes to heat pump subsidies needed in order to deploy the required number of heat pumps to meet the UK's climate targets, aligning with both national and global targets, importantly, decarbonising the UK's domestic sector by 2050. Furthermore, decarbonisation will occur through exploring technological advancements, policy frameworks and cost trajectories necessary for widespread adoption. Also, the role of potential government interventions such as proportional capital subsidies and electricity price subsidies were explored, alongside the current fixed capital subsidy (the Boiler Upgrade Scheme) in overcoming current market barriers. The findings suggest that with strong policy support and industry collaboration, heat pumps can play a transformative role in decarbonising the UK's heating sector. The transition will take time to implement as upgrades to the national grid will simultaneously need to occur. Data found has shown that with the current £7.5k Boiler Upgrade Scheme air source heat pump deployment is predicted to be 15% in 2030 and 86% in 2050 both a natural gas price of 50 £/MWh which is the current, rounded-up value. It is anticipated that the initial investments into heat pumps will pay-off with long-term benefits for society, the environment and the UK's economy.

1 Introduction

The overall objective for this report is exploring how policies would impact heat pump (HP) deployment. HPs are focused on despite there being other domestic heating candidates available because of the strong emphasis on HPs in the UK's 2021 Towards Net Zero report [6]. Heating accounts for almost $\frac{1}{3}$ of the UK's total carbon emissions [19]. The UK government has set an ambitious target of installing 600,000 HPs annually by 2028 with hopes to increase the annual installation rate to 1.7 million by 2035 [6]. This is part of the plan to reach the Paris Agreement target of limiting the increase in the global average temperature to below 2°C above pre-industrial levels and ideally, pursue efforts to further limit the temperature increase to 1.5°C [25]. Also, the UK has made a legally-binding pledge to reach net zero emissions by 2050, alongside 106 other countries that have made similar pledges, therefore to meet this, other types of heating technologies need to be deployed, replacing gas boilers [14]. An intermediary target to this goal is that all of the UK's electricity will come from clean sources and that all new heating appliances installed will be low carbon technologies by 2035. Both types of HPs are more efficient than gas boilers and they are carbon neutral if the electricity used is decarbonised, which will be the case if the UK's clean energy by 2035 target is met. To encourage the uptake of more HPs, it is expected that a HP will cost the same to buy and run as a gas boiler in this decade [6]. Additionally, in 2035, it is expected that there will be 13 million low carbon homes which includes 7-11 million homes with HPs [6]. This means that changes to current subsidies and levies need to be considered to further incentivise the population to install HPs to meet the ambitious targets set.

Being a slower adopter of HPs compared to the rest of Europe, the UK can be guided from the successful implementation of HPs in Norway, Finland and Sweden. In Norway, 60% of buildings are equipped with HPs [20], with Sweden and Finland at over 40%, showcasing examples that HPs can work effectively in colder climates [21]. In October 2023, the UK government further increased the subsidy for HPs via the Boiler Upgrade Scheme (BUS) from £5k for air source heat pumps (ASHPs) and £6k for ground source heat pumps (GSHPs) to £7.5k for both HP types to aid the adoption [8]. In comparison to other countries, the UK has one of the more generous subsidies available worldwide. However, a few

countries offer over £12k as a subsidy for example, France (for GSHPs), Germany and Lithuania, and also, Switzerland offers subsidies of up to £18.5k for switching from an electric boiler to a GSHP [4]. Alongside the BUS, the UK government has created other initiatives that aid in the deployment of HPs including the Public Sector Decarbonisation Scheme (PSDS), the Heat Pump Investment Accelerator Competition (HPIAC) and the Clean Heat Market Mechanism (CHMM) (set to launch in April 2025) which creates financial incentives for HP manufacturers and drives economies of scale.

Additionally, the government has committed to providing a further £6b in public funding from 2025-2028 including £450m for the BUS in England and Wales [7]. All new buildings from 2025 will require no further retrofit as regulations such as the Future Homes Standard and the Future Buildings Standard enforce that new builds will be zero carbon ready [18]. Another initiative that the government could implement would be to add a carbon price to energy retailers on gas. In the present, there are no carbon prices applied to heating homes, however, adding a carbon price would yield further carbon emission reductions through charging suppliers and consequently, consumers higher prices. In implementing this, supporting the lowest-income households in society would need to be taken into consideration. The UK Emissions Trading Scheme (ETS) which applies to energy-intensive industries, power generation and the aviation sector, could be one way to instil the carbon price aforementioned. It follows the 'cap and trade' approach where a maximum cap is set on the total emissions per year. Importantly, every year the overall cap is reduced which means the cost of releasing the same quantity of emissions increases as time progresses [11]. Adding the domestic heating sector to this scheme could assist in its decarbonisation. A tax on natural gas could play an effective role in addressing the carbon price gap between electricity and gas. Options for how to use the revenue generated from a carbon tax could be explored to rebalance the electricity to gas price ratio so it is lower than four [18].

Nordic countries were early adopters of HPs and Norway ensured to make HPs cost competitive to shift their residents away from the use of fossil fuel heating after the 1973 oil crisis [16]. As well, Norwegians have well-insulated homes even when temperatures drop to minus degrees in the winters. In 2024, there is progress seen in the UK's adoption of the 'HP revolution' with the total number of all-time certified HP installations surpassing 250,000 as found by Microgeneration Certificate Scheme (MCS), the UK's quality mark for small-scale renewable energy installations [23]. Critics state that even with the BUS, switching to a HP from a gas boiler could entail higher energy bills due to the larger price of electricity needed to run the HPs. It is has been found that a HP with a Seasonal Coefficient of Performance (SCOP) greater than three will match the running costs of an 85% efficient gas boiler [1]. A study comparing the running costs of HPs with an SCOP of 2.9 to an A-rated gas boiler concluded that the operating costs of the HP would be £20 more a year [1].

Regarding, the type of HP one should decide upon, it depends on factors including the type of property (retrofit or new-build) the HP lifetime, consumer preference and cost. A large majority of homeowners favour ASHPs if choosing to shift to a HP due to the lower initial investment costs. On the other hand, GSHPs are more efficient. A HP operates and their efficiency compared to gas boilers stems from the utilisation of a refrigeration cycle. The outside fan of a HP draws the air over a coil containing a refrigerant. The refrigerant absorbs heat from the outside air, causing it to evaporate into a gas. The gas is passed through a compressor to convert it into a high pressure, high temperature gas. This then passes through a heat exchanger, heat is transferred to water which then circulates through radiators in the household. The refrigerant condenses back into a liquid and is passed through an expansion valve to return it back to its original state. GSHPs work in a similar way except they absorb thermal energy from the ground instead of the air [10]. This means they yield a higher SCOP than ASHPs due to the ground temperatures remaining more stable than the air temperatures, which fluctuate throughout the seasons. This makes them more cost efficient, as less electricity is needed for the same energy demand, giving them a lower carbon footprint. Also, GSHPs have spatial requirements, needing on average of 2.5 times the square meterage of the house [2]. In many cases, this makes ASHPs the more practically viable option. This report will explore three key domestic heating technologies including, gas boilers, ASHPs and GSHPs. Data has been produced to critically evaluate the three against each other and to compare which technology is more suitable to adapt on a large scale.

Further action to implement HPs nationally is needed to be carried out to ensure this transition occurs successfully in the next 25 years to meet net zero by 2050, as pledged by the UK in 2019 [14]. The initial transition will be the hardest, as the demand for electricity will rise rapidly and could likely put strain on the national grid. Subsequently, energy companies

will need to swiftly adapt to a larger proportion of their customers having higher electricity demands and potential added grid connection costs to help pay for the overall upgrade costs.

2 Methodology

Providing a network with optimised systems is complex although, yields more accurate results. Integrating two Python models together is valuable, as combining different optimised models leads to an overall more sophisticated solution. The two Python models utilised are; a Household-level Optimisation and a Heat energy System Optimisation (HeSO) model. Figure 1 shows an overview of the two models and their interaction, including key input data and outputs.

2.1 Household-level Optimisation Model

The Household model's function is to optimise the technological design and operation of the HPs, it uses non-linear programming optimisation to do this. The optimised HPs and existing gas boiler data make up the domestic heating candidates considered in this report and are inputted into the HeSO model for different buildings [22].

For the detailed design of HPs, different heat transfer equations, refrigerants, compressor and heat exchanger models are consulted. All the heat exchangers are modelled as liquid to liquid however, as the ASHP heat exchanger is air to liquid, this cannot be modelled the same as the GSHP heat exchanger and this has been accounted for by applying a fine-tuned 5 K penalty to the ASHP heat exchanger. On the gas side of the heat exchanger of the ASHP unit, the heat transfer coefficients are lower and therefore, heat transfer is not as efficient compared to a liquid to liquid heat exchanger. Also, air temperatures are less stable than ground temperatures, which adds to their difference in performance. Apart from this, both HP types are modelled similarly. Gas boiler data has been gathered and inputted into the model [17]. Gas boilers are not specifically optimised in this model as they are a mature technology and therefore, their design has previously been optimised and a standard exists in industry.
Data sourced from the Cambridge Housing Model [5], on the types of households that make up the 28 million in the UK and their corresponding heat energy demands is fed into a clustering algorithm. From this clustering algorithm (k-medoids), five representative buildings are chosen as shown in Table 1 and their annual heat demands. The five representative buildings are scaled in proportion to the number of each of those types that represent the overall UK housing stock. As only a small percentage of new houses are built annually in the UK, new houses are not considered in this model and also, most new build properties are not eligible for the UK BUS [12].

Building Number	Annual Heat Demand in 2020 (kWh)	Number of Buildings in the UK Stock (million)
1	7,100	11.3
2	11,800	10
3	17,000	4.7
4	26,500	1.5
5	39,100	0.3

Table 1: Overview of the representative buildings considered in the HeSO model

2.2 HeSO Model

The HeSO model's function is to optimise the deployment of different heating technologies and power in different sectors to meet government decarbonisation targets by 2050.
The HP penetration is assessed in different scenarios varying the following inputs: natural gas price, grid upgrade cost, proportional capital expenditure (CAPEX) subsidy, fixed CAPEX subsidy and electricity price subsidy. Policies are examined via the subsidy variable inputs (proportional CAPEX, fixed CAPEX and electricity price). The grid upgrade cost or grid cost factor (GCF) refers to the costs associated with upgrading the national electricity grid to cope with the increase in electricity demand due to shifting from gas boilers to electrically powered HPs. In the present, the national grid is undertaking the 'Great Grid Upgrade' this will aid in a smoother transition to 100% clean energy by 2035 [9]. Increasing the number of transmission lines, pylons, transformers and substations are examples of how the grid is being expanded.

The HeSO model provides optimal solutions for the technology deployment from 2020 to 2050 using gas boilers, HPs and other clean energy technologies, including hydrogen boilers, in the hydrogen, electricity and heating (commercial, industrial and domestic) sectors. The lifetime of a HP is assumed to be 20 years and the lifetime of a gas boiler is 15 years. In addition, at the start of the model, 2020, it is assumed out of the total 28 million households that $\frac{1}{4}$ have gas boilers that are new or 0 years old, $\frac{1}{4}$ have 5 year old gas boilers, $\frac{1}{4}$ 10 year old gas boilers and $\frac{1}{4}$ have

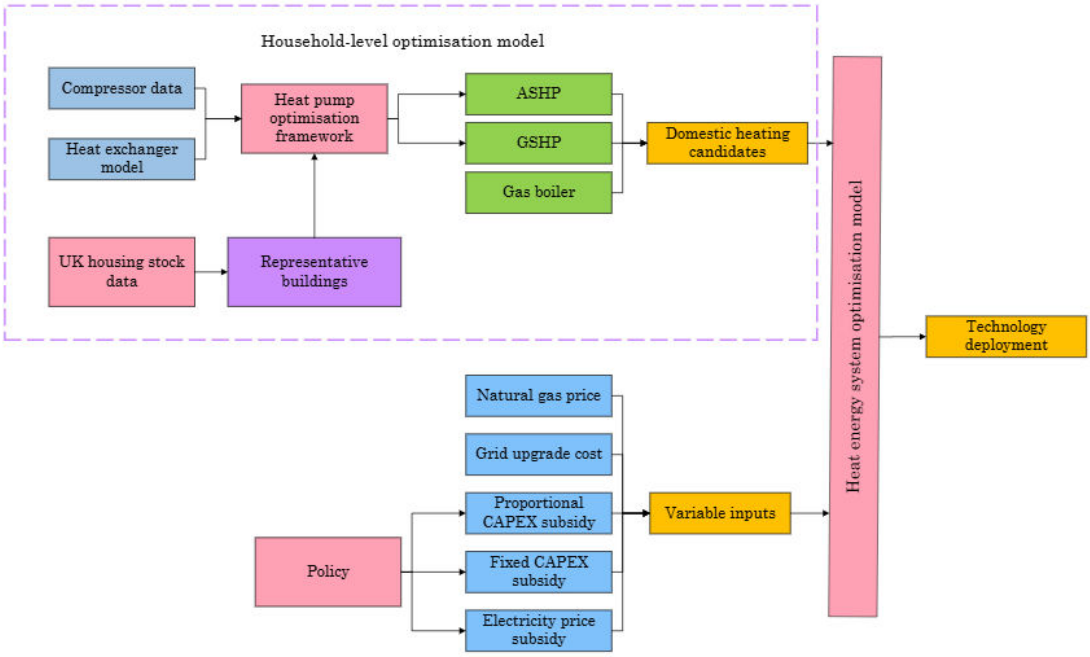


Figure 1: Flow diagram showing the integration of the household-level optimisation model and the HeSO model

15 year old or gas boilers at the end of their lifetime [12]. Also, the user can choose a number of typical days which the model uses to represent the average heat demand of the days in a year, usually a number between 4 to 12 is used. In this study, six typical days were used. In addition, the model accounts for extraordinary circumstances by having four additional extreme days which represent the maximum electricity demand, the maximum domestic heat demand, the maximum commercial heat demand and the day with the lowest average availability of renewables.

The HeSO model uses linear programming optimisation to minimise the cost of transitioning the national energy system while meeting energy demands and policies. The model assumes that the total UK heat demand will decrease linearly by 12% until 2050 as buildings are expected to increase in energy efficiency. In addition, the model accounts for a linear reduction in permissible carbon dioxide emissions every five years up until net zero in 2050 [12]. Each household has up to two heating technologies (a primary and a back-up heating technology) and a thermal energy store, only practical technology combinations are considered in the model. Practical technology combinations always include at least one technology but there is the option to have 'no technology' as the second option. The heating technologies include: gas boilers, electric boilers, hydrogen boilers, ASHPs and GSHPs.

3 Results and Discussion

Natural gas price, grid upgrade cost, proportional CAPEX subsidy, fixed CAPEX subsidy and electricity price subsidy are varied in the HeSO model to understand the policies required for the targeted deployment of HPs. Due to the negligible deployment percentages of electric boilers and hydrogen boilers being deployed, these technologies have not been plotted. Although the current UK HP subsidy is fixed, regardless of HP type, switching to a proportional CAPEX subsidy was investigated to compare and contrast the impact of this on domestic heating technology deployment. Subsidies were explored as opposed to taxes on gas prices due to wanting to add an incentive to attract customers to buy a HP instead of adding a penalty on the usage of gas for heating that may not necessarily improve the uptake of HPs. Gas prices have been varied from 0 £/MWh to 200 £/MWh in intervals of 25 £/MWh for each plot. Electricity price subsidies have been varied from 0 £/MWh to 150 £/MWh in intervals of 25 £/MWh for each plot. Fixed CAPEX subsidies have been varied from £0k to £10k in intervals of £2.5k. The proportion of CAPEX the consumer pays has been varied from 0 to 80% in 20% intervals.

Figure 2 consists of sub-plots with one subplot for each of the three technologies investigated: gas boilers, ASHPs and GSHPs. It displays the gas price against fixed subsidy values with each square of the heat map coloured ac-

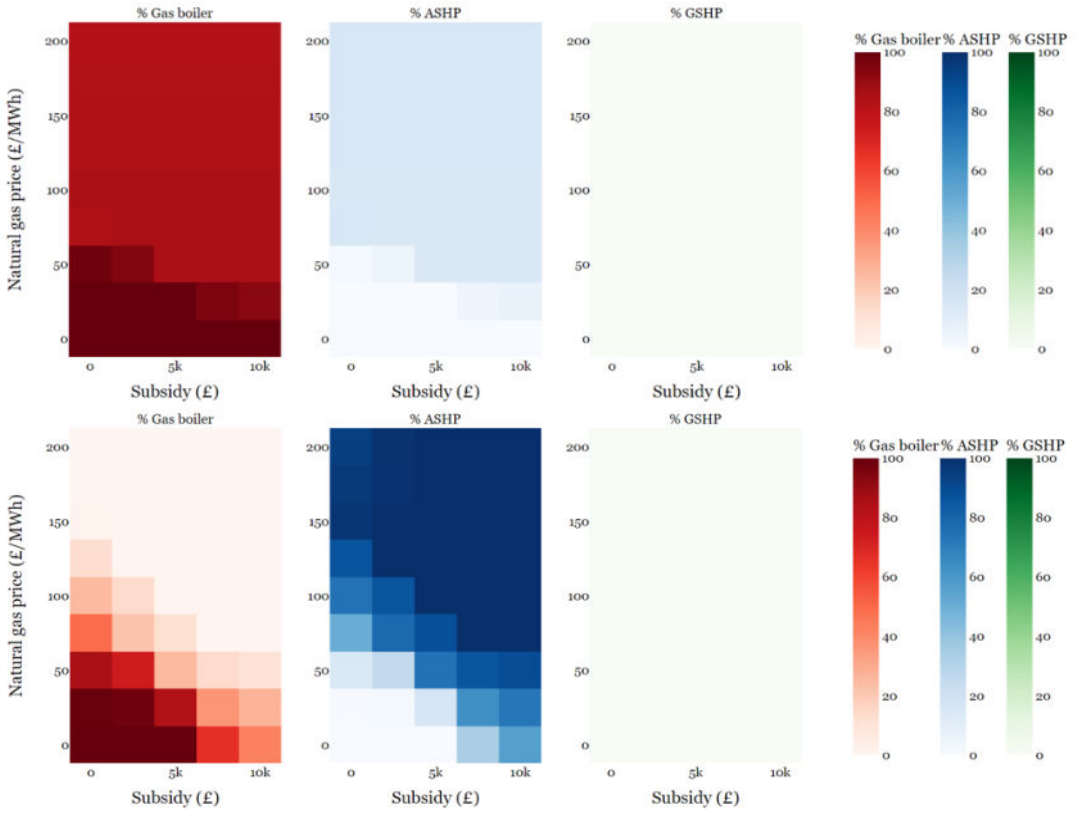


Figure 2: Natural gas price against fixed subsidy values for domestic heating technology deployment in 2030 and 2050

cording to the strength of the deployment of each specific technology, with darker shades of a colour representing a higher deployment percentage of that technology as shown in the legends of the plots. There is a stark contrast between 2030 and 2050 for the predicted deployment of HPs and gas boilers. In 2030, there is a much stronger deployment of gas boilers at 85% compared to 15% of ASHPs being deployed at the current, rounded-up gas price of 50 £/MWh (the current UK gas price as of November 2024 is 40.14 £/MWh (118 GBp/thm) [3]) and the fixed £7.5k subsidy. At gas prices between 0 £/MWh and 25 £/MWh and fixed subsidies between £0k and £5k, there is less than a 1% deployment of ASHPs and 99% deployment of gas boilers. The lower the available subsidies for HPs, the greater the reliance on gas boilers, especially, when natural gas prices are low which is equivalent to the current UK domestic heating scene where 73% of households rely solely on gas boilers for central heating [24]. Also, this suggests it is not economically viable in 2030 to switch from a gas boiler to a HP in the majority of cases and the transition will mainly occur within the 20 year period between 2030 and 2050. In the timescales, it is not required to replace existing gas boilers before the end of their lifetimes and another factor for the difference in deployment between the two plots is that it takes time for transitions to occur on a nation-wide scale. Furthermore, there are build-rate constraints, putting an upper limit on the number of HPs that can be installed each year as the infrastructure to manufacture the HPs is still developing. The upper limit of the number of HPs that can be installed each year will increase over time as the manufacturing capacity is increased. Whereas, in 2050, there is a definite increase in the proportion of ASHPs deployed.

At fixed subsidy, ASHPs are substantially more dominant than GSHPs. This is because the difference in price between the two technologies remains constant for the same fixed CAPEX subsidy. This outweighs their difference in efficiency and therefore, electricity consumption, meaning ASHPs are still by a large amount, the affordable option. In 2050, at £0k subsidy, and a natural gas price of 75 £/MWh, the percentage deployment of ASHPs exceeded 50% and for all further subsidies at this gas price the percentage deployment was higher than this. Key anchor points for the transition to ASHPs dominating are at a gas price of 100 £/MWh, fixed subsidy of £0k (ASHP deployment is 75%), at a gas price of 50 £/MWh, fixed subsidy of £5k

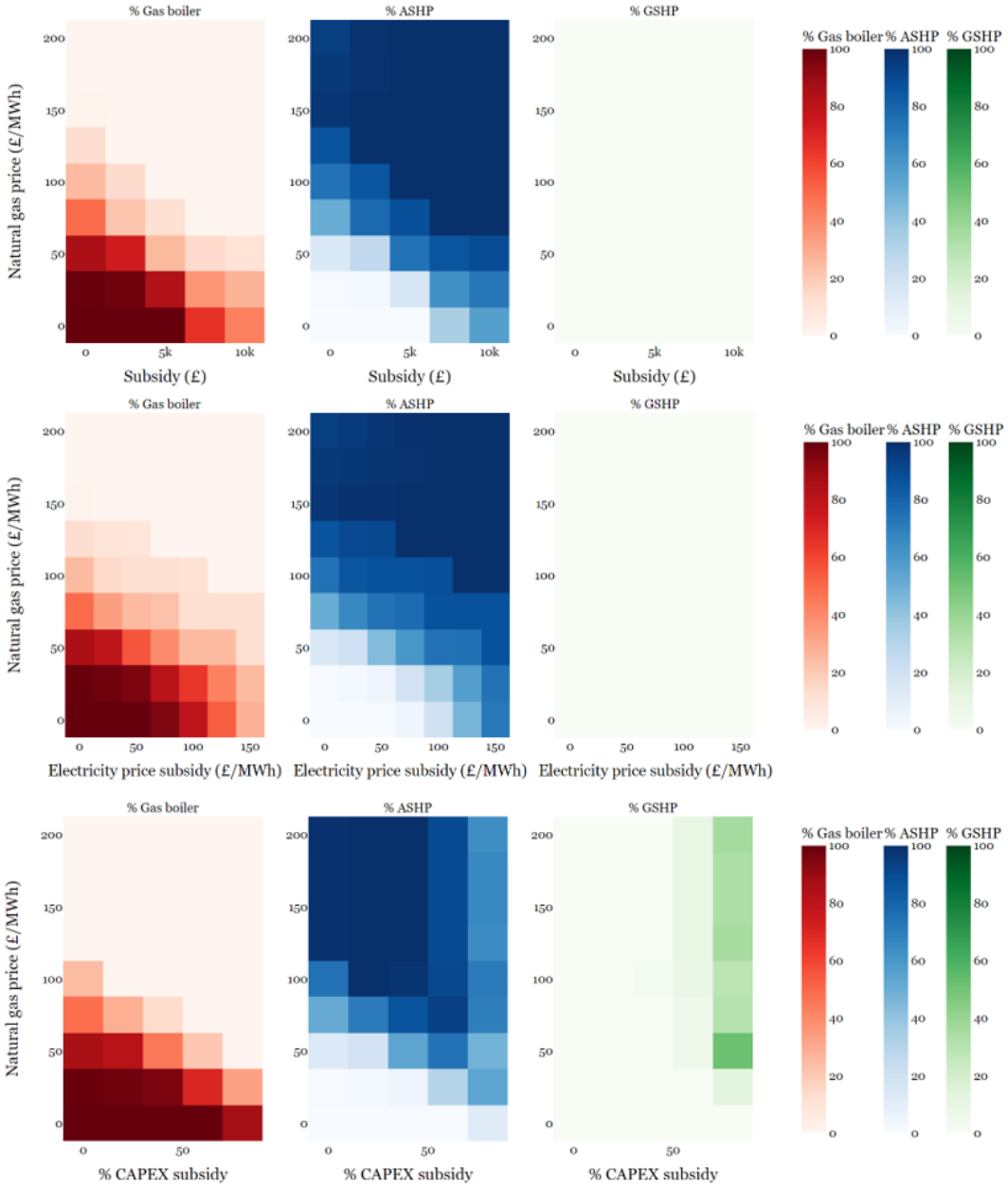


Figure 3: Natural gas price against fixed subsidy values, natural gas price against electricity price subsidy and natural gas price against proportional CAPEX subsidy for domestic heating technology deployment in 2050

(ASHP deployment is 75%) and at a gas price of 25 £/MWh, fixed subsidy of £7.5k (ASHP deployment is 63%) as shown in Figure 2.

The relationship between the natural gas price and the electricity price subsidy is displayed by the middle plot Figure 3. It is assumed that electricity price subsidy would be provided by the government to reduce the huge difference between electricity and gas costs, in the UK, currently, electricity costs can be 4 times or more expensive than gas costs [13]. At high natural gas price and high electricity price subsidy, a greater percentage deployment of ASHPs is observed. At a natural gas price of 150 £/MWh to 200 £/MWh, the percentage deployment of ASHPs is greater than 94% at every square in the region. The transition points between a majority of gas boilers deployed to a majority of HPs being deployed occurs at a gas price of 50 £/MWh, electricity price subsidy of 75 £/MWh (ASHP deployment is 60%), a gas price of 75 £/MWh and electricity price subsidy of 25 £/MWh (ASHP deployment is 66%) and at a gas price of 100 £/MWh and electricity price subsidy of 0 £/MWh (ASHP deployment is 75%).

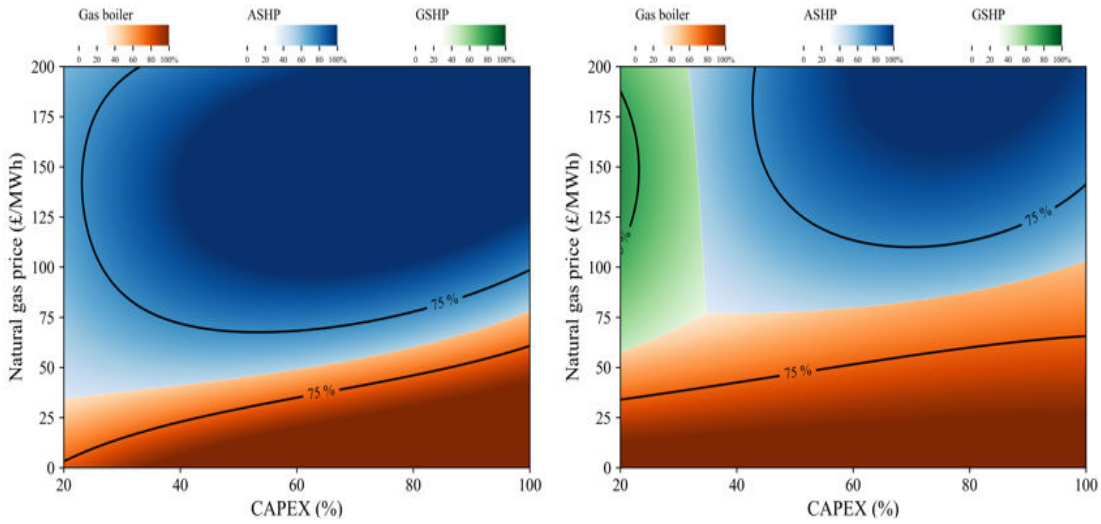


Figure 4: Natural gas price against CAPEX percentage graphs representing the competitive landscape of the domestic heating sector at grid cost factors of 0 and 750 m£/GW

Grid Cost Factor	CAPEX(%)	Natural Gas Price (£/MWh)
0	N/A	N/A
250	23.03	58.13
500	30.55	55.83
750	34.77	77.50

Table 2: Triple points for competitive landscape graphs

Grid Cost Factor	CAPEX(%)	Natural Gas Price (£/MWh)
0	100	78.42
250	100	84.21
500	100	92.31
750	100	103.28

Table 3: Boundary end points for competitive landscape graphs

Given that this case does not include any reduction in CAPEX for the HPs, ASHPs strongly dominate over GSHPs, and the deployment of GSHPs is negligible. The lower efficiency of ASHPs means they use more electricity per unit of energy produced than GSHPs, giving a higher operating expenditure (OPEX). As the electricity price subsidy rises, it decreases the operating costs of both HPs and the difference in cost between the two HPs. The great difference in CAPEX between ASHPs and GSHPs outweighs any savings from using a GSHP with a slightly lower OPEX than an ASHP.

Heat maps are produced for the competitive landscape to compare the three heating technologies as shown in Figure 4. Instead of a sub-plot for each heating technology like above in Figures 2 and 3, the sub-plots have been merged together. Only the dominant heating technology and its percentage deployment are visible and are plotted for specific gas prices and percentages of CAPEX that the consumer would pay. The black lines represent the boundary where the dominant heating technology has an adoption of 75% at a particular natural gas price and proportional CAPEX subsidy. The GCF was varied at 0, 250, 500 and 750 m£/GW.

The results highlight that GSHPs become more competitive with increasing GCF due to the addition of the green section on the plot for a GCF of 750 m£/GW. Using GSHPs decreases the peak electricity demand compared to ASHPs as they are more energy efficient, lessening the effect of the high grid cost. This is because if the customer was charged a higher electricity price, this favours purchasing a more efficient heating technology, which has a higher SCOP, and will not consume as much electricity compared to a less efficient technology and will reduce the total electricity demand on the grid at peak times. At higher grid costs and lower proportional CAPEX subsidies, the gas boiler deployment increases as the OPEX is less expensive than that of a HP and it becomes a more favoured option hence, a

larger red section of the graph is exhibited.

The triple point is defined as where each technology has an equal deployment of $\frac{1}{3}$. The locations of the triple points at each GCF tested can be viewed in Table 2. These points occur at higher CAPEX percentages for increasing GCFs. This is due to an increase in the deployment of GSHPs being visible at a lower proportional CAPEX subsidy. The boundary points at 100% proportional CAPEX subsidy between the two dominant technologies of ASHPs and gas boilers are shown in Table 3. At higher GCFs, the boundary point values increment due to a higher proportion of gas boilers being deployed for an economical advantage, to reduce the total electricity demand and reduce the pressure on the national grid.

4 Conclusions

This study analyses the future competitive landscape of gas boilers and HPs, with HPs as the main route to pave the way to net zero in the UK. Results show that the transition to HPs will not be instantaneous and will need to occur gradually, with the turning point for most households likely being between 2030 and 2050 with ASHP deployment predicted to be 15% in 2030 and 86% in 2050 at the rounded-up gas price of 50 £/MWh and at the fixed £7.5k BUS subsidy. Factors including waiting until existing gas boilers reach the end of their lifetimes, the building of public trust in HP technologies and the government providing adequate subsidies will affect how quickly the UK transitions to HP. It could be worthwhile to increase the BUS value in the short-term to accelerate the uptake of HPs significantly to improve the chances of meeting the 2028 HP deployment target. In 20 years, the subsidy could be re-evaluated as the data shows ASHPs will dominate at lower subsidies for example, £5k.

Despite the fact that GSHPs are more efficient than ASHPs, under most conditions, this does not offset the higher capital costs, concluding that in many cases ASHPs are the more favourable HP type. Reasons why a consumer may opt for an ASHP over a GSHP include that ASHPs have a lower spatial requirement, an easier installation process and are less expensive compared to GSHPs [15]. However, GSHPs have the potential to be favoured when substantial proportional subsidies are available and therefore, currently, the government is not promoting the choice of a more energy efficient heating system with the BUS to meet the 2050 net zero carbon emissions target.

Given the conclusion that fixed CAPEX subsidies favour ASHPs, it is recommended that the BUS is reconsidered so that more efficient heating systems, are not being overlooked. The results show that proportional CAPEX subsidies give deployment of both GSHPs and ASHPs at GCF of 750 m£/GW, therefore a proportional subsidy should be further looked into. To meet decarbonisation targets, the UK domestic heating sector should have a higher proportion of HPs compared to the current level. With the addition of an electricity price subsidy of 75 £/MWh and at a rounded-up, current natural gas price of 50 £/MWh, it is predicted that ASHP deployment would be the dominant heating technology at 60%. In addition, for the transition to HPs to be successful, each domestic user will need information from HP installers on which HP is best suited to their property's needs.

Natural gas prices are not constant, they fluctuate over time and can be affected significantly by global events. This means some residents struggle to pay their gas bills at times when the natural gas price is extremely high. Therefore, HPs are not a viable, affordable option for these people because of the high upfront costs. Subsidies have been heavily focused on over taxes in this report, as if the government enforces additional costs on residents to switch over to a HP, there may be societal backlash and could cause barriers which would not favour the transition to HPs and thus, hinder progress towards the net zero goals.

5 Outlook

Given a longer timescale for this project, there are a few areas that would be valuable to look in to. It would be beneficial to research further into the psychology behind choosing a heating technology, as in practice, the assumption that the user will choose a HP for their property rationally is not always true. There are various factors involved in the decision-making process including social pressures, spatial and financial constraints and the type of property the homeowner. Additionally, improvements to the property that would need to be considered for a HP to perform well include insulation, window glazing and switching to lower temperature radiators (to suit the settings of a HP to operate for longer periods but at lower temperatures compared to traditional radiators heated with gas). All of this needs to be assessed so a decision can confidently be made by the homeowner on the best heating technology to deploy. Therefore, more research on this would provide insight into what else the government could do

to accelerate the transition to HPs. Following from this, there would be benefit from looking into the present level of understanding of the general public on HPs, and then creating more public awareness schemes to improve peoples' knowledge and confidence in them. In countries where HP deployment has been successful, such as Nordic countries, it would be useful to learn the strategies that have worked in their public awareness schemes in order to tailor it to the UK to also make the promotion of HPs successful and ensure information is accessible and available.

This study focuses on ASHPs, GSHPs and gas boilers and exploring the effect of policies on HP deployment. To get a full picture of how the domestic heating sector can be decarbonised, a wider range of lesser-promoted heating technologies, such as hydrogen, pellet and biomass boilers and solar thermal collectors as additional heating candidates should be researched in further detail to determine whether they are also successful on a large scale so the full picture of the energy system is accounted for.

6 Acknowledgements

We would like to thank our supervisors, Matthias Mersch and Dr Paul Sapin for their help, valuable insights and contribution to the project. Also, we would like to thank Christos Markides for overseeing and supervising the project.

References

[1] J. Ambrose. Are heat pumps the future or just a lot of hot air? https://www.theguardian.com/environment/2024/sep/28/heat-pumps-future-hot-air-labour-home-upgrade, 2024. Date Accessed: 7th December 2024.

[2] Ollie Creevy. How much space is required for a ground source heat pump? https://heat-pumps.org.uk/how-much-space-is-required-for-a-ground-source-heat-pump/, 2022. Date Accessed: 11th November 2024.

[3] Trading Economics. Uk natural gas | 2021 data | 2022 forecast | 2020 historical | price | quote | chart. https://tradingeconomics.com/commodity/uk-natural-gas, 2023. Date Accessed: 22nd November 2024.

[4] European Heat Pump Association (EHPA). Subsidies for residential heat pumps in europe. https://www.ehpa.org/news-and-resources/publications/subsidies-for-residential-heat-pumps-in-europe/, 2023. Date Accessed: 22nd November 2024.

[5] Department for Business Energy Industrial Strategy (BEIS). Cambridge housing model and user guide. https://www.gov.uk/government/publications/cambridge-housing-model-and-user-guide, 2015. Date Accessed: 12th December 2024.

[6] UK Government. Uk net zero strategy: Build back greener. https://assets.publishing.service.gov.uk/media/6194dfa4d3bf7f0555071b1b/net-zero-strategy-beis.pdf, 2021. Date Accessed: 4th December 2024.

[7] UK Government. Heat pump investment roadmap leading the way to net zero. https://assets.publishing.service.gov.uk/media/649d690406179b000c3f751c/heat-pumps-investment-roadmap.pdf, 2023. Date Accessed: 11th December 2024.

[8] Gov.UK. Changes to the boiler upgrade scheme. https://www.gov.uk/government/publications/boiler-upgrade-scheme-changes-to-grant-levels/changes-to-the-boiler-upgrade-scheme-october-2023, n.d. Date Accessed: 7th December 2024.

[9] The National Grid. The great grid upgrade. https://www.nationalgrid.com/the-great-grid-upgrade, n.d. Date Accessed: 30th November 2024.

[10] h2x Engineering. How do air source heat pumps work? https://www.h2xengineering.com/blogs/air-source-heat-pumps/, n.d. Date Accessed: 10th December 2024.

[11] Energy Advice Hub. The uk emissions trading scheme: frequently asked questions. https://energyadvicehub.org/the-uk-emissions-trading-scheme-frequently-asked-questions/, 2022. Date Accessed: 28th November 2024.

[12] M. Mersch. Multi-scale energy system optimisation for efficient, affordable and secure net-zero transitions, 2024. Date Accessed: 5th December 2024.

[13] J. Minihane. Which is cheaper: gas or electric? https://www.moneysupermarket.com/gas-and-electricity/is-gas-or-electric-cheaper/, 2024. Date Accessed: 1st December 2024.

[14] United Nations. For a livable climate: Net-zero commitments must be backed by credible action. https://www.un.org/en/climatechange/net-zero-coalition, 2024. Date Accessed: 3rd December 2024.

[15] E.ON Next. Guide to ashps gshps. https://www.eonnext.com/air-source-heat-pumps/guides/ground-source-vs-air-source, 2020. Date Accessed: 7th December 2024.

[16] A. Niranjan. 'you can walk around in a t-shirt': how norway brought heat pumps in from the cold. https://www.theguardian.com/environment/2023/nov/23/norway-heat-pumps-cold-heating, 2023. Date Accessed: 5th December 2024.

[17] A.V. et al Olympios. Library of price and performance data of domestic and commercial technologies for low-carbon energy systems. https://doi.org/10.5281/zenodo.5758943, 2021. Date Accessed: 3rd December 2024.

[18] LSE Grantham Research Institute on Climate Change, the Environment, and Catapult Energy Systems. Extending the uk emissions trading scheme to heating and road transport fuels: What role can it play in decarbonising the uk economy? https://www.lse.ac.uk/granthaminstitute/wp-content/uploads/2024/05/Extending-the-UK-ETS_What-role-can-it-play-in-decarbonising-domestic-heating-and-road-transport.pdf, 2024. Date Accessed: 17th November 2024.

[19] Parliament.uk. Decarbonising heat in homes. https://publications.parliament.uk/pa/cm5802/cmselect/cmbeis/1038/report.html, 2021. Date Accessed: 27th November 2024.

[20] Project Heating Solutions (PHS). Why has norway succeeded with air source heat pumps. https://phscom.co.uk/why-has-norway-succeeded-with-air-source-heat-pumps/, 2024. Date Accessed: 29th November 2024.

[21] J. Rosenow. How heat pumps became a nordic success story. https://www.carbonbrief.org/guest-post-how-heat-pumps-became-a-nordic-success-story/, 2023. Date Accessed: 1st December 2024.

[22] P. et al. Sapin. Design and operational optimisation of an integrated thermal energy storage ground-source heat pump with time-varying electricity prices. https://heatpumpingtechnologies.org/publications/paper-no-715-design-and-operational-optimisation-of-an-integrated-thermal-energy-storage-ground-source-heat-pump-with-time-varying-electricity-prices-14th-iea-heat-pump-conference-chicago-us/, 2024. Date Accessed: 23rd November 2024.

[23] MCS (Microgeneration Certification Scheme). Uk reaches 250,000 certified heat pump installations. https://mcscertified.com/uk-reaches-250000-certified-heat-pump-installations/, 2024. Date Accessed: 5th December 2024.

[24] I. Stewart. Constituency data: Central heating. https://commonslibrary.parliament.uk/constituency-data-central-heating-2021-census/, 2024. Date Accessed: 9th December 2024.

[25] UNFCCC. The paris agreement. https://unfccc.int/process-and-meetings/the-paris-agreement, 2015. Date Accessed: 30th November 2024.

Selective Electrooxidation of Glycerol to Lactic Acid Using Platinum-Based Multi-component Catalysts

Ayce I Ates and John P Sim
Department of Chemical Engineering, Imperial College London, U.K.

Abstract
The electrooxidation of glycerol presents a promising route for the sustainable valorisation of crude glycerol, a byproduct of biodiesel production, into high-value chemicals such as lactic acid (LA), and hydrogen gas under mild conditions, specifically ambient pressure and temperature at 60°C in alkaline electrolyte. This study investigates the catalytic performance of Platinum (Pt) based multi-functional catalysts, specifically Pt/γ-Al_2O_3, Pt/ZrO_2, and Pt/H_2SO_4-treated ZrO_2, in selective glycerol electrooxidation (GEOX). Catalysts were evaluated using cyclic voltammetry (CV) and chronoamperometry (CA) in three-electrode setup, followed by chronopotentiometry (CP) under membrane electrode assembly (MEA). Following CV and CA, Pt/H_2SO_4-treated ZrO_2 was identified as the most effective catalyst, achieving a peak current density of 38 mA cm^{-2} in CA, high LA selectivity, and enhanced stability. MEA experiments demonstrated lower energy consumption and sustained performance of Pt/H_2SO_4-treated ZrO_2 over 1 hour, compared to untreated Pt/ZrO_2. The results further confirmed the co-production of hydrogen at reduced energy costs due to lower operating voltage compared to water electrolysis, which typically operates at potentials above 1.5V vs SHE. This work highlights the potential of advanced catalytic materials for glycerol valorisation by integrating GEOX into sustainable processes, potentially reducing reliance on conventional, less efficient methods for hydrogen production.

Keywords: Glycerol Electrooxidation, Platinum-Based Catalysts, Lactic Acid, Membrane Electrode Assembly

Introduction
With climate change becoming a prevailing issue in the coming years, many are looking into more sustainable energy sources like biodiesel. However, its production generates significant quantities of glycerol as a byproduct (0.1 kg of crude glycerol for every 1.0 kg of biodiesel (Huber, et al., 2006)), which currently holds limited commercial value due to its oversupply and low demand. GEOX presents a promising route, enabling the conversion of glycerol under mild conditions into valuable compounds, including glyceric acid (GLA), glycolic acid (GCA), formic acid (FA), tartronic acid (TA), oxalic acid (OA), acetic acid (AA), lactic acid, and other products. LA is especially desirable, with a commercial price of 1285 USD/MT, significantly higher than the price of crude glycerol at 295 USD/MT (Imarc Group, 2024). LA is a high value product with many applications in food and beverages, personal care, pharmaceuticals, and plastic manufacturing industries (Albuquerque, et al., 2017). There is growing demand for LA due to the rising market for polylactic acid (PLA), a bio-based polymer increasingly used in medical and packaging applications, providing a sustainable alternative to petrochemical-derived plastics (Li & Harrington, 2021).

Current production of LA from glycerol have been through hydrothermal methods. The focus of this study is to explore suitable catalytic materials for GEOX. Electrooxidation offers distinct advantages over traditional chemical oxidation routes, including operation under ambient pressures and moderate temperatures, which reduces energy input requirements and minimises safety concerns (Luo, et al., 2024). Moreover, electrooxidation provides selective oxidation pathways that can be controlled by adjusting electrode composition, catalyst surface, electrode potential and electrolyte conditions, influencing the selectivity and yield of LA.

Another benefit of GEOX is the production of H_2 gas at the cathode, which can be utilised for hydrogenation reactions or to produce heat and electricity. Hydrogen generation in this manner could support conventional methods, such as water electrolysis, by replacing the oxygen evolution reaction (OER) with the production of high-value liquid products, such as LA, at the anode, instead O_2, improving the efficiency and economic value associated with it. By coupling GEOX at the anode with the hydrogen evolution reaction (HER) at the cathode, the process could enable the co-production of H_2 at significantly reduced thermodynamic input. Furthermore, GEOX can be performed at a much lower potential compared to the high applied potential of water electrolysis (around 1.5V), resulting in considerable energy savings and enhancing the overall sustainability of hydrogen production (Kim, et al., 2017).

Background
Reaction mechanism
Previous research has largely focused on the exploration of various catalyst compositions and reaction mechanisms to optimize the production of target products such as LA and minimise byproducts. Strong alkaline conditions have also been known to encourage the Cannizzaro reaction which favours the formation of carboxylic acids,

such as LA (Li & Harrington, 2021). Figure 1 illustrates the two main pathways in GEOX. To achieve high LA selectivity, the selected catalyst should favour the thermal pathway over the electrochemical pathway. This can be accomplished by choosing catalysts with stronger adsorption towards the key intermediates (DHA and pyruvaldehyde) and those that strongly promote the dehydration step as well as the Cannizzaro reaction.

Platinum (Pt)

In the context of GEOX, Pt has been widely recognized as one of the most active catalysts. Its effectiveness stems from its relatively low d-band centre at approximately 2.25 eV, which strongly favours the initial deprotonation step. Pt was also selected for its ability to facilitate glycerol oxidation at relatively low overpotentials, with Pt-OH adsorption occurring at just 0.5–0.6 V under alkaline conditions. (Li & Harrington, 2021).

Electrooxidation of glycerol follows two primary reaction pathways, O-adsorption and C-adsorption. In the O-adsorption pathway, strong alkaline conditions significantly influence the initial deprotonation step. Once glycerol is chemisorbed onto the platinum surface, a second deprotonation step leads to the formation of DHA/Glyceraldehyde.

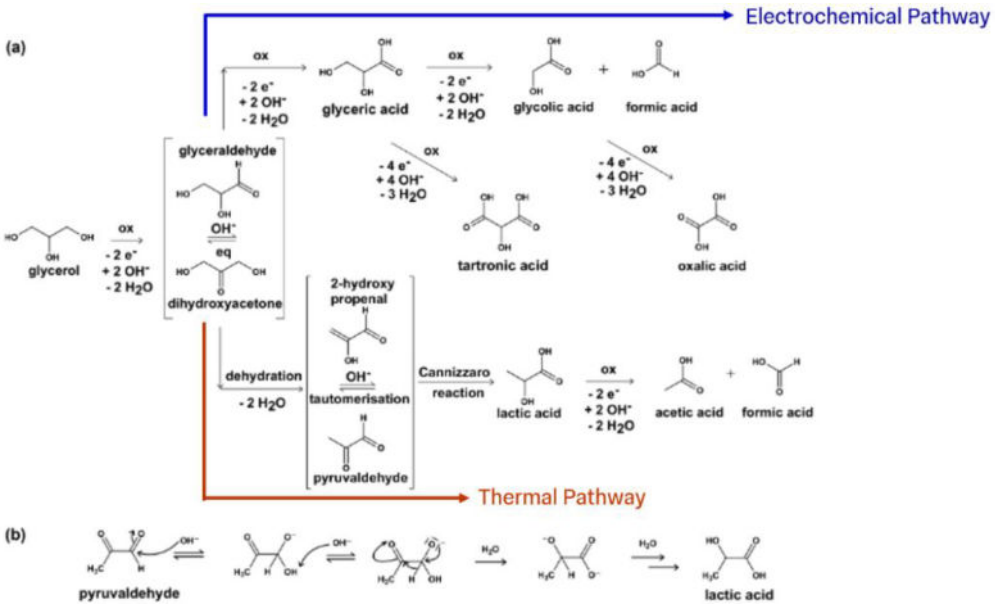


Figure 1. (a) GEOX reactions, illustrating the formation of key intermediates, and their subsequent transformations into products. (b) Mechanism of pyruvaldehyde conversion into lactic acid via the Cannizzaro reaction, driven by hydroxide-assisted tautomerization and hydration steps (Dai, et al., 2017).

Aluminium oxide (γ-Al_2O_3)

Multi-component catalysts have shown significant promise in optimizing GEOX. A 2024 study by Luo et al. reported the use of a tandem Pt/C-γ-Al_2O_3 multi-component catalyst for the electrochemical conversion of glycerol to LA, in which γ-Al_2O_3 provided acidic sites for catalysing DHA to pyruvaldehyde, achieving a glycerol-to-LA selectivity of 61.3%, among the highest performances reported. Furthermore, Luo et al. tested the metal oxides with different surface density of acidic sites, namely CeO_2, TiO_2 and, γ-Al_2O_3. γ-Al_2O_3 was highlighted as the most effective metal oxide when paired with Pt/C due to its high density of acidic sites, which correlates with the highest observed LA yield of 60.2%, hence a linear relationship between LA yield and acidic site density was recorded. Density functional theory (DFT) calculations further revealed that the strong binding affinity of γ-Al_2O_3 to the DHA intermediate is key to facilitating the dehydration pathway essential for selective LA production. This behaviour stems from γ-Al_2O_3's ability to chemisorb DHA through terminal hydroxyl groups at its acidic sites, favouring selective product formation and suppressing competing oxidation reactions (Luo, et al., 2024).

Zirconium Oxide (ZrO_2)

ZrO_2 is another metal oxide that has attracted attention due to its amphoteric nature and hydrothermal stability. Its highly polymorphic structure and ability to incorporate foreign cations allow for fine-tuning of its acid-base properties. Albuquerque et al. investigated the relationship between the acid-base properties of Pt/ZrO_2-based catalysts and their activity in the Cannizzaro reaction of pyruvaldehyde to LA. Their study highlighted that ZrO_2's ability to fine-tune its acid-base properties, through polymorphic adjustments, significantly impacts LA yield and selectivity. Albuquerque et al. suggested that the base sites were not the only sites involved in the Cannizzaro reaction and identified a linear relationship between catalyst activity and the total concentration of acid sites, suggesting the conversion of pyruvaldehyde into LA was predominantly catalysed by the Lewis

acid sites. Coordinatively unsaturated Zr^{4+} cations were identified as Lewis active sites, enabling the adsorption of pyruvaldehyde through keto and aldehyde carbonyl groups. Further experiments revealed that a balanced acid-base profile, particularly a sufficient concentration of medium-strength Lewis acid sites, was optimal for promoting the Cannizzaro reaction. These findings emphasize that the catalytic performance of Pt/ZrO_2 is intrinsically linked to its surface basicity and acidity (Albuquerque, et al., 2017).

H_2SO_4-treated ZrO_2

Building upon the foundational role of surface acidity in catalytic performance, Miranda et al. demonstrated that modifying ZrO_2 with sulphate ions significantly enhances its stability and reactivity by increasing both Brønsted and Lewis acid sites. Monoclinic and tetragonal ZrO_2 catalysts were sulphated by immersion in 1 M H_2SO_4 at 50°C, followed by drying and calcination at 400°C. The sulphation process promoted the formation of bidentate bridging sulphate groups on the ZrO_2 surface, as confirmed by FTIR and XPS analyses. These groups induce a strong electron-withdrawing effect, leading to enhanced Lewis acidity of the zirconium cations, while simultaneously generating new Brønsted acid sites through protonation of surface hydroxyl groups, as confirmed by pyridine adsorption values.

Additionally, the introduction of sulphate ions stabilises the tetragonal phase of ZrO_2, which is attributed to the suppression of phase transitions and the higher concentration of sulphate groups retained after calcination. The tetragonal phase sulphated zirconia exhibits a high density of Brønsted acid sites, which are more thermally stable and active than those in the monoclinic phase, enabling improved catalytic performance under prolonged reaction times.

Moreover, the role of increased surface acidity through the presence of Brønsted acid sites, enhances the catalyst's ability to facilitate dehydration reactions by donating protons to reactants, promoting the removal of water from molecules such as alcohols. The increased density of these acid sites provides a greater number of active sites for protonation, thereby accelerating the overall rate of dehydration reactions. Furthermore, the stronger acidity imparted by the sulphation process enhances the catalyst's ability to cleave C-OH bonds, improving dehydration efficiency (Miranda, et al., 2015). In the case of glycerol oxidation, this increased acidity can aid in the selective conversion to products like LA, while also maintaining the structural integrity of the catalyst under harsh reaction conditions.

Methodology

Ink Preparation

The preparation of catalyst inks for both the anode and cathode was a critical initial step in ensuring effective electrochemical performance during glycerol electrolysis. To create the cathode ink, 6 mg of 60 wt% Pt/C (manufacturer: Thermo Scientific) catalyst was dispersed in a mixture of deionized water (4200 µL) and ethanol (1584 µL) within a glass vial. After a ten-minute bath sonication, 216 µL of Nafion solution was added, followed by 10 minutes of pulse-mode probe sonication (5 seconds on, 5 seconds off). This approach served two purposes: the bath sonication gently encouraged particle dispersion while the probe sonication exerted higher mechanical energy, effectively disrupting any residual aggregates and resulting in a stable, uniformly dispersed ink, maximising the active catalytic surface area.

For the anode ink preparation, the same method as cathode ink was used, with modifications to accommodate the dual-catalyst system of 3 mg Pt/C and 9 mg of metal oxide (manufacturer: Sigma-Aldrich). The metal oxides used in the experiments were Al_2O_3, ZrO_2 and H_2SO_4- treated ZrO_2.

Material Characterisation

X-ray Diffraction **(**XRD) was conducted to determine the crystalline structure and phase composition of the catalysts ZrO_2 and H_2SO_4-treated ZrO_2. Samples were scanned over a 2θ range of 10°–80° at a step size of 0.02°.

Transmission Electron Microscopy **(**TEM) analysis was performed to examine the morphology and particle size of ZrO_2 and H_2SO_4-treated ZrO_2.

Brunauer-Emmett-Teller **(**BET) analysis was conducted to determine the specific surface area.

Electrode Preparation

The electrode substrates for both the anode and cathode were prepared using carbon paper (Freudenberg H23, sourced from Fuelcell Store). The anode was cut to a dimension of 3 × 3 cm for MEA and 1 × 1.5 cm for 3-electrode cell assembly, while the cathode was prepared as 2.5 × 2.5 cm for MEA. During the spray-coating process, the cut carbon paper sheets were positioned atop a heating plate set at 60 °C to facilitate the evaporation of the water-ethanol solvent from the ink. The prepared inks were then gently and uniformly sprayed across the surface of the carbon paper, with adjustments made to the spray speed to maintain an even layer without over-saturation.

Three-electrode Cell Assembly

To investigate the electrochemical performance of the glycerol oxidation, experiments were conducted in a three-electrode cell configuration. The cell was prepared using a 1M glycerol and 1M NaOH solution as the electrolyte. A platinum mesh

counter electrode and Hg/HgO reference electrode were used with the prepared carbon paper anode serving as the working electrode. The electrodes were connected to the potentiostat.

Once assembled, the cell was mounted on a magnetic stirrer to facilitate uniform mixing of the electrolyte. The stirrer remained active throughout the experiment, operating at a rotation rate of 400 rpm. To remove dissolved oxygen and prevent unwanted side reactions, the electrolyte solution was purged with nitrogen gas for 15 minutes prior to measurement.

Membrane Electrode Assembly

An overnight membrane activation in 1M NaOH was conducted, matching the electrolysis environment (1M glycerol, 1M NaOH), to improve the ion conductivity and stability of the Nafion membrane. The electrode stack assembly sequence began with the anode bipolar plate, followed by the anode electrode, membrane, separator, and a second membrane layer, concluding with the cathode electrode and cathode bipolar plate. All active electrode surfaces faced inward toward the membrane to maximise catalytic contact with the electrolyte during operation. The assembly was secured with screws, ensuring the thermocouple holes aligned correctly on each side of the stack to facilitate temperature monitoring.

The anolyte consisted of a 1M glycerol, 1M NaOH solution. A 10 mL aliquot of this prepared solution was transferred to the anolyte compartment. The catholyte was a 1M NaOH solution. 10 mL of this solution was added to the catholyte compartment. With the MEA fully assembled and both anolyte and catholyte in place, the pump was turned on to initiate circulation. The temperature was also elevated to 60°C. Once the system temperature stabilised at 60°C, electrical connections were made between the electrode cables and the MEA screws. The anode was connected as the working electrode, while the cathode served as both the counter and reference electrodes.

Electrochemical Measurements: **3-Electrode Cell**

Electrochemical measurements for the glycerol oxidation reaction were performed using an AUTOLAB potentiostat, where the anode served as the working electrode. The working electrode was connected to the potentiostat, where cyclic voltammetry and chronoamperometry were employed to characterise the electrochemical behaviour of the glycerol oxidation.

CV was first used to explore the electrochemical behaviour of glycerol oxidation. A potential range of 0.2V to 1.0V versus the reference electrode was applied to the working electrode, allowing for the observation of oxidation peaks. This technique provides valuable insights into the reaction's onset potential, and helps identify peak currents, which indicate the catalytic activity of the anode material toward glycerol oxidation.

CA was employed to further investigate the stability and kinetics of glycerol oxidation under a constant applied potential. Unlike CV, where the potential is swept, CA maintains a fixed potential, and the resulting current is monitored over time. This technique is crucial for examining the reaction's long-term stability and the electrode's activity under steady-state conditions. For this experiment, the potential was held at a value identified from the CV analysis (0.7V), which corresponds to the active glycerol oxidation region. The current response was recorded, and its decay over time was analysed to understand the electrode's catalytic efficiency and any potential deactivation or passivation effects.

Electrochemical Measurements: **MEA**

Electrochemical testing was conducted using an AUTOLAB potentiostat in galvanostatic mode. Chronopotentiometry was employed to investigate the electrochemical stability and efficiency of glycerol oxidation. Operating at higher current densities poses a significant risk of catalyst poisoning, which can compromise long-term performance. To mitigate this issue, a constant current density of 11.11 mA cm^{-2} (equivalent to 100 mA) was applied, and the resulting potential was recorded over a one-hour period. The expected potential range during electrolysis was 0.4–0.7V, characteristic of glycerol oxidation. This range served as a reliable indicator that the intended glycerol oxidation was occurring on the anode. In contrast, should the reaction have been dominated by water oxidation, the observed potential would have risen significantly to 1.6–1.7V.

High-Performance Liquid Chromatography (HPLC) Analysis

Following the electrochemical tests, HPLC was used to analyse the products generated from glycerol oxidation in both the anolyte and catholyte compartments. Given that the MEA membrane cannot entirely prevent the crossover of liquid products from the anolyte to the catholyte, this analysis was crucial for accurately quantifying reaction products in each compartment.

To ensure compatibility with the acidic nature of the HPLC columns, all samples—initially at an alkaline pH due to the NaOH electrolyte—were diluted with 0.55M H_2SO_4 prior to analysis. This acidification step stabilised the samples and prevented any pH-induced degradation of the HPLC columns, which are sensitive to highly basic conditions.

To assess both initial and reaction-induced product levels, six types of samples were analysed for MEA (2-7) and four samples for three-electrode cell (3-4*2), for each electrochemical experiment:

1. **Fresh DI Water**: 400µL deionized water

2. **Fresh Catholyte**: 200µL 0.55M H_2SO_4 and 200µL 1M NaOH solution.
3. **Fresh Anolyte**: 200 µL 0.55M H_2SO_4 and 200µL of a mixture of 1M glycerol and 1M NaOH.
4. **Anolyte Post-Electrochemistry**: 200µL 0.55M H_2SO_4 was added to 200µL of the collected anolyte.
5. **Catholyte Post-Electrochemistry**: 200µL 0.55M H_2SO_4 was added to 200µL of the collected catholyte.
6. **Anolyte Wash Post-Electrochemistry**: The anode compartment was washed with 1M NaOH solution after emptying.
7. **Catholyte Wash Post-Electrochemistry**: The cathode compartment was washed with 1M NaOH solution after emptying.

The samples were then placed in HPLC vials and subjected to HPLC analysis. The approach provided a comprehensive profile of glycerol oxidation products across both compartments, revealing any crossover between anolyte and catholyte.

Performance Metrics

The following calculations were performed to evaluate the performance of glycerol oxidation based on glycerol conversion (conv), product yield, selectivity, Faradaic efficiency (FE), and carbon balance (CB):

$$conv = \frac{mole\ of\ GLY\ consumed}{mole\ of\ initial\ GLY} \times 100\% \quad (1)$$

$$yield = \frac{mole\ of\ product\ formed}{mole\ of\ GLY\ consumed} \times 100\% \quad (2)$$

$$selectivity = \frac{mole\ of\ product\ X}{mole\ of\ all\ products} \times 100\% \quad (3)$$

Faradaic efficiency measures the proportion of the total charge passed that contributes to the formation of the desired product, expressed as a percentage.

$$FE = \frac{mole\ of\ product\ formed \times nF}{total\ charge\ passed} \quad (4)$$

where F is the Faradaic constant (96,485 C mol^{-1}) and n is the electron transfer number.

Carbon balance assesses the extent of glycerol oxidation by comparing the initial carbon input to the carbon present in the final products.

$$CB = \frac{3M_{C3} + 2M_{C2} + M_{C1} + 3M_{Gf}}{3M_{Gi}} \quad (5)$$

where M_{Gi} and M_{Gf} are initial and final moles of glycerol, M_{C1}, M_{C2} and M_{C3} are the moles of C_1, C_2 and C_3 products, respectively.

Results

Cyclic Voltammetry

CV was conducted in an applied potential range of 0-1V at room temperature using three-electrode setup to compare the catalytic activity of the metal oxides. The initial oxidation potential for GEOX is highest for Pt/Al_2O_3 at approximately 0.5V, while Pt/ZrO_2 and Pt/H_2SO_4-treated ZrO_2 both have similar onset potentials around 0.4V.

The peak current density, normalised by the area of the electrode, represents the maximum catalytic activity of each material and is highest for Pt/H_2SO_4-treated ZrO_2, reaching approximately 38 mA cm^{-2} at 0.82V. Pt/ZrO_2 follows with a peak current density of about 36 mA cm^{-2} at 0.83V, and Pt/Al_2O_3 has the lowest peak at around 12 mA cm^{-2} at 0.81V.

CV curves reveal a distinct oxidative peak during the forward scan, and a secondary oxidative peak in the reverse scan. Anomalies are observed in the reverse scan for both Pt/H_2SO_4-treated ZrO_2 and Pt/ZrO_2.

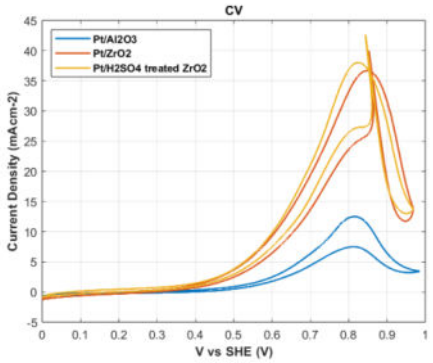


Figure 2. CV curves of Pt-based Al_2O_3, ZrO_2, and H_2SO_4-treated ZrO_2.

Chronoamperometry

As shown in Figure 3, Pt/H_2SO_4-treated ZrO_2 exhibits the highest initial current density, followed by Pt/ZrO_2 and Pt/Al_2O_3. In terms of stability, Pt/H_2SO_4-treated ZrO_2 displays a more gradual decline in current density over 15 minutes, with a 61.65% drop. By contrast, Pt/ZrO_2 and Pt/Al_2O_3 experience more substantial decreases, with 80.52% and 88.03% drops, respectively. The current drops are summarised in Table 1.

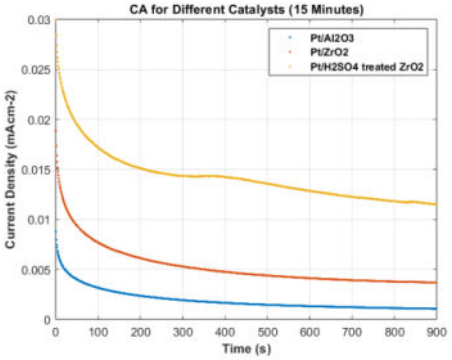


Figure 3. CA curves of Pt-based Al_2O_3, ZrO_2, and H_2SO_4-treated ZrO_2 for 15 minutes.

Table 1. Current drop during the first 15 minutes of CA.

Electrode	Current drop in first 15 minutes [mA]	Percentage drop in first 15 minutes [mA/mA]
Pt/Al_2O_3	7.75	88.03%
Pt/ZrO_2	15.18	80.52%
Pt/H_2SO_4-treated ZrO_2	18.47	61.65%

CA was conducted in four 15-minute intervals, each followed by a CV scan. As shown in Figure 4, the initial current density at the beginning of each CA interval appears slightly higher than the previous one, which suggests that the CV scans help regenerate the electrode surface by freeing up active sites.

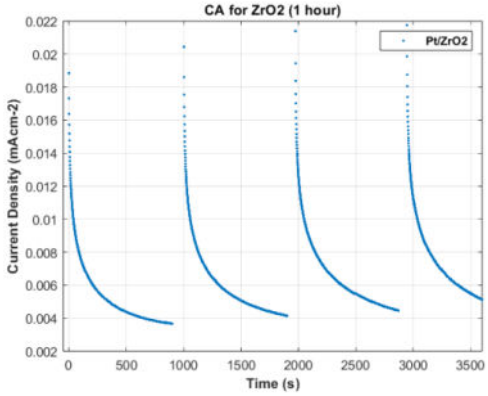


Figure 4. CA curves of Pt-based ZrO_2 catalyst over repeated 15-minute cycles for 1 hour.

Chronopotentiometry

CP measurements were conducted in MEA setup to compare the stability and performance of Pt/ZrO_2 and Pt/H_2SO_4-treated ZrO_2 as catalysts under a constant current of 100 mA. The recorded cell voltage versus time profiles are shown in Figure 5.

Pt/H_2SO_4-treated ZrO_2 exhibits a consistently lower cell voltage throughout the test duration, starting at approximately 0.52V and gradually increasing to around 0.67V, representing a 28.8% rise. Whereas untreated Pt/ZrO_2 starts at a higher initial cell voltage of about 0.60V and rises steadily to nearly 0.72V, with a slightly lower percentage change of 16.13% over the same duration.

Moreover, energy consumption was calculated to be 242.34J for Pt/ZrO_2 and 223.1J for Pt/H_2SO_4-treated ZrO_2, using the formula below.

$$W = \int E(t)I(t)dt \qquad (6)$$

Where W is the energy consumption, $E(t)$ is voltage specific to the energy of interest, and $I(t)$ is the current applied (Han, et al., 2015).

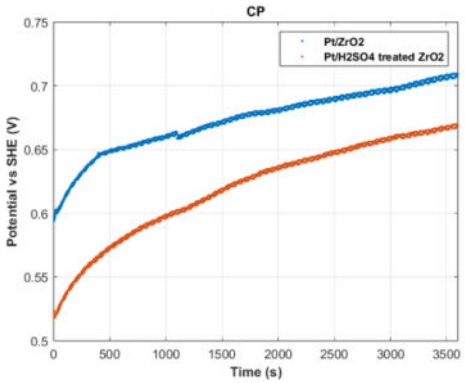


Figure 5. CP curves of ZrO_2 and H_2SO_4-treated ZrO_2 under a constant current of 100mA.

Performance Metrics

In MEA setup, performance metrics were analysed to identify the effectiveness of catalysts for LA production. Pt/H_2SO_4-treated ZrO_2 demonstrated a higher selectivity for LA at 76.03%, compared to 70.46% of Pt/ZrO_2. Additionally, it achieved a greater glycerol conversion rate of 13.74%, slightly exceeding the 12.91% of Pt/ZrO_2.

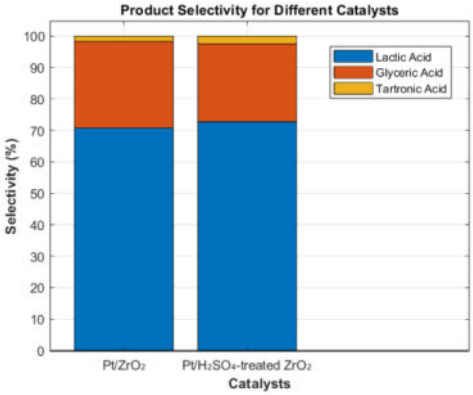


Figure 6. Selectivity towards LA, glyceric acid and tartronic acid in MEA setup.

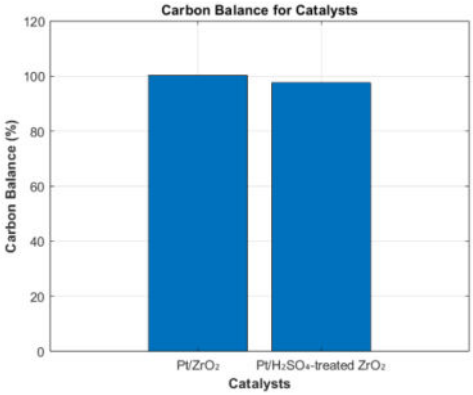


Figure 7. Carbon balance of Pt/ZrO_2 and Pt/H_2SO_4-treated ZrO_2 in MEA setup.

However, Pt/ZrO_2 exhibited a superior LA product yield of 71.7%, in contrast to 59.9% for Pt/H_2SO_4-treated ZrO_2. This is further supported by the Faradaic efficiency towards LA, which was higher for Pt/ZrO_2 at 49%, compared to 44.1% for the

H_2SO_4-treated catalyst. In terms of carbon balance, Pt/ZrO_2 achieved a score of nearly 1.0, slightly higher than the 0.949 achieved by Pt/H_2SO_4-treated ZrO_2.

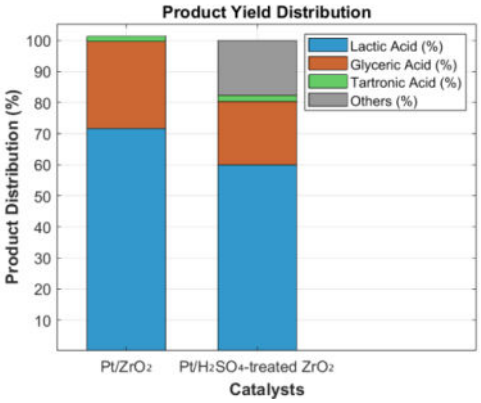


Figure 8. Product yield distribution of Pt/ZrO_2 and Pt/H_2SO_4-treated ZrO_2 in MEA setup.

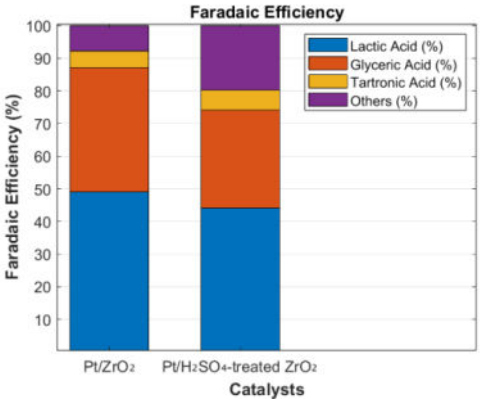


Figure 9. Faradaic efficiency of Pt/ZrO_2 and Pt/H_2SO_4-treated ZrO_2 in MEA setup.

Material Characteristics

XRD

The XRD analysis reveals that ZrO_2 and H_2SO_4-treated ZrO_2 possess varying crystalline structures. In Pt/ZrO_2, the most intense peaks are observed at 2θ = 30.13°, 35.36°, 50.40°, 60.2°, 62.83°, and 74.5° correspond to the tetragonal phase of zirconia, associated with the (101), (110), (112), (200), (211), (202), and (220) planes, respectively. Additionally, peaks at 2θ = 24.50°, 28.36°, 31.4°, 34.3°, 41.55°, 45.59°, 54.10°, 55.54°, and 65.75° provide evidence of the monoclinic phase of zirconia, as seen in Figure 10 (Kumar et al. 2017). There is a mild peak observed at 30° that represents the tetragonal phase in H_2SO_4-treated ZrO_2 indicating a less crystalline structure.

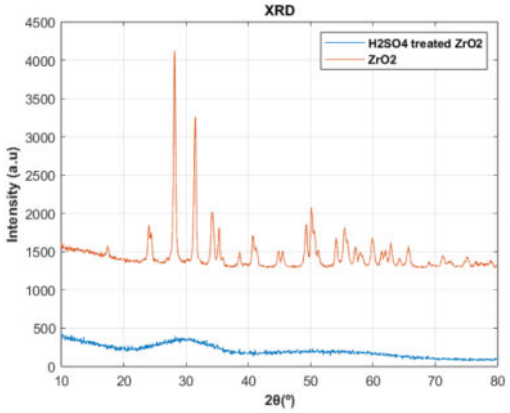


Figure 10. XRD patterns of ZrO_2 and H_2SO_4-treated ZrO_2, highlighting the structural differences between the two.

TEM

The TEM images of ZrO_2 and H_2SO_4-treated ZrO_2 reveal notable differences in morphology and particle characteristics. For untreated ZrO_2, the particles exhibit a relatively uniform morphology, forming smaller aggregated clusters with particle sizes around 50 nm. In contrast, H_2SO_4-treated ZrO_2 shows larger, more irregular aggregates with less uniform particle distribution, ranging from 60 to 150 nm, as seen in Figure 11. Al_2O_3, on the other hand, displays a cloud-like morphology with less defined particle boundaries and an amorphous structure, as observed in Figure 12.

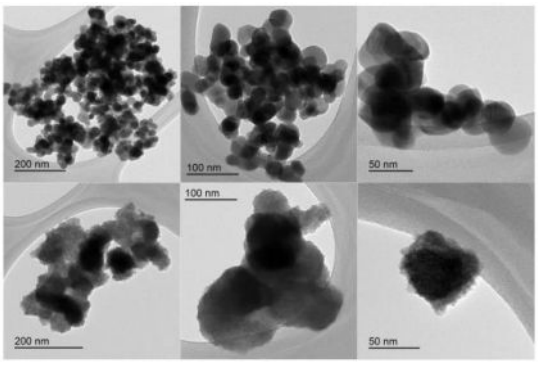


Figure 11. TEM images of ZrO_2 (top row) and H_2SO_4-treated ZrO_2 (bottom row) at 200, 100, 50 nm magnifications.

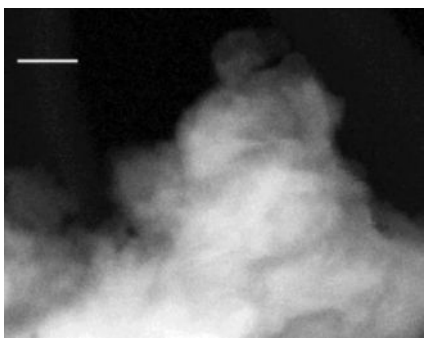


Figure 12. TEM image of Al_2O_3 at 50 nm magnification.

BET

The BET surface area analysis reveals significant variations in specific surface areas of catalysts, as summarized in Table 2, with Al_2O_3 exhibiting the highest specific surface area, indicating higher surface accessibility.

Table 2. BET surface area (SBET) of Pt/Al_2O_3, H_2SO_4-treated Pt/ZrO_2, and Pt/ZrO_2.

Material	Al_2O_3	H_2SO_4-treated ZrO_2	ZrO_2
SBET (m^2/g)	64	40	23.29

Discussion

Before starting MEA, CV and CA were performed in a three-electrode setup to determine which multi-component catalyst are promising for GEOX.

CV

The main concerning factors in CV are GEOX onset potential and the peak current density. The lower GEOX potential for Pt/ZrO_2, and Pt/H_2SO_4-treated ZrO_2 indicates a more energetically favourable reaction as a smaller energy input is required to initiate the oxidation process. This aligns with findings in the literature where ZrO_2-based catalysts are noted for their amphoteric nature and the ability to enhance catalytic activity through acid-base interactions and stabilised active sites (Miranda, et al., 2015).

Pt/H_2SO_4-treated ZrO_2 and Pt/ZrO_2 have significantly higher peak current densities compared to Pt/Al_2O_3. This suggests they are more effective in promoting GEOX. This can imply that the Pt/H_2SO_4-treated ZrO_2 and Pt/ZrO_2 catalysts have a higher density of active sites or improved electron transfer kinetics compared to Pt/Al_2O_3. The highest peak current density of Pt/H_2SO_4-treated ZrO_2 can suggest that the sulphation process enhances the density of active sites (Miranda, et al., 2015). Furthermore, sulphate ions can adsorb onto Pt surfaces, altering the potential distribution in the interfacial region and impacting the electronic structure of the electrode, which can inhibit the formation of active-site-blocking high-valence Pt oxides, thereby enabling more efficient glycerol oxidation (Cheng & Shi, 2019).

The shape of the CV graphs for all three metal oxides shows a distinctive oxidative peak in both the forward scan and the reverse scan. The oxidation peak in the forward scan is derived from the oxidation of glycerol, while the peak in the backward scan is due to the oxidation of the incompletely oxidized carbonaceous species generated from the forward scan (Ruan, et al., 2013). More specifically, the first oxidative peak is likely the C-adsorption of glycerol onto the Pt surface, as O-adsorption only occurs at higher overpotentials (Li & Harrington, 2021). Therefore, as the reverse scan frees up active sites at high overpotentials, O-adsorption can occur leading to a second oxidation peak in the reverse scan. The drop in current density at both oxidative peaks can be attributed to the blockage of active sites caused by the coverage of reactive species, poisoning species, and the deactivation of the catalyst due to platinum oxide formation.

The anomalies observed in the reverse scan for Pt/H_2SO_4-treated ZrO_2 and Pt/ZrO_2 are likely due to computational errors caused by iR overcorrection in the AUTOLAB system. A partial iR compensation factor of 0.85 was applied, as this is the standard for water electrolysis to mitigate overcorrection. However, its applicability to GEOX remains uncertain, as the choice of this partial compensation percentage is entirely theoretical and has not been specifically studied for this reaction. Consequently, this may have resulted in overcorrected data (Zheng, 2023).

CA

The durability of the catalyst is an important parameter that needs to be investigated. CA was conducted to observe the stability of the metal oxide at the common GEOX potential of 0.7V.

Initially, the current densities decreased rapidly for all the catalysts, due to the generation of carbonaceous intermediates occupying the active sites of the catalysts (Song, et al., 2015). The percentage drop in current density was measured in the first 15 minutes to study the stability of the material. Pt/H_2SO_4-treated ZrO_2 was the most stable with only a 61.65% drop. Pt/ZrO_2 also outperformed Pt/Al_2O_3 in terms of stability, which was likely due to the hydrothermal stability properties of Zirconia. The stability of Pt/H_2SO_4-treated ZrO_2 can be explained by enhanced surface acidity and sulphate groups stabilising the tetragonal phase of ZrO_2. The stabilised phase prevents phase transitions and reduces active site poisoning, thereby maintaining the activity of the catalyst under prolonged electrochemical conditions (Miranda, et al., 2015).

As the formation of LA from DHA is a thermal pathway, the three-electrode setup which operated at room temperature did not show high LA concentrations. Moving forward to the MEA, Pt/ZrO_2 and Pt/H_2SO_4-treated ZrO_2 were the only ones tested due to their favourable current densities, low GEOX potential, and stability.

CP

Both catalysts maintained high activity after 1 hour of CP, with ending potentials remaining below 0.72V. This highlights their suitability for crude glycerol applications, as they exhibited minimal potential increases. Notably, Pt/ZrO_2 demonstrated a smaller potential rise, suggesting that over extended operation, it could achieve lower overall energy consumption.

Performance Metrics

Pt/ZrO_2 achieved a perfect carbon balance of 1, indicating minimal to no loss of carbon-containing products due to undetected compounds. This is further validated by its total product yield of 100% and an impressive overall Faradaic Efficiency (FE) of 91.9%, underscoring its capability to utilise the majority of the electrons from glycerol oxidation effectively. On the other hand, Pt/H_2SO_4-treated ZrO_2 demonstrated a lower carbon balance of 0.976, accompanied by a reduced total product yield of 82.32% and an overall FE of 80.1%, suggesting that while sulfation enhances catalytic activity, it may also alter reaction pathways by promoting undesired side reactions (Miranda, et al., 2015). The use of Pt/H_2SO_4-treated ZrO_2 featured a higher presence of undetected products. In GEOX, there is a possibility of the formation of products that HPLC was unable to detect such as, CO_2 and CO_3^- (Jeffery & Camara, 2010). This suggests Pt/H_2SO_4-treated ZrO_2 is likely more successful in catalysing the scission of the C-C bond, leading to a higher yield of undesirable products.

Conclusion

This study highlights the potential of incorporating metal oxides onto Pt-based multi-component catalyst system for selective glycerol electrooxidation to LA. In a three-electrode setup, Pt/ZrO_2 and Pt/H_2SO_4-treated ZrO_2 outperformed Pt/Al_2O_3 in CV by achieving higher peak current densities and lower onset potentials, indicative of more energetically favourable and efficient reactions. CA further validated the superiority of Pt/ZrO_2 and Pt/H_2SO_4-treated ZrO_2, showing better stability under a constant potential of 0.7V and maintaining higher current densities throughout.

In the MEA setup, only Pt/ZrO_2 and Pt/H_2SO_4-treated ZrO_2 were examined. CP tests were conducted at a steady current density of 11 mA cm^{-2} for 1 hour. Pt/H_2SO_4-treated ZrO_2 exhibited a lower initial and final cell voltage, compared to Pt/ZrO_2, suggesting greater energy efficiency. However, Pt/H_2SO_4-treated ZrO_2 showed a higher percentage increase in cell voltage over time, indicating potential long-term stability concerns. Although the overall energy consumption of the potentiostat was lower for Pt/H_2SO_4-treated ZrO_2 within the one-hour test duration, this trend could shift if the experiment were extended for a longer duration. However, highly stable catalysts are crucial to utilise crude glycerol for long-term applications, as impurities may poison the catalyst.

Performance metrics calculated from HPLC results of the CP tests indicated no definitive overall superior catalyst. Pt/H_2SO_4-treated ZrO_2 demonstrated slightly higher selectivity for LA and overall glycerol conversion, while Pt/ZrO_2 excelled in FE towards LA, LA yield, and carbon balance. Both catalysts, however, were promising multi-component materials for selective GEOX towards LA. Under MEA setup at a relatively low temperature of 60°C, they achieved high LA selectivity of over 70%, successfully overcoming the thermal reaction pathway.

Outlook

Future research should delve deeper into the correlation between acidic site concentration and catalytic performance metrics using advanced characterisation techniques such as NH_3-TPD (Temperature-Programmed Desorption). Ion chromatography and Nuclear Magnetic Resonance (NMR) can be utilised to identify potential product losses due to carbonate formation.

Performing DFT on multicomponent catalyst surfaces is a valuable approach for future studies to comprehensively model the reaction pathways of both the purely electrochemical and the mixed electrochemical-chemical routes toward LA production. This method can provide valuable insights into the adsorption energies of various intermediates and help identify the electrode potentials that are most favourable for promoting the LA pathway. Complementing DFT with in situ FTIR spectroscopy can further validate the theoretical predictions by experimentally identifying the intermediates present at different stages of the reaction, thereby offering a more complete understanding of the catalytic process.

Moreover, gas chromatography could be employed to quantify hydrogen yield, exploring its viability as a co-product for industrial applications or as a sustainable energy source, thereby enhancing the overall energy efficiency of the system. Furthermore, other metal oxides with high acidic site concentrations, such as $Pt/ZrO_2/WO_3$ or Pt/TiO_2, should be investigated.

Acknowledgements

We would like to express our gratitude to Hanzhi Ye for providing invaluable support and feedback throughout the duration of the project.

References

Albuquerque, E. M., Borges, L. E. P., Fraga, M. A. & Sievers, C., 2017. *Relationship between Acid–Base Properties and the Activity of ZrO2-Based Catalysts for the Cannizzaro Reaction of Pyruvaldehyde to Lactic Acid.* [Online] Available at: https://chemistry-europe.onlinelibrary.wiley.com/doi/epdf/10.1002/cctc.201700305

Benaissa, M., Santiesteban, J. G., Diaz, G. & José-Yacamán, M., 1996. *Surface atomic resolution microscopy of sulfated-zirconia and Fe/Mn promoted sulfated zirconia.* [Online] Available at:

https://link.springer.com/article/10.1557/PROC-466-267?utm_source=chatgpt.com
Cheng, R. & Shi, M., 2019. *Sulfate-induced electrochemical instability in the transpassive region during the electrooxidation of Na2S on Pt.* [Online]
Available at: https://link.springer.com/article/10.1007/s10008-019-04234-4?utm_source=chatgpt.com
Dai, C. et al., 2017. *Electrochemical production of lactic acid from glycerol oxidation catalyzed by AuPt nanoparticles.* [Online]
Available at: https://www.sciencedirect.com/science/article/pii/S0021951717303615
Garcia, A. C. et al., 2016. *Strong Impact of Platinum Surface Structure on Primary and Secondary Alcohol Oxidation during Electro-Oxidation of Glycerol.* [Online]
Available at: https://pubs.acs.org/doi/10.1021/acscatal.6b00709
Han, L., Karthikeyan, K. G. & Gregory, K. B., 2015. *Energy Consumption and Recovery in Capacitive Deionization Using Nanoporous Activated Carbon Electrodes.* [Online]
Available at: https://iopscience.iop.org/article/10.1149/2.0431512jes/meta
Huang, B. et al., 2024. *Anode-Electrolyte Interfacial Acidity Regulation Enhances Electrocatalytic Performances of Alcohol Oxidations.* [Online]
Available at: https://onlinelibrary.wiley.com/doi/10.1002/anie.202409419
Huber, G. W., Iborra, S. & Corma, A., 2006. *Synthesis of Transportation Fuels from Biomass: Chemistry, Catalysts, and Engineering.* [Online]
Available at: https://pubs.acs.org/doi/10.1021/cr068360d
Imarc Group, 2024. *Lactic Acid Prices, Trend, Chart, Demand, Market Analysis, News, Historical and Forecast Data Report 2024 Edition,* s.l.: s.n.
Jeffery, D. Z. & Camara, G. A., 2010. *The formation of carbon dioxide during glycerol electrooxidation in alkaline media: First spectroscopic evidences.* [Online]
Available at: https://www.sciencedirect.com/science/article/pii/S1388248110002432
Kim, H. J. et al., 2017. *Coproducing Value-Added Chemicals and Hydrogen with Electrocatalytic Glycerol Oxidation Technology: Experimental and Techno-Economic Investigations.* [Online]
Available at: https://pubs.acs.org/doi/abs/10.1021/acssuschemeng.7b00868
Li, T. & Harrington, D. A., 2021. *An Overview of Glycerol Electrooxidation Mechanisms on Pt, Pd and Au.* [Online]
Available at: https://chemistry-europe.onlinelibrary.wiley.com/doi/10.1002/cssc.202002669
Li, Y. et al., 2024. *PtAu alloying-modulated hydroxyl and substrate adsorption for glycerol electrooxidation to C3 products.* [Online]
Available at: https://pubs.rsc.org/en/content/articlelanding/2024/ee/d4ee00485j
Luo, H. et al., 2024. *Selective glycerol to lactic acid conversion via a tandem effect between platinum and metal oxides with abundant acid groups.* [Online]
Available at: https://chemrxiv.org/engage/chemrxiv/article-details/65d392c166c1381729ffc134
Miranda, C. D., Ramírez, A. E., Gaona Jurado, S. & Vera, C. R., 2015. *Superficial effects and catalytic activity of ZrO2–SO42− as a function of the crystal structure.* [Online]
Available at: https://www.sciencedirect.com/science/article/pii/S1381116914005755?fr=RR-2&ref=pdf_download&rr=8e7b76035b2794ea
Ruan, L. et al., 2013. *Biomimetic Synthesis of an Ultrathin Platinum Nanowire Network with a High Twin Density for Enhanced Electrocatalytic Activity and Durability.* [Online]
Available at: https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.201304658
Song, P., Feng, J.-J., Guo, F.-Y. & Wang, A.-J., 2015. *Simple polyol synthesis of porous coral-like palladium–silver alloy nanostructures with enhanced electrocatalytic activity for glycerol oxidation reaction.* [Online]
Available at: https://pubs.rsc.org/en/content/articlelanding/2015/ta/c5ta03163j
Tang, Z. et al., 2019. *Pt/ZrO2 Prepared by Atomic Trapping: An Efficient Catalyst for the Conversion of Glycerol to Lactic Acid with Concomitant Transfer Hydrogenation of Cyclohexene.* [Online]
Available at: https://pubs.acs.org/doi/full/10.1021/acscatal.9b02139
Zheng, W., 2023. *iR Compensation for Electrocatalysis Studies: Considerations and Recommendations.* [Online]
Available at: https://pubs.acs.org/doi/10.1021/acsenergylett.3c00366?ref=pdf

Techno-Economic and Environmental Analysis of Liquid Direct Air Capture with Green and Grey Fuel Integration

Bingqi Li, Yuqi He
Department of Chemical Engineering, Imperial College London, U.K.

Abstract
Direct air capture (DAC) is a critical negative emission technology (NET) to combat global warming by removing carbon dioxide (CO_2) directly from ambient air. This study modelled the design of the L-DAC system powered by alternating fuel sources and layout in ASPEN Plus V14, providing insight into its performance in terms of process economic feasibility and CO_2 emissions. Three scenarios were investigated by altering the fuel supply: a green scenario utilising biomethane and green hydrogen fuel mixture derived from renewable energy sources, and two grey scenarios employing fuel mixtures of natural gas with either grey hydrogen or ammonia, varying in molar basis. In addition, heat exchange networks (HEN) were implemented, utilising natural gas and grey hydrogen as fuel distinctively. A natural gas-powered baseline scenario indicated a levelised cost of DAC (LCOD) of 125.84 \$/t-CO_2 with a carbon ratio of 3.66, which served as a benchmark for comparisons. By varying fuel compositions, an optimum green scenario operating at 30 %mol green hydrogen-biomethane mixture was identified to reduce the total CapEx by 10% compared to the baseline to 6.31 MM\$/yr, with a carbon ratio of 4.11. However, the higher cost of reactants yielded a LCOD of 167.10 \$/t-CO_2. In the grey energy scenario, a 30%mol grey hydrogen-natural gas mixture was found to have improved operating results, outperforming the baseline scenario with a decrease in Scope 1 carbon emissions. Finally, in heat integration scenarios where heat from the calciner preheats the fuel inlet stream to 800 °C, natural gas demonstrated the lowest LCOD of 121.85 \$/t-CO_2 among all scenarios and achieved fuel savings of 6066.62 t/year.

Keyword: Direct Air Capture, Process Simulation, Green Energy, Heat Exchange Network, Levelised Cost of DAC, Carbon Emissions

1. Introduction

Since the industrial revolution, the average global surface temperature has risen considerably due to the increasing atmospheric concentration of carbon dioxide (CO_2)—one of the primary components of anthropogenic greenhouse gases (GHGs)—resulting from the negligent continuous release of the gas into the atmosphere. This release of emissions has led to severe consequences: namely global warming and climate change. The Intergovernmental Panel on Climate Change (IPCC) reports that the CO_2 concentration in the atmosphere has increased from 280 parts per million (ppm) of pre-industrial times to 410 ppm in 2019 [1]. The overarching goal of the 196 United Nations Framework Convention on Climate Change (UNFCCC) members is to uniformly reduce GHG emissions swiftly and fairly in alignment with the Paris Agreement, an international treaty which aims to maintain the average global temperature at most 2 °C above pre-industrial levels by 2100, with hopes to ultimately limit temperatures to just 1.5 °C above their original levels [2]. Achieving this requires cutting anthropogenic CO_2 emissions as well as adopting aggressive carbon dioxide removal (CDR) technologies to offset GHG emissions, allowing a net-zero target to be reached.

Direct air capture (DAC) is one of the effective negative emission technologies (NETs) that has been used to mitigate the emission of CO_2 in the atmosphere since it was first introduced by Professor Klaus Lackner in 1999 [3]. To achieve negative emissions, CO_2 is captured straight from the atmosphere by passing ambient air through a sorbent. Unlike carbon capture technologies which generally carry out their processes at the point-source of emission, DAC is able to counteract the CO_2 produced by other major emission sectors such as transportation and agriculture [4]. This outlines the importance of DAC in eliminating historical emissions in the atmosphere. DAC has many additional benefits. Along with hopes that it can be used to achieve net-zero carbon emissions by 2050, DAC can also be coupled with a sequestration process to store CO_2 permanently through underground injection, often referred to direct air capture and storage (DACCS) [5]. Additionally, DAC presents a wide range of commercial opportunities in the carbon market, offering a source of profit to offset its expenses.

The two advanced technologies of commercially deployed DAC are liquid DAC (L-DAC) and solid DAC (S-DAC), where liquid solvent and solid sorbent are used to capture the CO_2 from the atmosphere respectively [6]. L-DAC is an absorption-based process that utilises aqueous alkali hydroxide solutions such as potassium hydroxide (KOH). The CO_2 in the bulk gas will react with aqueous solvent to form carbonates, leaving the outlet gas stream CO_2-free. The carbonates are subsequently calcined to release pure CO_2. The calcination process requires a high temperature of 900°C to liberate stored CO_2 in the carbonate salt. In contrast, S-DAC is an adsorption-based technique that captures CO_2 utilising a variety of solid sorbent such as amine-functionalised sorbent, and the regeneration process requires a low heating temperature of around 100°C under vacuum conditions.

As global efforts to mitigate climate change intensify, CDR technologies are becoming increasingly vital. Among these, L-DAC offers a mature and promising approach by using liquid solvents to absorb CO_2 directly from the atmosphere. However, one of the major challenges associated with L-DAC is the energy-intensive regeneration process, which requires a temperature of 900°C in the calciner. This high temperature can only be achieved through oxy-fuel combustion of natural gas, introducing emissions that counteract the negative-emission goals of DAC.

1.1. Project Aim

The primary objective of this research is to assess the techno-economic and environmental impacts of transitioning from natural gas to alternative energy sources for powering the calciner in L-DAC processes. Key economic metrics such as capital expenditure (CapEx), operating expenditure (OpEx), and levelised cost of DAC (LCOD) are considered for economic viability and deployment feasibility. Concomitantly, carbon ratio and emission factors are key environmental performance parameters for CO_2 emissions, constituting the carbon footprint of the process. Hydrogen and ammonia are potential fuels that produce zero CO_2 emissions during combustion. However, their carbon footprints vary greatly depending on production methods. Comparing green and grey energy scenarios allows for a clear evaluation of their economic and environmental impacts.

The study begins by thoroughly analysing a baseline scenario in ASPEN Plus V14, using natural gas (methane) as the primary energy source. The baseline scenario serves as a benchmark for comparison with the following proposed scenarios:

1. Green energy (renewable derived) scenario:
    - Fuel mixture of green hydrogen and biomethane (Green scenario)
2. Grey energy (conventional) scenario:
    - Fuel mixture of grey hydrogen and natural gas (Grey H_2 scenario)
    - Fuel mixture of ammonia and natural gas (NH_3 scenario)
3. Heat integration scenario:
    - Preheat fuel inlet to 20°C below autoignition temperature (517°C & 480°C for natural gas and grey hydrogen respectively)
    - Preheat fuel inlet to 800°C

The composition of the fuel mixture will be varied by mole fractions to find an optimal operating configuration. Additionally, a heat integration scenario with incorporate an additional heat exchanger unit to optimise energy recovery and reduce fuel consumption. This is done with the goal of enhancing energy recovery by maximising the utility of waste sensible heat around the calciner.

2. Background

Over the years, energy consumption has always been the key focus in L-DAC processes. In 2018, Keith et al. conducted a techno-economic analysis for a 1 Mt-CO_2/year L-DAC plant in a continuous process using an aqueous KOH sorbent coupled to a calcium caustic recovery loop, aiming to deliver CO_2 at 15 MPa. The process required either 8.81 GJ of natural gas or 5.25 GJ of gas and 366 kWh of electricity, per ton of CO_2 captured. This on-site combustion of natural gas to meet energy requirements produced an additional 0.48 Mt-CO_2/year, thereby reducing the plant's net emission CO_2 capture capacity [7]. This highlights the importance of reducing CO_2 emissions from natural gas combustion in L-DAC operations.

The regeneration of the solvent is the most energy-dense process in L-DAC. McQueen et al. has reported that the solvent processes normally have thermal energy requirements ranging from 5.25 to 8.1 GJ/tCO_2, primarily depending upon the extent of heat integration from the calcination process, and electricity demands of 1.3–1.8 GJ/tCO_2 [8]. On top of that, An et al. estimated that for the regeneration cycle of the L-DAC process, approximately 60% of the total energy is needed to detach the absorbed CO_2 from $CaCO_3$ to recycle solvent for successive cycles of CO_2 capture [9]. With this in mind, Han et al. began research on methods of reducing the temperature of calcination reactions in order to reduce energy consumption, aiming to develop an in-depth understanding of the decomposition mechanism of $CaCO_3$ [10]. However, many studies have not investigated the CO_2 capture performance and cycle stability of sorbents at lower calcination temperatures. In this light, a feasible solution to improve net CO_2 capture capacity is to explore alternative energy sources that meet the process's energy demands.

When it comes to replacing conventional energy sources, renewable energy has the capacity to provide the demanded energy while also supporting environmental sustainability. For example, Prats-Salvato et al. proposed a solar-powered L-DAC approach and developed a model to assess the influence of location and plant capacity on capture costs. It was evident that solar thermal energy (STE) is a mature yet sustainable green technology, capable of providing heat at temperatures above 1000°C. The continuous process of calcination is altered to intermittent solar calcination, which runs only during daytime. According to their simulation, the heat requirement of the solar calciner is equivalent to 5.44 GJ/t-CO_2 [11]. Location factor significantly influences solar calcination feasibility; insufficient solar radiation on a given day could render the process inoperable.

Another viable green energy source is offshore wind energy. In 2022, Ishaq et al. proposed a scenario to cover the energy demand for all designed configurations of the L-DAC process using offshore wind turbines with 15 MW power capacity. From the energy estimation, 160 wind turbines are required to satisfy the energy demand for the operation of a 1 Mt-CO_2/year L-DAC plant [12]. However, renewable energies are inherently fluctuating and intermittent, highly dependent on location and environmental conditions, which can impact the performance of L-DAC. A greener energy source with an uninterrupted supply is therefore preferred for such continuous processes.

3. Methodology

3.1. Process Overview

The green energy-powered L-DAC is a modified version of the process published by Carbon Engineering Ltd, composed of two closed chemical reaction loops and consisting of four major unit operations including an air contactor, pellet reactor, slaker and calciner [13]. Figure S1 presents the ASPEN Plus process flowsheet.

3.1.1. Air Contactor

The air contractor is the heart of DAC technology. The first step of the first chemical reaction loop occurs within the device, which captures CO_2 by bringing ambient air with an ultra-dilute CO_2 concentration of approximately 391 ppm into contact with an aqueous basic solution, KOH. The CO_2 molecules are chemically bound by this non-toxic solution and trapped as potassium carbonate salts in the liquid solution as shown in Equation 1.

$$2KOH_{(aq)} + CO_{2(g)} \rightarrow K_2CO_{3(aq)} + H_2O_{(l)}\ \Delta H = -95.8\ kJmol^{-1} (Eq.1)$$

It employs a crossflow configuration where ambient air flows horizontally through structured plastic packing, while the alkali capture solution flows downward as a 50 mm thin film. This design maximises the contact area for efficient CO_2 absorption.

3.1.2. Pellet Reactor

As the aqueous carbonate salt forms, it is continuously transported into the pellet reactor, which is adapted from water treatment technology developed by Royal HaskoningDHV (RHDHV) [7]. Within the pellet reactor, a causticisation reaction regenerates the aqueous basic solution KOH, simultaneously precipitating calcium carbonate into pellet form as shown in Equation 2. This is the second step of the first chemical reaction loop.

$$K_2CO_{3(aq)} + Ca(OH)_{2(s)} \rightarrow CaCO_{3(s)} + 2KOH_{(aq)}$$
$$\Delta H = -5.8\ kJmol^{-1} \quad (Eq.2)$$

The calcium carbonate pellets are removed by passing through a 20-mesh shaking screen, designed to give a particle size smaller than 0.85 mm. This size optimization is crucial for efficient heat transfer during downstream processing in the calciner unit.

3.1.3. Calciner

The second chemical reaction loop uses a sequence of high-temperature operating units to recover CO_2 from carbonate salts and regenerate alkali capture fluid. The loop begins in the calciner, where the calcium carbonate pellets are transported. In the calciner, CO_2 is liberated from calcium carbonate through thermal decomposition. This endothermic reaction described in Equation 3 occurs at 900°C.

$$CaCO_{3(s)} \rightarrow CaO_{(s)} + CO_{2(g)}\ \Delta H = +178.3\ kJmol^{-1} \quad (Eq.3)$$

A high-temperature fluidized bed reactor is used as the calciner. Oxygen, acting as the fluidizing gas, is supplied to the bottom of the calciner through a distribution plate, while natural gas is injected directly into the fluidized bed just above the plate for combustion. This process produces a gaseous mixture primarily composed of CO_2 and water.

3.1.4. Slaker

The calcium oxide from the calciner is then sent to the steam slaker to be rehydrated to close the chemical reaction loop shown in Equation 4. The slaker is a turbulent fluid bed reactor using steam as fluidise medium which operates at 300°C. After rehydration, the calcium oxide is recycled back into the pellet reactor to restart the chemical loop.

$$CaO_{(s)} + H_2O_{(l)} \rightarrow Ca(OH)_{2(s)}\ \Delta H = -63.9\ kJmol^{-1} \quad (Eq.4)$$

3.2. Heat Duty Specifications

The aim of this study is to supply the same amount of heat required by the calciner for the CO_2 liberation process, and for the release of CO_2 to be compressed to a high pressure. The calciner in the optimised baseline model has a heat duty of -5.79 MW, based on the Lower Heating Value (LHV) of methane of -50.04 MJ/kg and its operational synergy. To maintain the optimal performance of the calciner at the required heat, the heat duty and operational parameters must be met. Thus, the alternative fuel must supply the same amount of heat to the calciner as in the baseline model. This approach avoids modifications to the calciner, ensuring consistency in the system's performance evaluation and facilitating a direct comparison of results. Overheating the calciner will result in decreased thermal efficiency and increased costs of operation, while insufficient heating will result in a lower conversion rate and carbon capture efficiency.

3.3. Calciner Reactant Flowrate Determination

In this analysis, only the thermodynamic and combustion characteristics are considered. Methane (CH_4) is the primary fuel in the baseline model. In the modified model, two fuel substitution scenarios are explored: a CH_4-H_2 mixture and a CH_4-NH_3 mixture, with varying %mol in each case. The combustions of CH_4 and NH_3 are exothermic reactions which generate heat to sustain the operation of the calciner. Equations 5-7. represent the combustion reactions for the selected fuel:

$$CH_{4(g)} + 2O_{2(g)} \rightarrow CO_{2(g)} + 2H_2O_{(g)}\ \Delta H = -890.36\ kJmol^{-1} (Eq.5)$$
$$2H_{2(g)} + O_{2(g)} \rightarrow 2H_2O_{(g)}\ \Delta H = -285.5\ kJmol^{-1} \quad (Eq.6)$$
$$4NH_{3(g)} + 3O_{2(g)} \rightarrow 2N_{2(g)} + 6H_2O_{(g)}\ \Delta H = -382.58\ kJmol^{-1} (Eq.7)$$

The activity level, represented by the flowrates of both fuel and oxygen inlets, must be precisely tuned to complement each other to provide sufficient heat at their minimum flowrate in turn to keep the operational cost at a minimum. To ensure the stable performance of the calciner, the required heat duty must be consistently achieved by adjusting the reactant flowrates entering the calciner.

3.4. Economic Analysis

The economic analysis matrix evaluates the potential and feasibility of successfully implementing L-DAC technology at an industrial scale. Key economic metrics to focus on include capital expenditure (CapEx), operational expenditures (OpEx), and levelised cost of DAC (LCOD).

3.4.1. Capital Expenditure

Capital expenditure (CapEx) represents the upfront investment required to purchase, install, and commission the L-DAC system. These costs include the construction of core processing, as well as supplementary equipment like the heat exchangers, compressors and separators. The plant is designed to operate for 25 years including a 5-year startup phase, with annual operational hours of 8640.

The plant capital expenditure was estimated with the ASPEN Economic analyser by first obtaining installed equipment costs through mapping equipment scale based on the specifications of the process unit. CapEx is subsequently determined by summing installation, offsite and contingency costs, as outlined in Equation 8-9, providing a comprehensive measure of the financial investment required for implementation.

$$\text{CapEx} = \text{Installed Cost} + \text{Offsite Cost} + \text{Contingency} \quad (Eq.8)$$
$$\text{Offsite Cost} = 0.4 \times \text{Installed Cost} \quad (Eq.9)$$
$$\text{Contingency} = 0.1 \times (\text{Installed Cost} + \text{Offsite Cost}) \quad (Eq.10)$$

3.4.2. Operating Expenditure

Operating expenditure (OpEx) is in general divided into two categories: fixed and variable. The fixed OpEx ($OpEx_{Fix}$) expenses independent of operation rate, including maintenance, labour, and rent sit at 4.5% of the CapEx. On the other hand, the variable OpEx ($OpEx_{Var}$) depends on the operation rate, this includes the cost of raw materials and utilities such as fuel, electricity, steam and cooling water. $OpEx_{Var}$ is calculated based on the activity level and utilities, with pricing data sourced from both ASPEN default price and reputable online databases summarised in Table S1.

Fuel prices typically fluctuate within defined ranges due to changes in supply and demand, market conditions, and geopolitical factors as such, it is critical to understand the operational price dynamics when fluctuations in fuel price persist. To evaluate the effects of these price variations, a sensitivity analysis is conducted to examine how deviations within the price range affect their relationship to the average, providing insights into cost sensitivity and resilience under varying economic conditions. Using the average fuel price and its extreme values within the price range, a nominalised price ratio can be calculated to account for price deviation, giving the ratio of the actual price to the average fuel price, as shown in Equation 11.

$$\text{Nominalised Price Ratio} = \frac{\text{Fuel Price at \%Deviation}}{\text{Average Price of Fuel}} \quad (Eq.11)$$

3.4.3. Levelised Cost of DAC

The levelised cost of DAC (LCOD) measures the average cost of the captured CO_2 by DAC over the plant's operational lifespan. The LCOD accounts for the capital and operating cost of process equipment, as well as the cost of sorbent, as shown in Equation 12-13.

$$\text{LCOD} = \frac{\sum_p \left(\text{CapEx}_p \times \text{ACCR} + \text{OpEx}_p\right)}{\text{Productivity}} \quad (Eq.12)$$

$$\text{ACCR} = \frac{\text{WACC} \times (1+\text{WACC})^{\text{LT}}}{(1+\text{WACC})^{\text{LT}} - 1} \quad (Eq.13)$$

Where CapEx, OpEx, ACCR and LT are the capital expenditure, operating expenses for process equipment (p), annual capital charge ratio and equipment lifetime respectively. A Weighted Average Capital Cost (WACC) of 4.4 is adopted, based on low-carbon investments as reported by Ameli et al [14].

3.5. Environmental Analysis

In the context of this study, natural gas is being substituted with alternative fuels that do not produce CO_2 during combustion, with the objective of significantly reducing the carbon footprint while maintaining the energy requirements essential for operations.

The combustion of conventional fuels often results in the emission of carbon monoxide (CO), carbon dioxide (CO_2), and nitrogen oxides (NO_x). In contrast, H_2 and NH_3 do not produce these substances during complete combustion, making them attractive options for reducing emissions in L-DAC systems. However, the environmental impact of these fuels extends beyond combustion emissions.

Emissions from combustion are categorised as Scope 1 emissions, which refer to direct emissions resulting from on-site combustion processes. Utilities used during the L-DAC process are classified as Scope 2 emissions, which encompass indirect emissions related to the energy used in operations. Scope 3 emission associated with the production of the fuel used to supply the calciner. All emission scopes must be considered when evaluating the environmental performance of the system and can be calculated using Equation 14.

$$\text{Emissions}\left(\text{kg}_{CO_{2e}}/\text{yr}\right) = \text{Activity Level}\left(\text{kg}_{\text{fuel}}/\text{yr}\right) \times \text{Emission Factor}\left(\text{kg}_{CO_{2e}}/\text{kg}_{\text{fuel}}\right) \quad (Eq.14)$$

Emission factors serve as key indicators of CO_2-equivalent emissions from various fuel types and sources and will be used alongside carbon ratio calculations in Equation. 15 to assess environmental performance. Emissions across all scopes will be compared for each scenario, as summarised in Table S2.

3.6. Heat Exchange Network Design

This study introduces heat integration into the L-DAC system by implementing a heat exchanger network (HEN) in two scenarios: one powered by natural gas and another powered by pure grey H_2. The method outlined by Douglas for HEN design is applied, beginning with a first law of thermodynamics analysis to identify the net heat duty using values obtained from the baseline model, spotting any energy surplus or deficit within the system, as shown in Equation 16.

$$Q_{overall} = \sum |Q_{cooling}| - \sum |Q_{heating}| \quad (Eq.16)$$

However, applying only the first law of thermodynamics is not sufficient as it does not take heat transferred across the stream into consideration. To address this limitation, a second law of thermodynamic analysis is conducted. When implementing a heat exchanger in the system, a positive temperature driving force must exist for any heat transfer to take place. As per the second law of thermodynamics, heat will always flow spontaneously from the hot stream to the cold stream. In this case, a minimum driving temperature difference (ΔT_{min}) of 10K is set for all streams connected to the same heat exchanger. To guarantee this, 10K is subtracted from all hot stream temperatures before carrying out calculations when determining the heating and cooling duties of each temperature interval, along with their accumulative enthalpies Q as shown in Equation 17.

$$Q = \left(\sum FC_{p,hot} - \sum FC_{p,cold}\right) \times \Delta T \quad (Eq.17)$$

Finally, the pinch point and the minimum heating and cooling energy required are computed by making an energy adjustment.

4. Results & Discussion

4.1. Baseline scenario analysis

The baseline model represents the standard configuration of the L-DAC system, using methane as the primary fuel and designed to capture CO_2 at an efficient and stable operational level, it serves as a benchmark for all scenario comparisons, ensuring consistency throughout the study.

Table 1 summarised key parameters derived from techno-economic and environmental analyses of the baseline scenario. The baseline model plant captures 1.12Mt-CO_2/year from ambient air, with an additional 0.32 Mt-CO_2/year captured directly from on-site natural gas combustion used to supply heat requirements for the continuous operation of the calciner. The LCOD of the

baseline is 125.84 $/t-$CO_2$. This LCOD falls in between the range of 94-232 $/t-$CO_2$ proposed by Carbon Engineering, proving the validity of the baseline model. The baseline model captures 130.71 t-CO_2/hr with a carbon capture efficiency of 86.7% and a carbon ratio of 3.66.

With the focus on the calciner, the baseline fuel – CH4, enters the calciner at 25℃ with a flowrate of 13.4 t/hr, while the oxygen inlet enters at 674 ℃ with a flowrate of 58.5 t/hr composed of 95.6% O_2 and 4.4% N_2 by mass to process 336.72 t/hr of CaCO3 mixtures. It is worth noting that the calciner is a fluidised bed reactor, operating with a minimum fluidisation velocity of 0.25 m/s. The model is designed to shut down if the oxygen inlet flowrate falls below 18.29 t/hr, hence a larger flowrate is required to guarantee operation. Similarly, compressor B32 is tested to cease operation if the sum of the flowrates of O_2 and N_2 drops below a critical threshold of 167.67 kmol/hr. The heat duty of the calciner is -5.79MW, which represents the thermal energy required to maintain steady operation.

Table 1. Summary of the economic and environmental evaluations of the baseline scenario simulation.

Parameter		Value	Unit
Total capital cost		7.01	MM$/yr
Total operating cost		148.84	MM$/yr
Total annual cost		155.85	MM$/yr
CO_2 capture cost		125.84	$/t-$CO_2$
Operational hours		8640	hr/yr
CO_2 productivity		1.43	Mt/yr
CO_2 Emission	**Scope 1**	36.85	t-CO_2/hr
	Scope 2	28.43	
	Scope 3	7.545	
Carbon ratio		3.66	-

4.2. Calciner Reactant Flowrate

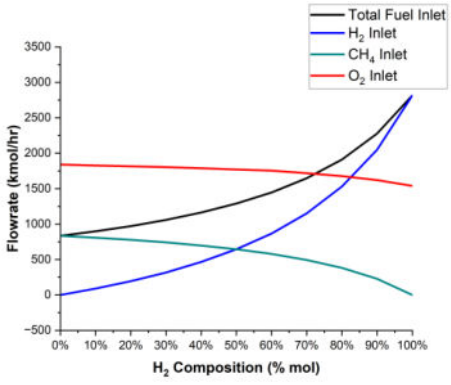


Figure 1. Flowrate of reactants for H_2 scenario

The calciner reactant flowrates corresponding to each 10% mol interval were determined through manual tuning, obtaining the targeted heat duty of -5.89MW with ± 20kW uncertainty.

In the H_2 scenario, as shown in Figure 1, the reactant flowrates are adjusted to ensure continuous operation of B32. Due to H_2's lower stoichiometric O_2 requirement, an increasing trend is expected. The exponential growth of the H_2 stream as %mol of H_2 increases reflects its deficient in energy release of combustion; achieving the same heat duty requires a flowrate of 2816.41 kmol/hr at 100 %mol H_2, compared to the baseline case of 835.27 kmol/hr. The corresponding O_2 inlet flowrate was reduced to 1538.7kmol/hr, compared to the baseline value of 1839.64 kmol/hr due to the decrease in reaction stoichiometry.

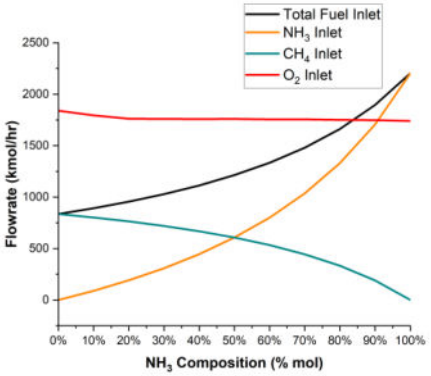


Figure 2. Flowrate of reactants for NH3 scenarios

In the NH_3 scenario, as shown in Figure 2, a similar exponentially increasing trend is also shown, as more NH_3 is required to compensate for its low enthalpy of combustion. To completely substitute CH_4, 2203.93kmol/hr of NH_3 is needed. The O_2 trend exhibits an initial decrease but hits a plateau as the NH_3 %mol increases in the system. This is due to the heat required to leverage the temperature of reactants feeding into the system as well as to heat up the N_2 produced during combustion. This demand for O_2 surpasses the stoichiometric decrease, resulting in a required oxygen inlet flowrate of 1740.21 kmol/hr at 100 %mol of NH_3.

4.3. Economic Analysis

4.3.1. Price Sensitivity Analysis

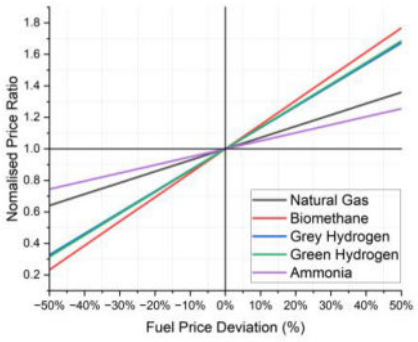


Figure 3. Normalised price ratio against fuel price deviation

In the L-DAC system, fuel expenditure is the most significant component in OpEx, accounting for 38-67% of the total OpEx. Fugure 3 illustrates the normalised price ratio related to the average price used in computing the OpEx subjected to price deviation. The gradient of Figure 3 indicates the extent of change in their actual market price relative to the average price when deviation occurs. One particular line with the steepest gradient – biomethane, would have a greater impact on OpEx, for having a price of 1.76 times higher than the average when the price deviates to its maximum. On the other hand, NH_3 has the most stable price variation as the price deviates – 1.25 times greater than the average price, meaning operating the plant with NH_3 is relatively less likely to be affected by the price fluctuation. Other fuels, such as natural gas, grey H_2, and green H_2, fall between these extremes.

This analysis highlights how even small deviations in fuel prices could lead to substantial changes in the actual market price of fuel, bringing uncertainty into the economic viability of implementing the L-DAC system. Recognising these dynamics is essential for maintaining cost-effective operation.

4.3.2. Capital Expenditure

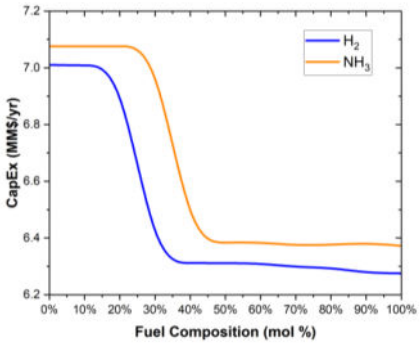


Figure 4.CapEx for both H_2 and NH_3 scenarios

A sudden drop in CapEx is observed at 20–30% fuel composition in the H_2 scenario and 30–40% in the NH_3 scenario as shown in Figure 4. This reduction is primarily attributed to the change in the installed cost of B32, which accounts for 55% of the total CapEx in the baseline scenario and is used to compress capture CO_2 to 15MPa. Increasing the %mol of substitute fuel reduces CO_2 emissions from combustion, resulting in less CO_2 being captured through on-site combustion, which in turn reduces both size and installation cost of B32. As a result, the CapEx for H_2 and NH_3 scenarios dropped to 6.31MM$/year and 6.38MM$/year respectively, representing a total reduction of 10% and 11% in CapEx in each scenario. The minor fluctuations after the drop are caused by slight variations in heat exchanger configurations.

The NH_3 scenario indicates a higher starting CapEx of 7.07 MM$/yr due to an additional unit of air separator being added, illustrated by an upward shift of the curve. The air separation unit is installed at a cost of 1.097MM$, as large quantities of N_2 are produced during the combustion reaction and need to be removed from the system to sustain the performance of DAC. This unit is critical for two reasons: first, to reduce the workload on compressor B32 at later stages of the process, and second, to prevent N_2 impurities in the compressed CO_2, ensuring the purity for the potential of selling the CO_2 as a product. Additionally, nitridation may occur in the system due to the presence of NH_3 and N_2 at 900℃. Additionally, nitridation may occur in the system due to the presence of NH_3 and N_2 at 900°C. Some steel materials are prone to nitriding in NH_3 environments at temperatures ranging from approximately 450°C to 700°C, which may cause deterioration in corrosion resistance and increase the cost of replacing the device [14].

4.3.3. Reactant Costs and OpEx

With the reactant flowrates determined in Section 4.2, the operational reactant costs can be easily computed, utilising the average price of fuel listed in Table S1. For O_2, a price of 0.085 $/kg is used in the calculations, based on its average production cost via cryogenic distillation [15]. The baseline reactants cost 56.60MM$/hr for operation. The quantity of substitute fuel increases exponentially to compensate for its deficiency in energy release during combustion, hence the fuel cost is expected to increase correspondingly, especially for NH_3 and green energy scenarios, as shown in Figure 5.

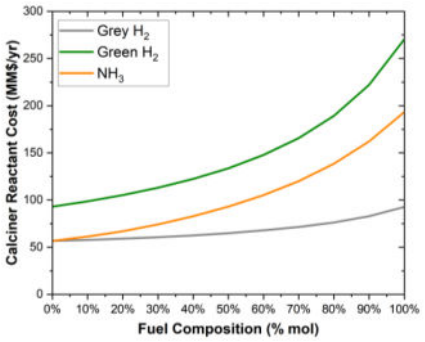


Figure 5. Calciner reactant Cost for all scenarios

In the grey energy scenarios, the NH_3 scenario experiences a greater cost increase than the H_2 scenario due to the higher price of NH_3. The reactant cost for H_2 scenario rises steadily from 56.60 MM$/yr at 0 %mol to 92.58 MM$/yr at 100 %mol, while the cost for NH_3 escalates sharply to 193.35 MM$/yr, starting at the same base cost. In the green energy scenario, a cost of 93.01MM$/yr using 100% biomethane shifted the curve upwards and increased to 270.41MM$/yr when fully powered by green H_2. This steep increase highlights the economic burden associated with green fuel, as a costly fuel option.

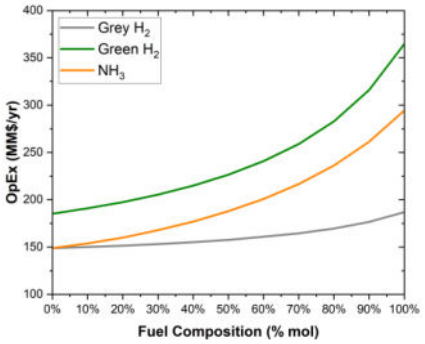


Figure 6. OpEx for all scenarios

The OpEx curve illustrated similar trends shown in Figure 6, as it primarily depends on the cost of fuel, which constitutes the largest share of operational expenses. Other components of OpEx include solvent costs, utility expenses, and miscellaneous operational costs, but fuel costs remain the dominant factor influencing the overall trend.

4.4. Levelised Cost of DAC

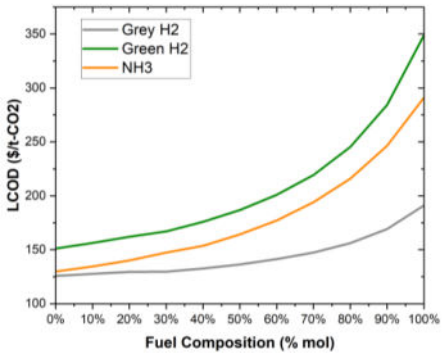


Figure 7. LCOD for all scenarios

The levelised costs of DAC (LCOD) illustrated in Figure 7 are computed following Equation 12-13. Although Figure 4 has shown a sudden decrease in CapEx with changes in fuel composition, the LCOD still exhibits an increasing trend for all scenarios, as the reduction in CapEx is insufficient to offset the rise in OpEx. In the NH_3 scenario, despite a higher CapEx compared to the baseline, the LCOD increased only slightly to 129.92 \$/t-CO_2. For the green energy scenario, the LCOD rose to 151.22 \$/t- CO_2, due to a high OpEx resulting from the higher cost of biomethane, and continues to rise across the fuel compositions, making it the most expensive option in all operating scenarios.

A noteworthy feature is the plateau observed in the grey H_2 scenario at a fuel composition range of 20-30 %mol, where the LCOD remains relatively stable at 129.46 \$/t- CO_2. Within this range, the increase in LCOD is limited to 0.25 \$/t- CO_2, suggesting an optimal operating point from an economic perspective, that minimises cost escalation while maintaining calciner performance. Green energy scenario exhibited a similar plateau around the same fuel composition, driven by a reduction in CapEx due to the decreased size and installation cost of the B32 compressor unit.

It is worth noting that the LCOD calculation in this study did not account for potential profits from selling the captured CO_2 as a product. The LOCD would decrease if CO_2 could be sold commercially, improving the economic viability of the process. Furthermore, improvements and advancements in technologies would further reduce the LCOD, making L-DAC more financially favourable and attractive for large-scale implementation.

4.5. Environmental Analysis

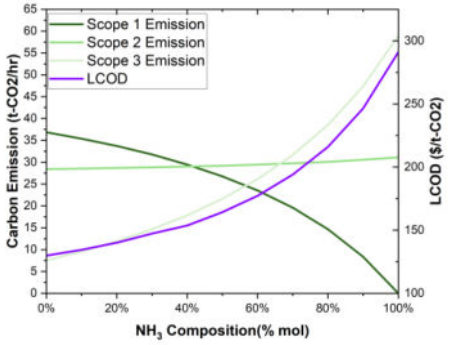


Figure 8. Carbon emission for NH_3 scenarios

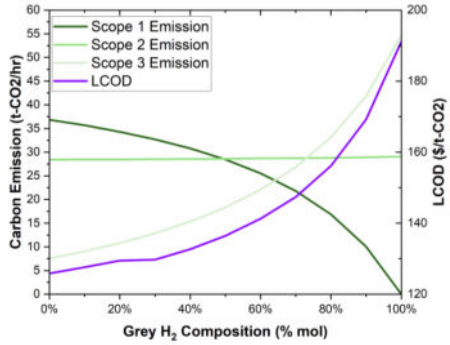


Figure 9. Carbon emission for grey H2 Scenarios

Emissions are categorised into 3 scopes, each analysed separately to prevent double counting emissions, as illustrated in Figure 8 and 9. In the grey energy scenarios, Scope 1 emissions only account for the emission of natural gas combustion, as other fuels do not emit CO_2 during the combustion process. Hence, in all grey energy cases, Scope 1 emissions decrease as the %mol of the substitute fuel increases, starting at 36 t-CO_2/hr and eventually reaching zero when natural gas is fully replaced.

Scope 2 emissions account for indirect CO_2 emissions from utilities, primarily electricity sourced from the national grid to operate the L-DAC. The utility usage for both grey and green H_2 scenarios is the same, resulting in 28.43 t-CO_2 /hr at 0 %mol of H_2, increasing slightly to 29.07 t-CO_2/hr when the system is fully powered by H_2. However, in the NH_3 case, a greater rate of increase in Scope 2 emissions is observed due to the additional utilities applied for the air separation unit. Starting at the same value of 28.43 t-CO_2/hr, Scope 2 emissions rise to 31.11 t-CO_2/hr when the system is fully powered by NH_3.

Scope 3 emissions account for the emissions associated with reactant production. In grey energy scenarios, carbon emissions increase exponentially as the fuel composition shifts, with NH_3 emissions exceeding those of grey H_2 due to the higher flowrate required, despite NH_3's lower emission factor compared to grey H_2. When fuel is completely substituted, Scope 3 carbon emissions for grey H_2 and NH_3 scenarios elevated to 54.66 and 59.95 t-CO_2/hr respectively, both starting from the baseline at 7.54 t-CO_2 powered by pure natural gas. This dramatic increase in carbon emission highlights the substantial environmental impact of grey energy sources, particularly when fuels are produced through non-environmentally friendly processes.

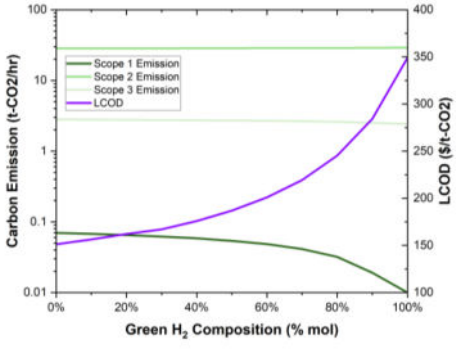


Figure 10. Carbon emission for green H_2 scenarios

The green energy scenario utilises biomethane as a substitute for natural gas, produced through anaerobic digestion. As its emission factor for combustion accounts for CO_2 sequestered during its production, these emissions are not counted in Scope 3 calculations. As a result, the green energy case has the least carbon emissions across all scenarios in all scopes of emissions, it is best to operate at 30%mol of green H_2 with Scope 1 emission of 0.061 t-CO_2/hr and Scope 3 emission of 2.75 t-CO_2/hr. This configuration balances reduced carbon emissions with a shallow increase in LCOD, achieving a balance between environmental and economic performance.

4.6. Heat Integration Scenarios

4.6.1. Thermodynamic Law Analysis

A heat surplus of -70165.775 kW obtained from the first law of thermodynamics analysis represents the minimum amount of energy required to be removed from the system.

Table 1. Heating, cooling and overall duties

Q_{cool} [kW]	Q_{heat} [kW]	$Q_{overall}$ [kW]
40766.7	-110932	-70165.775

After performing the second law of thermodynamics analysis, a pinch point of 1163.15K was found from the cascade diagram shown in Figure S4, with a minimum cooling duty of -70165.775 kW.

4.6.2. Heat Exchange Network

In the network design process, heat exchange is conducted through stream matching. While matching the hot steam and cold steam, regions should be split into two parts - above the pinch and below the pinch. From analysing the baseline scenario, all streams are below the pinch and thus only cold utility can be used. The cooling utilities in the system should be equal to the value of the calculated minimum cooling duty as stated in Section 4.6.1.

Preheating the fuel increases its energy content, thus leading to a decrease in the required fuel flowrates. To ensure safety, the first preheat temperatures were selected to be 20°C below the auto-ignition point of CH_4 and H_2, which is 517 °C and 480 °C respectively. A second preheat temperature of 800°C was selected to further investigate the effect on fuel consumption and the utility required. An extra heat exchanger was implemented that connects the outlet of the calciner with an outlet temperature of 900°C identified as the hot stream, and the fuel inlet at 25°C as the cold stream.

In the natural gas scenario, 5.45 MW and 9.61 MW of heat were transferred from the hot stream to the cold stream for the selected preheat temperatures respectively. The detailed integrated heat network is shown in Figure S2 and S3. The flowrates of the reactants were readjusted to meet the heat demand for the calciner. The flowrate of reactants decreased as expected for the reactant energy content leverage. A summary of flowrates of reactants is provided in Table S4. The fuel consumption of natural gas has respectively decreased to 810.42 kmol/hr and 791.5 kmol/hr compared to the baseline case at two preheated temperatures, which also led to a reduction in oxygen inlet by approximately 2.97% and 4.98% for each case to 1787 kmol/hr and 1748 kmol/hr.

When L-DAC is powered by pure grey H_2 without HEN, 2823.4 kmol/hr of grey H_2 and 1541.4 kmol/hr of oxygen are required. Introducing an additional heat exchanger to preheat the H_2 inlet to 480°C and 800°C reduces the fuel flowrate by 5.6% and 9.07%, respectively, to 2666.5 kmol/hr and 2567.2 kmol/hr., Correspondingly, the oxygen inlet flowrates decrease to 1465 kmol/hr and 1416.5 kmol/hr at these temperatures.

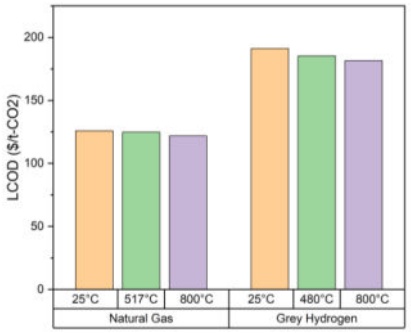


Figure 11. LCOD for grey H_2 and natural gas in heat integration scenarios

As illustrated in Figure 11, the LCOD exhibits the same decreasing trend as the reactant flowrates. Natural gas shows minimal variation in the overall trend; however, the lowest LCOD across all scenarios is achieved at the 800°C preheat temperature, reaching 121.85 $/t-$CO_2$. In contrast, grey H_2 exhibits significantly higher LCOD values compared to natural gas and shows greater variation across preheat scenarios. Achieving a LCOD of 181.52 $/t-$CO_2$, at 800°C preheat temperature.

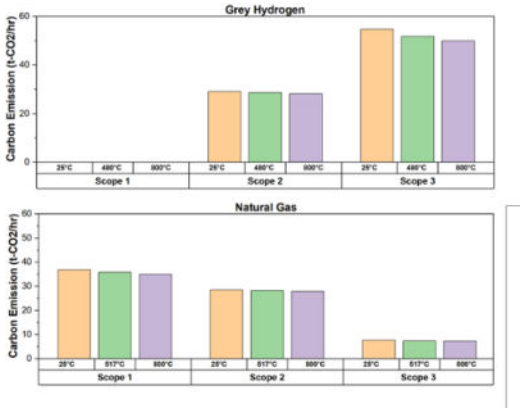


Figure 12. Carbon emission for grey H_2 and natural gas

With preheated temperature, carbon emissions for both fuels decrease slightly across all scopes, as demonstrated in Figure 12. Scope 1 emissions are higher for natural gas combustion, which emits CO_2, compared to grey H_2, which produces only water. The utility used for fuels at the two preheated temperatures is very similar, with minimal variation between the temperatures that can be neglected. As less utility is required for the increasing preheat temperatures, Scope 2 emissions decrease accordingly. For Scope 3, the emissions from grey H_2 are significantly higher than those from natural gas, this is due to the significantly higher flowrate of pure grey H_2 required to meet the calciner's heat demand, resulting in substantially greater emissions for its production compared to natural gas.

4.7. Comparisons of Scenarios

Table 2. A summary of preferred operating conditions

Parameter		Baseline	Grey H_2 (30% mol)	Green H_2 (30%mol)	Heat Integration (CH_4 with 800°C)	Unit
Total capital cost		7.01	6.31	6.31	6.27	MM$/yr
Total operating cost		148.84	153.05	205.42	144.26	MM$/yr
Total annual cost		155.85	159.36	211.73	150.53	MM$/yr
CO_2 capture cost		125.84	129.71	167.103	121.84	$/t-$CO_2$
Operational hours		8640	8640	8640	8640	hr/yr
CO_2 productivity		1.43	1.40	1.40	1.43	Mt/yr
CO_2 Emission	**Scope 1**	36.85	32.70	0.061	34.92	t-CO_2 /hr
	Scope 2	28.43	28.53	28.53	27.90	
	Scope 3	7.545	12.86	2.757	7.16	
Carbon ratio		3.66	4.11	4.11	4.09	-

The preferred operating conditions are summarized in Table 2, showcasing three scenarios that outperform the baseline scenario from an economic, environmental, or combined perspective.

From the grey energy scenarios, grey H_2 at 30 %mol is selected to be a potential operating condition as the relative reduction in CapEx and improvement in the carbon ratio. With the advancement in production technology, improvement of energy efficiency and scaling of the H_2 production facility, both LCOD and carbon emissions are expected to decrease, making grey H_2 an optimistic option. Conversely, the use of NH_3 in the L-DAC system is not recommended. Its performance in both economic and environmental perspectives falls short, offering a limited advantage over other proposed scenarios. While it may have a marginally lower scope 1 emissions compared to the baseline, the singular benefit does not outweigh its numerous drawbacks.

The most environmentally favourable scenario is the green scenario operating with green H_2 at 30 %mol. Regardless of its higher LCOD, the CO_2 emissions are the far lowest, hence it is considered as an optimal operating condition when minimising carbon emissions is the primary objective.

Another scenario involves heat integration in the natural gas-powered configuration, with preheating to 800°C. With its lowered LCOD and carbon emissions compared to the baseline, this is the optimum operating condition when both economic and environmental factors are taken into consideration.

4.8. Challenges and Outlook

Throughout the study, full combustion of fuel is assumed in the simulation to ensure the flowrates of reactants are both kept at a minimum to lower the operating cost. However, in practice, this poses a threat of incomplete combustion, especially for NH_3, which could emit significant amounts of nitrate oxide and nitrous oxide. These compounds contribute to environmental concerns such as acid rain and ozone depletion [16]. Full life cycle analysis is omitted in this study, neglecting environmental costs or benefits associated with fuel production, distribution and end-use. While green fuels are low-emissions alternatives, their adoption at industrial-scale is hindered by high costs, limited availability, and current technological constraints. Along with lower energy density requiring higher consumption rates, it creates a significant cost burden for operating the L-DAC system.

Future advancements in technology and emission mitigation strategies hold significant potential to address these challenges and limitations. Innovations in technologies for fuel production, coupled with stronger governmental policies forcing industries to decarbonise their operation with more integration with renewable sources, the prices will fall further. From an environmental perspective, according to the IEA, the emissions associated with H_2 production are expected to decrease to 6-7 kg CO_2/ton by 2030 and 0.8-0.9 kg CO_2/ton by 2050 in the NZE Scenario [17]. This reduction in emissions, along with advancements in renewable energy integration, would contribute to significant OpEx reductions, making L-DAC plants more sustainable and economically viable in the long term.

5. Conclusion

A novel design for the L-DAC plant integrated with green and grey energy sources has been proposed to explicitly decrease CO_2 emissions via on-site natural gas combustion, in supplying the required heat for the calciner. Three preferred scenarios are proposed with the aim of reducing from either an economic or environmental perspective. The most outperforming scenario – heat integration for natural gas with a preheating temperature of 800, is able to achieve the lowest LCOD at 121.84 $/t-$CO_2$ and carbon emissions. However, this analytical-based scenario proposes safety concerns as the preheat temperature exceeds its own autoignition point. If heat integration were to be implemented in reality, preheating to 20 lower than its autoignition temperature is the safest choice to apply, though their economic and environmental performance will be compromised.

Based on the findings of this study, there is sufficient evidence and significant potential to optimise heat integration strategies further to enhance overall energy efficiency. The results clearly demonstrate that implementing heat integration scenarios can improve both economic and environmental metrics, making it a viable approach to enhance the performance of the proposed scenarios. A comprehensive heat analysis of the entire process could be conducted to identify additional opportunities for optimisation. Further work could focus on implementing heat integration onto the proposed mixed fuel scenarios, such as the grey H_2 and green energy scenarios, to further enhance their performance.

6. References

[1] IPCC, “Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of teh Intergovernmental Panel on Climate Change,” 2023.

[2] United Nations, “The Paris Agreement,” [Online]. Available: https://unfccc.int/process-and-meetings/the-paris-agreement.

[3] K. Z. H. Lacker, “Carbon Dioxide Extraction from Air: Is It An Option?,” 1999.

[4] M. e. a. Habib, “Flexible carbon capture using MOF fixed bed reactors at an NGCC plant, SSRN.,” 2023.

[5] IEA, “Executive summary - direct air capture 2022 - analysis, IEA,” [Online]. Available: https://www.iea.org/reports/direct-air-capture-2022/executive-summary.

[6] Climeworks, “Climeworks' third indurstry snapshot,” 2023. [Online]. Available: https://climeworks.com/news/industry-snapshot-by-climeworks-jul23.

[7] D. e. a. Keith, “A Process for Capturing CO2 from the Atmosphere,” 2018.

[8] N. e. a. McQueen, “A review of direct air capture (DAC): scaling up commercial technologies and innovating for the future,” 2021.

[9] K. e. a. An, “ A comprehensive review on regeneration strategies for direct air capture,” 2023.

[10] R. e. a. Han, “Progress in reducing calcination reaction temperature of Calcium-Looping CO2 capture technology: A critical review,” 2022.

[11] E. P.-S. e. al., “Solar-Powered Direct Air Capture: Techno-Economic and Environmental Assessment,” 2024.

[12] H. a. C. C. Ishaq, “Potential of offshore wind energy for direct air capture,” 2022.

[13] Carbon Engineering, “What is Direct Air Capture and how DAC Works,” 2020. [Online]. Available: https://carbonengineering.com/our-technology/ .

[14] N. e. a. Ameli, “Higher cost of finance exacerbates a climate investment trap in developing economies,” 2021.

[15] O. C. C. A. C. R. a. P.-M. Sofía García-Luna, “Oxygen production routes assessment for oxy-fuel combustion,” 2022.

[16] Zero Carbon Shipping, “Managing Emissions from Ammonia-Fueled Vessels,” 2023. [Online]. Available: https://cms.zerocarbonshipping.com/media/uploads/documents/Ammonia-emissions-reduction-position-paper_v4.pdf.

[17] IEA, “Towards hydrogen definitions based on their emissions intensity,” 2023. [Online]. Available: https://www.iea.org/reports/towards-hydrogen-definitions-based-on-their-emissions-intensity.

Life Cycle Impact Assessment on the Green Hydrogen Supply Chain: Patagonia – Rotterdam

Rahul Limbachia, Jack Worthington
Department of Chemical Engineering, Imperial College London, U.K.

Abstract
Regions with significant renewable energy generation potential, such as Patagonia, offer a promising pathway to decarbonising global energy sectors through intercontinental hydrogen transportation. Greenhouse gas emissions of such energy transport is a key area of concern, hence the need to quantify and compare green hydrogen supply chains to current energy systems. Whilst numerous Life Cycle Impact Assessments (LCIAs) focusing on hydrogen production have been conducted, comprehensive assessments covering the entire supply chain, including shipping and leakage considerations, are scarce. This study quantifies the entire lifecycle emissions associated with green hydrogen production via wind energy, inclusive of intercontinental shipping from Patagonia (Chile) to Rotterdam (Netherlands), encompassing all cradle-to-gate stages. Key emission contributors were identified and varied through sensitivity and uncertainty analyses to drive recommendations for future supply chain optimisation. The results of the LCIA demonstrate that green hydrogen strongly outperforms alternative fuels, such as grey hydrogen and LNG, on an environmental basis, with a base case scenario placing the emissions intensity of an intercontinental green hydrogen supply chain at 1.82 $kgCO_2e/kgH_2$ delivered. The sensitivity analysis highlights transportation distance, leakages, and hydrogen production method as key sensitivities to the overall emissions intensity. The uncertainty analysis presents a range of 1.17-2.34 $kgCO_2e/kgH_2$, with a mean of 1.70 $kgCO_2/kgH_2$, placing the base case slightly above average. Upon economic comparison to counterfactual LNG and grey hydrogen supply chains, green hydrogen exhibits emissions avoidance costs of \$0.234/$kgCO_2e$ and \$0.412/$kgCO_2e$ respectively, outlining the necessity for further optimisation and refinement of green hydrogen supply chains.

Key Words
Green Hydrogen, Liquid Hydrogen, Life Cycle Impact Assessment, Life Cycle Inventory, Supply Chain, GHG Emissions, GWP, CO_2e, Sensitivity Studies.

1 Introduction

Mounting pressures to decarbonise global energy systems have intensified the search for sustainable alternatives to conventional energy production and transportation methods. Carbon-intensive fuels and energy vectors, such as liquefied natural gas (LNG), must be gradually phased out, whilst existing processes reliant on non-renewable energy sources require a transition toward cleaner fuels. Hydrogen (H_2) offers a promising alternative fuel substitute, with a dual advantage as both a renewable energy storage medium, and a pathway for intercontinental renewable energy transport. This potential has positioned hydrogen as a critical solution for enabling the energy transition.

LNG is currently the dominant energy vector that hydrogen seeks to replace, largely due to their comparable roles in storage and transport applications. Whilst LNG is a cleaner alternative to conventional fossil fuels like coal and oil, the drive to decarbonise global energy sectors necessitates further innovation. Hydrogen offers a significant advantage in this regard, as its environmental impact depends heavily on its primary energy source. When derived from renewable energy (green hydrogen), it boasts near-zero emissions, positioning it as a superior alternative to LNG for sustainable energy systems.

Current hydrogen supply chains initiate with energy production, with the primary energy source being either non-renewable (grey), natural gas (blue), or renewables (green). This energy powers the electrolysis process (with Alkaline and SOFC electrolysers) to produce gaseous hydrogen (GH_2). This hydrogen is then converted to liquid form (LH_2) within a liquefaction plant and stored temporarily, transported to its required location, and finally stored again, before being regasified upon entering domestic pipeline networks for industrial use. However, each stage of the supply chain introduces its own challenges in terms of energy use, emissions, and infrastructure requirements.

Despite the promising environmental benefits of hydrogen, its widespread implementation faces significant technical and economic challenges. The liquefaction process is highly energy-intensive, requiring cooling to -253˚C – far below the -162˚C needed for LNG [1]. Furthermore, the low boiling point of hydrogen can lead to boil-off losses during storage and intercontinental transportation, undermining resource efficiency and increasing emissions over long-distance routes through leakages. Infrastructure development for hydrogen supply chains is further hindered by the need for highly specialised and insulated storage tanks to reduce thermal losses, as well as the low energy density of LH_2 necessitating large transport volumes.

Economically, the lack of established infrastructure for cryogenic (<-150˚C) [2] storage and shipping poses significant barriers, requiring substantial upfront capital investment that limits the initial scalability of LH_2 supply chains. Maintaining the extremely low temperatures needed for LH_2 transportation and storage results in high operational costs, which add to the overall financial burden. Inefficiencies throughout the supply chain from cradle-to-gate intensify both technical and economic challenges, hindering the feasibility of hydrogen as a competitive energy vector.

To critically assess the viability of hydrogen as a sustainable energy vector despite these challenges, it is essential to address the limitations of existing research, particularly those related to its emissions across the supply chain. Whilst the environmental aspects of domestic green hydrogen production are well understood, there remains a significant gap and scarcity in assessments quantifying the overall carbon dioxide equivalent (CO_2e) impact of the green hydrogen supply chain, inclusive of intercontinental transportation.

One key limitation in existing research is that many of the available studies in this field rely upon outdated global warming potential (GWP) metrics, which have since been updated in recent scientific findings [3-17]. With the ongoing refinement of GWP values, particularly in the latest IPCC AR6 report [18], it has become clear that the contributions of certain potent gases, like methane and hydrogen, to climate change have been underestimated in earlier studies. For instance, a study by Sayer et al. [9], included CO_2e calculations based on IPCC assessment reports from early 2022. Older studies making use of now outdated GWP values may have miscalculated the true supply chain CO_2e impact, highlighting the need for updated research.

Many studies also fail to provide a complete cradle-to-gate analysis, often omitting critical components such as liquefied hydrogen storage and regasification [3-10, 12-17, 19-26]. A full analysis must encompass every stage, or risk significantly impacting results from omissions. For instance, Weidner et al. [10] fails to account for the emissions associated with shipping and regasification, whilst Sayer et al. [9] omits the emissions-intensive primary energy production and storage stages.

Failing to account for hydrogen leakages across the supply chain is also an issue that persists in many studies [3-9, 11-17, 19, 22-24, 27-29]. Hydrogen, though not a direct greenhouse gas (GHG), has been found to have significant indirect global warming effects when released into the atmosphere [30]. A cradle-to-gate supply chain study by Li et al. [11] fails to consider hydrogen leakage across each stage, this omission has likely led to an underestimation of overall CO_2e intensity, and a misrepresentation of the true environmental impact of green hydrogen.

Finally, another significant limitation in existing research is the lack of a relevant sensitivity analysis capturing key uncertainties and sensitivities, such as transportation distance and hydrogen production method split (Alkaline vs SOFC) [3-11, 13, 15, 17, 19-22, 28]. A study by Bhandari et al. [3], fails to perform this analysis, which reduces the robustness and reliability of their conclusions on CO_2e intensity. Further studies by Ozawa et al. [13] and Wulf et al. [17], fail to account for variability in GWP values and hydrogen leakage factors. Conversely, research by Krishnan et al. [27] explores the high impact of varying construction material requirements on CO_2e intensity, underscoring the need for relevant sensitivity analyses in future studies.

Accurate CO_2e accounting is critical for assessing the environmental benefits of green hydrogen, since its potential to be a sustainable alternative fuel hinges on minimising its CO_2e intensity. Studies that fail to analyse a complete supply chain, neglect hydrogen leakage effects, and omit a relevant sensitivity analysis, risk overstating the environmental advantages of green hydrogen.

The primary objective of this study is to build off current research performed on hydrogen supply chain optimisation [31] as shown in Figure 1, in order to produce estimates for the emissions (CO_2e) intensity of a green hydrogen supply chain through a Life Cycle Impact Assessment (LCIA), inclusive of intercontinental shipping. We aim to compare this value to a counterfactual LNG and grey hydrogen supply chain to determine the avoided CO_2e and cost benefits of transitioning to green hydrogen. This study will also explore key sensitivities, uncertainties, and economic implications to the supply chain, including the impact of transportation distance and hydrogen production method, to identify factors with the greatest influence on the overall carbon intensity and to enable counterfactual comparisons.

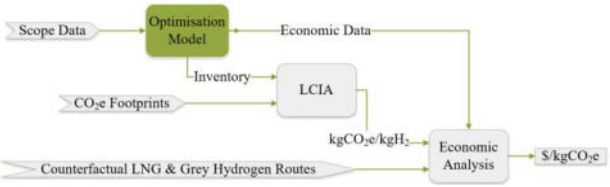


Figure 1. Optimisation model integration flowchart. The model inputs are represented by chevrons, with the research of this study being represented by grey boxes.

The remainder of this paper is structured as follows according to the phases of a complete Life Cycle Impact Assessment. Section 2 defines the methods and scope of this study, as well as the assumptions and methodology behind the inventory assessment and various analyses, Section 3 outlines the results of the assessment focusing on the supply chain CO_2e intensity, and Section 4 presents relevant sensitivity, uncertainty, and economic analyses with a discussion of the results.

2 Methods

It is crucial that the CO_2e emissions of the green hydrogen supply chain are accurately, and scrupulously, accounted for using novel knowledge and emerging technologies. This analysis considers the entire cradle-to-gate supply chain, from primary energy production, to entering domestic pipelines. This includes all stages of primary energy production, electrolysis, liquefaction, origin and destination storage, intercontinental shipping, and regasification, ensuring comprehensive accounting of carbon emissions throughout the supply chain, see Figure 2. Emissions associated with the distribution and usage of the hydrogen post-entering the pipeline are not considered within this study.

Allowing for the integration of the optimisation model into this study, the supply chain is defined as a route from Patagonia (Chile) to Rotterdam (Netherlands), where the gaseous hydrogen will enter domestic pipeline infrastructure.

The renewable primary energy source will be in the form of wind energy, as conditions in the Punta Arenas region of Patagonia are favourable for this method of energy production. This renewable energy will enter the domestic grid and supply an electrolysis facility located close to the Port of Valparaíso, where a combination of alkaline and SOFC electrolysers produce gaseous hydrogen. The gaseous hydrogen will then undergo liquefaction at a separate port facility, located near storage tanks for the liquefied hydrogen awaiting transportation. Liquefied hydrogen and boil-off fuelled tanker ships will transport the hydrogen to the Port of Rotterdam, where it will be held temporarily in storage, before undergoing regasification at a terminal. This study will consider the effect of hydrogen leakages, as well as all Scope 1 (direct), Scope 2 (energy-related) & Scope 3 (indirect, supply chain-related) emissions of each stage across the supply chain.

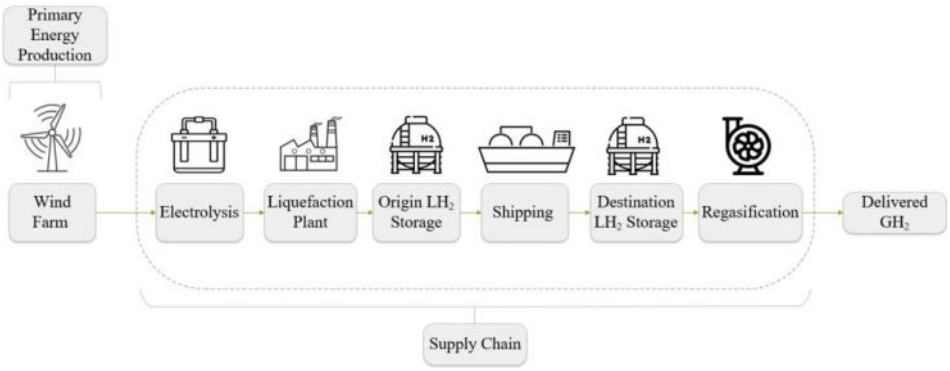


Figure 2. Schematic representation of the green hydrogen supply chain, illustrating the key stages from renewable energy production via wind energy, through electrolysis and liquefaction, to intercontinental shipping and regasification with storage, for delivery of gaseous hydrogen.



Figure 3. Detailed scaled breakdown of the energy flows, losses, and inefficiencies across the green hydrogen supply chain. Losses through each stage are highlighted, with wind energy conversion and electrolysis combined. Specific energy values are based on the output of the optimisation model [31].

2.1 Model Integration

Further to the research of Aldren et al. [31], this study has utilised an existing supply chain optimisation model in order to generate key inputs for the LCIA. The model has simulated energy flows throughout the supply chain, with detailed parameters for each stage, allowing for the characterisation of processes. The energy flows provide the basis for calculating the size and capacity of the equipment required at each stage. Figure 3 depicts the energy flows through each stage, with operational losses and product usage accounted for.

2.2 Life Cycle Inventory

The following inventory section outlines the key inputs and processes involved in the green hydrogen supply chain, forming the basis for the LCIA.

From the optimisation model, it was indicated that the primary energy for the entire supply chain is to be generated via a wind farm in the Punta Arenas region of Patagonia. A base plant with a total capacity of 4.755 GW contains a total of 1441 turbines, each with a capacity of 3.3 MW. In order to calculate the overall emissions of the plant, including construction, operational and end-of-life emissions, it was decided that lifetime emissions-to-air data for a single 3.3 MW commercial turbine model [32] would be scaled to reach the required energy output. Relevant GWP values were assigned to each organic and inorganic emission to generate a lifetime CO_2e emission value for the plant, which was then normalised to a per kilogram delivered GH_2 basis. This method ensured that environmental impacts associated with primary energy production are accurately and entirely accounted for. This approach has been structurally replicated for all other stages to ensure per-stage CO_2e impacts are proportionately represented.

To develop the inventory for the electrolyser stage, materials and process data from established alkaline [33] and solid oxide fuel cell (SOFC) [34] electrolyser models were taken and scaled to the required throughput. Both variants of electrolyser were assumed to be constructed within the same facility, with the electricity supply being exclusively provided via primary energy. The alkaline and SOFC

electrolysers would operate at 70% and 85% conversion efficiency respectively and would have maximum outputs of 58.26 t/h and 31.85 t/h, in line with the models. As a result of the initial optimisation model, the electrolysis facility would be required to produce 60.62 t/h of GH_2, and thus a production method split between alkaline and SOFC electrolysis is necessary. As a base case it was decided that alkaline electrolysis would be responsible for 50 t/h of GH_2 with the remainder being produced via SOFC electrolysis, due to the current wider adoption and large scale alkaline electrolysis within industry. The scaled construction material and operational resource requirements were then assigned relevant GWP values, and both an overall and a normalised CO_2e emission value was calculated.

To enable international transportation, the gaseous hydrogen produced by the electrolysers must be liquefied. The liquefaction and regasification plant designs were adapted from an IdealHY study [35], which outlined a base plant with a 50 t/d capacity and a specified materials inventory. To meet the requirement of the optimisation model, the base plant was scaled to a capacity of 57 t/h of liquid and gaseous hydrogen production. Construction and engineering materials were scaled using a calculated factor of 27.75, reflecting the proportionally increased capacity and material requirements of additional turbines. However, concrete requirements were adjusted from this scaling, as its primary use in the base plant was as foundational material for the already highly oversized facility footprint. This scaling approach ensured that the material inputs for the plant were accurately represented. The liquefaction plant was first evaluated and its emissions intensity calculated.

Similarly, the regasification plant was also evaluated, and its energy output was factored into the overall CO_2e balance. The energy produced from the regasification process, which occurs when the LH_2 is converted back into gaseous form, results in a negative CO_2e value based on the CO_2e intensity of the primary energy supply. This is due to the energy released during the expansion of the liquid hydrogen, which offsets the CO_2e emissions associated with the materials and energy inputs required for the process.

This study incorporates LH_2 boil-off gas as a shipping fuel to improve the overall CO_2e efficiency of intercontinental transportation, a consideration often overlooked in alternative assessments. The CO_2e inventory for the shipping stage was developed based on two key factors: the shipping distance, which influences transit times and emissions during transport, and the construction materials for the vessels themselves. Longer shipping distances necessitate additional ships, increasing material CO_2e emissions, in order to maintain an uninterrupted supply chain, introducing a logistical scaling factor with stepwise emissions impacts. It was assumed the ships would travel at design speed for the entire voyage, allowing transportation distance to be specified in hours. The ship design, derived from Project Jamila [36], reflects an LH_2 powered transport vessel optimised to align with load requirements, and minimise embodied CO_2e emissions from construction material requirements. After scaling, it was found that 4 ships would be required to facilitate a continuous supply chain across the 360-hour route from the Port of Valparaíso to the Port of Rotterdam.

The storage of LH_2 represents a crucial component of the supply chain, occurring both before and after the transportation phase. Maintaining the cryogenic state of LH_2 requires specialised storage tanks designed to withstand cryogenic temperatures and minimise leakages, a significant contributor to CO_2e emissions. For this study, the required storage tanks are modelled on the designs presented in Project Jamila [36], chosen for their compatibility with the load requirements and operational demands of the supply chain. To account for the CO_2e impact of storage, the inventory includes two 5000-ton cryogenic storage tanks on either side of the transportation stage, collectively capable of storing 200,000m^3 of LH_2.

Based on the findings of Sand et al. [30], it is vital to consider the CO_2e impact of any possible leakages throughout the supply chain. Utilising literature data [37], a base case assumption of 3% leakage was applied across all stages of the process, to account for boil-off losses and general equipment leakage. Using this leakage rate with the GWP of hydrogen [30], the normalised CO_2e contributions from hydrogen leakages were calculated and incorporated into the overall emissions inventory. The inclusion of this parameter ensures a more comprehensive and realistic representation of the overall emissions intensity.

2.3 Analysis of Results

In this study, comprehensive sensitivity, uncertainty, and economic analyses are undertaken to evaluate the robustness and financial implications of the base case results. Detailed frameworks for the sensitivity and uncertainty analyses are provided in the supplementary document (see Section 8). The economic analysis employs the avoidance cost equation (1) to enable a comparison between counterfactual supply chains.

3 Results

Following the outlined methodology, the life cycle inventory was assessed across the key supply chain stages, along with the impacts of hydrogen leakage, to calculate the associated CO_2e per kilogram of delivered gaseous hydrogen. The results indicate that primary energy production is the dominant contributor to the overall lifecycle emissions, overshadowing other stages of the supply chain. The total lifecycle emissions intensity was determined to be 1.82 $kgCO_2e$ per kilogram of hydrogen delivered.

Figure 4 illustrates the emissions intensity of both the energy production and supply chain stages of the green hydrogen production lifecycle. The data indicated that emissions from primary energy production (0.86 $kgCO_2e/kgH_2$), are only slightly lower than the combined emissions associated with

the subsequent supply chain (0.96 $kgCO_2e/kgH_2$), since the energy generated is expected to cover the energy demands of the entire supply chain.

Component	Delivered Basis ($kgCO_2e/kgH_2$)
Wind Farm	0.86
Electrolysis (Alkaline)	0.19
Electrolysis (SOFC)	0.02
Liquefaction Plant	0.03
Origin LH_2 Storage	0.07
Shipping	0.26
Regasification Terminal	-0.01
Destination LH_2 Storage	0.07
Hydrogen Leakage	0.33
Overall Intensity	1.82

Table 1. Breakdown of emissions intensity for each stage in the green hydrogen supply chain.

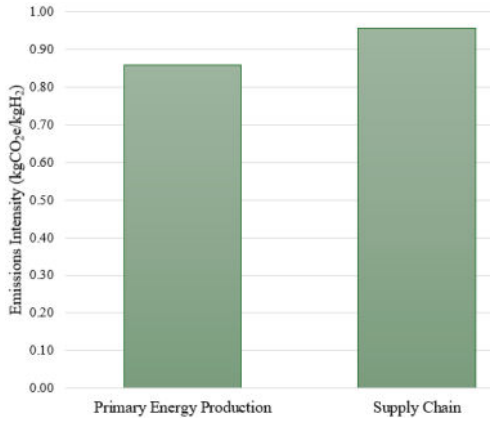


Figure 4. Comparison of emissions intensity contributions between primary energy production and supply chain stages.

Figure 5 presents the breakdown of supply chain stage emissions, illustrating the dominance of hydrogen leakage as the largest contributor at 0.33 $kgCO_2e/kgH_2$. This value reflects losses from boil-off management and fugitive emissions throughout the supply chain. Shipping follows as the second-largest source of emissions at 0.26 $kgCO_2e/kgH_2$, driven by the large material requirements needed for the construction of the vessels. Combined contributions from alkaline and SOFC electrolysis total 0.21 $kgCO_2e/kgH_2$. Liquid hydrogen storage accounts for a cumulative 0.14 $kgCO_2e/kgH_2$, primarily as a result of the construction materials required. The regasification stage, though part of the supply chain, was omitted from the diagram and contributes -0.01 $kgCO_2e/kgH_2$ due to its net-positive impact on the overall emissions intensity, as detailed in Section 2.2. The figure provides an insight into the proportional contributions of each supply chain component, highlighting potential areas of particular significance in future emissions reduction strategies.

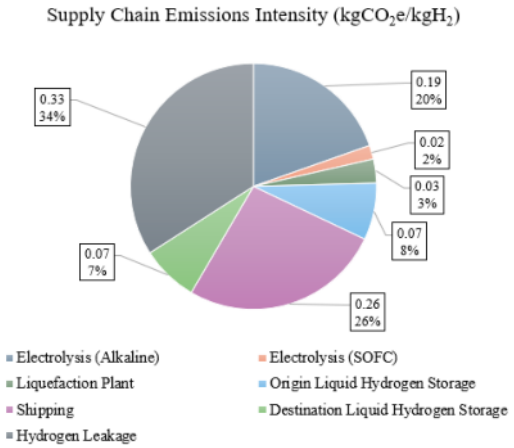


Figure 5. Breakdown of emissions intensity contributions across the green hydrogen supply chain stages, highlighting the dominant impacts of leakages, electrolysis, and shipping.

4 Discussion

4.1 Analysis of Base Case Results

The results highlight the proportional contributions of each stage of the supply chain and their importance in determining the overall emissions intensity of green hydrogen. Specifically, the dominance of primary energy emissions underscores the need for improved turbine designs with minimised material requirements. Whilst wind energy production has inherently near-zero emissions during operation, large-scale emissions arise from the manufacturing and installation of turbines, as well as their maintenance and replacement over time. In this case, the capacity of the wind farm was designed to meet the energy demands of the entire subsequent supply chain, meaning all energy usage emissions have been solely attributed to primary energy production, hence the significant contribution.

Within the supply chain stages, hydrogen leakage emerged as the greatest contributor to emissions, which has a disproportionately high impact relative to its scale. With an emissions intensity above that of the entire electrolysis process, minimising leakages serves as a critical area for improvement within supply chains. Small amounts of hydrogen leakage can undermine the overall lifecycle CO_2e advantages of using hydrogen, highlighting the importance of advancements in storage technologies to reduce boil-off losses, and facility infrastructure to reduce gaseous hydrogen leaks.

An additional significant contributor to overall emissions is shipping, with emissions mostly driven by the extensive quantities of construction materials required for each of the four the vessels. The necessity to transport the LH_2 over vast distances on the Chile-Netherlands route has meant large LH_2 fuel tanks must be constructed on the ship, requiring large quantities of high-strength and energy-intensive aluminium.

The findings suggest that improvements in ship design, low-emission material selection, and operational efficiency could significantly reduce overall CO_2e emissions.

The combined electrolysis facility also contributes substantially to overall emissions despite its energy usage being accounted for previously. These emissions are primarily a result of the emissions-intensive extraction and manufacturing of the materials used within the electrolysis processes. For instance, 42% of emissions associated with the alkaline electrolysis process were directly attributable to the heavily manufactured materials required within the electrolysers themselves. Further advances in electrolysis technologies should aim to reduce the usage of unsustainable materials.

The emissions associated with both origin and destination LH_2 storage are primarily driven by the construction materials required when building the storage infrastructure. These emissions stem from the high energy demands of producing and processing materials such as cryogenic-grade stainless steel and the advanced aluminium manufacturing processes required for the insulation. Although these stages contribute less to the overall emissions intensity compared to hydrogen production and shipping, the results emphasise the necessity to explore alternative low-emission construction materials or construction methods to lower their overall impact.

A unique finding is the negative emissions attributed to regasification, achieved through energy recovery during the expansion of LH_2 to its gaseous form. Although its impact is relatively small when compared to its neighbouring stages, the results demonstrate an opportunity to further offset the emissions of alternative stages of the supply chain through innovative energy recovery strategies.

The findings reveal several key variables that significantly influence the overall CO_2e intensity of the supply chain. To better understand the relative importance of these variables, a sensitivity analysis is conducted to identify priority areas for emissions reductions.

4.2 Sensitivity Analysis

The one value at a time (OVAT) sensitivity analysis explores the effect on overall CO_2e intensity of altering individual parameters. By systematically varying parameters, the influence of each factor can be quantified, see Figure 6. Transportation distance and ship quantity requirements have been varied to represent realistic ranges for intercontinental shipping. The bounds reflect differences in shipping scenarios, including short local distances versus long international supply chains, where the overall distance determines the number of ships required for the route to be continuous within a supply chain. The distribution of hydrogen production method has been varied as a percentage of sole alkaline electrolysis production. The bounds reflect the limits of alkaline hydrogen production for the electrolysis plant to meet the required gaseous hydrogen output. The remaining parameters have been varied according to their uncertainties from their base case values.

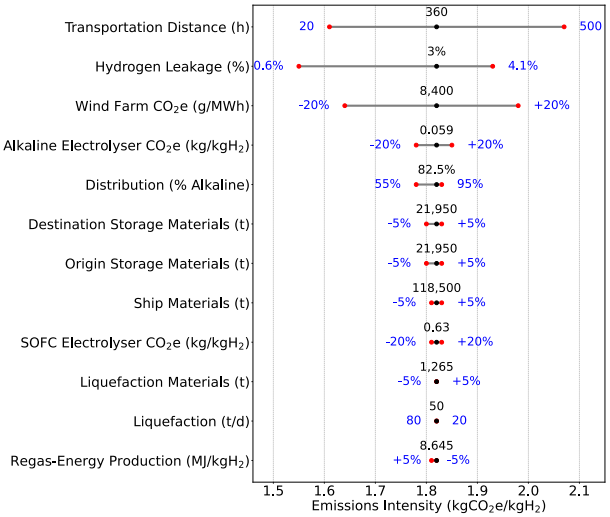


Figure 6. OVAT sensitivity analysis results demonstrating the impact of varying each parameter individually on the overall CO_2e intensity of the green hydrogen supply chain.

The variation of transportation distance produces the greatest change in overall emissions intensity, with an increase of 13.73% when shipping distance is increased to 500 hours from 360 hours, and a decrease of 11.54% when decreased to 20 hours. Shipping distances directly influence the number of required ships included within the inventory.

The distribution of hydrogen production method (alkaline vs SOFC) also induces substantial variation in the overall CO_2e intensity. SOFC electrolysis demonstrates approximately half the emissions intensity of alkaline electrolysis per kilogram of GH_2 produced, resulting in noticeable reductions in total supply chain emissions when SOFC production is prioritised. However, the outputs of the initial optimisation model revealed capacity limitations for each electrolyser type. If these constraints could be addressed, expanding SOFC production would be significantly more advantageous, offering greater emissions savings compared to alkaline electrolysis.

Predictably, hydrogen leakage significantly influences the overall emissions intensity, with its impact increasing proportionally as the leakage percentage rises. The omission of leakage effects within previous studies (See Section 1) has often led to a considerable underestimation of the true CO_2e intensity. The results of this analysis emphasise the importance of accurate hydrogen leakage accounting in subsequent life cycle impact assessments, in order to ensure valid emissions results. Furthermore, minimising hydrogen leakage should be prioritised, as it has been found to be a key determinant of the supply chain's environmental impact. As hydrogen leakage can contribute to substantial indirect warming effects [30], its reduction is essential for optimising the environmental benefits of green hydrogen.

Upon varying the total wind farm CO_2e according to its uncertainty [32], a ±20% variation resulted in a variation of approximately +8.79% and -9.90% of the overall CO_2e value. This significant impact can be attributed to the scale of the wind farm and its extensive material requirements. As the

primary energy source for the entire supply chain, its variation can influence overall emissions substantially, as high material and manufacturing demands for components introduce high embodied and released CO_2e emissions.

The remaining parameters exhibited minimal influence on the overall emissions intensity, as the associated stages contributed only marginally to the total lifecycle emissions. This is largely due to the relatively small proportion of CO_2e emissions originating from these stages compared to the dominant contributors, such as primary energy production, hydrogen leakage, transportation, and electrolysis. Parameters such as storage material requirements or variations in regasification energy production had limited impact, reflecting their secondary roles within the overall supply chain.

These findings suggest that optimisation efforts should focus primarily on the key sensitivities as dominant contributors, as adjusting parameters for low-impact stages is unlikely to significantly affect the viability of green hydrogen. Building on the insights of the sensitivity analysis, an uncertainty analysis was also conducted to capture the combined effects of parameter variation and uncertainty. This approach provides a probabilistic distribution of the overall emissions intensity, identifying the most likely value and its associated range, thereby offering an insight into the potential variability within the system.

4.3 Uncertainty Analysis

To quantify the impact of parameter uncertainties on the overall emissions intensity, a Monte Carlo simulation was conducted. The process involved 10,000 iterations of the life cycle inventory model, each using randomly sampled parameter values within their predefined uncertainty bounds. The resulting data was compiled into a histogram and fitted with a normal distribution curve, as shown in Figure 7.

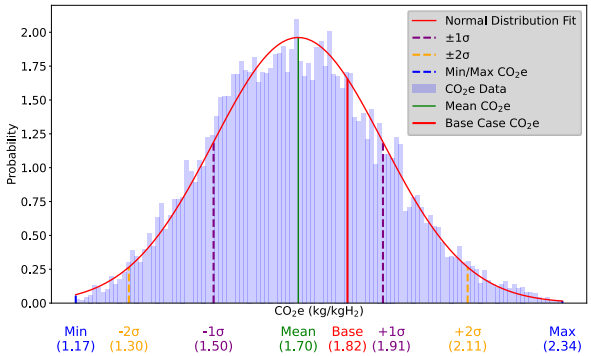


Figure 7. Results of the uncertainty analysis. 10,000 model iterations were compiled into a histogram and fitted with a normal distribution. Standard deviation percentiles have been indicated from the mean, with the base case value also being highlighted.

The mean CO_2e intensity was found to be 1.70 $kgCO_2e/kgH_2$, slightly lower than the base-case value of 1.82 kg CO_2e/kgH_2. This indicates that the base case represents a near-average scenario, slightly overestimating the emissions compared to the probabilistic mean. The distribution highlights the variability in the model, with emissions values ranging from 1.17-2.34 $kgCO_2e/kgH_2$.

Approximately 68% of the CO_2e values fall within one standard deviation (±1σ) of the mean (1.50-1.91 $kgCO_2e/kgH_2$). Similarly, approximately 95% of the values lie within two standard deviations (±2σ) of the mean (1.30-2.11 kg CO_2e/kgH_2). These ranges emphasise that whilst the base case estimate is slightly non-conservative, the actual emissions intensity for the supply chain is highly likely to fall within a small range of the base estimate, providing a reliable basis for decision making in hydrogen supply chain planning.

4.4 Comparison to Literature

The Renewable Energy Directive II (RED II) [38] mandates that Renewable Fuels of Non-Biological Origin (RFNBOs), such as hydrogen produced via electrolysis, must achieve a minimum of a 70% reduction in GHG emissions relative to fossil fuel-derived alternatives. Based on a benchmark value of 3.38 $kgCO_2e$ per kilogram hydrogen delivered, this study represents nearly a 50% reduction in the RED II target, highlighting significant progress toward meeting regulatory goals.

When compared to existing literature, which reports overall emissions intensities for electrolysis-based hydrogen ranging from 1.15-3.6 $kgCO_2e/kgH_2$ [39], depending on factors such as primary energy source, hydrogen production method, and GWP estimates, the base case estimate from this study initially falls on the lower end of the range. This relatively low value is primarily attributable to the exclusive use of wind energy for power generation and energy supply, effectively eliminating emissions associated with non-renewable energy sources. Furthermore, this study ensures compliance with the additional RED II criteria, including the requirement for energy vectors to be produced from fully renewable energy sources.

4.5 Economic Analysis

The concept of CO_2e avoidance cost serves as a vital metric for assessing the economic efficiency of GHG mitigation strategies. To enable the comparison of different energy vectors, a per unit of energy basis has been adopted. The Levelised Cost of Production (LCP) quantifies the production cost per unit of energy released by the hydrogen or LNG, expressed in $/MJ. The Emissions Intensity of Production (EI) represents the GHG emissions produced per MJ of product as referenced throughout the study, measured in $kgCO_2e/MJ$. The Avoidance Cost (AC), calculated in $/$kgCO_2e$ avoided, evaluates the cost-effectiveness of avoiding one kilogram of CO_2e by transitioning from a reference (higher emissions) technology to a novel (lower emissions) process, making it a critical tool for comparing the viability of different decarbonisation pathways. This metric aligns directly with the objectives of this study, providing a framework to assess the economic and environmental trade-offs of adopting green hydrogen production and its associated supply chain stages. The formula for calculating CO_2e avoidance cost is shown below:

$$AC_{CO_2e} = \frac{(LCP)_{New} - (LCP)_{Existing}}{(EI)_{Existing} - (EI)_{New}} \quad (1)$$

4.5.1 Comparison to a Grey Hydrogen Supply Chain

The emissions intensity of green hydrogen produced in this study (0.015 $kgCO_2e/MJ$) represents a substantial improvement compared to conventional grey hydrogen production (0.078 $kgCO_2/MJ$) [40]. This demonstrates an 80.64% reduction in CO_2e emissions, highlighting the significant environmental advantages of transitioning to green hydrogen. Such reductions align closely with the decarbonisation targets of RED II. The vast contrast in emissions intensities is driven by the fundamental differences in energy sources.

The avoidance cost for transitioning from grey to green hydrogen is calculated to be \$0.412 per $kgCO_2e$ mitigated. This cost is influenced by the significantly higher LCP for green hydrogen, at \$0.039 per MJ [31], compared to grey hydrogen, which has an LCP of \$0.013 per MJ [40]. Whilst green hydrogen production involves higher upfront costs, the substantial reduction in emissions reinforces its value as a key solution for decarbonisation compared to grey hydrogen.

4.5.2 Comparison to an LNG Supply Chain

When comparing the emissions intensity of existing LNG supply chain routes to that of green hydrogen, this study reveals a significant environmental advantage. LNG supply chain routes report an emissions intensity of 0.141 $kgCO_2e/MJ$ [41], whereas green hydrogen production from this study achieves an emissions intensity of 0.015 $kgCO_2e/MJ$ – an 89.36% reduction. Despite the environmental benefit, the LCP for green hydrogen is \$0.039/MJ, considerably higher than the \$0.010/MJ reported for LNG supply chains [42]. Similar to grey hydrogen, LNG exhibits a higher emissions intensity, but a substantially lower LCP compared to green hydrogen. However, the closer emissions intensity between LNG and green hydrogen results in a higher avoidance cost for transitioning.

This avoidance cost of \$0.234 per $kgCO_2e$ reflects the financial trade-offs required to achieve meaningful emissions reductions. The economic considerations associated with transitioning to green hydrogen highlight the necessity of implementing supportive policy measures, such as carbon pricing and production subsidies, to bridge the cost gap between green hydrogen and more cost-effective yet carbon-intensive alternatives like LNG and grey hydrogen. These mechanisms are critical to accelerating the widespread adoption of green hydrogen as a viable decarbonisation solution.

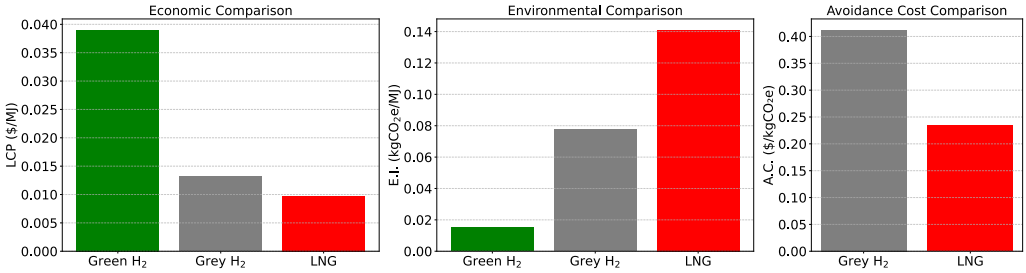


Figure 8. Comparison of Economic, Environmental, and Avoidance Costs for green hydrogen, grey hydrogen, and LNG. The avoidance costs are shown relative to green hydrogen. The graphs illustrate that whilst green hydrogen offers the greatest environmental benefits, it is the least cost-effective energy vector option.

5 Conclusions

This study successfully fulfils its primary objectives by providing a comprehensive estimate of the emissions intensity of a green hydrogen supply chain. Through a detailed Life Cycle Impact Assessment, the emissions intensity of the green hydrogen supply chain was determined to be 1.82 $kgCO_2e/kgH_2$ or 0.0152 $kgCO_2e/MJ$, a significant improvement over the 0.141 $kgCO_2e/MJ$ benchmark for LNG supply chains, and the 0.078 $kgCO_2e/MJ$ associated with grey hydrogen. These results reinforce the substantial environmental benefits of transitioning to green hydrogen, achieving emissions reductions of 89.36% and 80.64%, respectively.

In addition to emissions quantification, this study evaluated the economic efficiency of GHG mitigation through CO_2e avoidance costs. Green hydrogen was found to have avoidance costs of \$0.412 and \$0.234 per $kgCO_2e$ when transitioning from grey hydrogen and LNG, respectively, suggesting that it would be more economically efficient to transition LNG routes first. Whilst these costs demonstrate financial trade-offs are required for significant decarbonisation, they also highlight the potential for green hydrogen to be incorporated into new net-zero policies, with measures such as carbon pricing and production subsidies.

The sensitivity analysis identified transportation distance, hydrogen production method, and hydrogen leakage as the most influential variables on the emissions intensity of the green hydrogen supply chain. The findings suggest that optimisation efforts should be focused on these dominant contributors, as their relative impacts hugely outweigh the impacts of smaller contributors, such as construction materials usage.

By integrating a Monte Carlo uncertainty analysis, this study mapped a normal distribution of CO_2e values that centred slightly below the base case value of 1.82 $kgCO_2e/kgH_2$, with a range of 1.17-2.34 $kgCO_2e/kgH_2$ reflecting parameter uncertainties. This probabilistic framework offered a robust understanding of the variability

within the system and supports future decision-making based on calculated uncertainty.

In summation, this study combines a comprehensive green hydrogen LCIA with sensitivity, uncertainty, and economic analyses to produce a robust evaluation of the environmental and economic implications of an intercontinental green hydrogen supply chain. By addressing the critical parameters affecting emissions intensity and contextualising these findings against current alternative fuels, this study demonstrates that green hydrogen supply chains and imports can be a viable and essential component of the energy transition.

6 Outlook & Future Work

This study presents a comprehensive LCIA of green hydrogen transport between Patagonia and Rotterdam. Building on this, future studies could investigate alternative hydrogen transport pathways to regions anticipated to experience significant hydrogen demand by 2050. Asia is expected to remain as the region with the largest hydrogen demand, particularly in China and India [43]. Conducting LCIAs on hydrogen transport routes from major production regions, such as Patagonia, to high-demand regions could provide valuable insights into environmental and economic implications. For example, economic benefits could arise from creating a hydrogen supply chain that originates in regions with established renewable energy infrastructure.

Further investigation into the role of policy instruments, such as carbon pricing, subsidies, and green hydrogen certification schemes, would be crucial for improving the competitiveness of green hydrogen. Performing a sensitivity analysis on these policy variables would help to identify the most impactful factors and inform effective policy development.

Finally, whilst this study focuses on a 100% green hydrogen supply chain, the reality of the transition to net-zero will involve a mix of green hydrogen and alternative energy vectors, as well as hydrogen produced via non-renewable energy sources. Developing a dynamic model that accounts for a blend of these vectors could help to facilitate a smoother transition by balancing the need to meet global demand, whilst progressively phasing out higher-emission fuels. Such models would be crucial in ensuring an efficient shift to a decarbonised energy economy.

7 Acknowledgements

We would like to thank Cameron Aldren for his invaluable support and guidance throughout this study.

8 Supplementary Information

Accompanying this study is a detailed supplementary document, which outlines the calculations supporting the life cycle impact assessment (LCIA), and provides the complete framework for the sensitivity and uncertainty analyses, including parameter bounds and the Monte Carlo simulation methodology.

9 References

[1] National Institute of Standards and Technology. (1997, August 19). *NIST Chemistry WebBook - SRD 69*. Nist.gov; National Institute of Standards and Technology. https://webbook.nist.gov/chemistry/

[2] Varbanov, P. S. (2013). Basic Process Integration Terminology. *Handbook of Process Integration (PI)*, 28–78. https://doi.org/10.1533/9780857097255.1.28

[3] Bhandari, R., Trudewind, C. A., & Zapp, P. (2014). Life cycle assessment of hydrogen production via electrolysis – a review. *Journal of Cleaner Production*, *85*, 151–163. https://doi.org/10.1016/j.jclepro.2013.07.048

[4] Cetinkaya, E., Dincer, I., & Naterer, G. F. (2012). Life cycle assessment of various hydrogen production methods. *International Journal of Hydrogen Energy*, *37*(3), 2071–2080. https://doi.org/10.1016/j.ijhydene.2011.10.064

[5] Dufour, J., Serrano, D. P., Gálvez, J. L., González, A., Soria, E., & Fierro, J. L. G. (2012). Life cycle assessment of alternatives for hydrogen production from renewable and fossil sources. *International Journal of Hydrogen Energy*, *37*(2), 1173–1183. https://doi.org/10.1016/j.ijhydene.2011.09.135

[6] Granovskii, M., Dincer, I., & Rosen, M. (2006). Life cycle assessment of hydrogen fuel cell and gasoline vehicles. *International Journal of Hydrogen Energy*, *31*(3), 337–352. https://doi.org/10.1016/j.ijhydene.2005.10.004

[7] Koroneos, C., Dompros, A., Roumbas, G., & Moussiopoulos, N. (2004). Life cycle assessment of hydrogen fuel production processes. *International Journal of Hydrogen Energy*, *29*(14), 1443–1450. https://doi.org/10.1016/j.ijhydene.2004.01.016

[8] Ji Hyun Lee, An, S., Cha, K., & Hur, T. (2010). Life cycle environmental and economic analyses of a hydrogen station with wind energy. *International Journal of Hydrogen Energy*, *35*(6), 2213–2225. https://doi.org/10.1016/j.ijhydene.2009.12.082

[9] Sayer, M., Amela Ajanovic, & Haas, R. (2024). Economic and environmental assessment of different hydrogen production and transportation modes. *International Journal of Hydrogen Energy*, *65*, 626–638. https://doi.org/10.1016/j.ijhydene.2024.04.073

[10] Weidner, T., Tulus, V., & Guillén-Gosálbez, G. (2022). Environmental sustainability assessment of large-scale hydrogen production using prospective life cycle analysis. *International Journal of Hydrogen Energy*. https://doi.org/10.1016/j.ijhydene.2022.11.044

[11] Li, J., Zhu, X., Djilali, N., Yang, Y., Ye, D., Chen, R., & Liao, Q. (2022). Comparative well-to-pump assessment of fueling pathways for zero-carbon transportation in China: Hydrogen economy or methanol economy? *Renewable and Sustainable Energy Reviews*, *169*, 112935. https://doi.org/10.1016/j.rser.2022.112935

[12] Tayarani, H., & Ramji, A. (2022). Life Cycle Assessment of Hydrogen Transportation Pathways via Pipelines and Truck Trailers: Implications as a Low Carbon Fuel. *Sustainability*, *14*(19), 12510. https://doi.org/10.3390/su141912510

[13] Ozawa, A., Kudoh, Y., Kitagawa, N., & Muramatsu, R. (2019). Life cycle CO2 emissions from power generation using hydrogen energy carriers. *International Journal of Hydrogen Energy*, *44*(21), 11219–11232. https://doi.org/10.1016/j.ijhydene.2019.02.230

[14] Ozawa, A., Inoue, M., Kitagawa, N., Muramatsu, R., Anzai, Y., Genchi, Y., & Kudoh, Y. (2017). Assessing Uncertainties of Well-To-Tank Greenhouse Gas Emissions from Hydrogen Supply Chains. *Sustainability*, *9*(7), 1101. https://doi.org/10.3390/su9071101

[15] Wulf, C., & Kaltschmitt, M. (2018). Hydrogen Supply Chains for Mobility—Environmental and Economic Assessment. *Sustainability*, *10*(6), 1699. https://doi.org/10.3390/su10061699

[16] Akhtar, M. S., Dickson, R., & Liu, J. J. (2021). Life Cycle Assessment of Inland Green Hydrogen Supply Chain Networks with Current Challenges and Future Prospects. *ACS Sustainable Chemistry & Engineering*, *9*(50), 17152–17163. https://doi.org/10.1021/acssuschemeng.1c06769

[17] Wulf, C., & Zapp, P. (2018). Assessment of system variations for hydrogen transport by liquid organic hydrogen carriers. *International Journal of Hydrogen Energy*, *43*(26), 11884–11895. https://doi.org/10.1016/j.ijhydene.2018.01.198

[18] IPCC. (2021). *Sixth Assessment Report — IPCC*. Ipcc.ch; IPCC. https://www.ipcc.ch/assessment-report/ar6/

[19] Noh, H., Kang, K., & Seo, Y. (2022). Environmental and energy efficiency assessments of offshore hydrogen supply chains utilizing compressed gaseous hydrogen, liquefied hydrogen, liquid organic hydrogen carriers and ammonia. *International Journal of Hydrogen Energy*. https://doi.org/10.1016/j.ijhydene.2022.11.085

[20] Kanz, O., Brüggemann, F., Ding, K., Bittkau, K., Rau, U., & Reinders, A. (2023). Life-cycle global warming impact of hydrogen transport through pipelines from Africa to Germany. *Sustainable Energy & Fuels*, *7*(13), 3014–3024. https://doi.org/10.1039/D3SE00281K

[21] Kodgire, P. (2024). Hydrogen - Impending Clean and Green Energy: Hydrogen Production Technologies Life Cycle Assessment Review. *Process Safety and Environmental Protection*. https://doi.org/10.1016/j.psep.2024.11.019

[22] Gerloff, N. (2021). Comparative Life-Cycle-Assessment analysis of three major water electrolysis technologies while applying various energy scenarios for a greener hydrogen production. *Journal of Energy Storage*, *43*, 102759. https://doi.org/10.1016/j.est.2021.102759

[23] Lee, D.-H. (2014). Development and environmental impact of hydrogen supply chain in Japan: Assessment by the CGE-LCA method in Japan with a discussion of the importance of biohydrogen. *International Journal of Hydrogen Energy*, *39*(33), 19294–19310. https://doi.org/10.1016/j.ijhydene.2014.05.142

[24] Wagd Ajeeb, Baptista, P., & Rui Costa Neto. (2024). Life cycle analysis of hydrogen production by different alkaline electrolyser technologies sourced with renewable energy. *Energy Conversion and Management*, *316*, 118840–118840. https://doi.org/10.1016/j.enconman.2024.118840

[25] Khan, A., Tarun Sitaraman, Haque, N., Leslie, G., Serkan Saydam, Rahman Daiyan (2024). Strategies for life cycle impact reduction of green hydrogen production – Influence of electrolyser value chain design. *International Journal of Hydrogen Energy*, *62*, 769–782. https://doi.org/10.1016/j.ijhydene.2024.01.081

[26] Perna, A., Jannelli, E., Di Micco, S., Romano, F., & Minutillo, M. (2023). Designing, sizing and economic feasibility of a green hydrogen supply chain for maritime transportation. *Energy Conversion and Management*, *278*, 116702. https://doi.org/10.1016/j.enconman.2023.116702

[27] Krishnan, S., Corona, B., Gert Jan Kramer, Junginger, M., & Koning, V. (2024). Prospective LCA of alkaline and PEM electrolyser systems. *International Journal of Hydrogen Energy*, *55*, 26–41. https://doi.org/10.1016/j.ijhydene.2023.10.192

[28] Ratner, S., Sergi, B. S.(2024). Eco-efficiency of hydrogen supply chains: NDEA-based approach. *International Journal of Hydrogen Energy*, *73*, 598–608. https://doi.org/10.1016/j.ijhydene.2024.06.053

[29] Rey, I., Barrio, V. L., & Agirre, I. (2024). Environmental assessment of a hydrogen supply chain using LOHC system with novel low-PGM catalysts: A life cycle approach. *International Journal of Hydrogen Energy*. https://doi.org/10.1016/j.ijhydene.2024.11.197

[30] Sand, M., Skeie, R. B., Sandstad, M., Krishnan, S., Myhre, G., Bryant, H., Derwent, R., Hauglustaine, D., Paulot, F., Prather, M., & Stevenson, D. (2023). A multi-model assessment of the Global Warming Potential of hydrogen. *Communications Earth & Environment*, *4*(1), 1–12. https://doi.org/10.1038/s43247-023-00857-8

[31] Aldren C., Shah N. and Hawkes A. (2024). Ammonia or Liquid Hydrogen: The True Cost of Shipping Low-Carbon Energy. *Article Submitted for Publication*

[32] Garrett, P. (2014). Electricity Production from an Onshore V126-3.3MW Wind Plant. *Vestas*. https://www.vestas.com/content/dam/vestas-com/global/en/sustainability/reports-and-ratings/lcas/LCAV12633MWfinal060614.pdf.coredownload.inline.pdf

[33] Krishnan, S., Corona, B., Gert Jan Kramer, Junginger, M., & Koning, V. (2024). Prospective LCA of alkaline and PEM electrolyser systems. *International Journal of Hydrogen Energy*, *55*, 26–41. https://doi.org/10.1016/j.ijhydene.2023.10.192

[34] Häfele, S., Hauck, M., & Dailly, J. (2016). Life cycle assessment of the manufacture and operation of solid oxide electrolyser components and stacks. *International Journal of Hydrogen Energy*, *41*(31), 13786–13796. https://doi.org/10.1016/j.ijhydene.2016.05.069

[35] Stolzenburg, & Mubbala, R. (2013). *Integrated Design for Demonstration of Efficient Liquefaction of Hydrogen (IDEALHY) Fuel Cells and Hydrogen Joint Undertaking (FCH JU) Grant Agreement Number 278177 Title: Hydrogen Liquefaction Report*. https://www.idealhy.eu/uploads/documents/IDEALHY_D3-16_Liquefaction_Report_web.pdf

[36] Alkhaledi, A. N., Sampath, S., & Pilidis, P. (2021). A hydrogen fuelled LH2 tanker ship design. *Ships and Offshore Structures*, 1–10. https://doi.org/10.1080/17445302.2021.1935626

[37] Cooper, J., Dubey, L., Bakkaloglu, S., & Hawkes, A. (2022). Hydrogen emissions from the hydrogen value chain-emissions profile and impact to global warming. *Science of the Total Environment*, *830*, 154624. https://doi.org/10.1016/j.scitotenv.2022.154624

[38] Erbach, G., & Svensson, S. (2021). *BRIEFING Towards climate neutrality*. https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/747085/EPRS_BRI(2023)747085_EN.pdf

[39] Kolb, S., Müller, J., Luna-Jaspe, N., & Karl, J. (2022). Renewable hydrogen imports for the German energy transition – A comparative life cycle assessment. *Journal of Cleaner Production*, *373*, 133289. https://doi.org/10.1016/j.jclepro.2022.133289

[40] Erbach, G., & Jensen, L. (2021). *BRIEFING Towards Climate Neutrality*. https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/689332/EPRS_BRI(2021)689332_EN.pdf

[41] Siemens (2015). *LNG and Coal Life Cycle Assessment of Greenhouse Gas Emissions Prepared for: Center for Liquefied Natural Gas*. https://www.ourenergypolicy.org/wp-content/uploads/2015/10/PACE_Report.pdf

[42] Nizami, M., Nugroho, R. I., Milati, K. H., Pratama, Y. W. (2022). Process and levelized cost assessment of high CO2-content natural gas for LNG production using membrane and CFZ CO2 separation integrated with CO2 sequestration. *Sustainable Energy Technologies and Assessments*, *49*, 101744. https://doi.org/10.1016/j.seta.2021.101744

[43] McKinsey & Company. (2024). *Global Energy Perspective 2023: Hydrogen outlook* https://www.mckinsey.com/industries/oil-and-gas/our-insights/global-energy-perspective-2023-hydrogen-outlook

An experimentally validated numerical model of a hybrid spectral splitting concentrating photovoltaic-thermal collector

Oluwakemi Akinnola and Kwak Minsun
Department of Chemical Engineering, Imperial College London, U.K.

Abstract: Much research has been done on developing hybrid photovoltaic-thermal collectors, as they become increasingly popular. However, there exists a large gap between experimental and simulated work in existing literature, with simulation-based work being more abundant. This study presents a model of a novel, hybrid spectral splitting concentrating photovoltaic thermal collector, validated through experimental data from a real-world collector. Ray optics, heat transfer and pipe flow modules in COMSOL Multiphysics were used to predict the behaviour of a real-life experimental setup created by TEC Mexico. The experimental setup consists of a parabolic trough concentrator that focuses radiation onto photovoltaic cells, attached to adjacent sides of a triangular receiver. Spectral splitting is achieved using a solid filter. Key findings demonstrate that the model achieves a thermal efficiency of 23% and an electrical efficiency of 5% at optimal conditions in the top channel. In the bottom channel, a thermal efficiency of 14% was achieved. Error analysis revealed a deviation of 1.6% in both efficiencies between experimental and simulated data in the top channel and 4% in the bottom channel, showing the model's reliability and accuracy in predicting the behaviour of the real-life experimental setup. This study bridges the gap between theory and practice through experimental data and reinforces the potential of SS CPV-T collectors for achieving higher solar energy utilisation efficiencies.

Keywords: CPV-T collector, Spectral splitting, Experimental Validation, Turbulent flow

1. Introduction

The sun is a renewable and clean source of energy that offers heat and light, which can be converted to electrical energy. This can be done using a photovoltaic (PV) cell: a nonmechanical device that converts sunlight directly into electricity [1]. Sunlight strikes the cell, which is made of semiconductors and the band gap of the semiconductor determines which wavelengths of electromagnetic radiation can be absorbed. If it matches the incident radiation, it absorbs part or all of this radiation, exciting electrons which flow, producing electricity [2].

2. Background

In a photovoltaic cell, only part of the radiation absorbed is converted to electricity. The rest heats the cell, reducing its efficiency. To combat this, in the early 1970s, scientists began modifying PV technology to harness this thermal energy and direct it where required [3]. Thermal absorbers were implemented to convert this waste heat into useful energy, resulting in a hybrid photovoltaic-thermal (PV-T) collector. The harvested heat however is limited by the operating temperature of the PV cells. It fails to exceed 50 ˚C, rendering this technology insufficient for industrial applications, which usually require temperatures well above 100 ˚C [4]. On their own, solar cells cannot convert more than 24% of incident solar energy into electricity [5]. However, using concentrators, an equivalent amount of radiation is harnessed with fewer receivers, resulting in reduced production costs. Several concentrators exist such as parabolic troughs, Linear Fresnel, compound parabolic and flat plate concentrators [6]. Parabolic troughs are the most developed, cheapest, and widely used concentrator for large-scale solar energy applications [7].

These concentrating technologies can exceed 100 ˚C, making them suitable for a range of industrial processes. Several types of receivers exist as well. Though a great advantage, the high temperatures offered by concentrators cause a drop in electrical efficiency of these collectors. This makes them unable to convert more than 46% of incident radiation into electrical energy [8]. Spectral splitting (SS) involves using solid or liquid optical filters to only allow radiation that matches the band gap of the semiconductor material to pass through to the PV cell. Splitting the solar spectrum allows useful radiation to be harvested without infrared radiation which heats the cells and reduces their efficiency. SS can be achieved through liquid (selective absorbing filters) and

solid filters (thin films, dichroic, and interference filters for instance) [8].

Low et al. [9] used interference filters to develop a method for optimizing spectrum allocation to maximize collector efficiency in a Sterling engine based CPV-T collector. Combined collector efficiencies as high as 36% were achieved through this method. Qui et al. [8] designed a hybrid wave-selecting filter-coated compound parabolic concentrator with a linear Fresnel reflector (LFR) to enhance solar energy utilization. They found that larger focal lengths enhanced optical and exergy efficiencies of 95% and 36% respectively.

Ma et al. [10] developed a nanofluid spectral splitting CPV-T collector which contained a triangular cooling receiver, whose overall efficiency was 7.9% higher than flat plate receivers. Mahmoud et al. [11] separated concentrated solar radiation into distinct spectral bands (colours) to optimize PV and thermal energy generation. A dish reflector with a two-axis solar tracking collector and an optical linear splitting device were used. The collector achieved a transmissivity of 76% in the 400–800 nm range and the yellow band, at 15 $mWcm^{-2}$ at noon, was most effective for power generation during most of the day. Arnaoutakis, G. and Katsaprakakis, D. [12] investigated the annual energy performance of a CSP power plant with a parabolic trough collector in Southern Crete, Greece. They found that 80% reflectivity of the splitter produced up to 89 MWh/month compared to a 90% reflectivity splitter which only yielded 64.116 GWh.

Dong et al. [13] explored the use of a novel trapezoidal fluid channel design to improve the cooling efficiency and overall energy output of PV-T collectors. The trapezoidal fluid channels reduced the PV cell temperature by an average of 13°C compared to a standard PV module. Maatallah et al. [14] employed a parabolic dish concentrator with a conical cavity receiver and a beam splitter filter to enhance both electrical and thermal energy outputs of CPV-T collectors. Four different pipe cross-sections: rectangular, semi-circular, semi-elliptical, and triangular, were investigated and the triangular design proved to be the most effective in heat dissipation, at 446 W and thermal power output at 12 kW.

Yang et al. [15] used a two-axis solar tracker to keep the PV-T collector aligned with the sun. A parabolic trough concentrator and a solid filter consisting of multiple alternating layers of high and low refractive index thin films were used in this study. Electrical and thermal conversion efficiencies of 16% and 21% were achieved respectively. Li et al. [16] proposed a non-confocal two-stage parabolic dish concentrating photovoltaic/thermal hybrid collector using spectral beam splitting. It achieved an optical efficiency of 81% and improved energy and exergy efficiencies by 18% and 10%, respectively. The collector demonstrates stability under variable conditions, optimizing solar energy use for electricity and heat. Still by Li et al. [17] in a separate study, the use of a Cassegrain concentrator was investigated and an optical efficiency of 79% was achieved.

Kumar et al. [18] improved PV-T efficiency using a Fresnel lens, nanofluid-based liquid spectrum filter, and dual-axis solar tracker. Results show a 29% increase in electrical and 7% in thermal efficiency, optimizing solar energy utilization. Saadelnour et al. [19] studied the influence of liquid absorption filter properties on the performance of coupled and decoupled hybrid CPV-T collectors. Thermal efficiencies of 50% and 42% were achieved respectively and research showed that higher specific heat capacity enhanced thermal efficiency. Wang et al. [20] designed and experimentally studied a fluid filter-based compound parabolic concentrator collector under actual outdoor environments. Results showed that silver nanoparticles and cobalt sulphate improve the spectral absorptivity of deionized water and yield a thermal efficiency of up to 56%. Variava et al. [21] evaluated the performance of four CPV-T collectors — Rectangular, Trapezoidal, Hexagonal Honeycomb, and Chevron Pattern — under different conditions to assess their efficiency. Rectangular channels showed the greatest efficiency at 49%.

Overall, most work done in literature is largely based on modelling or experimental work. The novelty of this study is characterised by the presence of a real SBS CPV-T collector. Experiments were conducted and subsequent data was collated, which has been used to validate the model proposed in this paper. The objectives in this study were to:

1. Analyse the influence of mass flowrate and inlet temperature of the collector's performance

2. Develop a characteristic curve of the collector
3. Compare the simulated results with experimental data obtained from TEC Mexico.

3. Methodology

The collector setup is discussed in this section, as well as the fundamental equations defining derived parameters. Optical, electrical, heat transfer and fluid mechanics models are also discussed in detail.

3.1. Collector configuration

The hybrid CPV-T collector examined in this study is illustrated in Figure 1.1. It uses a commercially available parabolic trough concentrator, produced by Inventive Power (a Mexican company) and consists of several key components discussed in the subsections below.

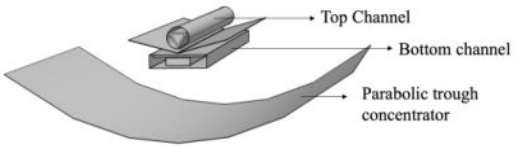


Figure 1.1: CPV-T collector setup

3.1.1 Optical concentrator

The optical concentrator, with its parabolic trough structure, focuses incoming light onto the primary receiver, resulting in a higher concentration of radiation. Its dimensions are summarised in Table 1.

3.1.2 Primary receiver

The primary receiver, also referred to as the top channel, is an equilateral triangle channel made of aluminium, which converts concentrated solar radiation into electrical energy. It is enclosed within a cylindrical glass envelope that traps long-wavelength radiation and reduces convective heat losses. The PV cells are mounted on adjacent sides of the triangular cooling channel. Water, the heat transfer fluid (HTF), flows through the channel and cools the cells. This enhances overall performance by reducing efficiency losses due to heating. Dimensions of the top channel are summarised in Table 2.

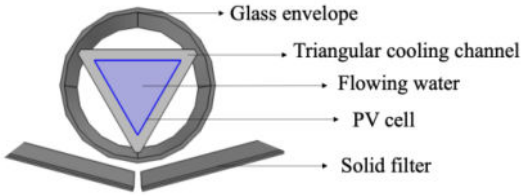


Figure 1.2 Schematic diagram of the top channel

3.1.3. Solid splitting filter

Spectral splitting is achieved using a solid filter, sandwiched between two rectangular Silica glass sheets beneath the primary receiver. It selectively transmits specific wavelengths of solar radiation while reflecting the rest, ultraviolet and mostly infrared, onto the secondary receiver. The filter used in this study is 3M^{TM} Chill. The wavelength-dependent transmittance of this filter was imputed into COMSOL and used to model spectral splitting, simulating the presence of the filter. The key dimensions of the solid splitting filter are summarised in Table 3.

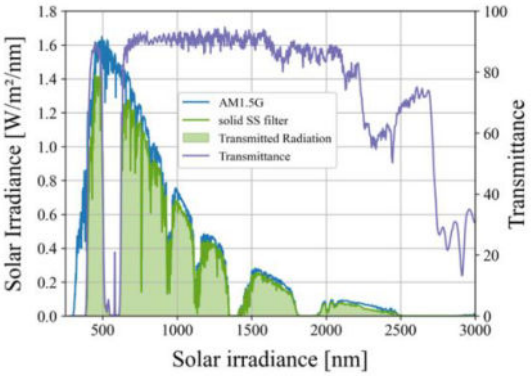


Figure 1.3: AM1.5G spectrum and transmittance profile of the solid SS filter

3.1.4 Secondary receiver

The secondary receiver, also referred to as the bottom channel, captures radiation reflected by the filter and transfers the resulting thermal energy to the HTF. It consists of a flat, rectangular aluminium sheet and a rectangular cooling channel, connected by two strips of thermal grease as shown in Figure 1.4. The aluminium sheet acts as an absorber, capturing radiation from the primary receiver and channelling it effectively to the HTF. The thermal grease enhances conductivity and ensures efficient thermal energy transfer. These are contained in a rectangular glass envelope.

The remaining space between the cover glass and the cooling channel contains trapped air, which facilitates convective heat loss, further optimizing the performance of the bottom channel. Like the primary receiver, water serves as the HTF, circulating through the channel constructed from thermally conductive aluminium to absorb heat. Its dimensions are detailed in Table 4.

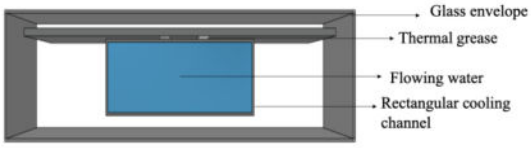


Figure 1.4 Schematic diagram of the bottom channel

3.2. Numerical methods

3.2.1 Optical model

The optical model describes how solar radiation is directed and processed within the collector. The local flux distribution on the SS filter and the primary receiver channel is estimated using the geometrical optics sub module in COMSOL. The optical properties of the components are summarised in Table 5, while the overall solar spectrum received by the PV cells is defined mathematically by equation 2.1.

$$G_{\mathrm{PV}}(\lambda) = G_{\mathrm{AM1.5}}(\lambda)\,\rho_{\mathrm{c}}\,\tau_{\mathrm{g}}\tau_{\mathrm{fl}}(\lambda) \qquad (2.1)$$

Where $G_{\mathrm{AM1.5}}(\lambda)$ and $G_{\mathrm{PV}}(\lambda)$ are the incident solar radiation spectrum and the solar spectrum received by the PV cells respectively. ρ_{c} is the reflectivity of the mirror, τ_{g} the transmittivity of the glass, and $\tau_{\mathrm{fl}}(\lambda)$ represents the transmissivity of the filter as a function of wavelength. The AM1.5G spectrum and the transmittance profile of the solid SS filter are shown in figure 1.3. The optical properties of all components are displayed in table 5.

1.1 Electrical model

The electrical efficiency (η_{el}) quantifies the ability of PV cells to convert incident solar irradiation into electrical energy. It is defined by equation 2.2 below:

$$\eta_{\mathrm{el}} = \eta_{\mathrm{el},0}[1 - \beta(T_{PV} - T_{\mathrm{std}})] \qquad (2.2)$$

where $\eta_{\mathrm{el},0}$ is the efficiency of the PV cells. β represents the temperature coefficient and $(T_{\mathrm{PV}} - T_{\mathrm{std}})$ represents the temperature difference between the PV cell operating temperature (T_{PV}) and the reference temperature (T_{std}).

3.2.2 Heat transfer and fluid dynamics model

Heat transfer and fluid dynamics in the collector are governed by the following assumptions to enable efficient modelling and analysis:

i. Steady state operation is assumed $(\frac{\partial T}{\partial t} = 0)$.
ii. The flow is assumed to be incompressible $(\frac{\partial \rho}{\partial x} = \frac{\partial \rho}{\partial y} = \frac{\partial \rho}{\partial z} = 0)$.
iii. Fully developed flow is assumed.
iv. Gravitational effects on the fluid and heat transfer are neglected.
v. Thermal expansion and thermoelastic damping within the solid components are also neglected.
vi. The material used in SS filter is modelled as a single-phase medium (solid), disregarding any phase-change effects.
vii. Thermal resistance between the PV solar cells and the metal cooling channel of the primary receiver is ignored, assuming ideal thermal coupling.

The steady-state heat conduction equation is applied to all solid domains, including the glass envelopes and aluminium cooling channels, to model the temperature distribution. The equation is expressed as:

$$\nabla \cdot (\boldsymbol{q} + \boldsymbol{q}_{\mathbf{r}}) = k\left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2}\right) + \nabla \cdot \boldsymbol{q}_{\mathbf{r}} = 0 \qquad (2.3)$$

Where $\boldsymbol{q}$ denotes the heat flux vector due to conduction, $\boldsymbol{q}_{\mathbf{r}}$ represents the radiative heat flux vector, applicable in materials like silica glass that allow radiation transfer, and k accounts for the thermal conductivity of the solid material.

To model heat transfer within the fluid, the COMSOL Multiphysics solves the steady-state energy equation. This equation is used to describe heat transfer in the HTF, absorption of solar energy by the solid SS filter, and

dissipation of energy due to viscous effects caused by turbulent flow. The governing equation is as follows:

$$\rho c_{\mathrm{p}} \boldsymbol{u} \cdot \nabla T + \nabla \cdot \boldsymbol{q} = Q_{\mathrm{p}} + Q_{\mathrm{vd}} \qquad (2.4)$$

Where $\rho c_{\mathrm{p}} \boldsymbol{u} \cdot \nabla T$ represents the convective heat transfer term, $\nabla \cdot \boldsymbol{q}$ represents the heat conduction term, Q_{p} refers to solar energy absorbed by the filter and Q_{vd} represents viscous dissipation, accounting for heat generated in turbulent flow.
The Reynolds number is a dimensionless parameter that characterises the flow regime in the triangular and rectangular cooling channels. It is calculated using the equation 2.5.

$$Re = \frac{\rho u D_{\mathrm{h}}}{\mu} \qquad (2.5)$$

Where μ is the dynamic viscosity of the fluid, ρ is the fluid density, u represents mean velocity of the fluid, and D_{h} is the hydraulic diameter. Hydraulic diameter is specific to the geometry of the pipe in which the fluid flows and is evaluated as follows:

$$D_{\mathrm{h}} = \frac{4A}{P} \qquad (2.6)$$

Where A is the cross-sectional of the pipe and P is the wetted perimeter, the boundary length in contact with the HTF. In this study, mass flowrates of the HTF were varied, giving rise to a range of Reynolds numbers. The turbulent regime was defined for Reynolds numbers greater than 1100 [23]. For mass flowrates corresponding to Reynold's numbers characteristic of the laminar regime, the standard laminar setting was used in COMSOL. The k-ε turbulence model, frequently used to predict fluid flow behaviour was chosen to simulate turbulent flow in the cooling channels.

The solar irradiance absorbed by the optical filter and PV solar cells is calculated under the boundary heat source terms as follows:

$$q_{\mathrm{PV}} = \int_{280}^{4000} G_{\mathrm{PV}}(\lambda)\, \mathrm{d}\lambda \qquad (2.7)$$

Where $G_{\mathrm{PV}}(\lambda)$ is the spectral irradiance reaching the PV cells, and q_{PV} is the total irradiance absorbed. These equations represent the sum of the contributions of irradiance across all wavelengths within the given range (280 nm – 4000 nm), quantifying the total energy absorbed by the PV cells.
The heat generated from the top channel is identical to the thermal energy dissipated from the PV cells and is calculated as thus:

$$q_{\mathrm{g}} = \alpha_{\mathrm{pv}}\, q_{\mathrm{pv}} - \eta_{\mathrm{el}}\, C_{\mathrm{geo}}\, G_{\mathrm{AM1.5}} \qquad (2.8)$$

Where α_{pv} is the PV cell absorptivity and η_{el} is the PV cell electrical efficiency. C_{geo} is the geometrical concentration ratio, defined as the ratio of the concentrator aperture area to the receiver area, and $G_{\mathrm{AM1.5}}$ represents the incident solar irradiance on the collector aperture.

The heat generated from the bottom channel is estimated using equation 2.8:

$$q_{\mathrm{g}} = \alpha_{\mathrm{abs}}\, \tau_{\mathrm{g}}\, \rho_{\mathrm{filter}} G_{\mathrm{AM1.5}} \qquad (2.9)$$

Where α_{abs} is the absorptivity of the bottom channel, τ_{g} is the transmissivity of the glass coverage, and ρ_{filter} is the reflectivity of the filter.
In the collector, convective and radiative heat losses occur. Convective heat loss to the ambient from the outer surface of the glass envelope is given by:

$$q_{\mathrm{conv}} = h_{\mathrm{w}}(T - T_{\mathrm{amb}}) \qquad (2.10)$$

Where h_{w} is the convective heat transfer coefficient, T is the surface temperature of the glass envelope, and T_{amb} is the ambient temperature.
Convective heat transfer coefficient of the primary and secondary receivers is calculated using equations 2.10 and 2.11 below:

$$h_{\mathrm{w}} = 4V_{\mathrm{wind}}^{0.58} D_{\mathrm{o.g}}^{-0.42} \qquad (2.11)$$

$$h_{\mathrm{w}} = 2.8 + 3V_{\mathrm{wind}} \qquad (2.12)$$

Where V_{wind} is the wind speed and $D_{\mathrm{o.g}}$ is the outer diameter of the primary receiver.
Radiative heat losses are modelled using a surface-to-surface radiation interface in COMSOL. To estimate the effective sky temperature, equation 2.12 was used:

$$T_{\mathrm{sky}} = 0.0552\, T_{\mathrm{amb}}^{1.5} \qquad (2.13)$$

Thermal efficiency for the primary and secondary cooling channels are estimated according to equations 2.14 below:

$$\eta_{\text{th}} = \frac{\dot{m}_{\text{HTF}}\, c_{\text{p}}\, (T_{\text{in}} - T_{\text{out}})}{G_{\text{AM1.5}}\, A_{\text{c}}} \qquad (2.14)$$

where $\dot{m}_{\text{HTF}}$ represents the mass flow rate of HTF, c_{p} is the specific heat capacity of HTF, T_{in} and T_{out} are the HTF inlet and outlet temperatures, and A_{c} is the aperture area of the concentrator.

The reduced temperature represents the temperature difference between the HTF and ambient conditions, normalised by incident solar radiation. It is an important metric that indicates a collector's heat retention ability and is calculated as follows:

$$T_{\text{r}} = \frac{\frac{T_{\text{in}} + T_{\text{out}}}{2} - T_{\text{amb}}}{G_{\text{AM1.5}}} \qquad (2.16)$$

Where T_{in} and T_{out} are the cooling channel inlet and outlet temperature, respectively.

4. Results and Discussion

The goal of this project is to develop a model that predicts the behaviour of a concentrating SBS CPV-T collector. This was done using COMSOL and the results were validated using experimental values obtained during operation of the real setup by TEC Mexico. Data analysis was performed with MATLAB. Values of model parameters are listed in the table below:

Parameter	Value	Unit
Irradiance	1000	Wm^{-2}
$T_{ambient}$	298	K
Concentration Ratio	10.2	
Wind Speed	1	ms^{-1}
Electrical efficiency	0.06	
T_{sky}	284	K

4.1. Effect of varied mass flowrate

First, equivalent Reynolds numbers for a range of flowrates, varying from 0.005 kgs^{-1} to 0.2 kg s^{-1}, were estimated. A consequent rise in thermal and electrical efficiency is observed in both channels. In the top channel however, beyond a flowrate of 0.05 kgs^{-1}, neither thermal nor electrical efficiency increases significantly. This is because between 0.005 kgs^{-1} and 0.05 kgs^{-1}, the flow regime is laminar, and convective heat transfer is poor. Above a mass flowrate of 0.05 kgs^{-1}, turbulent profile is established, enhancing convective heat transfer. Hence, higher mass flowrates produce minimal changes in results. This trend aligns with a study by Variava et al., in which the k-ε turbulence model was used to show that turbulent flow enhances heat transfer in collectors. The same trend is observed in the bottom channel, though its turbulent regime was realised at a flowrate of 0.1 kg s^{-1}. These results are represented graphically in Figures 4.1 and 4.2 below.

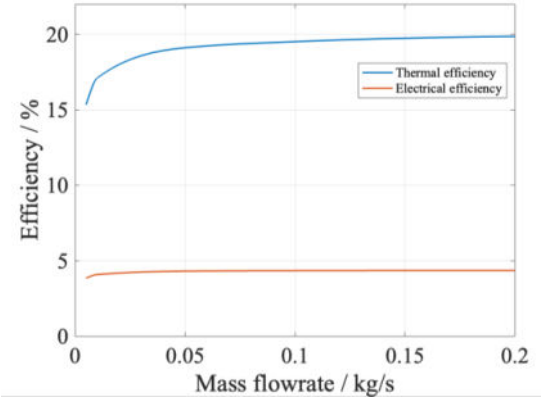


Figure 4.1. Effect of varying mass flowrate on top channel thermal and electrical efficiencies. As mass flowrate increases efficiency increases.

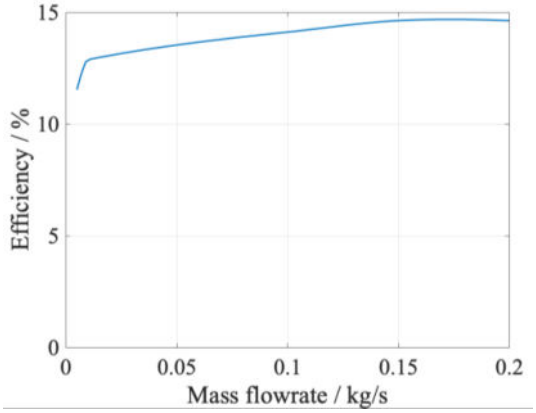


Figure 4.2. Effect of varied mass flowrate on bottom channel thermal efficiency. A rise in mass flowrate yields greater thermal efficiency.

4.2. Effect of varied inlet water temperature

The influence of inlet temperature on thermal and electrical efficiency was then analysed for flowrates that marked the beginning of the

turbulent regimes in both channels. The simulations were run at this value because greater values of mass flowrate did not yield significantly greater thermal or electrical efficiencies. Inlet temperature was varied and the results, displayed in Figures 4.3 and 4.4 show an indirect relationship between inlet temperature and efficiency. This is expected because an increase in inlet temperature raises the average temperature of the channel, effectively heating it. Hence, the collector loses heat to its surroundings, reducing thermal efficiency and the PV cells degrade, impacting electrical efficiency.

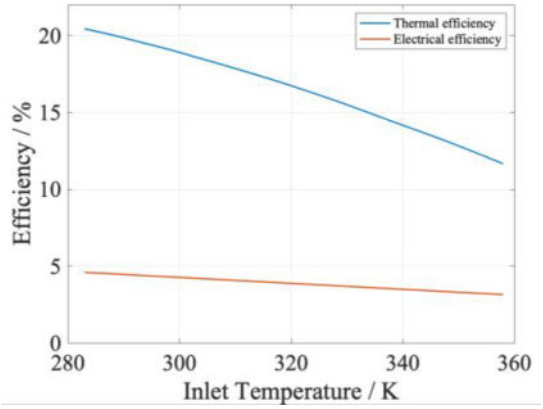


Figure 4.3. Effect of varied inlet temperature on top channel efficiencies. A rise in inlet temperature causes reductions in thermal and electrical efficiency.

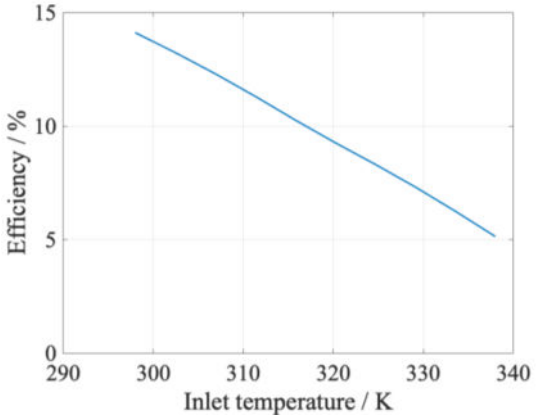


Figure 4.4. Effect of varied inlet temperature on bottom channel thermal efficiency. A rise in inlet temperature causes reductions in thermal and electrical efficiency.

4.3. Model validation

To validate the model, simulated values were combined with experimental values and plotted against reduced temperature. The resulting graphs are shown in Figures 4.5 and 4.6. Figure 4.5 represents the characteristic curve of the collector, a graphical representation of its behaviour. As inlet temperature was previously increased, a reduction in reduced temperature. Larger values of reduced temperature indicate greater convective and radiative thermal losses from the collector to its surroundings. Hence, the reduction in thermal efficiency. Similar results were obtained in a study by Huang et al.. Thermal and electrical efficiency both decreased for increasing reduced temperature. These effects on electrical efficiency were less pronounced due to decoupling of the thermal and electrical outputs [25].

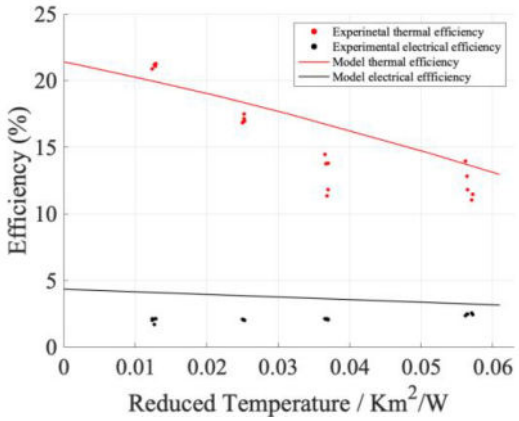


Figure 4.5. Validation of top channel model using experimental data. As reduced temperature increases, thermal and electrical efficiency both decrease.

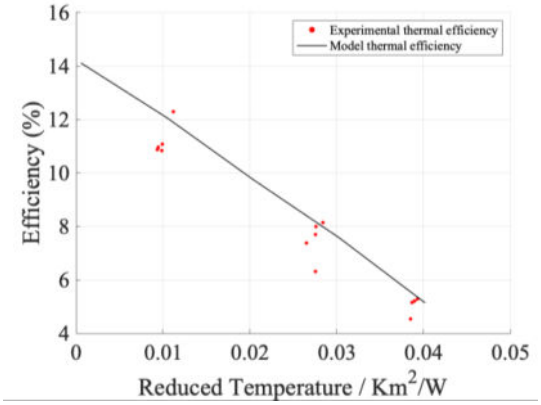


Figure 4.6. Bottom channel model validation with experimental data.

4.4. Error analysis and mesh independence study

The deviations between experimental and similar data were estimated by calculating their root mean square values. For the top channel electrical and thermal efficiencies saw a 1.6%

deviation between experiments and simulated data. In the bottom channel, this value was 4%. The accuracy of COMSOL was analysed through a mesh independence study: an investigation of the effect of number of elements on the accuracy of derived values. The coarse mesh contains the least number of elements and runs a simulation in 26 minutes. However, the finer mesh, with 10 times more elements, takes over two hours to run. Increasing the size of the mesh resulted in less than 0.2% deviation between obtained values. Hence in this study, the fine mesh was used, striking a balance between accuracy and computational time.

Mesh	No. of elements	T_{out}	Time / mins	Deviation / %
Coarse	21321	306.97	26	-
Normal	38361	306.7	43	0.08
Fine	80104	306.58	71	0.13
Finer	245197	306.42	125	0.18

Table R.1. Results of mesh independence study. As mesh density increases, computational time increases, offering less than 0.2% deviation in outlet temperature results.

5. Conclusion

In this study, a comprehensive optical, thermal and electrical model which predicts the behaviour of a real SBS CPV-T collector was developed. Mass flowrate and inlet temperature of the HTF were key parameters varied to develop the model. Via simulations performed using COMSOL Multiphysics, maximum thermal efficiencies of 23% and 14% in the top and the bottom channels were obtained respectively. A maximum electrical efficiency of 5% was achieved. An error analysis revealed that the simulated results and experimental data showed a deviation of 1.6% for both efficiencies in the top channel and 4% in the bottom channel. Hence, the model was successfully validated, and the characteristic curve of the collector was created. The alignment between simulation results and experimental data highlights the accuracy and reliability of the proposed model.

6. Outlook

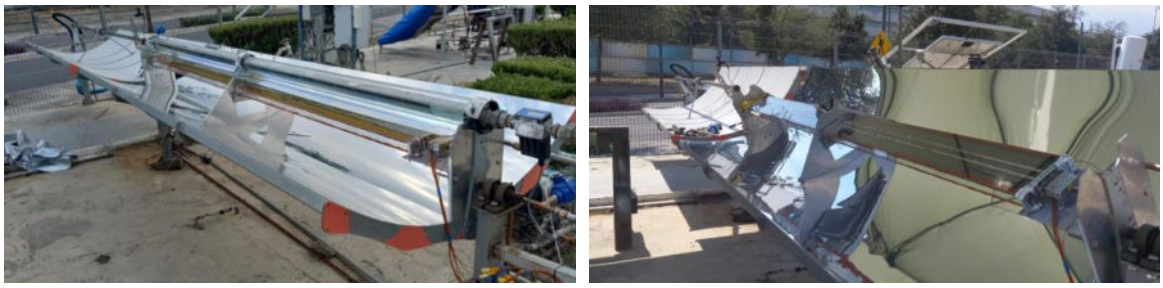
There are several ways to increase thermal and electrical efficiencies achieved in this study. One of these is exploring the use of different filters. The 3M chill filter transmits wavelengths in the 500-700 nm range. However, it can be observed in Figure 1.3 that it also transmits majority of the infrared spectrum, meaning incident radiation also heats the PV cells. A filter that only transits the full visible spectrum while completely reflecting all IR radiation would be ideal. Much shading on the experimental setup is seen in Figure A.2. Reducing this would boost electrical efficiency. In the bottom channel, thermal contact can also be improved. Two strips of thermal grease were used to connect the absorber to the cooling channel, not allowing for good thermal contact. Enhanced thermal contact allows the heat transfer fluid to absorb more heat, yielding greater thermal efficiencies.

Acknowledgements

We recognise and appreciate TEC Mexico for providing experimental data and all information regarding test rig setup. We also especially thank Dr Chandan Pandey and Dr Sandesh Chougule for their guidance and unwavering support throughout this project.

Appendices

Figures A.1 and A.2 Experimental setup of collector in Mexico



References

1. U.S. Energy Information Administration (2023). *Photovoltaics and Electricity - U.S. Energy Information Administration (EIA)*. [online] Eia.gov. Available at: https://www.eia.gov/energyexplained/solar/photovoltaics-and-electricity.php [Accessed 9 Nov. 2024].

2. Solar Energy Technologies Office (2022). *Solar Photovoltaic Cell Basics*. [online] Energy.gov. Available at: https://www.energy.gov/eere/solar/solar-photovoltaic-cell-basics [Accessed 11 Nov. 2024].

3. Xia, X., Wei, W., Yu, B., Li, N. and Ji, J. (2023). Experiment and numerical investigation on a spectral splitting PV/T system for electrical energy and thermal output. *Energy*, [online] 288, p.129911. doi:https://doi.org/10.1016/j.energy.2023.129911

4. An, W., Li, J., Ni, J., Taylor, R.A. and Zhu, T. (2017). Analysis of a temperature dependent optical window for nanofluid-based spectral splitting in PV/T power generation applications. *Energy Conversion and Management*, 151, pp.23–31. doi:https://doi.org/10.1016/j.enconman.2017.08.080.

5. Khelifa, A., Abdelgaied, M., Attia, M.E.H., Harby, K. and Arıcı, M. (2024). Development and optimization of dual-objective PVT modules for a smart residential building using inclined fins. *Journal of Building Engineering*, [online] 97, p.110893. doi:https://doi.org/10.1016/j.jobe.2024.110893

6. F. Najafi Roudbari, Ehsani, H., Amiri, S.R., A. Samadani, Shabani, S. and A. Khodadad (2024). Advances in photovoltaic thermal systems: A comprehensive review of CPVT and PVT technologies. *Solar Energy Materials and Solar Cells*, 276, pp.113070–113070. doi:https://doi.org/10.1016/j.solmat.2024.113070

7. Ahmad, A., Prakash, O., Kausher, R., Kumar, G., Pandey, S. and Hasnain, S.M.M. (2024). Parabolic trough solar collectors: A sustainable and efficient energy source. *Materials Science for Energy Technologies*, [online] 7, pp.99–106. doi:https://doi.org/10.1016/j.mset.2023.08.002

8. Qiu, H., Liu, H., Xia, Q., Lin, Z. and Chen, C. (2024). A spectral splitting CPV/T hybrid system based on wave-selecting filter coated compound parabolic concentrator and linear Fresnel reflector concentrator. *Renewable Energy*, 226, p.120403. doi:https://doi.org/10.1016/j.renene.2024.120403

9. Low, C.F.Y., Chan, P.Y. and Tan, W.S. (2024). Mathematical model to optimize solar spectrum allocation for a Stirling engine-based concentrated photovoltaic thermal system. *Renewable Energy*, [online] 223, p.120067. doi:https://doi.org/10.1016/j.renene.2024.120067.

10. Ma, Y., Han, X., Chen, Z. and Khosa, A.A. (2023). Performance analysis of a nanofluid spectral splitting concentrating PV/T system with triangular receiver based on MCRT-FVM coupled method. *Applied Thermal Engineering*, [online] 239, p.122096. doi:https://doi.org/10.1016/j.applthermaleng.2023.122096

11. Ben Amara, M., Rdhaounia, E. and Balghouthi, M. (2023). Experimental study of a concentrating solar spectrum splitting system for hybrid solar energy conversion. *Renewable Energy*, [online] 221, p.119745. doi:https://doi.org/10.1016/j.renene.2023.119745

12. Arnaoutakis, G.E. and Katsaprakakis, D.A. (2024). Energy Yield of Spectral Splitting Concentrated Solar Power Photovoltaic Systems. *Energies*, 17(3), pp.556–556. doi:https://doi.org/10.3390/en17030556

13. Dong, S., Long, H., Guan, J., Jiang, L., Zhuang, C., Gao, Y. and Di, Y. (2023). Performance investigation of a hybrid PV/T collector with a novel trapezoidal fluid channel. *Energy*, [online] 288, p.129594. doi:https://doi.org/10.1016/j.energy.2023.129594

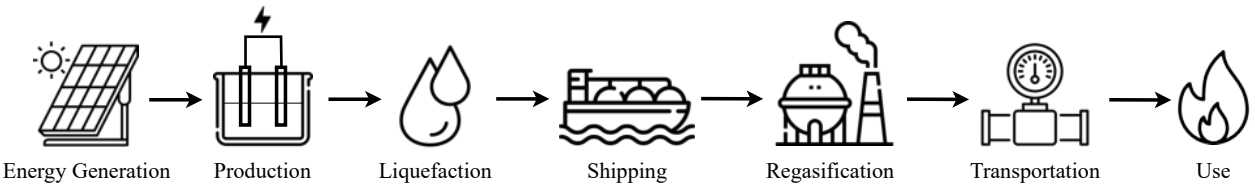
14. Taher Maatallah, Ahlem Houcine, Saeed, F., Khan, S. and Ali, S. (2024). Simulated Performance Analysis of a Hybrid Water-Cooled Photovoltaic/Parabolic Dish Concentrator Coupled with Conical Cavity Receiver. *Sustainability*, [online] 16(2),pp.544–544. doi:https://doi.org/10.3390/su16020544
15. Yang, N., Liu, Q., Wang, B., Xuan, J. and Xing, L. (2024). A novel PV/T system integrated with spectral beam filter: A comprehensive optical and thermodynamic study. *Applied Thermal Engineering*, [online] 257, p.124457. doi:https://doi.org/10.1016/j.applthermaleng.2024.124457
16. Li, J., Yang, Z., Wang, Y., Dong, Q., Qi, S., Huang, C., Wang, X. and Lin, R. (2023). A novel non-confocal two-stage dish concentrating photovoltaic/thermal hybrid system utilizing spectral beam splitting technology: Optical and thermal performance investigations. *Renewable Energy*, 206, pp.609–622. doi:https://doi.org/10.1016/j.renene.2023.02.078
17. Li, J., Yang, Z., Ge, Y., Wang, Y., Dong, Q., Wang, X. and Lin, R. (2024). Performance study of photovoltaic-thermochemical hybrid system with Cassegrain concentrator and spectral splitting integration. *Energy*, [online] 292, p.130611. doi:https://doi.org/10.1016/j.energy.2024.130611
18. Kumar, S., Thakur, R., Kumar, S., Lee, D. and Kumar, R. (2024). Impact of liquid spectrum filter and solar tracker on the overall effectiveness of a photovoltaic thermal system: An experimental investigation. *Renewable Energy*, 226, pp.120390–120390. doi:https://doi.org/10.1016/j.renene.2024.120390.
19. Abdueljabbar Adam, S., Ju, X., Lin, J., Xu, C., I. M. Abdelmagid, T., Ahmed, I.A.A. and Abd El-Samie, M.M. (2024). Influence of material properties of liquid absorption filters for concentrated photovoltaic/thermal hybrid systems. *Applied Thermal Engineering*, [online] 252, p.123639. doi:https://doi.org/10.1016/j.applthermaleng.2024.123639.
20. Wang, G., Chen, X., Lin, J., Xu, C. and Ju, X. (2023). Experimental study on a concentrating solar photovoltaic/thermal system using different fluid spectral beam filters. *Journal of Cleaner Production*, [online] 429, p.139549. doi:https://doi.org/10.1016/j.jclepro.2023.139549.
21. Tu, L., Ma, Y., Han, X. and Zhu, M. (2023). Numerical analysis of a modified combined volumetric filtered concentrating PV/T system using MCRT and FVM coupled method. *Journal of Cleaner Production*, [online] 410, p.137296. doi:https://doi.org/10.1016/j.jclepro.2023.137296
22. Variava, J.M., Ratnadhariya, J.K., Haque Siddiqui, M.I. and Sadasivuni, K.K. (2024). 3D numerical model of a concentrated photovoltaic thermal (CPV/T) system for thermal and electrical performance optimization. *Case Studies in Thermal Engineering*, [online] 61, p.104823. doi:https://doi.org/10.1016/j.csite.2024.104823
23. Richard C. Cope and Richard W. Hanks *Industrial & Engineering Chemistry Fundamentals* **1972** *11* (1), 106-117 DOI: 10.1021/i160041a016
24. National Institute of Standards and Technology (2018). *Thermophysical Properties of Fluid Systems*. [online] Nist.gov. Available at: https://webbook.nist.gov/chemistry/fluid/
25. Huang, G., Wang, K., Curt, S.R., Franchetti, B., Pesmazoglou, I. and Markides, C.N. (2021). On the performance of concentrating fluid-based spectral-splitting hybrid PV-thermal (PV-T) solar collectors. *Renewable Energy*, 174, pp.590–605. doi:https://doi.org/10.1016/j.renene.2021.04.070.

Design and Comparison of Dual-Use Terminals for Cold Energy Recovery during Liquid Hydrogen Regasification

Callum Wootton and Hephzibah Olaleye
Department of Chemical Engineering, Imperial College London, U.K.

Abstract

With hydrogen gaining attention as a potential key part of the energy transition, the sustainability of the hydrogen supply chain comes under critical focus. Liquid Hydrogen (LH_2) is a prospective hydrogen carrier, but requires significant energy during liquefaction. This study explores the potential to enhance the sustainability of the hydrogen supply chain by recovering and repurposing the cold energy from LH_2 regasification. This paper aims to determine a practical, economically viable use for this cold energy that also has a positive impact on the sustainability of the supply chain. Two dual-use terminals were investigated, where either CO_2 or N_2 is liquefied alongside the regasification of LH_2. The two terminals were modelled on Aspen Plus before completing an energy, economic and carbon analysis; comparing both terminals to each other, and a LH_2 regasification terminal without cold energy recovery. This analysis was completed across three design bases of terminal throughput. The designed LH_2 import / CO_2 export terminal reused 75.8% of cold energy and possessed a greater added Net Present Value (NPV) of up to $\$0.91/kg_{H_2}$. This added NPV, however, was highly dependant on the revenue generated from CO_2. The LH_2 import / N_2 export terminal reused only 21.1% of cold energy and possessed a lower added Net Present Value (NPV) of only $\$0.41/kg_{H_2}$. Both terminals add value compared typical a single use LH_2 regasification terminal, making them economically viable additions to the hydrogen supply chain.



1 Introduction

As greenhouse gas emissions targets become more prevalent, it is anticipated that the large volumes of crude oil and liquefied natural gas (LNG) that are currently traded will drastically decrease. High-quality renewable energy is not always available; therefore, the cost of renewable energy generation will differ between countries. This cost differential may be sufficient to stimulate energy trade between these nations, thus invoking a need for a vector to effectively store this energy in a transportable format. The hydrogen economy has been put forward to take this role.

A key challenge with liquid hydrogen is the significant energy penalty required to liquefy it at -250°C. Upon delivery, the liquid hydrogen is regasified, such that it can be injected into a pipeline. The regasification process shows potential to recover some of the cold energy from the liquid hydrogen, either utilising it to do work in a thermodynamic engine or to liquefy other materials that can be returned to the liquefaction process at the origin (thus reducing the energy penalty of the liquefaction).

Several authors have proposed processes that focus on the recovery (and repurposing) of cold energy. For example, research by Mun et al.[1] proposes combining Brayton and Rankine cycles in series and parallel configurations, and in doing so, cold energy is more efficiently recovered across wide temperature ranges during LH_2 regasification. Although this paper proposes three configurations and compares energy efficiency via energy analysis, it does not discuss how shipping modifies the use of the terminal.

It was found that by repurposing cold energy recovered in LH_2 regasification for power generation, the need for carbon-emitting power generating methods was diminished. The economic viability of the proposed processes was determined via assessing capital and operating costs as well as profits from electricity and carbon credits. The Net Present Value (NPV) of the proposed processes was also found and compared. The paper concludes that profits from selling electricity and carbon credits was proportional to the amount of power the process produces. The study also explores how profits from selling electricity and carbon credits vary and add to the NPV.

Another paper that researches the use of cold energy recovered from the regasification of LH_2, by Naquash et al.,[2] proposes a Brayton cycle using helium as its working fluid, to produce liquefied natural gas (LNG), natural gas liquids (NGL) and liquid Helium (LHe) from natural gas (NG). An energy analysis was conducted to compare the energy efficiency of the initial and optimised designs of the proposed process. This analysis found that by adjusting parameters such as refrigerant flowrates, compressor pressure ratios etc, additional energy savings could be made. An economic analysis was conducted that analysed individual process units to determine their impacts on overall costs; it was determined that high costs could be attributed to high energy consuming compressors and cryogenic heat exchangers which were increased in number to optimise the process. The impact of shipping shaping the use of the terminal was not discussed, instead this paper focuses on the terminal itself as an independent

entity rather than part of the overall hydrogen supply chain. An environmental analysis was not conducted therefore does not highlight how the terminal will impact the environment and the overall sustainability of the supply chain.

Finally, research by Kim et al[3] investigated the integration of the regasification of Liquid Hydrogen (LH_2) with Liquid Air Energy Storage (LAES). Within this process, cold energy could be recovered from the LH_2 and transferred to the compressed air and an organic Rankine cycle (ORC) that uses methane and propane as its working fluids. Three different configurations were proposed: a LH_2 regasification process and LAES system in conjunction with compressors (WC) (LH_2-LAES-WC), a LH_2 regasification process and LAES system without the use of compressors (WOC) (LH_2-LAES-WOC) and finally LH_2 regasification process and LAES system in conjunction with the ORC (LH_2-LAES-WORC). An energy analysis was performed, during which the specified energy consumption (SEC) was calculated to compare the energy efficiency of the three proposed processes and the LAES alone. It was found that compared to the LAES the terminals exhibited a decrease in SEC due to the utilization of the cold energy derived from liquid hydrogen, this improving energy efficiency. Following this analysis, an economic analysis was conducted where the NPV was calculated to determine the economic viability of each process. These NPVs are compared and the LH-LAES-WORC was found to have the highest NPV showing potential for positive returns on investment.

Many LH_2 regasification processes that reuse cold energy have been proposed and optimised. However, there is a lack of holistic comparison of economic viability between different terminals[1,2,3] . Additionally, the literature focuses on optimising energy recovery, but rarely provides context to the environmental benefit of this[2,].[3] Finally, the practicality of a LH_2 regasification terminal and how shipping modifies the use of the terminal is not considered[1,2,3] . This paper aims to address gaps in the literature around these points.

The goal of this project is to determine practical, economically viable uses for this cold energy, to design a flowsheet for a regasification terminal and to derive a capital cost for such a terminal. The proposed terminals discussed in this paper aim to be practical in the sense that they consider the whole hydrogen supply chain, and how shipping modifies the use of the terminal. The added NPV of each terminal will be investigated to determine their economic feasibility. Finally, the emissions of each terminal will be evaluated.

2 Terminal Descriptions

Six terminals were proposed in a pre-selection stage, and evaluated based on the following factors: process and economic viability, environmental impact, energy savings, and the simplicity of the process. The terminals that were considered included: Liquid Air Energy Storage; a dual-use terminal liquefying LNG, LPG, CO_2 or N_2; and a thermodynamic engine.

Both LAES[3] and Thermodynamic Cycles[1] have been explored in previous work, and thus were not chosen to consider further. A dual-use Liquefied Natural Gas (LNG) or Liquefied Petroleum Gas (LPG) terminal provides a potentially economically viable option but were not chosen as their use may not comply with future carbon emission targets. The LH_2/CO_2 process aims to liquefy CO_2, for the purpose of Carbon Capture and Storage (CCS) - a critical advantage in the context of current net-zero carbon goals. This facilitates a reduction in carbon tax or the sale of carbon credits, generating an additional revenue stream for the terminal. The LH_2/N_2 process presents the opportunity to ship liquefied N_2 back to the H_2 liquefaction facility to reduce the energy required to liquefy H_2 - saving in utility costs and improving circularity in the supply chain.

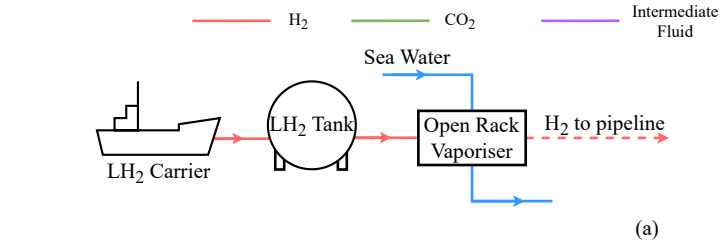
After this initial evaluation, **dual-use LH_2/CO_2** and **LH_2/N_2** terminals were chosen to investigate further. For the purposes of this study, the terminal is assumed to be situated in the port of Rotterdam, due to it's position as a major global port and part of the future European Hydrogen Backbone[4].[5] See Supplementary Information for detailed terminal Process Flow Diagrams.

2.1 LH_2 Regasification

LH_2 is shipped at 1 bar and -254°C. Storage vessels must be carefully designed to avoid boil-off, due to LH_2's extreme conditions. This is generally achieved with a stainless steel and aluminium tank, with multiple layers of insulation[6] and refrigeration.[7] For stationary tanks, a spherical tank is the most practical design due to its low surface area to volume ratio. However, tanks on the LH_2 carrier will have a type-C tank[8] as its aspect ratio is better suited to the geometry of a ship. Once delivered to the port, it is unloaded via unloading arms into a storage vessel. Similar to an LNG port, a boil-off gas header is present to recondense any vaporised LH_2. LH_2 passes through a high pressure pump, and enters an open-rack vaporiser using sea water. It then enters the H_2 pipeline tie-in at 70 bar and 10°C. This is the **base case** - where there is no attempt at cold energy recovery and is depicted in Figure 1a.

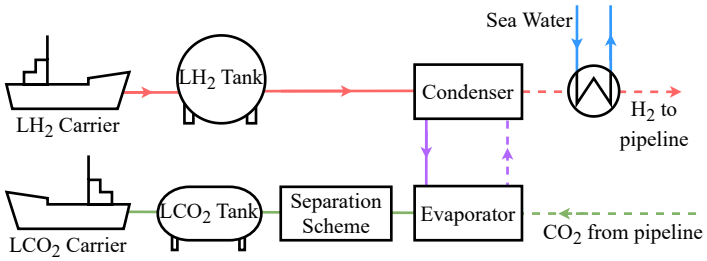
Nomenclature

CAPEX	Capital Costs	CCS	Carbon Capture & Storage	C_{AO}	Annual Operating Cost
C_c	Capital Cost	Dep	Annual depreciation	EoS	Equation of State
FEED	Front-End Engineering Design	GWP	Global Warming Potential	LCA	Life Cycle Analysis
LH_2	Liquid Hydrogen	MSE	Mean-Squared Error	MTPA	Million Tonnes Per Annum
NPV	Net Present Value	ODP	Ozone Depletion Potential	OPEX	Operating Costs
o-H_2	Ortho-Hydrogen	ORV	Open-Rack Vaporiser	PCC	Post Combustion Capture
p-H_2	Para-Hydrogen	Pf_{tot}	Annual profit	PL	Plant life in years
Q	Heat duty	r	Annual discount rate	SEC	Specific Energy Consumption
ΔH	Enthalpy change	$\dot{m}$	Mass flow rate	φ	Tax rate



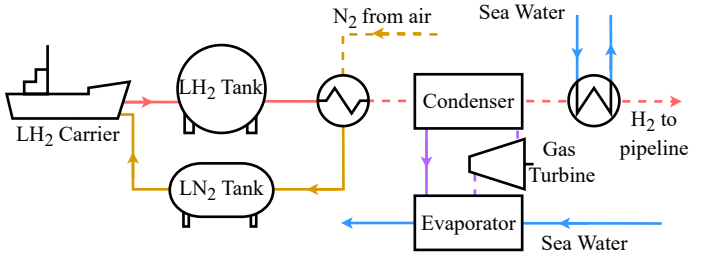
(a)

	LH_2 in	H_2 out
Flowrate [tons/hr]	114	114
Temperature (°C)	-254	10
Pressure [bar]	1	70
Phase	Liquid	Vapour



(b)

	LH_2 in	H_2 out	CO_2 in	LCO_2 out
Flowrate [tons/hr]	114	114	799	799
Temperature (°C)	-254	10	50	-55
Pressure [bar]	1	70	6	6
Phase	Liquid	Vapour	Vapour	Liquid



(c)

	LH_2 in	H_2 out	N_2 in	LN_2 out
Flowrate [tons/hr]	114	114	114	114
Temperature (°C)	-254	10	20	-196
Pressure [bar]	1	70	1	1
Phase	Liquid	Vapour	Vapour	Liquid

Figure 1: Block flow diagrams and stream tables for a) the base case (using of an open-rack vaporiser to regasify H_2) b) a dual-use LH_2 import / CO_2 export terminal c) a dual-use LH_2 import / N_2 export terminal

2.2 Dual-use LH_2 import / CO_2 export terminal

The dual-use LH_2 import / CO_2 export terminal is depicted in Figure 1b. After LH_2 is transported to the port, it is regasified and heated to -65°C in a heat exchanger, contacted with an organic refrigerant in an intermediate loop. In turn, this refrigerant is simultaneously used to liquefy CO_2. An intermediate loop is employed to minimise the formation of dry ice that may occur in a heat exchanger if CO_2 contacts with LH_2 at very low temperatures.

The CO_2 is to be sourced from post-combustion capture (PCC) facilities near to the port. Subsequently, the liquefied CO_2 enters a separation scheme since its required purity for shipping is above 99%.[9] It is noteworthy that in the case of a pipeline being used for offshore sequestration of CO_2, the purity requirements are less strict (95%).[9] For the purposes of shipping, two flash vessels in series can usually achieve the required purity. LCO_2 is then stored in a bullet-shaped vessel - to be unloaded onto a ship to be exported. The chosen shipping conditions were that of a low pressure model (-55°C, 6 barg).[10] H_2 then enters a trim heater to bring it to its desired outlet temperature of 10°C.

The CO_2 product stream from a typical PCC process is 40-50°C at 1-2 bar, after being cooled for downstream compression.[11] Under the assumption that CO_2 is sourced from nearby PCC processes at the port of Rotterdam, these streams would likely be compressed for short-distance pipeline transportation before being expanded to 6 bar for use in this terminal.

During this paper, pipeline transportation of CO_2 was not considered. A UK CO_2 Shipping study found that the break-even distance - where it was more economically viable for pipeline transportation as opposed to shipping transportation - was 500km for a basis of 5 Million Tonnes per Annum (MTPA) of CO_2.[10] Therefore, due to shipping transportation's greater flexibility compared to the use of a pipeline, CO_2 shipping was used as a basis for this study.

2.3 Dual-use LH_2 import / N_2 export terminal

The dual-use LH_2 import / N_2 export terminal is depicted in Figure 1c. After the LH_2 is transported to the port, it is regasified and heated to -220°C in a heat exchanger, contacted with gaseous N_2. Since H_2 still has a considerable amount of recoverable sensible heat, it is utilised as a heat sink in a Rankine Cycle, using sea water as a heat source. It exits the condenser at -45°C before being heated in a trim heater to 10°C. H_2 is then injected into the pipeline tie-in at the desired conditions.

Within the Rankine cycle, the working fluid is liquefied in the condenser by the H_2 before absorbing sensible heat from sea water to be vaporised. The working fluid's pressure is increased while in a liquid phase by the pump before being expanded in a gas turbine, which in turn produces electricity.

The N_2 is sourced from the air, and for this analysis, is assumed to be at atmospheric pressure and temperature (1.013 bar, 20°C). Once the LH_2 carrier has been unloaded, the LN_2 is loaded from the LN_2 storage tank onto the ship. Due to the fact that the storage vessel on the LH_2 carrier will be a specific design for carrying LH_2, it is assumed that the mass transported in the vessel must not exceed its design limits. Therefore, a 1:1 mass ratio of H_2:N_2 was chosen to investigate. The ship returns to the initial liquefaction port, where the LN_2 is used to reduce the utilities in the H_2 liquefaction process.

Previous work has investigated shipping solid-phase N_2[12]

back to the initial liquefaction port for the purposes of reducing the energy required in H_2 liquefaction. This study, however, focuses only on liquid phase transportation - due to potential limitations around shipping solid N_2 in a LH_2 carrier.

3 Methodology

The proposed terminals were simulated in Aspen Plus V11, using the GERG2008 EoS[13] to calculate thermodynamic properties of each stream. The following assumptions were made in the process simulation:

1. No heat is transferred between the process units and the environment.
2. Sea water is readily available at 20°C, and must not be returned to the sea below 15°C.
3. LH_2 is 100% pure.
4. Send-out of H_2 is at steady-state.
5. Desired H_2 tie-in conditions are 70 bar, 10°C.
6. The assessed terminal throughputs were 0.1, 0.5 & 1 MTPA of LH_2.
7. Isentropic efficiencies of process units are 0.8.
8. Pumps must not contain vapour.
9. Turbines must not contain liquid.
10. The temperature difference between hot and cold streams in heat exchangers must be at least 3°C.

The temperature of sea water in Rotterdam fluctuates between 6°C in winter months, and 20°C in summer months. In an actual terminal, process parameters would be adjusted to account for this - maintaining the terminal's operation.

3.1 Thermodynamic Model

Accurately predicting the properties of liquid H_2 is essential for the design and analysis of a LH_2 regasification terminal. Thus, to choose a suitable model to conduct the basis of analysis within Aspen Plus, a Mean-Squared Error (MSE) study was completed. The smaller a model's MSE value, the less its predicted values deviate from experimental data - making it better suited to modelling LH_2.

The following equations of state (EoS) were suggested[14] : Soave-Redlich-Kwong (SRK EoS),[15] Lee-Kesler-Plöcker (LKP EoS),[16] Peng Robinson (PR EoS)[17] and GERG2008.[18] These models were compared against experimental data sourced from the National Institute of Standards and Technology (NIST).[19]

Following the Mean-Squared Error analysis, the GERG2008 model proved suitable for modelling key properties of LH_2, namely specific heat capacity and enthalpy (see Supplementary Information for detailed results). One drawback of GERG2008 is the fact it can only model 21 components.[20] This however, proved to be inconsequential for the purposes of this study.

3.2 Conversion between para and ortho hydrogen

Two isomeric forms of hydrogen occur, para-hydrogen (p-H_2) and ortho-hydrogen (o-H_2), depending on the nuclear spin states of each proton in the molecule.[21] In a liquid state, H_2 consists of 99.8% p-H_2, whereas at ambient temperatures the mixture is 25% p-H_2 and 75% o-H_2. The enthalpy of conversion from p-H_2 to o-H_2 is considerable at 525 kJ/kg, larger than the latent heat of vaporisation of H_2 at 1 atm (448 kJ/kg). Thus, this endothermic conversion is an important consideration for a regasification terminal. The equilibrium ratio of o-H_2 to p-H_2 is a function of temperature,[21] and gives the relationship shown in Figure 2. This relationship was implemented within Aspen Plus, to determine the additional duty that LH_2 receives during regasification.

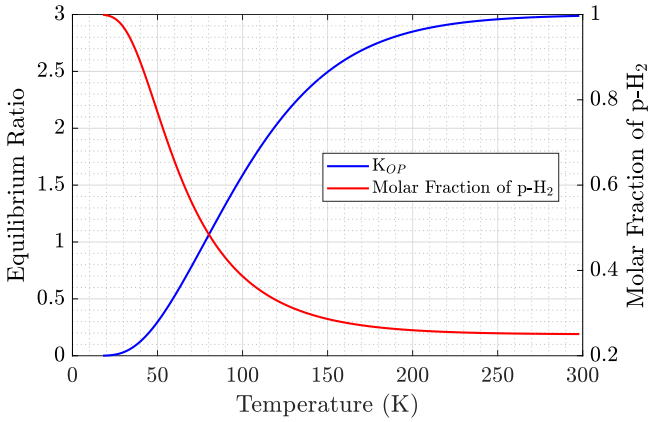


Figure 2: Equilibrium Ratio and Molar Fraction of p-H_2 vs Temperature

3.3 Intermediate Working Fluids

For the LH_2/CO_2 terminal, an intermediate loop is used to avoid directly contacting the gaseous CO_2 and LH_2. Whereas, in the LH_2/N_2 terminal, a organic rankine cycle is employed to generate electricity from the LH_2 regasification process. Thus for both terminals, a choice of intermediate fluid must be made.

The following properties dictated the choice of the intermediate fluid:

1. Must be completely vaporised at least 3°C below the heat source stream's outlet temperature
2. Must be completely liquefied at least 3°C above the heat sink stream's outlet temperature
3. Global Warming Potential (GWP) less than 50
4. Ozone Depletion Potential (ODP) equal to 0
5. Can be modelled in GERG2008 EoS

3.3.1 Intermediate fluid in the LH_2/CO_2 terminal

The choice of intermediate fluid in the LH_2/CO_2 terminal is dictated by the boiling point of the CO_2 stream. At 6 bar, the boiling point of CO_2 is -53°C. Due to the aforementioned constraint that a temperature difference of 3°C between hot and cold streams must be present: the intermediate must be completely vapourised below -56°C.

Ethane (R170) at 3.7 bar proved a suitable fluid, with a GWP of 5.5, an ODP of 0[22] and a boiling point of -60°C.[19]

3.3.2 Working fluid in the LH_2/N_2 terminal

The choice of working fluid in the LH_2/N_2 terminal is dictated by the temperature of the sea water stream. It is assumed that sea water cannot be dispensed back into the sea below temperatures of 15°C, due to environmental concerns. Therefore, the working fluid must be completely vapourised below 12°C.

Propane (R290) proved a suitable fluid, with a GWP of 0.02, an ODP of 0[22] and a boiling point of -42°C.[23] Propane should leave the turbine at 1 bar, and exit the pump at 6.8 bar.

3.4 Operability Sensitivity Analysis

The inlet and outlet conditions of each terminal is described in stream tables in Figure 1. To ensure operability of the terminal across a range of inlet compositions, the process units were evaluated at different compositions.

The LH_2 is assumed to be pure, due to the fact that impurities would be solidified in the LH_2 carrier at -254°C. Since, the N_2 is sourced from the air - its purity is fairly constant at 78%, thus an operability study of various compositions was not deemed necessary. The composition of CO_2, on the other hand, cannot be assumed as the composition of flue gas entering a PCC process may change - thus changing the composition of the CO_2 product stream. Therefore, a sensitivity analysis of the terminal was completed with three different compositions (shown in Table 4); ensuring its functionality in the case of varied inlet compositions.

Cases used in this study include: a design case, with 100% purity of CO_2; a typical case, which is the composition generally expected from the captured carbon stream of a PCC process; and a 'maximum impurities' case - with a higher molar fraction of impurities.[24]

3.5 Energy Analysis

The aim of integrating the LH_2 regasification terminal with CO_2 or N_2 liquefaction is to capture and utilise cold energy that would have otherwise been lost, thus improving energy efficiency.

Therefore, an energy balance was completed on each process unit (such as compressors, pumps and expanders) to determine the enthalpy change of each for each individual component (Equation 1). The heat duties for heat exchangers and condensers used in design were also found (using Equation 2). With these parameters, a comprehensive assessment of each terminal's energy efficiency can be developed. The energy recovered in each terminal is found using equation Equation 3, where i is either CO_2 or N_2 for their respective integrated terminals.

$$\Delta H_i = \dot{m}(h_i^{out} - h_i^{in}) \tag{1}$$

$$Q = \dot{m}_{hot}(h_{hot}^{out} - h_{hot}^{in}) = \dot{m}_{cold}(h_{cold}^{out} - h_{cold}^{in}) \tag{2}$$

$$\text{Energy Recovered} = \frac{\Delta H_{CO_2}}{\Delta H_{H_2}} \times 100\% \tag{3}$$

$$\dot{W}_{pump} = \dot{m} \cdot (h_{pump}^{out} - h_{pump}^{in}) \tag{4}$$

$$\dot{W}_{comp} = \dot{m} \cdot (h_{comp}^{out} - h_{comp}^{in}) \tag{5}$$

3.6 Economic Analysis

A dual-use regasification terminal is only practically feasible if it is economically viable, therefore an economic analysis was conducted on the chosen terminals as well as the base case. Within this study, the CAPEX, OPEX and NPV were calculated assuming a terminal throughput of 0.1, 0.5 & 1 MTPA of H_2.

3.6.1 Operating Costs

The OPEX of each terminal was calculated by summing the yearly utility and operating costs of each process unit. Equations 6 and 7, show cost of utilities for pumps and compressors, respectively. The additional operational costs associated with maintenance of the terminal was assumed to be 2% of the terminal's capital cost yearly. OPEX values were also retrieved from Aspen Plus' economic analyser to validate these calculations.

$$\text{OPEX}_{pump} = \dot{W}_{pump}\ [\text{kW}] \cdot \$0.179/\text{kWh} \cdot 8000\ \frac{hours}{year} \tag{6}$$

$$\text{OPEX}_{comp} = \dot{W}_{comp}\ [\text{kW}] \cdot \$0.179/\text{kWh} \cdot 8000\ \frac{hours}{year} \tag{7}$$

3.6.2 Capital Costs

One of the initial aims of the study was to estimate the additional capital cost of each terminal, compared to the base case. The capital cost of a terminal includes, but is not limited to: the costs of land acquisition and preparation, marine facilities, storage vessels, utilities and safety systems, administrative buildings as well as the sum of installed costs of individual process units. When comparing terminals to one another, the significant difference in capital costs stem from the differing process units and storage vessels. Thus, when calculating CAPEX for terminal comparison, only the installed costs of process units and storage vessels was considered. The CAPEX values are likely accurate within +/-40%, due to the pre-FEED nature of this study.

To calculate the installed costs of process units, various methods were implemented, including methods outlined by the Royal Society of Chemistry (2005), US Energy Department (2002), and Douglas (1988) (See Supplementary Information). The Marshall and Swift Equipment Cost Index[25] was employed to obtain more accurate values from these methods - to account for inflation since the dates of publication. CAPEX values were also obtained from Aspen Plus' economic analyser. The larger value between the calculated installed cost and the value given by the economic analyser was used for estimates to give a worst-case scenario value of CAPEX.

The terminal storage tanks were sized on a basis of 100% capacity of their respective shipping carriers. The liquid level in each tank is a transient value, designed to fluctuate to accommodate the loading/unloading schedules associated with shipping. It is assumed a period of 36 hours is required for loading and an additional 36 hours for loading at the port. Thus, no more than 121 shipments can occur per year before additional jetties and port facilities must be built.

LH_2 carriers with a capacity of 150,000 m^3 are currently in development,[26] therefore this size was used as a basis for the LH_2 storage tank. A 2020 study predicted a 50,000 m^3 LH_2 storage tank would cost $320 million,[27] but also that a 150,000 m^3 LH_2 carrier would cost only $92 million more. Since information surrounding the cost of large scale (50,000m^3) LH_2 tanks is sparse - as none have been built - $320 million was used as a basis for the cost of a 50,000 m^3 LH_2 storage tank. Since the required storage capacity is 100% of the LH_2 carrier's capacity, each terminal requires 3 LH_2 storage tanks.

There are current efforts to build LCO_2 carriers with up to a 70,000 m^3 capacity,[28] therefore this capacity was chosen as a basis for the terminal's LCO_2 storage tank. The capital cost of this 70,000 m^3 LCO_2 storage tank was assumed to be $70 million - based on an assumption of $1,000 per m^3.[29]

Since a 1:1 mass ratio of LH_2:LN_2 was chosen for this

study, a 15,000 m^3 LN_2 tank was used as a basis - due to the differences in density between LN_2 and LH_2.[19] The LN_2 storage tank was assumed to be similar to that of an LNG tank, since their temperatures are similar (-196°C vs -163°C) at around atmospheric pressure. Therefore, the capital cost of this 15,000 m^3 LN_2 storage tank was assumed to be $50 million.[30]

3.6.3 Net Present Value Analysis

In order to determine the additional value brought to the regasification process by building a dual-use terminal, the NPV was calculated using Equation 8. The NPV provides a holistic view of an asset's value over its lifespan. It encompasses: the revenue, operational costs, capital cost, annual depreciation and discount rates, tax rates as well as the asset's lifespan. Parameters used in the calculations of the NPV are listed in Table 1.

$$\text{NPV} = -C_C + \sum_{y=1}^{PL} \frac{(Pf_{tot} - C_{AO} - Dep) \times (1-\varphi) + Dep}{(1+r)^y} \quad (8)$$

Table 1: Parameters used during the Economic Analysis

Parameter	Value	Ref
Usable energy from H_2	33.33 kWh/kg	
Price of electricity	$0.179/kWh	
Annual depreciation, Dep	4% of C_C	[31]
Tax rate, φ	21%	[32]
Terminal Lifespan, PL	25 years	
Annual Discount Rate, r	3.45%	[33]
M&S Cost Index	2400	[25]
Plant Operating Time	8000 hrs/yr	

The added profit of a terminal is calculated with Equations 9, 10 and 11.

$$\text{Pf}_{tot} = \text{Pf}_{CO_2} + \text{Pf}_{elec} \quad (9)$$

$$\text{Pf}_{CO_2} = \text{TPA of } CO_2 \cdot \text{Profit Generated (\$/tonne of } CO_2) \quad (10)$$

$$\text{Pf}_{elec} = \dot{W}_{turbine}[\frac{kWh}{yr}] \cdot \$0.179/\text{kWh} \quad (11)$$

3.7 Carbon Analysis

Hydrogen is becoming an increasingly more popular energy vector due to its environmental benefits, such as not being a large emitter of greenhouse gases . This benefit is most notable in the production of green hydrogen[34] which is generated using clean electricity from renewable sources e.g. solar and wind power to electrolyse water, therefore emits virtually no carbon emissions. Because of its prevalence in the effort to reducing carbon emissions from the energy sector, making the hydrogen supply chain more sustainable is critical.

A carbon analysis was performed on the designed terminals, so that their impacts on the environment and sustainability of the hydrogen supply chain can be evaluated. It is important to note that the analysis within this study does not constitute a Life Cycle Analysis and does not consider hydrogen slip, but rather gives an overview of each terminal's impact on carbon emissions.

The Specific Energy Consumption (SEC), is a measure that can be used to compare the power requirement of each process. The SEC is calculated using Equation 12.

$$SEC = \frac{\text{Energy Used [kWh]}}{\text{Product amount [kg]}} \quad (12)$$

During 2023 in Rotterdam, where the terminal is expected to be built, an average of 0.421 kg_{CO_2}[35] was emitted per kWh of electricity generated. Using this value alongside the SEC, the expected carbon emissions from each terminal can be determined.

4 Results and Discussion

4.1 Energy Analysis

The performance of each terminal was analysed using the methods described in Section 3.5. The energy recovery of both terminals is shown by Sankey Diagrams in Figure 3c.

4.1.1 LH_2/CO_2 Terminal Energy Analysis

The flowsheet depicted in Figure 3a was implemented in Aspen Plus to complete an energy analysis of the process. Following the operability sensitivity analysis, this flowsheet's separation scheme proved suitable for all 3 compositions discussed in Section 3.4.

The terminal enabled a mass ratio of 7:1 of LH_2 regasified to CO_2 liquefied - resulting in an energy recovery of 75.8%. The resulting 983 kJ/kg of extractable cold energy is lost, due to the fact that LH_2 cannot be heated above -63°C in the condenser or violate previous constraints.

A thermodynamic cycle was considered to recover some of this lost energy. Using the LH_2 at -63°C as a heat sink and sea water at 20°C as a heat source. It was decided that the net work obtained from this cycle was not sufficient to warrant the additional capital costs, and such this was not considered as a viable inclusion to the terminal. Using low-grade heat, usually available within a large port, as a heat source could potentially enable the use of a thermodynamic engine to recover some of this energy; however, this was not investigated.

4.1.2 LH_2/N_2 Terminal Energy Analysis

The flowsheet depicted in Figure 3b was implemented in Aspen Plus to complete an energy analysis of the process. As mentioned in Section 2.3, a 1:1 mass ratio of LH_2 regasified to N_2 liquefied was assumed for this study, resulting in an energy recovery of 21.1% - comprised of 10.4% from the liquefaction of N_2 and 10.7% as electricity produced in the gas turbine. Despite a somewhat optimised Rankine cycle: using process parameters obtained in previous work,[1] only 15% of the 'cold energy' transferred to the working fluid is converted into electricity using this method. This could be potentially be improved with the use of low-grade heat,[36] as it would increase the heat source's - thus improving the thermodynamic engine's efficiency.

In order to justify this terminal's suitability for reusing the cold energy from LH_2 - the amount of energy saved as a result of liquefying N_2 and shipping it back to the original H_2 liquefaction port must be calculated.

H_2 requires 10-13 kWh/kg[37] in practice to be liquefied, however this value is expected to drop as low as 7 kWh/kg[38] as liquefaction technologies advance. Assuming the conditions in Table 6, the use of N_2 to reduce

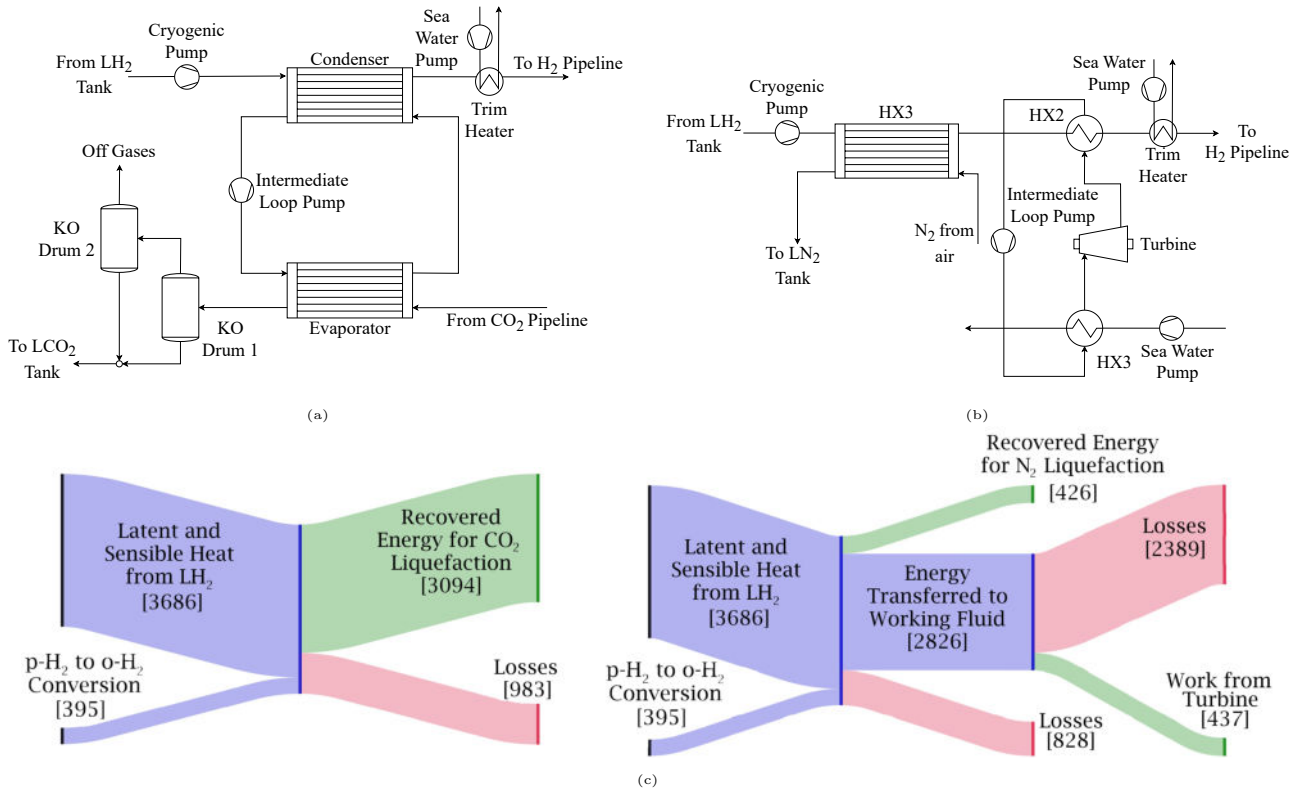


Figure 3: a = Process Flow Diagram of the LH_2/CO_2 Terminal implemented in Aspen Plus. b = Process Flow Diagram of the LH_2/N_2 implemented in Aspen Plus. c = Sankey Diagram showing the energy recovery from a) the LH_2 import / CO_2 export Terminal and b) the LH_2 import / N_2 export Terminal. Values in brackets are kJ/kg of H_2

energy during liquefaction saves only 1.19% of the initial liquefaction energy requirement.

$$\text{Energy Saved} = \frac{\dot{m}[C_p\Delta T + h_{LV}]}{10\ \text{kWh/kg}} = \frac{0.119\ \text{kWh/kg}}{10\ \text{kWh/kg}} = 1.19\%\ \text{Reduction in Energy} \quad (13)$$

This makes the case of a LH_2/N_2 terminal quite difficult to argue for. Instead of using liquid N_2, solidifying and shipping N_2 at a 6.43:1 mass ratio of LH_2:N_2 was found to save 25.4% of the initial liquefaction cost.[12] However, for the purposes of this study, it was assumed that the LH_2 carrier must carry liquids and cannot carry above the weight of the initial LH_2 cargo.

From these analyses we can conclude that more 'cold energy' is reused when using the LH_2/CO_2 terminal. However, this information should not be used in isolation to determine the terminals' value - an economic analysis must also be completed to assess their economic viability.

4.2 Economic Analysis

See Supplementary Information for values calculated during the Economic Analysis.

4.2.1 LH_2/CO_2 Terminal Economic Analysis

Using the methods outlined in Section 3.6, the CAPEX, OPEX and NPV of the integrated LH_2 import / CO_2 export terminal was compared against a non-integrated LH_2 import / CO_2 export terminal as well as the base case LH_2 regasification terminal. The 'Maximum Impurities' case was assumed when estimating the capital and operational costs of process units, as equipment should be sized to handle the worst-case scenario.

As shown in Tables 7 and 8, compared to two separate CO_2 liquefaction and LH_2 regasification terminals, an integrated terminal saves $28 million per MTPA of CO_2 in OPEX. This large reduction can be explained by the huge decrease in energy required to liquefy CO_2. The integrated LH_2/CO_2 terminal requires an additional $0.08-0.09 per kg of H_2 in OPEX, compared to the base case LH_2 regasification terminal. This is due to additional pumping and cooling costs as well as the operational costs associated with a larger number of process units. See Figure 4b for a comparison of additional OPEX compared to the base case at different design bases.

Building an integrated LH_2/CO_2 terminal compared to two separate terminals saves $136 million in capital costs for a basis of 1 MTPA of LH_2. This is due to the fact that large compression units are not necessary for an integrated terminal. Assuming the same 1 MTPA basis, the integrated LH_2/CO_2 terminal requires an additional $78 million in CAPEX, compared to the base case LH_2 regasification terminal. This is due to the fact that the base case terminal does not require storage capacity for CO_2, which is the largest contributer to the additional CAPEX. See Figure 4a for a comparison of capital cost at various design bases.

The added NPV of an integrated LH_2 import / CO_2 export terminal is largely dependant on the profit generated from CO_2, either derived from carbon credits or a reduction in carbon tax. A sensitivity analysis of added NPV for different values of H_2 throughput (MTPA) and different values for the profit generated per tonne of CO_2[39] is shown in Figure 4d. As expected, a larger terminal throughput yields a greater return on investment. When comparing the dual-terminal against building separate

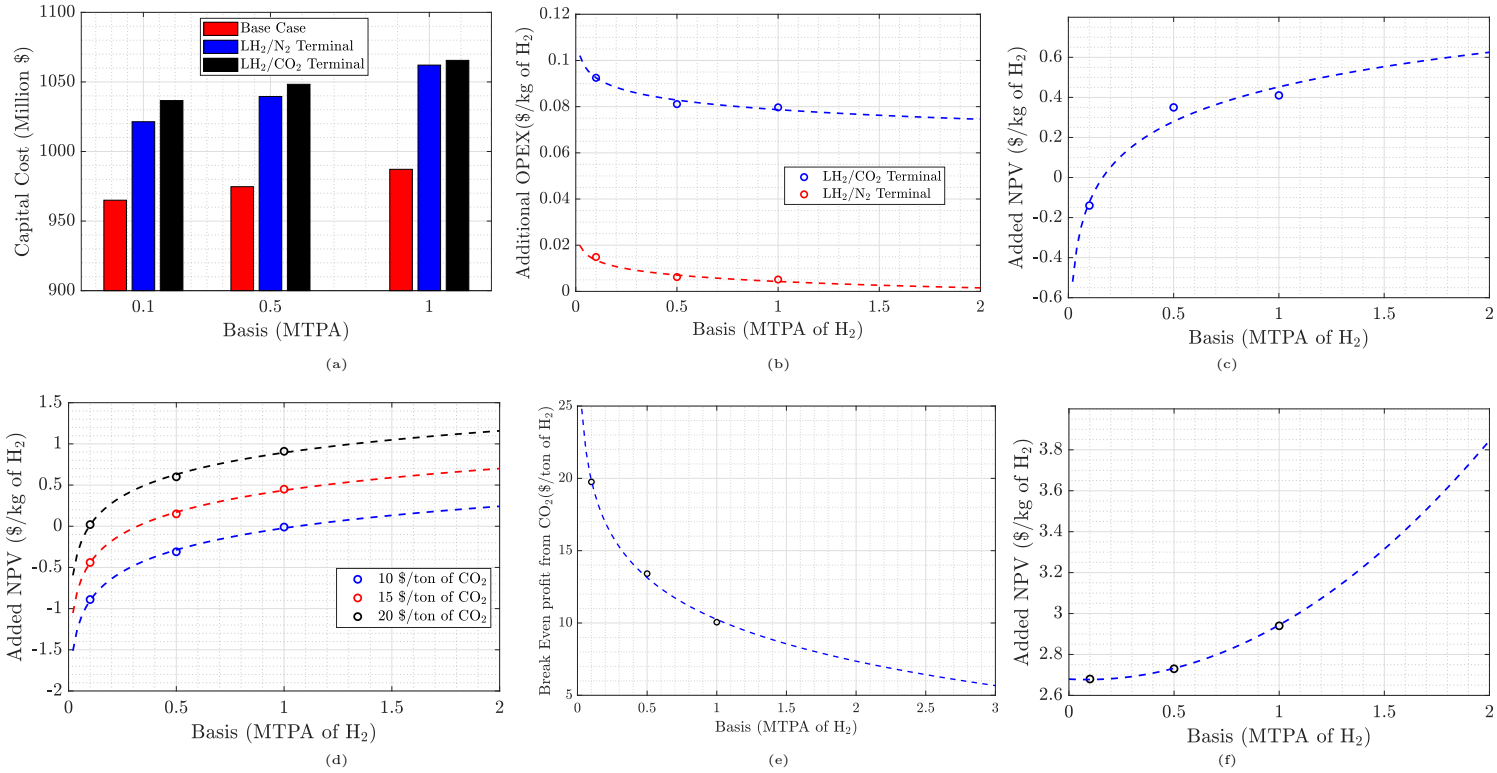


Figure 4: The X axis of all plots is the design basis of LH_2 throughput (MTPA of H_2) a = Capital Cost of Base Case, LH_2/CO_2 and LH_2/N_2terminals. b = Added OPEX/kg of H_2 for LH_2/CO_2 and LH_2/N_2 terminals compared against the base case c = Added NPV per kg of H_2 for LH_2/N_2 terminal d = Added NPV per kg of H_2 for different values of profit generated from CO_2 ($/tonne$_{CO_2}$) - LH_2/CO_2 terminal vs Base Case. e = Minimum required profit generated per tonne of CO_2 ($/tonne$_{CO_2}$) to add value to the LH_2/CO_2 terminal when compared against the base case. f = Added NPV per kg of H_2 for the integrated LH_2/CO_2 when compared against two separate CO_2 and LH_2 terminals

LH_2 regasification and CO_2 liquefaction terminals, the added NPV per kg of H_2 is even more substantial as shown in Figure 4f.

The trends in Figures 4d, 4e and 4f are not valid above a basis of 2 MTPA, as this throughput would require additional storage capacity and marine facilities.

4.2.2 LH_2/N_2 Terminal Economic Analysis

As with the LH_2/CO_2 terminal, an economic analysis was conducted on the LH_2/N_2 terminal using methods described in section 3.6. The CAPEX, OPEX and NPV of the integrated LH_2 import / N_2 export terminal was compared to the base case LH_2 regasification terminal. The installation costs for this integrated terminal is described in Table 13 and can be compared to the base case installation costs described in Table 10.

As shown in Figure 4b, compared to the base case a LH_2/N_2 terminal requires an additional \$0.005-0.02 per kg of H_2 in OPEX. This increase is due to the extra operational costs of processing units within the Rankine cycle and the N_2 tank.

Assuming a basis of 1 MTPA, the LH_2/N_2 terminal requires an additional \$76 million in CAPEX, compared to the base case LH_2 regasification terminal. This is due to the fact that the base case terminal does not have a Rankine cycle or a N_2 tank. See Figure 4a for a CAPEX comparison at various design bases.

The added NPV of an integrated LH_2 import / N_2 export terminal stems from the electricity generated in the Rankine cycle, and the energy saved in the initial liquefaction process by loading the LN_2 onto the LHN_2 carrier. The volatility of this NPV will thus be determined by electricity costs rather than the price of CO_2. A sensitivity analysis of added NPV for different values of H_2 throughput (MTPA) is shown in Figure 4c. As with the LH_2/CO_2 terminal, a larger terminal throughput yields a greater return on investment. To break even, a basis of at least 0.15 MTPA of LH_2 is required - otherwise the added capital costs and operational costs of additional process units and storage vessels outweigh the reduction in utilities.

From these analyses we can conclude that: for profit values larger than \$15 per tonne of CO_2, the added NPV of the LH_2/CO_2 surpassed that of the LH_2/N_2 terminal. However, depending on the carbon market's volatility - a LH_2/N_2 may be a more secure investment, since a reduction in energy requirements will always add value.

4.3 Carbon Analysis

Using methods described in Section 3.7, the SEC of each terminal and the subsequent carbon emissions per terminal is calculated and presented in Table 14.

The LH_2/CO_2 process produces an additional 0.0219 kg CO_2eq per kg of LH_2 in emissions. Despite this, the terminal has a positive impact on reducing carbon emissions because it involves the liquefaction of CO_2, which can then be stored rather than emitted into the atmosphere. This does not fall under the scope of the terminal's emissions reduction - however, providing energy for the liquefaction of CO_2 to facilitate CCS is a hugely beneficial feature of the terminal

The LH_2/N_2 terminal can achieve carbon neutrality, i.e. produces 0 kg CO_2eq per year, if power generated in the

gas turbine is used to power process units in the terminal itself, and the remaining is sold for profit. If all the electricity is sold - and none used within the process - the LH_2/N_2 terminal produces an additional 0.000213 kg CO_2eq per kg of LH_2 in emissions.

The LH_2/N_2 process also improves the sustainability of the hydrogen supply chain by utilising the LN_2 produced in the regasification process during LH_2 liquefaction, to reduce the energy requirement. However, for a 1:1 mass ratio of LH_2:N_2, this contribution to the energy demand is minimal - only 1.2% for a 1:1 mass basis of LH_2:LN_2. Moreover, shipping the nitrogen back to the original port will result in more carbon emissions, due to the fuel consumption required to ship the nitrogen back to the original port, thus negating some of the environmental benefits of repurposing of LN_2. Returning to the importance of enhancing the sustainability of the hydrogen supply chain, this additional carbon output undermines the benefits of this terminal. The terminal can be improved by using a larger ratio of N_2:H_2, or by using a solid-phase N_2, but this was not investigated during this study.

5 Conclusion

The purpose of this study was to design a practical, economically viable terminal that reuses cold energy released during the regasification of LH_2. The terminal should consider how shipping hydrogen between ports affects the process as well as being both economically viable and energy efficient. The terminal should also have a positive effect on the environment.

After an initial pre-selection stage of six different terminal proposals, the LH_2/CO_2 and the LH_2/N_2 terminals were chosen for further analysis. These terminals were simulated using Aspen Plus V11, under relevant shipping and pipeline injection conditions. For both terminals an intermediate fluid was required: the LH_2/CO_2 terminal used Ethane (R170) as an intermediate fluid to prevent direct contact between the hydrogen and CO_2 streams, whereas the LH_2/N_2 terminal employs a Rankine cycle to generate electricity from the regasification process, and uses propane (R290) as its intermediate fluid. The choice of intermediate fluid was dictated by properties discussed in Section 3.3. Both terminals were compared to the base case hydrogen regasification terminal and each other - to assess their cold energy recovery, economic viability and carbon emissions.

The LH_2/CO_2 terminal proved to recover more cold energy - aligning with this paper's aim to propose an energy efficient process. The terminal recovered up to 75.8% of cold energy released during the regasification of LH_2, utilising the energy for the liquefaction of CO_2. The LH_2/N_2 terminal recovers only 21%, which is then utilised for both the liquefaction of N_2 and power generation. While this study utilised optimised process parameters obtained from literature,[1] a methodical optimisation scheme such as Quasi-Newton may obtain better-performing values for each process - potentially improving energy recovery for each terminal.

Both terminals proved to be more economically viable than the base case hydrogen regasification terminal. The LH_2/CO_2 terminal's added NPV was highly dependant on the profit generated from CO_2. The revenue produced from CO_2 is sourced from the sale of carbon credits or a reduction in carbon tax. For profit values larger than \$15 per tonne of CO_2, the added NPV of the LH_2/CO_2 surpassed that of the LH_2/N_2 terminal. The added NPV of the LH_2/N_2 terminal was observed only above a basis of 0.15 MTPA of LH_2. The additional capital cost of process units and storage vessels for the LH_2/CO_2 and LH_2/N_2 terminals were \$78 and \$76 million, respectively - assuming a basis of 1 MTPA of LH_2. The CAPEX values are likely accurate within +/-40%, due to the pre-FEED nature of this study. The values associated with storage vessels are estimated[27] - due to the fact that large scale LH_2 tanks have not been built. More precise values could not be determined within time constraints of this project. This is a source of inaccuracy within this study, and further work should look to increase the reliability of these estimates.

The environmental impact, in the scope of carbon emissions under assumptions stated previously, of the supply chain is improved by both dual-use terminals. Using the average energy blend in the Netherlands during 2023, it was calculated that the LH_2/CO_2 terminal emits an additional 0.0219 kg CO_2eq per kg of LH_2, while simultaneously liquefying CO_2 for CCS. The LH_2/N_2 terminal produces a sufficient amount of energy to allow the terminal to be self-sufficient - thus having no energy from carbon-emitting sources. However, if the electricity is sold instead, the terminal emits an additional 0.000213 kg CO_2eq per kg of LH_2. The ability to use LN_2 to reduce the energy required in H_2 liquefaction also reduces the need for carbon emitting power sources, promoting a more circular supply chain. To fully understand how each terminal impacts the environment, a cradle-to-grave life cycle assessment should be completed - involving essential considerations like hydrogen permeation,[40] which can lead to ozone depletion.

In conclusion, the proposed processes highlight a significant step forward in improving circularity in the hydrogen supply chain. Each terminal effectively and practically reuses cold energy during LH_2 regasification, adding value to the overall hydrogen supply chain. Further exploration into terminal feasibility and process/marine safety can also be accomplished with a WITNESS[41] model.

Acknowledgments

This work was supported by Cameron Aldren, whose contributions and insight were vital to the completion of this project.

References

[1] Mun H, Park S, Lee I. Liquid hydrogen cold energy recovery to enhance sustainability: Optimal design of dual-stage power generation cycles. Energy. 2023;284:129229. Available from: https://www.sciencedirect.com/science/article/pii/S0360544223026233.

[2] Naquash A, Riaz A, Qyyum MA, Aziz M, Assareh E, Lee M. Liquid hydrogen storage and regasification process integrated with LNG, NGL, and liquid helium production. Renewable Energy. 2023;213:165-75. Available from: https://www.sciencedirect.com/science/article/pii/S0960148123007553.

[3] Zhang L, Ye K, Wang Y, Han W, Xie M, Chen L. Performance analysis of a hybrid system combining cryogenic separation carbon capture and liquid air energy storage (CS-LAES). Energy. 2024;290(C). Available from: https://ideas.repec.org/a/eee/energy/v290y2024ics0360544223032619.html.

[4] Gasunie. Hydrogen Network Netherlands; 2024. Accessed: 2024-11-20. Available from: https://www.gasunie.nl/en/projects/hydrogen-network-netherlands.

[5] Kountouris I, Bramstoft R, Madsen T, et al. A unified European hydrogen infrastructure planning to support the rapid scale-up of hydrogen production. Nature Communications. 2024;15:5517. Available from: https://doi.org/10.1038/s41467-024-49867-w.

[6] Sass JP, Cyr WWS, Barrett TM, Baumgartner RG, Lott JW, Fesmire JE. GLASS BUBBLES INSULATION FOR LIQUID HYDROGEN STORAGE TANKS. AIP Conference Proceedings. 2010 04;1218(1):772-9. Available from: https://doi.org/10.1063/1.3422430.

[7] Notardonato WU, Swanger AM, Fesmire JE, Jumper KM, Johnson WL, Tomsik TM. Zero boil-off methods for large-scale liquid hydrogen tanks using integrated refrigeration and storage. IOP Conference Series: Materials Science and Engineering. 2017 dec;278(1):012012. Available from: https://dx.doi.org/10.1088/1757-899X/278/1/012012.

[8] Liu Y, Zhou P, Jeong B, Wang H. Design and optimization of a type-C tank for liquid hydrogen marine transport. International Journal of Hydrogen Energy. 2023;48(89):34885-96. Available from: https://www.sciencedirect.com/science/article/pii/S0360319923023753.

[9] CCUS Network. CO_2 Specifications for Transport: TG3 Briefing; 2024. Accessed: 2024-11-20. Available from: "https://www.ccusnetwork.eu/sites/default/files/TG3_Briefing-CO\textsubscript{2}-Specifications-for-Transport.pdf".

[10] Capture UC, (UKCCSRC) SRC. The Role of CO_2 Shipping; 2020. Accessed: 2024-11-20. Available from: https://ukccsrc.ac.uk/wp-content/uploads/2020/06/The-role-of-CO\textsubscript{2}-Shipping-KO.pdf.

[11] Wang N, Wang D, Krook-Riekkola A, Ji X. MEA-based CO_2 capture: a study focuses on MEA concentrations and process parameters. Frontiers in Energy Research. 2023;11.

[12] Hunt JD, Marin Montanari P, Nieto Hummes D, Taghavi M, Zakeri B, Joel Romero O, et al. Solid air hydrogen liquefaction, the missing link of the hydrogen economy. International Journal of Hydrogen Energy. 2023;48(75):29198-208.

[13] Kunz O, Wagner W. The GERG-2008 Wide-Range Equation of State for Natural Gases and Other Mixtures: An Expansion of GERG-2004. Journal of Chemical & Engineering Data. 2012;57(11):3032-91.

[14] Charalampidi AM. Evaluation of thermodynamic models for pure hydrogen and hydrogen-containing binary mixtures; 2022. Available from: https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/3027678.

[15] Redlich O, Kwong JNS. On the Thermodynamics of Solutions. V. An Equation of State. Fugacities of Gaseous Solutions. Chemical Reviews. 1949;44(1):233-44. PMID: 18125401. Available from: https://doi.org/10.1021/cr60137a013.

[16] Lee BI, Kesler MG. A generalized thermodynamic correlation based on three-parameter corresponding states. AIChE Journal. 1975;21(3):510-27. Available from: https://aiche.onlinelibrary.wiley.com/doi/abs/10.1002/aic.690210313.

[17] Peng DY, Robinson DB. A New Two-Constant Equation of State. Industrial and Engineering Chemistry: Fundamentals. 1976;15:59-64.

[18] Kunz O, Wagner W. The GERG-2008 Wide-Range Equation of State for Natural Gases and Other Mixtures: An Expansion of GERG-2004. Journal of Chemical & Engineering Data. 2012;57(11):3032-91. Available from: https://doi.org/10.1021/je300655b.

[19] of Standards NI, Technology. Thermophysical Properties of Fluid Systems; 2018. Available from: https://webbook.nist.gov/chemistry/fluid/.

[20] CONVAL. GERG-2008; 2024. Accessed: 2024-11-20. Available from: https://conval.de/en/conval/gerg-2008-0.

[21] Ghafri SZA, Munro S, Cardella U, Funke T, Notardonato W, Trusler JPM, et al. Hydrogen liquefaction: a review of the fundamental physics, engineering practice and future opportunities. Energy & Environmental Science. 2022;15(7):2690-731.

[22] Razzaq MA, Ahamed JU, Hossain MAM. A REVIEW ON HYDROCARBON (HCs) AS AN ALTERNATIVE REFRIGERANT: BASED ON THERMODYNAMIC AND ENVIRONMENTAL APPROACH. 11. 2019:86-96. Available from: https://www.researchgate.net/publication/332785950_A_REVIEW_ON_HYDROCARBON_HCs_AS_AN_ALTERNATIVE_REFRIGERANT_BASED_ON_THERMODYNAMIC_AND_ENVIRONMENTAL_APPROACH.

[23] Stausholm T. IPCC Includes GWPs for Hydrocarbons in New Report; 2021. Available from: https://naturalrefrigerants.com/ipcc-includes-gwps-for-hydrocarbons-in-new-report/.

[24] Adamu A, Russo-Abegão F, Boodhoo K. Process intensification technologies for CO_2 capture and conversion – a review. BMC Chemical Engineering. 2020;2(1).

[25] Marshall, Swift. Marshall and Swift Equipment Cost Index (Q12015); 2015. Available from: https://scholars.unh.edu/cost/6.

[26] Collins L. Oil giant TotalEnergies to co-develop the world's largest liquid-hydrogen carrier vessel; 2023. Available from: https://www.hydrogeninsight.com/innovation/oil-giant-totalenergies-to-co-develop-the-world-s-largest-liquid-hydrogen-carrier-vessel/2-1-1443046?zephr_sso_ott=XvHEzl.

[27] International Energy Agency (IEA). The Future of Hydrogen. Paris; 2019. Licence: CC BY 4.0. Available from: https://www.iea.org/reports/the-future-of-hydrogen.

[28] Baggio T, Taylor E. Liquid CO_2 carriage by sea: Market movements; 2022. Available from: https://www.clarksons.com/home/news-and-insights/2022/liquid-CO\textsubscript{2}-carriage-by-sea-market-movements/.

[29] Thunder Said Energy. Storage Tank Costs: Storing Oil, Energy, Water, and Chemicals; n.d. Accessed: 2024-11-26. Available from: https://thundersaidenergy.com/downloads/storage-tank-costs-storing-oil-energy-water-and-chemicals/.

[30] M Habib Chusnul Fikri, Jooned Hendrarsakti, Kriyo Sambodho, Frengki Mohamad Felayati, Nilam Sari Octaviani, Mohamad Jeffry Giranza, Gregorius Andrico Hutomo. Estimating Capital Cost of Small Scale LNG Carrier; 2018. Accessed: 2024-12-02. Available from: https://www.scitepress.org/Papers/2018/85421/85421.pdf.

[31] Taxes in Europe Database v2; 2015. Accessed: 2024-11-20. Available from: https://ec.europa.eu/taxation_customs/tedb/legacy/taxDetail.html?id=436/1424159279&taxType=CIT.

[32] Deloitte; 2022. Available from: https://dits.deloitte.com/#TaxGuides.

[33] Reference and discount rates; n.d. Available from: https://competition-policy.ec.europa.eu/state-aid/legislation/reference-discount-rates-and-recovery-interest-rates/reference-and-discount-rates_en.

[34] Grid N. The hydrogen colour spectrum. Available from: https://www.nationalgrid.com/stories/energy-explained/hydrogen-colour-spectrum.

[35] Nowtricity. Current emissions in Netherlands. Available from: https://www.nowtricity.com/country/netherlands/.

[36] Sun Z, Lai J, Wang S, Wang T. Thermodynamic optimization and comparative study of different ORC configurations utilizing the exergies of LNG and low grade heat of different temperatures. Energy. 2018;147:688-700. Available from: https://www.sciencedirect.com/science/article/pii/S0360544218301038.

[37] U S Department of Energy HP. Energy Requirements for Hydrogen Gas Compression and Liquefaction as Related to Vehicle Storage Needs; 2009. Accessed: 2024-11-20. Available from: https://www.hydrogen.energy.gov/docs/hydrogenprogramlibraries/pdfs/9013_energy_requirements_for_hydrogen_gas_compression.pdf?Status=Master.

[38] Barclay JA. Active Magnetic Regenerative Liquefier; 2009. Available from: www.hydrogen.energy.gov.

[39] BloombergNEF. Carbon Credits Face Biggest Test Yet, Could Reach $238/Ton in 2050, According to BloombergNEF Report; 2024. Available from: https://about.bnef.com/blog/carbon-credits-face-biggest-test-yet-could-reach-238-ton-in-2050-according-to-bloombergnef-report/.

[40] Topics S. Hydrogen Permeability; n.d. Accessed: 2024-12-10. Available from: https://www.sciencedirect.com/topics/engineering/hydrogen-permeability.

[41] Group L. WITNESS Simulation Software; 2024. White paper introducing WITNESS 12, focusing on simulation software features, optimization, and sustainability modeling. Available from: https://www.lanner.com.

[42] Fluids - Latent Heat of Evaporation; 2019. Available from: https://www.engineeringtoolbox.com/fluids-evaporation-latent-heat-d_147.html.

Characterisation of structural and kinetic properties of DETA-functionalised polymeric resin adsorbents for direct air capture of CO_2

Dimitrios Stathogiannopoulos and Paul J. Popovich
Department of Chemical Engineering, Imperial College London, U.K.

Abstract

To mitigate the impact of anthropogenic climate change, 107 nations have pledged to achieve net-zero emissions by 2050. Achieving this goal requires the use of direct air capture (DAC) technologies, with one of the most promising being adsorption by amine functionalised polymeric resins. Despite the existence of benchmark materials, few have been optimised for DAC of CO_2, and key performance characteristics, namely kinetic parameters, remain underreported in literature. This project consists of the characterisation of structural properties – potentially impactful to the adsorption kinetics – of a novel amine-functionalised resin. Pore size distributions (PSDs) and pore volumes across the full range of pore widths were obtained via N_2 sorption at 77 K and mercury intrusion porosimetry (MIP). The PSDs were used to evaluate the material surface area, and together with the parallel pore model, the effective straight pore diffusivity. Microporous samples displayed high surface areas (162 to 774 m^2/g) and low diffusivities (5.4×10^{-6} to 8.9×10^{-6} m^2/s). Conversely, macroporous samples displayed low surface areas (4 to 8 m^2/g) and high diffusivities (5.0×10^{-5} m^2/s). To aid the understanding of the trade-off between these properties, a mathematical model was developed. Synthetic PSDs were constructed using both unimodal and bimodal Gaussian distributions, and the relationships between the distribution parameters, the diffusivity and the surface area were investigated. The data collected in this project will be used to identify whether the rate-limiting kinetic mechanism is mass transfer or chemisorption and to optimise the polymer synthesis method to improve adsorption performance.

1. Introduction

As of June 2024, 107 sovereign nations, collectively responsible for approximately 82% of the world's CO_2 emissions, have pledged to achieve net-zero before or by 2050 (United Nations Environment Programme, 2024). However, as indicated by the International Energy Agency (IEA), these commitments are unlikely to be met in the absence of drastic near-term policies. Among the targets set by the IEA in its Net Zero by 2050 roadmap was the use of Direct Air Capture (DAC). The IEA predicts that for the net zero target to be met, DAC will need to account for the removal of 980 Mt/year of CO_2 by 2050 (IEA, 2022). Currently, there are 27 commissioned DAC plants in the world amounting to the removal of approximately 0.01 Mt CO_2/year (Budinis, 2024). An additional 130 plants are currently under planning and expected to go operational in the next couple of decades while the first 1 Mt CO_2/year facility is due to come on-line in the US in the next 2-3 years (IEA 2022). Clearly, DAC technologies will need to play a much larger role in the effort to limit the elevated greenhouse effect in years to come, and improvements in current methods are expected to be of critical importance.

Carbon dioxide can be separated from gaseous mixtures in multiple ways, including the use of various liquid solvents, solid adsorbents, electrochemical cells or membranes. In electrochemical cells, CO_2 is captured through the application of a voltage, while membranes achieve separation by size exclusion. These techniques could constitute promising solutions in the future since they can offer relatively low-energy and potentially low-cost carbon sequestration. However, their scalability and effectiveness under atmospheric CO_2 concentrations are still very much under investigation. Liquid absorbents like amine-based compounds such as monoethanolamine (MEA) or alkaline solutions like KOH have proven to be effective in capturing CO_2 from concentrated streams. In fact, liquid absorption has been the benchmark carbon-capture technique in industrial settings. Nevertheless, the much lower concentration of CO_2 in the atmosphere (400 ppm) compared to industrial effluent streams, in combination with the high energy requirements for regeneration makes their application in DAC sub-optimal. On the other hand, adsorption on solid materials can offer effective and highly selective CO_2 adsorption at atmospheric concentrations while requiring a considerably lower energy input for regeneration and consequently, lower operational costs (Yafiee et al., 2024).

Polymers, silicas, zeolites and metal-organic frameworks (MOFs) are among the solids that effectively adsorb CO_2 physically (Low et al, 2023a). Functionalisation of adsorbents with amines has been seen to greatly increase their CO_2 adsorption (chemisorption) capacity through the formation of carbamic acid or alkylammonium carbamate (Alfonso et al., 2019). The formation of covalent bonds leads to higher affinity of the adsorbent for the adsorbate gas (i.e. CO_2), as well as improved selectivity relative to other atmospheric gases (e.g. N_2). Lewatit VP OC 1065 (referred to as Lewatit) is the current standard DAC material. It is a divinylbenzene crosslinked macroporous polymeric resin functionalised with benzylamine with an equilibrium CO_2 uptake (at DAC conditions – 0.04kPa CO_2 partial pressure and 298K) of $0.95 mmol \cdot g^{-1}$(Low et al., 2023b). Despite its popularity in DAC research, Lewatit has not been specifically designed for this application and therefore it is unlikely that the structure and chemistry is optimised for the given conditions.

This study therefore investigates similar amine-functionalised polymeric adsorbents with different pore structures to achieve a better understanding of their effect on the adsorption performance.

One critical aspect of adsorbent characteristics which is rarely investigated at this stage and for this class of adsorbents is the rate of CO_2 uptake. Kinetics is a key indicator of adsorbent performance. If kinetics are seen to be favourable (i.e., rapid) the adsorption unit can handle higher throughputs and operate under shorter cycle times, culminating to more efficient, energy- and cost-effective processes, especially when combined with a large equilibrium CO_2 uptake. Lewatit's CO_2 adsorption kinetics at DAC conditions have been investigated rather sparsely (Low et al., 2024) (Wu et al., 2022). This also applies to the wider range of amine-functionalised polymeric DAC adsorbents, indicating the presence of a knowledge gap in that field. The overall kinetics of adsorption depend on two factors; i) mass transfer (i.e., diffusion) to the site of chemisorption and ii) the rate of chemisorption itself. A clearer understanding of the relationship between adsorbent properties, such as structural and textural characteristics and chemical properties, is essential to improving the adsorption rate of DAC materials.

A few recent studies have attempted to explore this relationship. Hua et al. (2024) demonstrated that bimodal pore networks (composed of mesopores and macropores) in polymer X5, provided both enough surface area for the functional groups (PEI) to be grafted on, and enough free volume to facilitate fast mass transfer. On the other hand, Wang et al. (2020) worked with more unimodally distributed porous materials (polystyrene resins functionalised with quaternary amines) and concluded that overall, cylindrical mesopores provide better conditions for CO_2 adsorption relative to macropores as indicated by the very high kinetics – equilibrium half-time of 2.9 minutes – achieved by one of their quaternised mesoporous resins at 400 ppm and 297 K. These were the fastest kinetics measured for DAC adsorbents although the material demonstrated a small equilibrium CO_2 uptake.

This study characterizes the structural and textural properties, such as total pore volume, pore size distribution (PSD), porosity, surface area, and effective straight pore diffusivity, of vinylbenzyl chloride-based adsorbents. The findings aim to contribute to a broader investigation into how these properties influence mass transfer kinetics, and ultimately, the overall adsorption kinetics of these materials. Samples of different porosities and PSDs were synthesized through variations in the synthesis mixture (fractional composition of monomer, crosslinker and porogen). Skeletal densities were measured using He pycnometry, and used to calculate bulk densities and porosities. Moreover, adsorption isotherms of N_2 at 77 K, along with mercury intrusion porosimetry (MIP) were used to yield the PSDs from which the surface areas and overall diffusivities of the materials were calculated. The analysis of empirical data was complemented by the development of a computational model, where synthetic PSDs (unimodal and bimodal) were generated leading to a clearer understanding of the effect of PSD on surface area and overall diffusivity. A first attempt at quantifying the equilibrium CO_2 uptake and overall kinetics of adsorption at 400 ppm and 303 K was also made using a thermogravimetric analyser (TGA).

2. Methods

2.1. Material Synthesis

The polymeric resins used in this study were synthesized by Tristan Spreng, a PhD student in the Petit Group. The production of the non-functionalised materials started by mixing divinylbenzene (cross-linker), vinylbenzyl chloride (monomer) and azobisisobutyronitrile (AIBN) in a toluene and isoamyl alcohol (porogens) mixture at 10 °C in an autoclave. Deionized water, gelatin and polydiallyldimethyl-ammonium chloride (PolyDDA) were added to the autoclave while increasing the temperature to 40 °C and stirring at 100 rpm. Subsequently, the stirring speed was increased to 1000 rpm and the temperature was raised to 75 °C. The mixture was held at these conditions for 8 hours for the suspension to be formed and the polymerization to take place. The mixture was then cooled to room temperature and the aqueous phase was filtered out using a filter-fitted needle before being washed three times with DI water. The non-aqueous phase was separated from the polymer beads by addition of acetone and constant stirring for 2 hours at 60 °C for 2 cycles. The product (R-Cl, non-functionalised polymeric resin) was dried in a 60 °C vacuum oven for 24 hours.

R-Cl was amine functionalised by refluxing with diethylenetriamine (DETA) at 90 °C for 24 h. The mixture was cooled to room temperature, aspirated, and washed with methanol (stirred for 15 minutes at 30 °C) thrice before being aspirated to remove any non-grafted DETA. The product (R-DETA, amine-functionalised polymeric resin) was then dried in a 60 °C vacuum oven for 24 h. A diagram summarizing the synthesis mechanism can be found in *Appendix A.1, Figure S.1.*

The synthesis mixture composition (i.e., fractions of monomer, cross-linker and porogen as well as the ratio between the two porogens – toluene and isoamyl alcohol) were varied in order to induce differences in the structural properties of the resultant polymers. A total of 8 batches were produced (i.e., 8 sets of unique synthesis conditions – see *Appendix A.1, Table S.1*). Each of the batches contained 3 non-functionalised/R-Cl samples and 3 functionalised/R-DETA samples. This was done to enhance the reliability of the results. Therefore, resins R-1, R-2 and R-3 are subsets of Batch 1, resins R-4, R-5 and R-6 are subsets of Batch 2 and so on. Hence, a total of 48 samples were produced (8 batches containing 3 R-Cl and 3 R-DETA samples each).

2.2. Structural and textural properties

2.2.1. Skeletal density

Helium pycnometry was used to quantify the skeletal density of all samples (both functionalised and non-functionalised). Approximately 50 – 400 mg of sample adsorbent was placed in the sample chamber of a Micromeritics AccuPyc II 1340 pycnometer and the exact mass was recorded. The sample was then inserted in the

machine and underwent 20 He purge cycles, followed by an additional 10 measurement cycles at approximately 1.34 bar (19.5 psig). The measured skeletal density for each material was averaged over the 10 runs and recorded.

2.2.2. N_2 adsorption isotherms at 77K

Structural properties of the adsorbents, in the 0.4 to 50 nm pore size region, were characterised using a volumetric analyser (3Flex, Micrometrics), through evaluation of the N_2 adsorption isotherms at 77 K. The adsorbent particles were first sieved to ensure that only particles of diameters between 0.2 mm and 2 mm were tested, aiming to minimize any potential particle size effect. Approximately 30 – 200 mg of adsorbent were used for each run. The sample was degassed ex-situ at 393 K and 0.002 kPa for at least 12 hours in a VacPrep Micromeritics sample preparation device. Then, the mass of the sample was recorded before an additional, 4 hour in-situ degas cycle at 393K and 0.00002 kPa.

2.2.3. Micro-mesopore PSD

The PSD in the micropore and mesopore range (0.4 – 50 nm) was determined using the Non-Local Density Functional Theory (NLDFT) model assuming slit pore geometries, and manually optimised to simultaneously minimize the root mean square (RMS) error of the fit and the roughness of the resultant distribution. The NLDFT model was used to increase accuracy and provide high-resolution estimations of the samples' PSDs in the micro-(diameter < 2 nm) and mesopore (2 – 50 nm) regions. The NLDFT model, assumes a structured phase near the pore walls, and a bulk fluid state near the pore centre with a smooth transition between the two states. This model was selected over other alternatives since it accounts directly for molecular interactions in confined pore spaces unlike the Barrett, Joyner and Halenda (BJH) and Dollimore-Heal (DH) models which are based on capillary condensation and assume bulk fluid properties throughout the cross-sectional area of the pore. This assumption introduces inaccuracies, especially at the micropore level (Micromeritics, n.d.).

2.2.4. Macropore PSD

The gas adsorption data fails to provide a detailed insight into the contributions of macropores to the total pore volume since even at high pressures the gas never fills the macropores completely. To characterise the structure of macropores (diameter > 50 nm), mercury intrusion porosimetry (MIP) was used. In this technique, the dry sample is first placed in the sample chamber, which is then filled with mercury. Pressure is increased incrementally to force the liquid into the pores, with larger pores filled first (i.e., at low pressures) and smaller ones filled last. This technique allows characterisation of pore volume for pores of widths ranging from 5 – 350,000 nm. MIP data was used only for pores larger than 45 nm, as N_2 sorption at 77 K is more reliable in the mesopore region (Wouters et al., 2016). Furthermore, upon analysis of the MIP data, it became apparent that all investigated materials demonstrated similar trends at pore sizes larger than 10^4 nm (*Appendix A.3*). This degree of uniformity was unexpected given the differences in the synthesis conditions and therefore was attributed to external factors (e.g., void space between adsorbent particles) rather than the actual pore structure. Furthermore, given that the minimum particle size tested was 0.2 mm in diameter, a cut-off in pore diameters of 0.01 mm (10^4 nm) seems appropriate.

Once collected, the PSD data from the 3Flex and MIP were combined, and the resultant comprehensive PSD was normalised over the entire range of pore widths. MIP data was collected by the PhD student, T. Spreng, as this technique requires special training. The code used to combine the gas-adsorption and MIP PSD data was developed by Dr Hassan Azzan, a member of the Petit Group, and works by interpolating the incremental PSD data generated by each technique and joining them at a user-defined threshold – in this case, 45 nm.

2.3. CO_2 adsorption kinetics

All kinetic CO_2 adsorption measurements were taken using a TA 550 thermogravimetric analyser. Approximately 30 mg of DETA-functionalised sample were loaded on the TGA crucible and the exact weight was recorded. The sample was subsequently loaded on the machine and was degassed in-situ at 120 °C for 3 hours using pure Helium (BOC CP grade). The temperature was lowered to 30°C, in an intermediate, equilibration step before CO_2 was introduced in the inlet gas stream. An 800 ppm CO_2 in Helium cylinder (BOC, research grade) was used as the CO_2 supply and was mixed with the pure He stream to achieve a concentration of approximately 400 ppm. The mass of the sample as CO_2 was adsorbed was then measured once every 0.2 seconds for 3 hours at 30 °C. The CO_2 concentration at the gas outlet of the TGA was verified using a LiCor Li-850 gas analyser. At all times, a total gas flowrate of 300ml/min was maintained.

2.4. Theoretical background

2.4.1. Effective Straight Pore Diffusivity – Parallel Pore Model

The parallel pore model was used to characterize the effective straight pore diffusivity of each material. The parallel pore model assumes the adsorbent material is composed of a collection of pores aligned in parallel to one another with no intersections. Diffusion through these pores occurs through various mechanisms, and the pore size is the factor that determines which of these mechanisms is limiting (i.e., the slowest one). The model is only valid for pores above 2 nm width, due to chemisorption and activated diffusion mechanisms dominating in the micropores. In macropores, where the mean free path (i.e., the mean distance covered by a molecule between collisions with other molecules or the walls of the container) of the adsorbate is much smaller than the pore diameter, intermolecular collisions are much more common than molecule-wall collisions. Hence, bulk molecular diffusivity (D_m) dominates. On the other hand, in smaller pores, the

mean free path of the molecules may be found to be greater than, or at least comparable to the pore diameter, making molecule-wall collisions more frequent than intermolecular ones. Therefore, Knudsen diffusivity (D_K) is limiting. The combined diffusivity through a pore of radius "r", $D(r)$, is calculated by combining the effects of Knudsen and molecular diffusion:

$$\frac{1}{D(r)} = \frac{1}{D_K(r)} + \frac{1}{D_m} \qquad (Eq.1)$$

where $D_K(r) = 97r\sqrt{\frac{T}{M}}$, with r being the pore radius, T being the temperature and M being the molecular weight of the diffusing species (i.e., CO_2) (Kärger et al., 2012). Molecular diffusivities are typically found in literature as the diffusion of one species through another. In this study, the molecular diffusivity of CO_2 (at concentration of 400 ppm) through Helium ($D_m = 5.74 \times 10^{-5}\ m^2/s$, (Winkelmann & Lechner, 2007)) was used to ensure continuity between 3Flex and TGA measurements.

Finally, the product of the normalised PSD function of the material, $f(r)$, and the calculated combined diffusivity term, $D(r)$, is integrated over the entire range of pore radii to yield the effective straight pore diffusivity of the material per the Parallel Pore Model:

$$\langle \varepsilon_p D_p \rangle = \int_{r_{min}}^{r_{max}} D(r) \cdot f(r) dr \qquad (Eq.2)$$

where $f(r)$ is given by:

$$f(r) = \frac{PSD(r)}{\int_{r_{min}}^{r_{max}} PSD(r) dr} \qquad (Eq.3)$$

This term can be divided by the tortuosity, τ (measure of deviation from straight pores) to yield the overall diffusivity, D (Kärger et al., 2012):

$$D = \frac{\langle \varepsilon_p D_p \rangle}{\tau} \qquad (Eq.4)$$

However, the only estimation of τ is given by the MIP which does not consider the effect of pores smaller than 5 nm. We decided to use $\langle \varepsilon_p D_p \rangle$ instead of D to compare the diffusivity of materials to remove the effect of inaccuracies in τ.

2.4.2. Surface Area Calculation

Polymeric resins are known to have non-uniform, irregular pore geometries, largely due to the inherent variability introduced by the polymerization reaction (i.e., when monomers are crosslinked to form irregular constructs) and reinforced by post-synthesis modifications like functionalisation. The structure of polymeric resins is highly dependent on the synthesis pathway but generally it can be visualized as a variety of irregular interconnected voids (Wang et al., 2017). Traditionally, the surface area is obtained via the BET method from N_2 sorption, but Low et al. (2023a) note that it introduces inaccuracies related to the quadrupole moment of N_2 and diffusional limitations of microporous materials at 77 K, so it was not used in this investigation. Instead, it was deemed appropriate to calculate the specific surface area of our samples by assuming different pore geometries (slit and cylindrical), aiming to establish a possible range of surface area estimates to account for the irregularities in the actual structure.

If the pores are assumed to be straight cylinders of length L and radius r, their volume V and (interior) surface area A are related according to:

$$\frac{A}{V} = \frac{2\pi r \times L}{\pi r^2 \times L} = \frac{2}{r} \qquad (Eq.5)$$

Similarly, if the pores are assumed to be a slit with width w (which is twice the radius r of the equivalent cylindrical pore), formed by two parallel square planes of side length L, essentially forming a cuboid of dimensions $L \times L \times w$, their volume V and (interior) surface area A are related according to:

$$\frac{A}{V} = \frac{2 \times L^2}{L^2 \times w} = \frac{2}{w} = \frac{1}{r} \qquad (Eq.6)$$

In both cases, the surface area can be obtained as a function of pore radius and volume:

$$A_{cylinder} = \frac{2V}{r}, \quad A_{slit} = \frac{V}{r} \qquad (Eq.7\ \&\ 8)$$

A discretised array of pore radii and corresponding pore volumes was obtained from the N_2 sorption and mercury intrusion porosimetry. Therefore, the total surface area was obtained by evaluating the area at each pore radius and summing over all pore radii. Notably, the cylindrical surface area is simply twice the slit surface area. Thus, an upper and lower estimate for the surface area for pores of a given characteristic width are provided.

2.5. Computational model construction

To simulate the PSD of a real material, a gaussian distribution was used. A Gaussian distribution can be fully characterised by its mean, and the standard deviation. The standard deviation was defined by a ratio to the mean pore width, as this relative value is more meaningful than the absolute value. Therefore, using a vector of mean pore widths (4 nm to 200 nm) and a vector of standard-deviation-to-mean-width ratios (0.05 to 0.5) a matrix of synthetic PSDs, one for every combination of mean pore width and standard deviation, was generated. For each, the curve was normalised to yield an area of unity in the positive region of pore widths, as negative pore width is non-physical, and the negative region of the distribution was therefore not included in any calculation. The corresponding effective diffusivity and surface area (assuming cylindrical pores) was calculated for each PSD using the same methods outlined in *Section 2.4*, but given that a continuous function is available, the integral rather than discrete forms of these equations were evaluated. To obtain a value for surface area, the normalised PSD must be multiplied by a specific pore volume. A value of 1.8 cm^3/g was used as a starting

point, based on our analysis of the material R-Cl-22. For diffusivity calculations, only the contribution of pores wider than 2 nm was considered, as configurational diffusion is assumed to occur in all smaller pores, and they therefore do not contribute to the effective straight pore diffusivity, as described in *Section 2.4.1*.

It was deemed worthwhile to investigate a multimodal distribution, as the distributions obtained experimentally often had several "peaks" at various pore widths. The method was similar to that described above, but instead of a single vector of mean widths, two were generated - one for a range of smaller pores (4 nm to 10 nm width) and one for larger pores (20 nm to 400 nm). The standard deviation was fixed at 10% of the mean width. Normal distributions were made for these values, and then added in pairs and normalised to create a bimodal distribution. Calculations for the specific surface area and effective diffusivity were then made as described in *Section 2.4.*

3. Results and Discussion

3.1. Experimental Section

3.1.1. N_2 Isotherms (77 K)

The observed N_2 adsorption isotherms of the 8 different batches of adsorbents (*Figure 1*) generally fell within the categories of Type II (micro- & macroporous) and Type III (non-porous or macroporous). Some samples also demonstrated elements of mesoporosity in the form of a noticeable hysteresis between the adsorption and desorption branches, indicating condensation in mesopores. Adsorbents 1, 7, 16 and 19 exhibited Type III isotherms (i.e., negligible uptake at small partial pressures, indicating the absence of micropores with a relatively small increase in uptake as p/p^o approaches 1). This, in combination with the small N_2 uptake (all were below 10 cm^3 STP/g) demonstrates that these materials are either non-porous or composed of a small number of large macropores. Material 5 exhibited a similarly modest N_2 uptake at 77 K. However, as observed in its isotherm, there is noticeable gradient increase at large partial pressures, indicating the presence of macropores.

Materials 10, 13 and 22 all contain both micropores and macropores, indicated by the large N_2 uptake at low partial pressures and the fast increase in volume of adsorbed N_2 as p/p^o approaches 1. Adsorbent 10 shows the highest initial uptake – indicating micropores contribute more to its total pore volume relative to R-DETA-13 and 22. R-DETA-22 has a slightly sharper knee (gradient transition from sharp to shallow roughly at $p/p^o = 0.05$) in comparison to the other two porous materials. This could be explained by a relatively smaller contribution of mesopores to the total pore volume. This is reinforced by the observation of the relatively smaller hysteresis between the adsorption and desorption branches in its isotherm relative to the others. Hence, the micropores (monolayer adsorption) are filled to a larger extent before multilayer adsorption in macropores begins, thus explaining the sharpness of the knee in the measured isotherm. On the other hand, in resins 10 and 13, which likely have larger mesopore contributions, there is a greater degree of overlap between the onset of monolayer and multilayer adsorption.

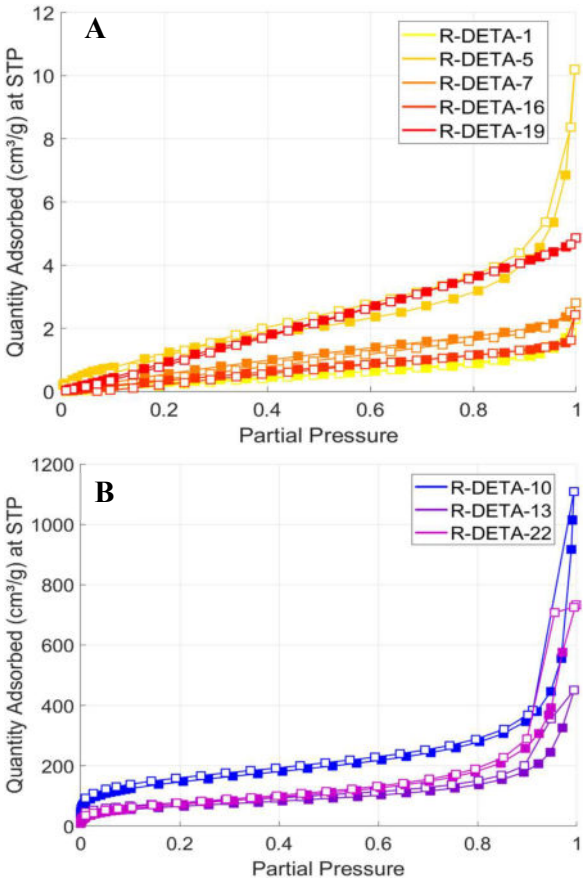


*Figure 1: N_2 adsorption isotherms at 77K for materials with **A)** Type III isotherms and **B)** Type II isotherms. Filled marks/squares represent adsorption isotherm datapoints while hollow marks/squares represent desorption isotherm datapoints*

Pore size distributions

The five most promising adsorbents were selected for MIP measurements. Theoretically, an effective DAC adsorbent would have both a large surface area, in order to maximize the amount amine-functionalised sites on the surface, and pores of sufficient size as to not hinder the rate of mass transfer to the active sites. Therefore, R-DETA-10, 13 and 22 were determined to be promising. The presence of micropores, alongside the large micropore and mesopore volume (derived by N_2 adsorption) and the rapid increase in adsorbed quantity at high partial pressures indicate that these materials are likely to possess the aforementioned qualities that characterise an effective adsorbent. R-DETA-5 was also considered a potentially important sample to characterise given its presumed macroporous structure evident from its N_2 adsorption isotherm. Lastly, one of the materials presumed to be non-porous, R-DETA-19 was also

selected for MIP testing. The rest of the samples were not characterised further. A more insightful investigation into their structure would be useful, however, given the aforementioned time constraints along with the hypothesis that even if they did display a macroporous structure, the lack of micro- and mesopores would result in an impractically low surface area led to the prioritisation of other materials.

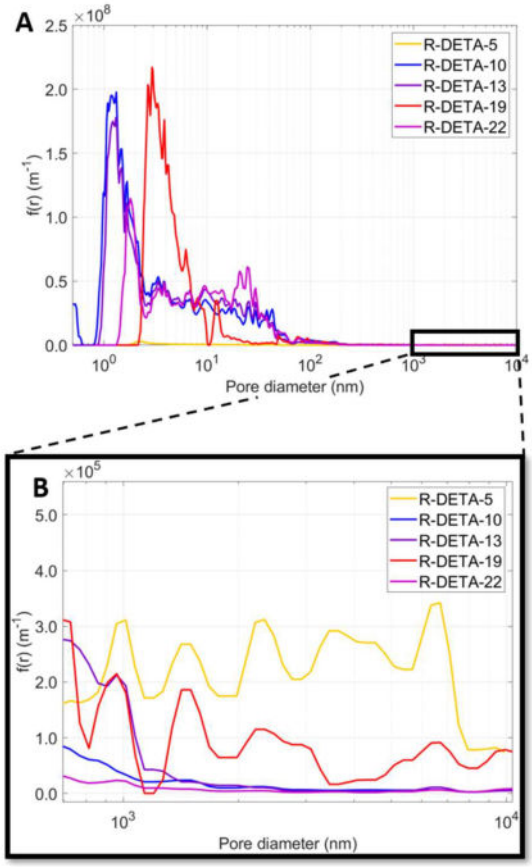


*Figure 2: **A)** Overlapped complete PSDs for materials characterised with N_2 adsorption at 77K and MIP **B)** Overlapped PSDs magnified in the range 7×10^2 to 1×10^4 nm.*

The complete PSDs (*Figure 2*), spanning a range of pore diameters between 0.4 nm and 10^4 nm, partially verify the trends observed in the materials' isotherms. Adsorbents R-DETA-10, R-DETA-13 and R-DETA-22 (in descending order) were found to possess the largest proportion of micropores and ultramicropores, hence explaining the high N_2 uptake at low partial pressures. However, these adsorbents also demonstrated very similar degrees of mesoporosity relative to one another, thus disproving the hypothesis that the sharpness of the knee observed in the R-DETA-22 isotherm was caused by a relatively lower mesopore contribution. Instead, inspection of the PSD reveals that large mesopores (> 20 nm) are more abundant in R-DETA-22 while R-DETA-13 and 22 have a greater number of small mesopores (2 – 10 nm), thus explaining the relative difference in the sharpness of the knee in their corresponding isotherms. R-DETA-19 exhibits a unimodal distribution, with most of its pores occurring within 2 to 20 nm. However, this is of little significance to the potential adsorbent performance of this material since its total pore volume was estimated at ~ 0.02 cm^3/g (see *Table 1*), thus validating the hypothesis regarding its non-porosity. Furthermore, the prediction made in *Section 3.1.1* regarding the dominance of macropore contributions to the pore volume of sample R-DETA-5 was also validated. On first inspection of the overlapped normalised PSDs presented in *Figure 2A*, the data for R-DETA-5 seems peculiar since there appears to be no clear peak while the PSD curve seems to stay close to the x-axis. This is because most of the contributions towards the total pore volume of this material are almost equally distributed among large macropores in the range of 10^3 – 10^4 nm of diameter (*Figure 2B*).

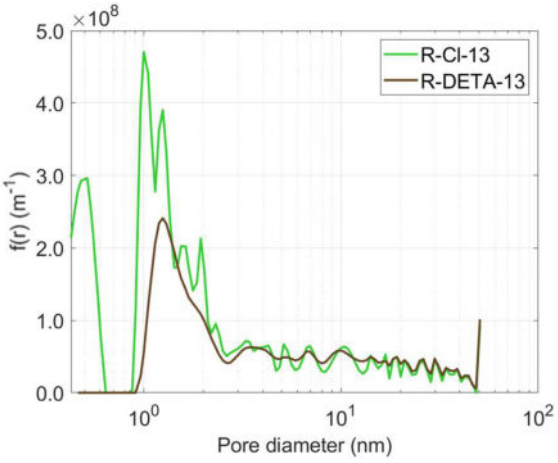


Figure 3: Effect of functionalisation on the micropore and mesopore region of the PSD of Resin 13

Next, the effect of functionalisation on the pore structure of the investigated materials was also investigated by plotting the PSDs of non-functionalised (R-Cl) samples on the same axes as their functionalised (R-DETA) counterparts. Generally, one can observe that the contribution of micropores and small mesopores (diameter < 15 nm) to the total pore volume is smaller for R-DETA samples. This can be explained by the concept of pore obstruction. When DETA is grafted on the surface of the adsorbent, the DETA molecules take up some of the void space in the pores. For micropores and ultramicropores, this may even lead to complete pore blockage, as observed in *Figure 3* showing complete elimination of ultramicropores for Resin 13 upon grafting with DETA. The loss of pore volume contributions from micropores and small mesopores results in a downward shift in the normalised PSD of functionalised materials as compared to the equivalent non-functionalised ones. Simultaneously, this obstruction signifies that the contributions of larger mesopores and macropores to the total pore volume will be proportionately larger for R-DETA samples than for R-Cl ones as seen in the overlapped

normalised PSD plots (especially for mesoporous materials like R-13). The plots showing the effect of functionalisation on the remaining materials can be found in *Appendix A.2*. It must be mentioned that MIP was not performed for the non-functionalised samples, thus the PSDs presented only cover the micropore and mesopore regions characterised by N_2 adsorption at 77 K. Nonetheless, the largest effect of functionalisation is noticed in the micropore range, where the difference between pore diameter and the size of the DETA molecules (approximately 1 nm according to Baur et al (2018)) is small.

3.1.3 Effective Straight Pore Diffusivity

The effective straight pore diffusivities of the investigated materials are presented in *Table 1*. Evidently, the materials which exhibited Type II isotherms (R-DETA-10, 13 & 22) have lower effective straight pore diffusivities in relation to samples which demonstrated Type III isotherm behaviour. This is because the prevalence of micropores in the materials exhibiting Type II isotherms means that Knudsen diffusion plays a larger role relative to when micropores are absent. On the other hand, R-DETA-5, which is almost exclusively macroporous, displays the largest effective straight pore diffusivity of the tested materials, only 8×10^{-6} m^2/s less than the molecular diffusivity term, D_m, used in effective straight pore diffusivity calculation (D_m for 400 ppm CO_2 in Helium). This indicates the dominance of this diffusion mechanism in the large pores of the sample over Knudsen diffusion. R-DETA-19 is mainly composed of mesopores and thus has an effective straight pore diffusivity term smaller than that of R-DETA-5 but larger than those of the micropore-containing samples. Of course, R-DETA-19 is not a considerably porous material overall as has been explained previously because of its small total pore volume. In general, for diffusivity metrics to have significance regarding the potential adsorption efficacy of a material they must be considered in parallel to the rest of the material's properties.

A visual representation of explanation behind the materials' observed effective straight pore diffusivity terms is given in *Figure 4*. As observed in part **A**, at low pore volumes the combined diffusivity, $D(r)$, is limited by the Knudsen diffusivity term, $D_K(r)$. As pore diameter increases, $D_K(r)$ increases linearly with it (the plot has a logarithmic x-axis and hence the trend looks exponential). However, the $D(r)$ does not increase at the same rate since at the point of intersection between $D_K(r)$ and D_m (455 nm) the molecular diffusivity becomes limiting. Therefore, at large pore diameters, the combined diffusivity asymptotically approaches D_m. As explained in *Section 2.4.1*, the effective straight pore diffusivity is equal to the integral of the product of the PSD function, $f(r)$ and the combined diffusivity $D(r)$ over the full range of r. Therefore, when $D(r)$ values are large, the effective straight pore diffusivity, $\langle \varepsilon_p \cdot D_p \rangle$, is also large. Part **B** of *Figure 4* shows the cumulative PSDs of the different materials. Macroporous materials like R-DETA-5 are mostly composed of pores within the diameter range where the combined diffusivity is large and limited by D_m, while materials with smaller pore diameters have smaller diffusivity terms due to Knudsen limitations.

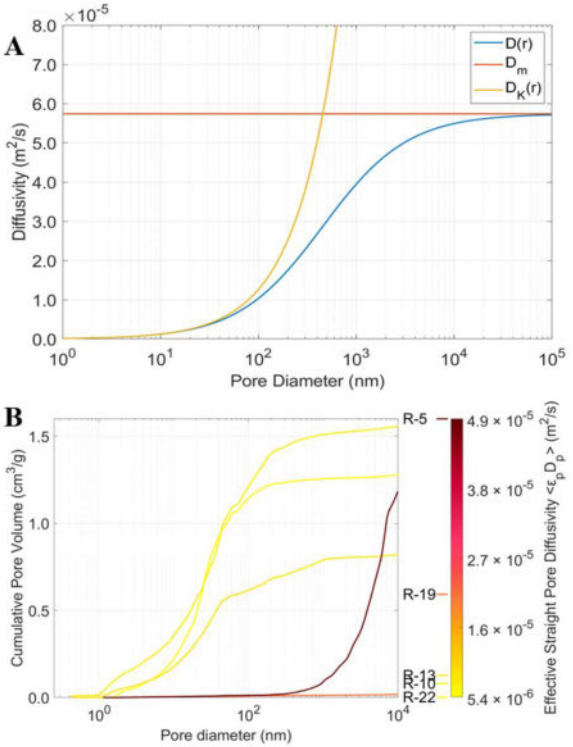


*Figure 4: **A)** Knudsen, molecular and combined diffusivities as a function of pore diameter **B)** Cumulative PSDs for investigated materials. Colour-coded according to effective straight pore diffusivity.*

Table 1: Summary of the structural and textural properties of investigated adsorbents

Adsorbent Sample	Total Pore Volume (cm^3/g)	Skeletal Density (g/cm^3)	Bulk Density (g/cm^3)	Porosity, ε_p	Effective Straight Pore Diffusivity, $\langle \varepsilon_p \cdot D_p \rangle$ (m^2/s)	Specific Surface Area (m^2/g)	
						Cylindrical geometry	Slit geometry
R-DETA-5	1.18	0.71	0.39	0.46	4.94×10^{-5}	8	4
R-DETA-10	1.56	0.88	0.37	0.58	7.60×10^{-6}	774	387
R-DETA-13	0.82	1.02	0.55	0.45	8.87×10^{-6}	325	162
R-DETA-19	0.02	1.08	1.06	0.02	2.17×10^{-5}	6	3
R-DETA-22	1.28	0.99	0.44	0.56	5.40×10^{-6}	376	188

3.1.4 Surface Area

As seen in *Table 1*, the investigated materials are characterised by a wide range of surface areas. The 3 materials that exhibited Type II isotherms have specific surface areas in the order of magnitude of 10^2 m^2/g, while the samples that exhibited Type III isotherms had estimated surface areas below 10 m^2/g. Of the 3 micro- and mesoporous materials, R-DETA-10 has the largest surface area (387 m^2/g – 774 m^2/g depending on the assumed pore geometry, considerably larger than the BET surface areas reported for the benchmarks Lewatit and Purolite – 31 and 28 m^2/g respectively). This is the result of the combination of a larger proportion of micropores and ultramicropores in its PSD and its large total pore volume relative to other materials. R-DETA-13 and 22 are characterised by similar specific surface areas, despite the fact that their PSDs do not have such a large degree of similarity. R-DETA-13 has a considerably larger micropore contribution compared to R-DETA-22. However, its total pore volume is less than two thirds that of R-DETA-13, thus the resultant specific surface area is lower than what would be expected if the materials had the same total pore volume. R-DETA-5 has an expectedly low specific surface area (4 m^2/g – 8 m^2/g) since its pore volume is attributed primarily to the presence of large macropores of diameters between 1,000 and 10,000 nm which have low surface area to volume ratios. R-DETA-19 has a similarly small specific surface are (3 m^2/g – 6 m^2/g), due to its very small total pore volume. Overall, it is observed in the experimental data, that materials with large diffusivities tend to have small specific surface areas and vice versa. This is because materials with a large proportion of micropores have large surface areas but small average individual pore volumes, resulting in predominantly Knudsen-controlled diffusion while macroporous materials are characterised by smaller surface areas and larger average individual pore volumes, where molecular diffusion dominates. However, in the experimental dataset, the described effect is somewhat warped by the differences in overall porosity of the materials (hence the similarities in surface area between R-DETA-13 & 22 and R-DETA-5 & 19). The computational model developed and presented in *Section 3.3* aims to address this limitation by focusing solely on the impact of PSD on the aforementioned properties.

3.1.5 Skeletal & Bulk Densities and Porosity

In theory, the skeletal densities were expected to be approximately the same for all tested materials since the backbone of each one is composed of vinylbenzyl chloride cross-linked with divinylbenzene. Some degree of variation may be attributed to the differences in the composition of the synthesis mixture that was used to produce the materials as well as the varying levels of DETA grafting due to differences in surface area. However, it is unlikely that these factors can lead to the degree of variation observed in *Table 1*. Instead, a potential explanation could be the presence of fully enclosed void pockets. Since these structures would not be connected to the outside environment, their presence would not be detected by He pycnometry and they would be considered in the final density measurement as part of the adsorbents skeletal structure. This could explain the very low (given the densities of the skeleton's components) skeletal density of R-DETA-5.

The bulk densities are by definition always smaller than the skeletal densities since they account for the pores of the tested polymers. As expected, materials with larger total pore volumes exhibit a larger difference between their skeletal and bulk densities relative to less porous materials. Lastly, porosity follows and equally intuitive trend, with more porous materials displaying large porosities (R-DETA-10 has the largest – 0.58) and less porous materials displaying smaller ones (R-DETA-19 is relatively non-porous so it has a porosity of 0.02).

Overall, of the investigated materials, R-DETA-10 and R-DETA-22 seem to be the most promising adsorbents. They have the largest porosities (0.58 and 0.56 respectively, considerably larger than the reported values for the benchmarks, Lewatit and Purolite – 0.338 and 0.442 respectively), large surface areas on which amines can be grafted while not having the lowest effective straight pore diffusivities. Of course, CO_2 uptake measurements (equilibrium uptake and kinetics of adsorption) are required to evaluate this hypothesis.

3.2. CO_2 uptake kinetics

Six runs of the flowrate study were carried out (twice each at 100, 200 and 300 mL/min), and 300 mL/min was chosen as the flowrate to carry out the kinetic measurements. Unfortunately, the dynamics of the change in CO_2 concentration when switching from pure He to the He/CO_2 mixture were too slow for the measurement to yield the kinetics intrinsic to the material. Further kinetic measurements were therefore abandoned in this study.

3.3. Computational model

3.3.1 Unimodal distribution

The standard deviation can be seen to have a negligible effect on both the diffusivity and surface area of the material. The mean pore width on the other hand, has a large effect on both observed properties. Across the investigated range, the diffusivity increases from 3.3×10^{-7} to 4.94×10^{-6} m^2/s. This difference can lead to an order of magnitude difference of the diffusive timescale when considering kinetics of CO_2 uptake. The surface area varies from 2270 to 3.6 m^2/g, a difference of 3 orders of magnitude. This could lead to a comparable difference in equilibrium CO_2 uptake, assuming it is proportional to the surface area (due to the number of DETA active sites). This data highlights the trade-off between equilibrium uptake and diffusive kinetics. If the chemical kinetics of the system were known, this would allow for the surface area to be optimised, by reducing the mean pore width to the point where the time scale of diffusion would match that of chemical reaction.

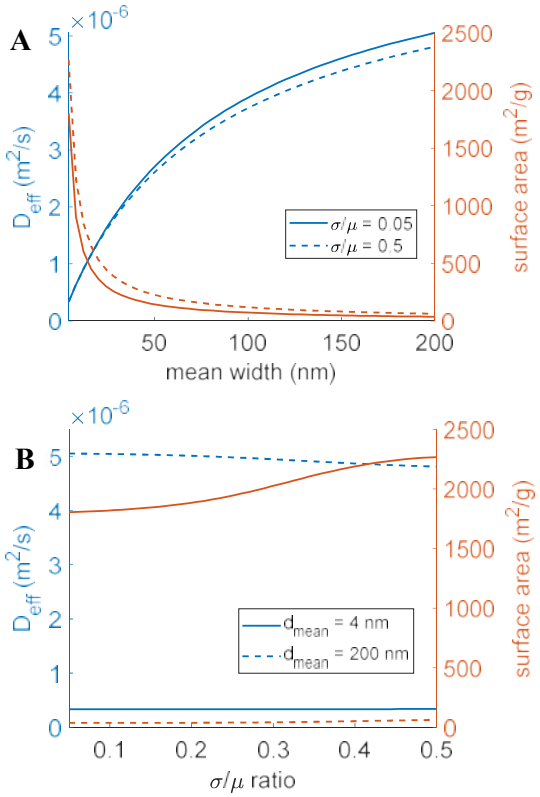


*Figure 5: **A)** Effective straight pore diffusivity and surface area as a function of mean pore width at different σ/μ ratios **B)** Effective straight pore diffusivity and surface area as a function of σ/μ ratio at different mean pore widths*

3.3.2 Bimodal distribution

Comparing the unimodal materials shown in *Figure 6,* one can observe that for the entire range of investigated pore widths, the surface area is at least an order of magnitude superior in the small-pored material, and vice-versa with diffusivity in the larger pored material. Importantly, the bimodal material manages to preserve a relatively high diffusivity, even at the lowest mean pore widths (4 and 20 nm), and a high surface area at the highest mean pore widths (10 and 400 nm). This finding suggests that if this could be controlled in synthesis, it would be desirable for the material to be made with both a significant micro-mesopore contribution and macropore contribution to the total pore volume. This would prevent the material form having a prohibitively low surface area or diffusivity at either extreme of the pore size range. The bimodal distribution presents a number of further parameters which could be investigated, such as the relative distance between the two means, or the ratio between the pore volume contribution of the small pores to the large pores, which is simply 1:1 in this model. By biasing the ratio towards small pores, the surface area could be improved at the cost of diffusivity, and vice-versa by biasing it towards large pores.

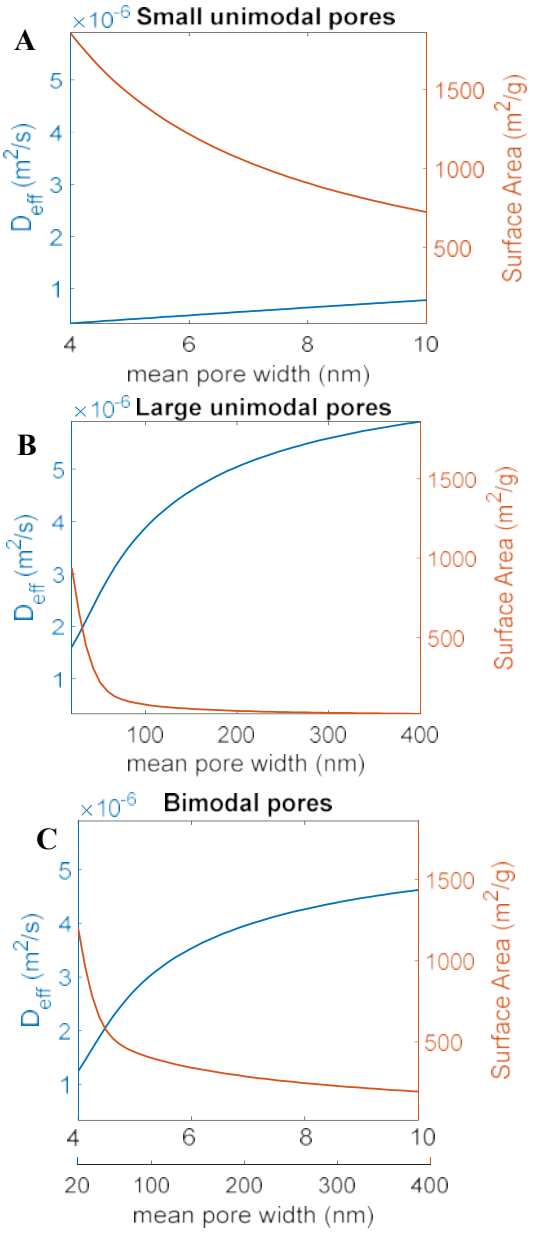


*Figure 6: Effective straight pore diffusivity and surface area as a function of mean pore width for **A)** a unimodally distributed material with small pores **B)** a unimodally distributed material with large pores **C)** a bimodally distributed material with both small and large pores*

4. Conclusion

The structural properties of novel DETA-functionalised porous organic polymer were successfully characterised. Three of the eight studied samples were found to have significant micro- and mesopore contribution to the total pore volume, with surface areas in the range of 300-800 m^2/g, substantially higher than the current benchmark,

Lewatit which has a tabulated BET area of 31 m^2/g. The samples with low micro- and mesoporosity presented surface areas in the range of 1-6 m^2/g. For the 5 samples on which MIP was performed and the full range of porosity was characterised, the effective straight pore diffusivities were found to be in the range of $5.4x10^{-6}$ to $5.6x10^{-5}$ m^2/s. Continuation of this work would include revising or finding an alternative method of measuring the intrinsic kinetics of CO_2 adsorption. Alternatively, an explicit study of the chemical kinetics of CO_2 capture on a DETA-functionalised polymer surface could be carried out. This data could then be used in combination with data collected in this research for the purposes of optimising the polymer synthesis, as planned in the PhD work of T. Spreng. By comparing the diffusive kinetics elucidated in this research to the kinetics found from TGA (or similar) studies, one can determine which of the mechanisms (chemical or diffusive) is rate-limiting and optimise the adsorbent accordingly. Furthermore, providing total pore volume, skeletal and bulk density and kinetic data at DAC conditions (400 ppm CO_2, 30 °C) could aid in development of process models, for which this necessary data is often under-reported. The mathematical model developed in this project also provides insight into design of porous media. In particular, it highlights how surface area can be maximised by aiming to generate pores as small as the diffusivity allows, for specific area to not be the rate limiting kinetic parameter.

5. Acknowledgements

We sincerely thank Tristan Spreng for his exceptional guidance, support, and dedication throughout this project, which exceeded all expectations. We also thank Dr. Reza Vakili for assistance during experimentation and Dr. Hasan Azzan for help with data analysis.

6. References

1. Afonso, R. et al. (2019) 'Unravelling the structure of chemisorbed co2 species in mesoporous Aminosilicas: A critical survey', *Environmental Science & Technology*, 53(5), pp. 2758–2767. doi:10.1021/acs.est.8b05978.

2. Baur, G.B., Spring, J. and Kiwi-Minsker, L. (2018) 'Amine functionalized activated carbon fibers as effective structured adsorbents for formaldehyde removal', *Adsorption*, 24(8), pp. 725–732. doi:10.1007/s10450-018-9974-x.

3. Budinis, S. (2024) *Direct air capture*, IEA. Available at: https://www.iea.org/energy-system/carbon-capture-utilisation-and-storage/direct-air-capture (Accessed: 01 December 2024).

4. Hua, J. et al. (2024) 'Direct air capture of CO2 by amine-impregnated resin: The effect of resin pore structure and humid conditions', *Carbon Capture Science & Technology*, 12, p. 100237. doi:10.1016/j.ccst.2024.100237.

5. IEA (2022) *Direct Air Capture 2022*, IEA, Paris. Available at: https://www.iea.org/reports/direct-air-capture-2022 (Licence: CC BY 4.0).

6. Kärger, J., Ruthven, D.M. and Theodorou, D.N. (2012) *Diffusion in nanoporous materials*. Hoboken: John Wiley & Sons.

7. Low, M.-Y. (Ashlyn) et al. (2023a) 'Analytical review of the current state of knowledge of adsorption materials and processes for direct air capture', *Chemical Engineering Research and Design*, 189, pp. 745–767. doi:10.1016/j.cherd.2022.11.040.

8. Low, M.-Y. (Ashlyn) et al. (2023b) 'Measurement of physicochemical properties and CO2, N2, AR, O2, and H2O unary adsorption isotherms of Purolite A110 and Lewatit VP OC 1065 for application in direct air capture', *Journal of Chemical Engineering & Data*, 68(12), pp. 3499–3511. doi:10.1021/acs.jced.3c00401.s003.

9. Low, M.-Y. (Ashlyn) et al. (2024) *Quantification of temperature-dependent CO2 adsorption kinetics in Lewatit VP OC 1065, Purolite A110, and Tifsix-3-NI for direct air capture* [Preprint]. doi:10.26434/chemrxiv-2024-3phh7.

10. Micromeritics (no date) *Whitepaper: An introduction to NLDFT*. Available at: https://www.micromeritics.com/Repository/Files/WP_An_Introduction_to_NLDFT.pdf (Accessed: November 2024).

11. United Nations Environment Programme (2024) *Emissions Gap Report 2024*. Available at: https://wedocs.unep.org/20.500.11822/46404 (Accessed: 11 December 2024).

12. Wang, B. et al. (2017) 'Macroporous materials: Microfluidic fabrication, functionalization and applications', *Chemical Society Reviews*, 46(3), pp. 855–914. doi:10.1039/c5cs00065c.

13. Wang, T. et al. (2020) 'Quaternary functionalized mesoporous adsorbents for ultra-high kinetics of CO2 capture from air', *Scientific Reports*, 10(1). doi:10.1038/s41598-020-77477-1.

14. Winkelmann, J. and Lechner, M.D. (2007) *Gases in gases, liquids and their mixtures*. Berlin, Heidelberg: Springer Berlin Heidelberg.

15. Wouters, S. et al. (2016) 'Comprehensive study of the macropore and mesopore size distributions in polymer monoliths using complementary physical characterization techniques and liquid chromatography', *Journal of Separation Science*, 39(23). doi:10.1002/jssc.201670231.

16. Wu, J. et al. (2022) 'The analysis and evaluation of direct air capture adsorbents on the material characterization level', *Chemical Engineering Journal*, 450, p. 137958. doi:10.1016/j.cej.2022.137958.

17. Yafiee, O.A. et al. (2024) 'Direct air capture (DAC) vs. Direct Ocean Capture (DOC)–a perspective on scale-up demonstrations and environmental relevance to sustain decarbonization', *Chemical Engineering Journal*, 497, p. 154421. doi:10.1016/j.cej.2024.154421.

Computational Modelling of Blood Clot Lysis: The Effect of Clot Heterogeneity on Lytic Efficacy

Suphapacha (Gots) Narongchartsopon, Yew Jueen Tan
Department of Chemical Engineering, Imperial College London, U.K.

ABSTRACT This study investigates the effects of clot heterogeneity on the dynamics of tissue plasminogen activator-mediated thrombolysis by developing a reduced-order computational model incorporating spatial and temporal complexities. To represent more realistic clot structures compared to current homogeneous models, our model simulates architecturally layered clots with denser platelet-rich shell and finer red blood cell-rich core that compositionally adapt according to lysis progression via a novel sigmoid-form hindrance function proposed. Results revealed up to a six-fold increase in predicted lysis time due to the addition of a second cellular component and a shell layer. The final heterogeneous model predicted a complete recanalisation time of 43 minutes, aligning with current clinical data. This highlights the importance of describing clot composition accurately in computational thrombolysis models. Future work should include the use of real-time imaging to inform mathematical representations of clot properties, and more rigorous validation of computational results against clinical data.

KEYWORDS acute ischemic stroke; thrombolysis; clot heterogeneity; tissue plasminogen activator; computational modelling; pharmacokinetics; pharmacodynamics

INTRODUCTION

Acute ischemic stroke (AIS) occurs when a blood clot obstructs blood flow to the brain, leading to cell death and potential long-term neurological damage [1]. AIS is a leading cause of disability worldwide and a significant contributor to mortality, especially in low- and middle-income countries, highlighting the need for effective and easily accessible interventions [2]. Current therapeutic approaches to restore cerebral perfusion primarily involve two methods: thrombolysis with tissue plasminogen activator (tPA) and mechanical thrombectomy.

However, the selection and efficacy of these strategies depend significantly on the characteristics of the clot. For instance, thrombectomy is more effective for large, fibrin-dense clots, whereas thrombolytic therapy with medication is the primary option for smaller clots. Despite its potential, tPA-mediated thrombolysis often exhibits great variability in clinical efficacy due to the heterogeneity of clot composition, structure, and spatial arrangement [3-8]. These differences influence clot permeability and response to fibrinolytic agents. Consequently, understanding the impact of these compositional and structural variations is essential to optimising therapeutic strategies and improve AIS patient outcomes.

This study investigates clot heterogeneity and its impact on thrombolytic therapy through two primary dimensions: spatial heterogeneity and temporal heterogeneity. Spatial heterogeneity itself encompasses two aspects:

1. Variability in clot composition – thrombi are composed of varying proportions of multiple components such as fibrin, red blood cell (RBC), and platelet [3-6].
2. Layered architecture – thrombi can display a dense, platelet-rich shell surrounding an RBC-rich core, with the outer shell acting as a barrier to thrombolysis while the core is more easily degraded [6-8].

The temporal heterogeneity refers to the dynamic redistribution of trapped clot components during lysis. These factors collectively highlight the need to account for clot heterogeneity in computational models to improve their predictive accuracy and clinical relevance.

BACKGROUND

The treatment of AIS through thrombolysis with tPA has been extensively studied, with both experimental work and computational models providing insights into the biochemical mechanisms involved. However, current computational models often oversimplify the representation of real-world thrombi by assuming uniform composition and pure fibrin clots. For example, Piebalgs et al. developed a three-dimensional computational model to simulate fibrinolysis with the assumption that the clot is composed of purely fibrin with no other cellular components [9]. Similarly, Gu et al. simulated fibrinolysis of a pure fibrin clot and assumed spatially homogeneous clot structure [10]. These simplified assumptions fail to account for the heterogeneity observed in vivo, which directly affect the rate of fibrinolysis and recanalisation outcomes. For instance, denser regions provide greater resistance to drug penetration and enzymatic activity, while coarser regions are more easily penetrated and more readily lysed owing to their relatively loose network [11].

Despite these documented findings, clot heterogeneity remains underexplored in the realm of computational modelling, which limits the translational applicability of model predictions in clinical context. Building upon the one-dimensional (1D) homogeneous clot model developed by Gu et al. in 2019, this study introduces spatial and temporal heterogeneities to more accurately represent AIS thrombi. By incorporating these complexities, the aim is to bridge the gap between idealised models and clinical observations, and thus provide more applicable insights for thrombolytic treatment of AIS.

METHODS

To model the systemic behaviour of key fibrinolytic proteins within the body, a two-compartment pharmacokinetics (PK) and pharmacodynamics (PD) model was used. This is referred to in short as the systemic PKPD model. The central compartment represents well-perfused environments like blood, liver, and kidneys, while the peripheral compartment represents poorly-perfused regions like muscle tissue [12, 13]. Only tPA was considered to exchange between these two compartments, as the changes in concentrations of other proteins were assumed to be negligible in comparison to the dosage of tPA administered. The only species that enters the central compartment externally is tPA through intravenous infusion, at a rate determined by the administration strategy. A typical dosing regimen was used, where 10% of a total 0.9 mg/kg dosage is delivered as an initial bolus within the first minute, while the remainder of the dosage is delivered via constant continuous infusion over the following 60 minutes [14].

For all species, elimination occurs from the central compartment at a rate inversely dependent on half-life. The elimination constant of tPA was calculated based on the presence of two disposition phases [15]. This process follows first-order kinetics with the rate of elimination being proportional to the current plasma concentration. The model assumed that each species is generated at a rate equal to the elimination rate at initial plasma concentrations. Considered reactions occur solely within the central compartment, and the overall rate of production or consumption of each species due to chemical reaction depend on the reactions they are involved in. The network of plasma phase reactions considered is illustrated below in Figure 1.

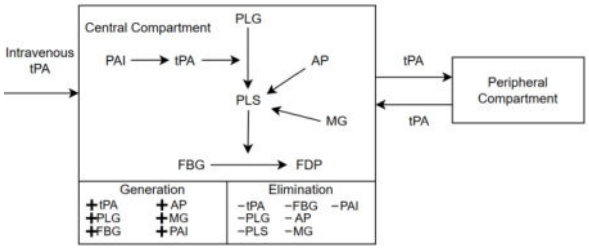


Figure 1. Two-compartment systemic PKPD model.

A brief overview of the reactions occurring is outlined. tPA activates plasminogen (PLG) to plasmin (PLS), but can be inhibited by plasminogen activator inhibitor type-1 (PAI). PLS is able to convert fibrinogen (FBG), a precursor to fibrin, into degradation products (FDP). PLS is inhibited by antiplasmin (AP) and α2-macroglobulin (MG) [16]. The kinetics governing these chemical reactions are outlined in Tables S1 and S3 of the Supplementary Information (SI), available as a separate file.

Performing mass balances on a control volume defined by the central compartment give rise to a set of ordinary differential equations (ODEs) described by Equations 1 and 2,

$$\frac{dC_{tPA,c}}{dt} = \left[\frac{I}{M_{W,tPA}V_{plasma}} + k_{p\to c}C_{tPA,p}\right] - \left[\left(k_{el,tPA} + k_{c\to p}\right)C_{tPA,c}\right] + G_{tPA} + R_{tPA}^{plasma} \quad (1)$$

$$\frac{dC_i}{dt} = [0] - \left[k_{el,i}C_i\right] + [G_i] + \left[R_i^{plasma}\right] \quad (2)$$

where i refers to each of the species reacting in the plasma phase, $i = \{PLG, PLS, FBG, AP, MG, PAI, PLS\text{-}AP\}$. I refers to the infusion rate of tPA, and $k_{p\to c}$ and $k_{c\to p}$ are the kinetic constants for the rate of exchange between each compartment. For variables not explicitly mentioned, their definitions can be found in Table S0. Each of the terms separated by square brackets represent the rate of inflow, outflow, generation, and production/consumption due to reaction respectively. In accordance with the stated assumption, for the peripheral compartment, there is only flow due to exchange of tPA between compartments.

$$\frac{dC_{tPA,p}}{dt} = \left[k_{c\to p}C_{tPA,c}\right] - \left[k_{p\to c}C_{tPA,p}\right] + [0] + [0] \quad (3)$$

These coupled ODEs were solved using MATLAB's in-built ODE solver *ode23s*, which was selected due to its ability to handle stiff differential equations effectively [17]. The stiffness can arise from the range of timescales involved in each reaction, where some occur much faster than others. *ode23s* employs a modified Rosenbrock formula that ensures computational efficiency while maintaining sufficient accuracy [18]. The resolved concentration profiles represent the behaviour of key proteins throughout the body once tPA is administered. Simplistically, this can be considered as the general behaviour in the red area (blood/plasma) in Figure 2.

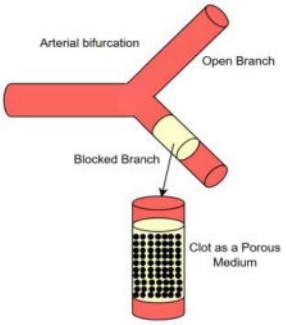


Figure 2. Schematic representation of the arterial bifurcation, with the behaviour in the occluded artery (blocked branch) represented by the local PD model.

Within the neighbourhood surrounding the clot, lysis dynamics are captured through the interplay between:

1. Convection – the flow of blood.
2. Diffusion – the transport of fibrinolytic species.
3. Reaction – the chemical interactions of proteins.

These three phenomena are captured through a 1D convection-diffusion-reaction equation, giving rise to the set of partial differential equations defined by Equation 4 to model the plasma phase behaviour of species flowing through the occluded artery with respect to space and time.

$$\frac{\partial c_i}{\partial t} = -\frac{\partial (U c_i)}{\partial x} + D_i \frac{\partial^2 c_i}{\partial x^2} + R_i^{tot} \tag{4}$$

The blood velocity, U, was calculated via Darcy's law,

$$U = \frac{1}{\varepsilon_{clot}} \frac{k_t}{\mu} \Delta P_x \tag{5}$$

which is dependent on the pressure drop across the clot, ΔP_x, as well as clot permeability, k_t, porosity, ε_{clot}, and blood viscosity, μ [19]. The values of parameters used in Equations 4 and 5 are given in Tables S7 and S9. The total reaction term for each species, R_i^{tot} is the sum of their reaction terms in each phase. All the reactions considered for each of the phases are illustrated in Figure 3.

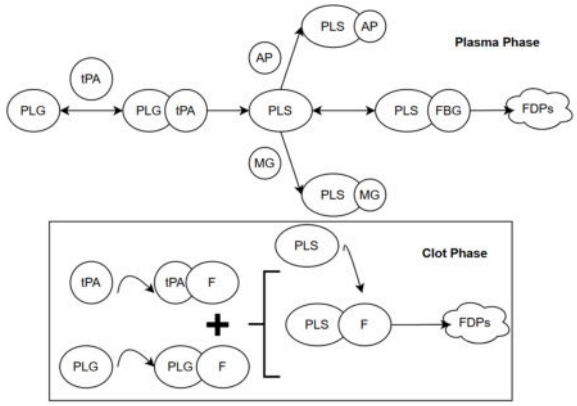


Figure 3. Mechanism of fibrinolysis.

The plasma phase here is described by the exact same reaction network as Figure 1. In addition, the presence of a fibrin network allows tPA, PLG and PLS to form bound phase species that behave according to the clot phase reaction kinetics.

$$\frac{\partial n_i}{\partial t} = R_i^{clot} \tag{6}$$

Details regarding kinetic constants and the exact reaction and rate equations are outlined in Tables S2 and S3. The local PD model refers to the coupling of the 1D flow, transport, and reaction model with the fibrinolysis mechanism, specifically focusing on the blocked branch.

Equations 4 and 6 were solved numerically according to the procedure outlined by Gu et al. [10]. The discretised spatial and temporal grid sizes chosen provide a balance between accuracy and computational time with respect to very fine mesh tested, allowing the process to be modelled in an efficient manner. The inlet (boundary) conditions to the local PD model were the solutions of the systemic PKPD model. The initial conditions for both models were the same, as outlined in Table S6. For further details regarding the mathematics of the numerical procedure, the reader is referred to the SI of [10].

Calculation of the Darcy velocity requires overall clot permeability to be known. Several methods to compute this permeability were initially explored, each requiring different combinations of equations. Considering a homogeneous clot, a base permeability was computed using one of the three equations, all of which depend purely on fibrin properties, namely the fibre radius, R_f, and the fibrin volume fraction ϕ_f. The first was Davies' model, while the second and third were those proposed by Jackson and James, and Clague et al. [20-24]. Davies' equation is shown in Equation 7,

$$k_{f,i} = \frac{4R_f^2}{70\phi_{f,i}^{1.5}\left(1+52\phi_{f,i}^{1.5}\right)} \tag{7}$$

where i refers to the specific layer within the clot, $i = \{s, c\}$, with s being the shell and c being the core. Davies' model was selected for calculation of fibrin permeability due to its ability to provide reasonable values of permeability across all ranges of fibrin volume fraction. The function is very well-behaved compared to, for instance, the non-linearities in the Jackson and James model which unrealistically predict negative permeabilities at higher volume fractions. This is further supported by the frequent usage of Davies' equation in many existing computational studies of thrombolysis [9, 10, 25-28]. Additionally, it seems to be in good agreement with the experimental data obtained by Wufsus et al. for the majority of compositions relevant to this study, as demonstrated by Figure 4 [21].

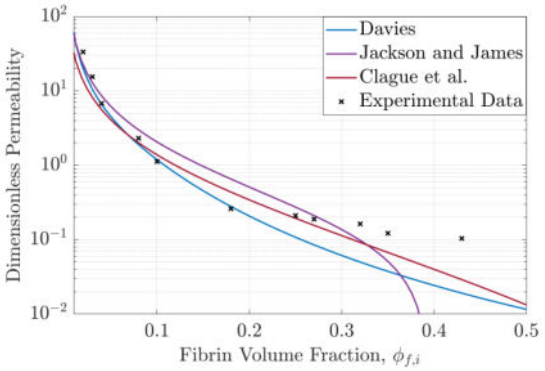


Figure 4. Log-linear plot of dimensionless fibrin permeability functions – Davies' equation (blue), Jackson and James' (purple), and Clague et al.'s (red) – and experimental data (cross) of Wufsus et al. [20-24].

To incorporate the effect of a second component on clot permeability, three additional models were considered. The first was Ethier's model, with the other two being an unweighted resistance in parallel (RiP) model and a fractional packing (FP) model [21, 29-31]. Ethier's model is given by Equation 8,

$$k_{t,i} = k_{f,i}\left(\frac{1-\phi_{j,i}}{1+\phi_{j,i}}\right) \quad (8)$$

where j refers to the type of second component considered, being either p for platelet or RBC for RBC. Ethier's model, which has been validated against experimental data for platelet-rich clots, was selected to be used in combination with Davies' equation [21]. The model effectively captures the permeability reduction in two-component porous materials with highly dissimilar radii [29]. Given that both RBC and platelet radii are significantly larger than that of the fibrin fibres modelled (0.1 µm) in both the core and shell, the equation was deemed suitable for application to both layers. Additionally, the function displays consistent shape and trends across all combinations of volume fractions of each component, and is able to provide reasonable predictions for permeability compared to the other models. For example, FP very quickly predicts impermeable clots as volume fraction is increased, as seen in Figure 5, and the clot essentially cannot be lysed.

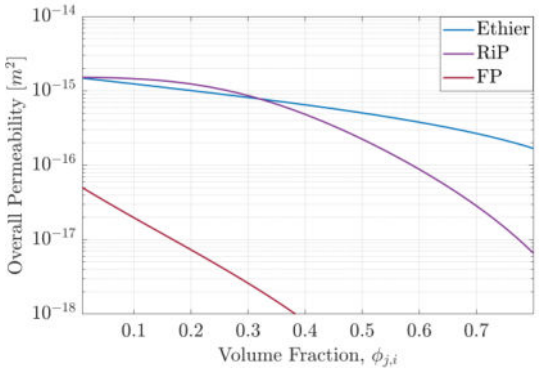


Figure 5. Log-linear plot of overall permeability functions against second component volume fraction, at constant fibrin permeability computed at a fibrin volume fraction of 0.01 using Davies' equation – Ethier's model (blue), RiP (purple), and FP (red) [21, 29-31].

The presence of a clot within the artery as well as the presence of layered architecture were modelled through separating the discretised artery into multiple segments. The differences between the shell and core segments were in their length and solids composition. Table 1 summarises the clot properties considered. These were selected based on values found in the literature [7, 8, 32, 33]. The fibrin volume fraction in the shell was purposely set towards the upper limit of reported values to highlight the consequences of a dense shell.

Table 1. Initial clot properties.

L_{clot} [mm]	L_{shell} [mm]	$\phi^0_{f,s}$	$\phi^0_{p,s}$	$\phi^0_{f,c}$	$\phi^0_{RBC,c}$
5	0.05	0.30	0.30	0.01	0.70

The dissolution of the clot was modelled through a decrease in the total number of binding sites within the clot, n_{tot}, over time. The lysis progression was quantified by a variable defined as the extent of lysis,

$$E_L(x,t) = 1 - \frac{n_{tot}}{n_{tot,0}} \quad (9)$$

for which the initial total number of binding sites, $n_{tot,0}$, was approximated using the procedure of Piebalgs et al. [9]. Since the extent of lysis is a variable in both space and time, a spatially averaged value can be calculated for each time step to allow for a more easily comparable visualisation of lysis progression with different model configurations.

$$\overline{E_L}(t) = 1 - \frac{\sum_i n_{tot}(x_i,t)}{\sum_i n_{tot,0}(x_i)} \quad (10)$$

This was heavily used throughout the study, and so a sample calculation for determining a core-averaged lysis rate is illustrated by Equation 11 to ensure clarity. Consider a core thickness and artery length according to those defined by Tables 1 and S8, with a spatial discretisation size of $\Delta x = 0.05$ mm between each cell x_i. The indices representing the core can be determined as $i = \{102,103,\ldots,199\}$, and averaging the lysis rate over these indices gives the following expression.

$$\overline{E_L}(t) = 1 - \frac{\sum_{i=102}^{199} n_{tot}(x_i,t)}{\sum_{i=102}^{199} n_{tot,0}(x_i)} \quad (11)$$

The same method applies for calculation of the front-shell, rear-shell, and clot-averaged lysis rates, but note that given the selected model conditions, the shell thickness was represented by just a single cell and thus the summation is not required for shell lysis rates.

As the clot dissolves, the fibrin composition changes over time according to the extent of lysis, such that at complete lysis, there is no fibrin left within the artery.

$$\phi_{f,i} = \phi^0_{f,i}\big(1 - E_L(x,t)\big) \quad (12)$$

When fibrin is cleaved by PLS, it makes sense that the trapped components are released and provide less obstruction to flow. Applying this concept to our model, we coined a hindrance function, which captures the extent of hindrance to flow provided by RBCs in the core. The initial RBC volume fraction was multiplied by the hindrance function to give a composition that updated over time as lysis progressed, resulting in a temporally heterogeneous apparent volume fraction that influenced subsequent permeability calculations and thus the clot resistance and lysis dynamics.

$$\phi_{RBC,c} = \phi^0_{RBC,c} H(x,t) \quad (13)$$

Inspired by the thrombolysis model of Gu et al. involving nanomedicine with mobile activated platelets, two forms of the hindrance function were investigated [34]. The first takes a hyperbolic-tangent-form analogous to their

mobility function, with parameter m adjustable to change the function shape.

$$H(x,t) = \begin{cases} \tanh\left[m\left(1-\frac{E_L}{E_{L,crit}}\right)\right], & E_L < E_{L,crit} \\ 0, & E_L \geq E_{L,crit} \end{cases} \quad (14)$$

The second function is an arbitrary function we propose, taking a sigmoid-form, with parameter k in addition to m, which enables a wider variety of function shapes to be modelled. This is represented by Equation 15.

$$H(x,t) = \begin{cases} 1 - \frac{1}{1+e^{-m\left(E_L - \frac{E_{L,crit}}{k}\right)}}, & E_L < E_{L,crit} \\ 0, & E_L \geq E_{L,crit} \end{cases} \quad (15)$$

The goal of our function was to more realistically capture the rate of change in the hindrance provided by RBCs as they are released – that is, a fast rate when the clot is most fully packed initially, which then slows down as more components escape due to reduced repulsive interactions, fewer collisions, and the likes. This logic is followed perfectly by the qualitative gradient profile of our function. Another potential benefit introduced by the sigmoid-form is that if the value of the function itself is known, the gradient can then be derived directly from the function value. As such, if RBCs were to be modelled as a flowing or reacting species involved in mass balances, numerical methods employed to solve these may be able to perform more efficiently. A comparison of the two functions is shown in Figure 6.

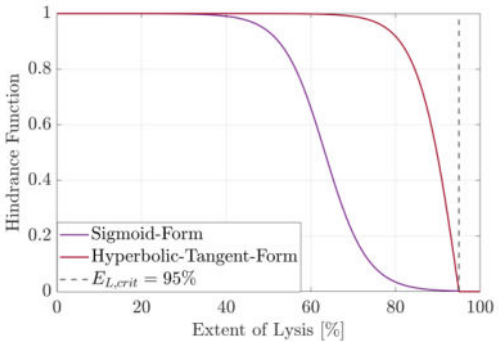


Figure 6. Shape of the hindrance function for the sigmoid-form with $m = 20$ and $k = 1.5$ (purple), and the hyperbolic-tangent-form with $m = 10$ (red), with a critical extent of lysis of 95% (grey dotted).

RESULTS

Effect of Multiple Components

The presence of an additional component reduced the overall clot permeability by a factor non-linearly related to the volume fraction of that component. The initial permeability of a homogeneous core with fibrin volume fraction of 0.01 was reduced by over 80% when an RBC volume fraction of 0.70 was considered. In a similar fashion for the shell, the permeability reduced by almost 50% when a platelet volume fraction of 0.30 was considered. These reductions and its comparative magnitudes are apparent in Figure 7.

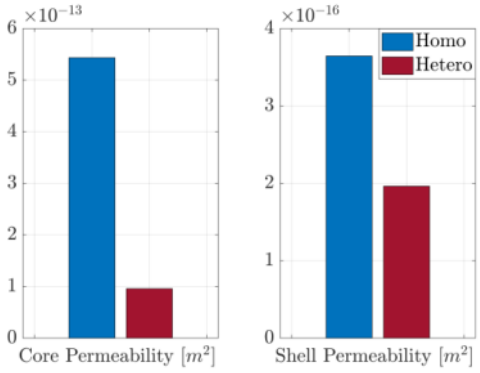


Figure 7. Initial permeability of a homogeneous (blue) pure fibrin core/shell (left/right) compared to a heterogeneous (red) core/shell (left/right) with fibrin and RBC/platelet (left/right).

To illustrate the effect of these changes in permeability, the profiles of the clot-averaged lysis rate were compared between a pure fibrin core and a core that also included RBCs. The less permeable heterogeneous core displayed a similarly shaped lysis profile, but acceleration of lysis began at a later moment in time. The time taken for the clot to dissolve was approximately 50% higher, going from just under 10 minutes to over 15 minutes. These effects are described by Figure 8.

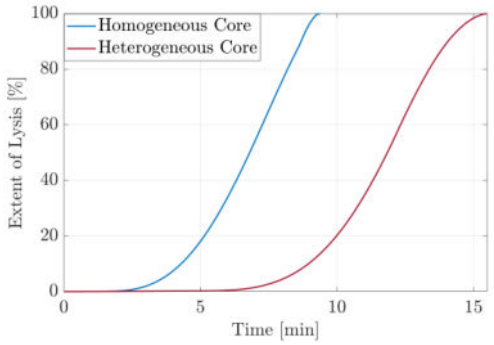


Figure 8. Clot-averaged lysis profile for a homogeneous pure fibrin core (blue) compared to a heterogeneous core with fibrin and RBC (red).

Effect of Layered Architecture

Building upon the results from multi-component consideration, the effects of a multi-component hard shell on lysis compared to a shell-less multi-component clot were explored, with various combinations in how the shell was spatially placed with respect to the core. The lysis profiles for different clot architecture configurations are shown in Figure 9.

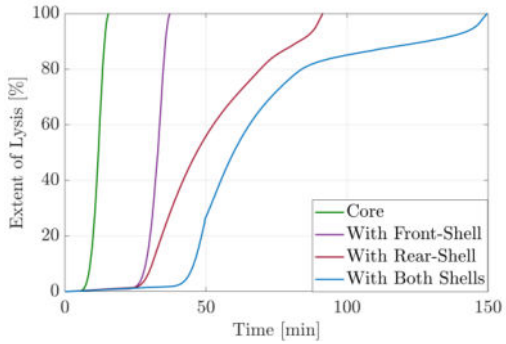


Figure 9. Clot-averaged lysis profiles for a core with fibrin and RBC (green), core with fibrin and RBC plus a front-shell with fibrin and platelet (purple), core with fibrin and RBC plus a rear-shell with fibrin and platelet (red), and a full architecture with heterogeneous compositions (blue).

The addition of a shell at the front of the core had a similar effect to the addition of a second component, whereby the onset of lysis acceleration was delayed but the shape of the curve remained unaffected. The front-shell provided significant resistance that increased the lysis time by over twice the time taken for an architecturally homogeneous clot to dissolve. When a shell was placed just behind the core, the lysis time increased almost six-fold compared to the core-only model, and over twice as long as the model with the shell placed at the front of the core. Finally, with the addition of a shell at both the front and rear ends of the core, the lysis time was approximately 150 minutes.

To gain a deeper understanding of how the presence of a shell as well as its position influences lysis dynamics, the lysis profiles within each layer were analysed. Figure 10 breaks down the clot-averaged lysis profile into sub-profiles to visualise the progression within each layer.

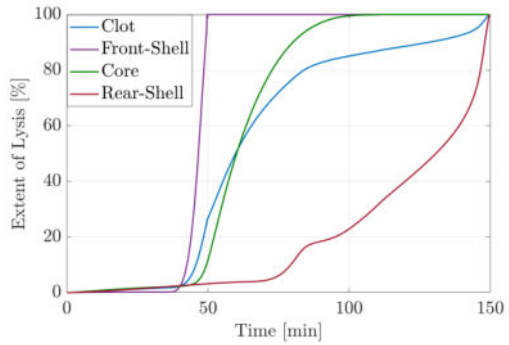


Figure 10. Lysis profiles of the front-shell (purple), averaged over the core (green), of the rear-shell (red), and averaged over the clot (blue).

The front-shell took around 50 minutes to fully dissolve before core lysis accelerated, while the rear-shell slowly dissolved and its lysis did not accelerate until the core was around 90% lysed at 80 minutes. The core fully dissolved at 100 minutes, and the rear-shell required an additional 50 minutes before the clot was fully lysed.

Effect of Dynamic Hindrance

The sigmoid-form hindrance function resulted in an RBC volume fraction profile that decreased at both an earlier onset and at a faster rate, as seen by the divergence between the two profiles in Figure 11. There was an order of magnitude reduction in RBC volume fraction within 30 minutes for the sigmoid-form, compared to over an hour for the hyperbolic-tangent-form. Both functions were able to capture an initial fast decrease before an eventual slowing down of the rate of decrease.

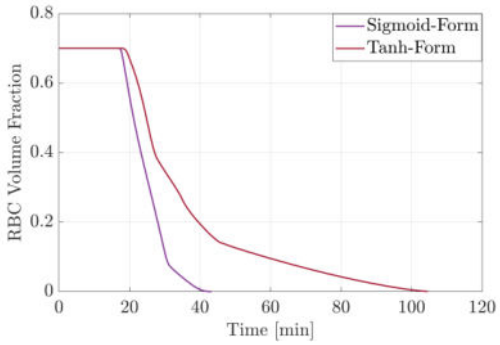


Figure 11. RBC volume fraction profile using sigmoid-form hindrance function (purple) compared to using hyperbolic-tangent-form (red).

The implications of this aspect of heterogeneity were explored by comparing the lysis profiles between a model using the sigmoid-form hindrance function and one with constant RBC volume fraction in time. The incorporation of our hindrance function predicted an over 70% lower lysis time compared to a model with temporally homogeneous RBC volume fractions. As seen in Figure 12, the lysis profile with a sigmoid-form reduction completed before 50 minutes, while the homogeneous model had not even reached a fifth of its lysis progression and continued for significantly longer. The times taken to dissolve each layer were all lower, especially the time scale for lysis of the rear-shell in comparison to other layers.

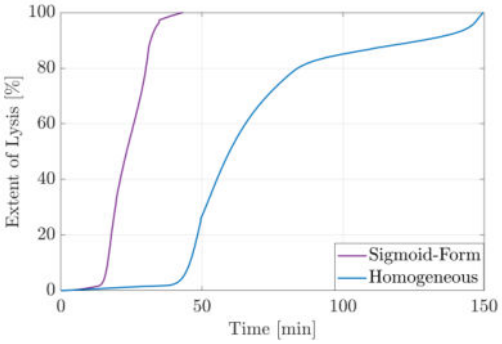


Figure 12. Clot-averaged lysis profile when the model incorporates a sigmoid-form hindrance function (purple) compared to one with a temporally homogeneous RBC volume fraction (blue).

Finally, Figure 13 captures the full effect on predicted lysis time of going from a fully homogeneous model to our

final heterogeneous model, whereby there was a four-fold increase from under 10 minutes to over 40 minutes.

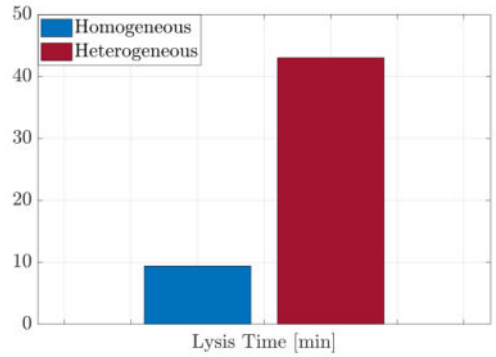


Figure 13. Comparison of predicted lysis time between the initial fully homogeneous model (blue) and the final model incorporating all of the discussed aspects of heterogeneity (red).

DISCUSSION

The results obtained demonstrate the profound impact of heterogeneity, especially multi-component heterogeneity and the role of the shell on clot lysis dynamics. The observed trends align with logical expectations and challenge the predictions of simplified homogeneous models.

Effect of Multiple Components

By observing the order of magnitude of each sub-plot within Figure 7, it is evident that fibrin composition is the primary determinant of clot permeability. This claim is supported by Figure 4, where both the computed and experimental permeability show a sharp order of magnitude decrease with an increase in fibrin volume fraction from 0.01 to just 0.03. Beyond the 0.10 to 0.20 range of volume fractions, the rate of reduction slows down. In comparison, achieving an order of magnitude reduction in overall permeability due to the second cellular component requires a volume fraction addition to fibrin of over 0.80. With these insights, the non-linear influence of each component on clot permeability is clear. This is as expected both in reality and mathematically, since each component varies in structure such as shape, size, and roughness. This means they should influence clot resistance to different extents and necessitates distinct mathematical formulations, which aligns with the approach of this study.

The introduction of additional components such as RBCs reduced clot permeability. This finding is consistent with several studies where component heterogeneity has been shown to hinder the efficacy of fibrinolytic proteins [35-37]. RBCs introduce heterogeneity within the fibrin network and significantly affect clot structure and viscoelastic properties in a concentration-dependent manner [35]. They alter fibrin architecture through interactions with FBG, resulting in thinner fibres which limit tPA penetration and reduce PLG activation at the binding sites, increasing lysis time by up to 22.7% for RBC concentrations of 40% (v/v) and less [36]. This logic physiologically explains the pronounced effect of incorporating our hindrance function, which is discussed in a later section. Even though our model does not mechanistically capture the discussed RBC-induced chemical and structural effects, Figure 8 indicates that our model was able to reflect the same outcome of increasing lytic resistance to a comparable extent. Similarly for the platelet-rich shell, the model predictions agree with claims in literature of hindered permeability due to platelet-driven mechanisms [37, 38]. Fibrin density is enhanced and platelet-derived phospholipids introduce diffusional barriers which limit penetration through the clot [37].

Effect of Layered Architecture

The addition of a multi-component shell introduced further resistance to fibrinolysis, particularly when shells were positioned in an uneasily accessible way such as behind the core. As seen in the comparison illustrated by Figure 9, the six-fold increase in lysis time for the rear-shell model compared to the core-only model highlights the importance of having direct access to the shell layer. The delayed acceleration of rear-shell lysis, observed only once core lysis is near complete, implies that fibrinolytic proteins must first penetrate the core and fully degrade its structure before they can exert their full force on the rear-shell. In comparison, the front-shell delayed the time taken for lysis rate to escalate but did not alter the overall lysis dynamics to the same extent as the rear-shell. This locational asymmetry in shell effects could be linked to the difference in concentration profiles of species at each location. In more realistic clot models, there exists a preferred path of blood flow due to clot and arterial geometry, whereby more accessible regions are impacted first by incoming fibrinolytic agents [9]. Thus, it is reasonable to suggest that the proximity of the front-shell to the artery entrance allows for the development of more favourable concentration profiles of active proteins due to its greater extent of exposure.

Taking a look at the Darcy velocity also provides additional insights into how variations in porosity within each layer influence transport dynamics. While the velocity is proportional to overall clot permeability, which decreases, albeit non-linearly, with higher solids content, it is also to some extent inversely dependent on voidage due to the nature of flow through porous and tortuous paths [19]. As voidage decreases, there is some velocity-increasing effects as suggested by the division by voidage in Equation 5. However, the effect of decreasing permeability ultimately dominates, leading to slower transport. This is important since the rate of thrombolysis does depend on the ability for tPA and associated fibrinolytic proteins to reach clot binding sites [20]. Considering the balance between transport and chemical reaction rates, the lower the porosity, the more this process is shifted towards diffusion-control, which decreases the likelihood of complete recanalisation [39, 40].

In addition, the complex influences of porosity are further complicated by how reaction rates are affected. Lower porosity hinders the kinetics of bound phase

plasminogen activation, but simultaneously increases the fibrin-binding efficiency of tPA, PLS, and PLG. This is more easily visualised with the rate equation for reaction B4 shown in Table S3 and the reaction source terms as seen in Equations S12-14. Fully understanding the interplay between transport- and reaction-influences of porosity could further explain the observed effects of heterogeneous solids content on lysis progression.

Effect of Dynamic Hindrance

The use of a sigmoid-form hindrance function revealed a fast and early reduction in RBC content within the core, which directly influenced lysis progression. The observed order-of-magnitude reduction in RBC volume fraction within 30 minutes in Figure 11 suggests that our hindrance function is able to represent a significant redistribution of RBCs during lysis. This was one of the objectives in introducing our hindrance function.

Furthermore, the comparison between the results using temporally heterogeneous and homogeneous RBC volume fractions in Figure 12 highlights the huge effect that considerations of time-dependent compositional changes have on clot models. Note that the kinks observed in some lysis profiles are due to the nature of spatially averaging diverse values across the heterogeneous clot. Regardless of whether the model is 100% accurate, it does represent AIS thrombi more realistically than homogeneous models by adding the structural and compositional complexities reported in literature. Thus, the large four-fold increase in lysis time, as illustrated in Figure 13, and deviations in lysis dynamics when transitioning from a fully homogeneous model to our final heterogeneous model underscores the significance of accounting for realistic clot structures. The final predicted time of 43 minutes does come to agreement with a study which found that complete or partial recanalisation occurred at an average of 45±20 minutes following the bolus delivery of tPA [41]. Another congruent reported value was 42±23 minutes for the same dosing regimen simulated [42]. Compared to our model, the homogeneous model heavily underestimates the time required for complete lysis, which may be a critical error since underprediction is more problematic than overprediction in terms of treatment planning.

Reflections

The initial objectives of understanding the effects of clot composition on thrombolysis were well met. The study went beyond initial expectations with the development of a model that captures heterogeneous core-shell structure and apparent cellular mobility. These outcomes allowed for direct comparisons between in vivo-derived clots, existing homogeneous models, and our developed heterogeneous model, revealing key effects of heterogeneity in lysis outcomes as discussed. This prompts further refinement of computational models to more accurately simulate realistic clot conditions. Relaxing existing assumptions would enable greater applicability of simulation insights to the design of therapeutic treatment.

Given the discussed consequences of simplifications, a critique of our model limitations must consider the major assumption of constant pressure drop across the clot in Equation 5. This overlooks realistic variations in arterial pressure that impact transport phenomena during thrombolysis as highlighted by [9]. While spatial distribution of the specified compositions was accounted for, real clots could have a core with regions of denser fibrin structure irregularly dispersed between loose networks. However, this is hard to capture since these irregularities itself can be patient-specific, and thus the intricacies in certain model predictions would not be generalisable to the average population. Due to the computational nature of this study, the absence of result stochasticity prevented any significant quantification of inherent error. Incorporating real-time imaging data would help validate the model, particularly the hindrance function, by examining RBC volume fraction profiles throughout lysis. Furthermore, the model could be enhanced by exploring more realistic geometries and extending its framework into higher-order dimensions. It is important to highlight that this study did not address safety concerns, such as elevated risk of intracerebral haemorrhage due to depleted FBG levels. The primary focus was on exploring the effects of clot heterogeneity under consistent treatment conditions, rather the implications of varied treatment. Future studies should incorporate safety considerations to fully understand the risks associated with thrombolytic therapy, which may be exacerbated by clot heterogeneity.

CONCLUSIONS

This study highlights the critical influence of clot heterogeneity on predicted thrombolysis outcomes, emphasising the limitations of simplified models in representing true lysis dynamics. The introduction of spatial and temporal heterogeneities, including multi-component compositions and layered core-shell architecture, revealed increased resistance to fibrinolysis and four-fold longer lysis times compared to predictions of homogeneous models. These insights suggest potential implications for thrombolytic therapy. The significantly extended lysis times associated with heterogeneous clots may require higher doses or more tactical thrombolytic strategies such as combination therapy to improve safety, efficacy, and ultimately recanalisation success rates in AIS scenarios [43]. Further computational research should aim to continue to add complexities to existing simplifications, such as those discussed in our reflections.

REFERENCES

1. Poomalai, G., Prabhakar, S. and Sirala Jagadesh, N. (2023). Functional Ability and Health Problems of Stroke Survivors: An Explorative Study. *Cureus*, [online] 15(1). doi: https://doi.org/10.7759/cureus.33375.
2. Saini, V., Guada, L. and Yavagal, D.R. (2021). Global Epidemiology of Stroke and Access to Acute Ischemic

Stroke Interventions. *Neurology*, 97(20 Supplement 2), pp.S6–S16. doi: https://doi.org/10.1212/WNL.0000000000012781.
3. Jolugbo, P. and Ariëns, R.A.S. (2021). Thrombus Composition and Efficacy of Thrombolysis and Thrombectomy in Acute Ischemic Stroke. *Stroke*, 52(3). doi: https://doi.org/10.1161/strokeaha.120.032810.
4. Staessens, S., Denorme, F., Francois, O., Desender, L., Dewaele, T., Vanacker, P., Deckmyn, H., Vanhoorelbeke, K., Andersson, T. and Meyer, S.F.D. (2020). Structural analysis of ischemic stroke thrombi: histological indications for therapy resistance. *Haematologica*, [online] 105(2), pp.498–507. doi: https://doi.org/10.3324/haematol.2019.219881.
5. Marder, V.J., Chute, D.J., Starkman, S., Abolian, A.M., Kidwell, C., Liebeskind, D., Ovbiagele, B., Vinuela, F., Duckwiler, G., Jahan, R., Vespa, P.M., Selco, S., Rajajee, V., Kim, D., Sanossian, N. and Saver, J.L. (2006). Analysis of Thrombi Retrieved From Cerebral Arteries of Patients With Acute Ischemic Stroke. *Stroke*, 37(8), pp.2086–2093. doi: https://doi.org/10.1161/01.str.0000230307.03438.94.
6. Dumitriu LaGrange, D., Reymond, P., Brina, O., Zboray, R., Neels, A., Wanke, I. and Lövblad, K.-O. (2023). Spatial heterogeneity of occlusive thrombus in acute ischemic stroke: A systematic review. *Journal of Neuroradiology*, [online] 50(3), pp.352–360. doi: https://doi.org/10.1016/j.neurad.2023.01.004.
7. Di Meglio, L., Desilles, J.-P., Ollivier, V., Nomenjanahary, M.S., Di Meglio, S., Deschildre, C., Loyau, S., Olivot, J.-M., Blanc, R., Piotin, M., Bouton, M.-C., Michel, J.-B., Jandrot-Perrus, M., Ho-Tin-Noé, B. and Mazighi, M. (2019). Acute ischemic stroke thrombi have an outer shell that impairs fibrinolysis. *Neurology*, [online] 93(18), pp.e1686–e1698. doi: https://doi.org/10.1212/wnl.0000000000008395.
8. Oana Madalina Mereuta, Agarwal, T., Sherief Ghozy, Dai, D., Santhosh Arul, Waleed Brinjikji, Kallmes, D.F. and Ram Kadirvel (2023). Shell Versus Core Architecture and Biology of Thrombi in Acute Ischemic Stroke: A Systematic Review. *Clinical and Applied Thrombosis/Hemostasis*, [online] 29. doi: https://doi.org/10.1177/10760296231213632.
9. Piebalgs, A., Gu, B., Roi, D., Lobotesis, K., Thom, S. and Xu, X.Y. (2018). Computational Simulations of Thrombolytic Therapy in Acute Ischaemic Stroke. *Scientific Reports*, [online] 8(1), p.15810. doi: https://doi.org/10.1038/s41598-018-34082-7.
10. Gu, B., Andris Piebalgs, Huang, Y., Longstaff, C., Hughes, A.D., Chen, R., McG, S.A. and Zheng, S. (2019). Mathematical Modelling of Intravenous Thrombolysis in Acute Ischaemic stroke: Effects of Dose Regimens on Levels of Fibrinolytic Proteins and Clot Lysis Time. 11(3), pp.111–111. doi: https://doi.org/10.3390/pharmaceutics11030111.
11. Dow, N., Coleman, J.R., Moore, H., Osborn, Z.T., Sackheim, A.M., Hennig, G., Saulius Butenas, Nelson, M.T., Moore, E.E. and Freeman, K. (2019). Dense and dangerous: The tissue plasminogen activator-resistant fibrinolysis shutdown phenotype is due to abnormal fibrin polymerization. *Journal of Trauma and Acute Care Surgery*, 88(2), pp.258–265. doi: https://doi.org/10.1097/ta.0000000000002554.
12. Zou, H., Banerjee, P., Leung, S.S.Y. and Yan, X. (2020). Application of Pharmacokinetic-Pharmacodynamic Modeling in Drug Delivery: Development and Challenges. *Frontiers in Pharmacology*, 11. doi: https://doi.org/10.3389/fphar.2020.00997.
13. Ahmed, T.A. (2015). Pharmacokinetics of Drugs Following IV Bolus, IV Infusion, and Oral Administration. *Basic Pharmacokinetic Concepts and Some Clinical Applications*. [online] doi: https://doi.org/10.5772/61573.
14. The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group (1995). Tissue Plasminogen Activator for Acute Ischemic Stroke. *New England Journal of Medicine*, 333(24), pp.1581–1588. doi: https://doi.org/10.1056/nejm199512143332401.
15. Tebbe, U., Tanswell, P., Seifried, E., Feuerer, W., Scholz, K.-H. and Herrmann, K.S. (1989). Single-bolus injection of recombinant tissue-type plasminogen activator in acute myocardial infarction. *The American Journal of Cardiology*, 64(8), pp.448–453. doi: https://doi.org/10.1016/0002-9149(89)90419-0.
16. Murray, V., Norrving, B., Sandercock, P.A.G., Terént, A., Wardlaw, J.M. and Wester, P. (2010). The molecular basis of thrombolysis and its clinical application in stroke. *Journal of Internal Medicine*, 267(2), pp.191–208. doi: https://doi.org/10.1111/j.1365-2796.2009.02205.x.
17. Mathworks.com. (2014). ODE with Single Solution Component. [online] Available at: https://uk.mathworks.com/help/matlab/ref/ode23s.html [Accessed 8 Dec. 2024].
18. Ernst Hairer and Wanner, G. (1991). *Solving Ordinary Differential Equations II*. doi: https://doi.org/10.1007/978-3-662-09947-6.
19. Basic theory. (2020). *Fluid Flow in Porous Media*, pp.47–67. doi: https://doi.org/10.1142/9789811219535_0002.
20. Diamond, S.L. (1999). Engineering Design of Optimal Strategies for Blood Clot Dissolution. *Annual Review of Biomedical Engineering*, 1(1), pp.427–461. doi: https://doi.org/10.1146/annurev.bioeng.1.1.427.
21. Wufsus, A.R., Macera, N.E. and Neeves, K.B. (2013). The Hydraulic Permeability of Blood Clots as a Function of Fibrin and Platelet Density. *Biophysical Journal*, 104(8), pp.1812–1823. doi: https://doi.org/10.1016/j.bpj.2013.02.055.
22. Davies, C.N. (1953). The Separation of Airborne Dust and Particles. *Proceedings/Proceedings - Institution of Mechanical Engineers*, 167(1b), pp.185–213. doi: https://doi.org/10.1177/002034835316701b13.

23. Jackson, G. and James, D.E. (1986). The permeability of fibrous porous media. *Canadian Journal of Chemical Engineering*, 64(3), pp.364–374. doi: https://doi.org/10.1002/cjce.5450640302.
24. Clague, D.S., Kandhai, B.D., Zhang, R. and P. M. A. Sloot (2000). Hydraulic permeability of (un)bounded fibrous media using the lattice Boltzmann method. *Physical review. E, Statistical physics, plasmas, fluids, and related interdisciplinary topics*, 61(1), pp.616–625. doi: https://doi.org/10.1103/physreve.61.616.
25. Diamond, S.L. and Anand, S. (1993). Inner clot diffusion and permeation during fibrinolysis. *Biophysical Journal*, 65(6), pp.2622–2643. doi: https://doi.org/10.1016/s0006-3495(93)81314-6.
26. Govindarajan, V., Rakesh, V., Reifman, J. and Mitrophanov, Alexander Y. (2016). Computational Study of Thrombus Formation and Clotting Factor Effects under Venous Flow Conditions. *Biophysical Journal*, 110(8), pp.1869–1885. doi: https://doi.org/10.1016/j.bpj.2016.03.010.
27. Xu, Z., Lioi, J., Mu, J., Kamocka, M.M., Liu, X., Chen, D.Z., Rosen, E.D. and Alber, M. (2010). A Multiscale Model of Venous Thrombus Formation with Surface-Mediated Control of Blood Coagulation Cascade. *Biophysical Journal*, 98(9), pp.1723–1732. doi: https://doi.org/10.1016/j.bpj.2009.12.4331.
28. Yang, Y., Gu, B. and Xiao Yun Xu (2023). In Silico Study of Different Thrombolytic Agents for Fibrinolysis in Acute Ischemic Stroke. *Pharmaceutics*, [online] 15(3), pp.797–797. doi: https://doi.org/10.3390/pharmaceutics15030797.
29. Ethier, C.R. (1991). Flow through mixed fibrous porous materials. *AIChE Journal*, 37(8), pp.1227–1236. doi: https://doi.org/10.1002/aic.690370812.
30. Mattern, K.J. and Deen, W.M. (2007). 'Mixing rules' for estimating the hydraulic permeability of fiber mixtures. *AIChE journal*, 54(1), pp.32–41. doi: https://doi.org/10.1002/aic.11350.
31. Koltermann, C.E. and Gorelick, S.M. (1995). Fractional packing model for hydraulic conductivity derived from sediment mixtures. *Water Resources Research*, 31(12), pp.3283–3297. doi: https://doi.org/10.1029/95wr02020.
32. Mereuta, O.M., Rossi, R., Douglas, A., Gil, S.M., Fitzgerald, S., Pandit, A., McCarthy, R., Gilvarry, M., Ceder, E., Dunker, D., Nordanstig, A., Redfors, P., Jood, K., Magoufis, G., Psychogios, K., Tsivgoulis, G., O'Hare, A., Power, S., Brennan, P. and Nagy, A. (2021). Characterization of the 'White' Appearing Clots that Cause Acute Ischemic Stroke. *Journal of Stroke and Cerebrovascular Diseases*, [online] 30(12), p.106127. doi: https://doi.org/10.1016/j.jstrokecerebrovasdis.2021.106127.
33. Khismatullin, R.R., Nagaswami, C., Shakirova, A.Z., Vrtková, A., Procházka, V., Gumulec, J., Mačák, J., Litvinov, R.I. and Weisel, J.W. (2020). Quantitative Morphology of Cerebral Thrombi Related to Intravital Contraction and Clinical Features of Ischemic Stroke. *Stroke*, 51(12), pp.3640–3650. doi: https://doi.org/10.1161/strokeaha.120.031559.
34. Gu, B., Huang, Y., Manchester, E.L., Hughes, A.D., McG, S.A., Chen, R. and Xu, X.Y. (2022). Multiphysics Modelling and Simulation of Thrombolysis via Activated Platelet-Targeted Nanomedicine. *Pharmaceutical Research*, 39(1), pp.41–56. doi: https://doi.org/10.1007/s11095-021-03161-2.
35. Gersh, K., Nagaswami, C. and Weisel, J. (2009). Fibrin network structure and clot mechanical properties are altered by incorporation of erythrocytes. *Thrombosis and Haemostasis*, 102(12), pp.1169–1175. doi: https://doi.org/10.1160/th09-03-0199.
36. Wohner, N., Péter Sótonyi, Raymund Machovich, László Szabó, Kiril Tenekedjiev, Silva, M.M.C.G., Longstaff, C. and Kolev, K. (2011). Lytic Resistance of Fibrin Containing Red Blood Cells. *Arteriosclerosis Thrombosis and Vascular Biology*, 31(10), pp.2306–2313. doi: https://doi.org/10.1161/atvbaha.111.229088.
37. Wohner, N. (2008). Role of Cellular Elements in Thrombus Formation and Dissolution. *Cardiovascular & Hematological Agents in Medicinal Chemistry*, 6(3), pp.224–228. doi: https://doi.org/10.2174/187152508784871972.
38. Collet, J.Ph., Montalescot, G., Lesty, C. and Weisel, J.W. (2002). A Structural and Dynamic Investigation of the Facilitating Effect of Glycoprotein IIb/IIIa Inhibitors in Dissolving Platelet-Rich Clots. *Circulation Research*, 90(4), pp.428–434. doi: https://doi.org/10.1161/hh0402.105095.
39. Disharoon, D., Trewyn, B.G., Herson, P.S., David and Neeves, K.B. (2021). Breaking the fibrinolytic speed limit with microwheel co-delivery of tissue plasminogen activator and plasminogen. 20(2), pp.486–497. doi: https://doi.org/10.1111/jth.15617.
40. Labiche, L.A., Malkoff, M. and Alexandrov, A.V. (2003). Residual Flow Signals Predict Complete Recanalisation in Stroke Patients Treated With TPA. *Journal of Neuroimaging*, 13(1), pp.28–33. doi: https://doi.org/10.1111/j.1552-6569.2003.tb00153.x.
41. Blacker, D. (2000). High Rate of Complete Recanalisation and Dramatic Clinical Recovery During tPA Infusion When Continuously Monitored With 2-MHz Transcranial Doppler Monitoring. 31(3), pp.610–614. doi: https://doi.org/10.1161/01.str.31.3.610.
42. Alexandrov, A.V., Burgin, W.S., Demchuk, A.M., El-Mitwalli, A. and Grotta, J.C. (2001). Speed of Intracranial Clot Lysis With Intravenous Tissue Plasminogen Activator Therapy. *Circulation*, 103(24), pp.2897–2902. doi: https://doi.org/10.1161/01.cir.103.24.2897.
43. Yang, Y., Gu, B. and Xu, X.Y. (2024). In silico study of combination thrombolytic therapy with alteplase and mutant pro-urokinase for fibrinolysis in ischemic stroke. *Computers in Biology and Medicine*, 171, pp.108141–108141. doi: https://doi.org/10.1016/j.compbiomed.2024.108141.

Optimisation of Feed Conditions of Catalysed Inverse Vulcanisation with Rapeseed Oil and Waste Cooking Oil

Xin Chen, Mariam Abdalla
Department of Chemical Engineering, Imperial College London, U.K.

Abstract

Sulphur is a byproduct of the oil and gas industry; storage of excess sulphur is a challenge. There are also large volumes of waste cooking oil, whose irresponsible disposal causes environmental damage, that could be used with sulphur to produce inverse vulcanisation polymers. Catalytic inverse vulcanisation, developed in 2019, was not fully explored with vegetable oils and there was limited investigation into waste cooking oil as a crosslinker. This paper aims fill this gap by making efforts to optimise the reaction feed conditions as well as compare the catalyst preferences and resulting polymer properties. The waste cooking oil will be compared to virgin rapeseed oil by ^{1}H-NMR, FTIR, and acid value. DOE was carried to scan the experimental space. Polysulphide products were characterised via FTIR, Raman, XRF, TGA and DSC. The oils were found to have differing free fatty acid contents and degree of unsaturation. Despite the impurities present in WCO, its performance was similar to the virgin oil. It was concluded that $ZnDEDC_2$ was a more suitable catalyst, displaying more consistent rate enhancement with increase catalyst loading improving rate. Increased sulphur to oil ratio produced more thermally stable polymers.

Keywords: Polymerisation, Sulphur, Waste Cooking Oil, Inverse Vulcanisation

Introduction

Sulphur, yellow solid in its elemental form, is a byproduct of desulphurisation in petroleum refineries and natural gas processing. It is a cheap and abundant resource. Worldwide production of Sulphur, in all its forms, had reached 82 million tons in 2023 (US G.S., 2024) and is expected to increase in the following decade due to further oil and gas exploration. There is an excess of production compared to demand, including as sulphuric acid and other forms, leading to challenges in storage of excess sulphur. To mitigate this excess; there is a need for alternative uses of sulphur, seen in sulphur concrete and polymer production.

Global consumption of vegetable oils in 2022 as food was 389 million tonnes according to the FAO of the United Nations 2024. After vegetable oils (VO) are used in cooking and frying, the once yellow liquid becomes darker and more viscous. Waste cooking oil (WCO) can have serious health implications, such as increased risks of liver disease and cancer, if reused. When discarded down the drain, WCO can build up in sewer pipes resulting in formation of fatbergs, for example the Whitechapel fatberg (BBC, 2017), and contamination of water sources.

Polysulphides can be produced from sulphur and organic crosslinkers through the process of inverse vulcanisation. These sulphur-rich polymers have many applications including as cathodes in Li-S batteries (Hoefling, A. et al., 2016), sorbents of Pd and Hg from water (Worthington, M.J.H. et al., 2017), controlled release coatings (Feng, G.D., et al., 2024) and, as films and adhesives.

This paper sets out to investigate optimal feed conditions for the catalytic inverse vulcanisation of WCO and how it compares with virgin vegetable oil (VO). The feed conditions investigated were catalysts mass loading and the sulphur to oil (S:oil) ratio. WCO and VO were compared as comonomers in catalytic inverse vulcanisation along with the properties of the resulting polysulphide and catalyst preferences. The virgin vegetable oil used in this study was rapeseed oil. It was initially hypothesised that the VO would perform better due to the potential of side reactions, caused by the presence of impurities in the WCO formed during the cooking process.

Background

The polymerisation of vegetable oils has been a growing area of research due to rising environmental concerns and a desire for more sustainable materials. Vegetable oils are good candidates as polymer feedstock as they are abundant, renewable, cheap, and do not produce volatile organic compounds. VOs consist of a mixture of triglycerides and may have a range of different fatty acids containing polymerisable functional groups. Substitution of VOs with WCO as polymer feedstock eliminates potential issues of food supply shortages caused by

diverting edible vegetable oils to polymer production.

When heated, sulphur rings can undergo ring opening polymerisation; but as this reaction is reversible, these polysulphides are unstable. The idea of producing sulphur-based polymers via direct copolymerisation with unsaturated organic comonomers was first developed by Chung et al. in 2011. This process was later named "inverse vulcanisation" due to the roles of sulphur and the organic comonomer in classical vulcanisation being reversed, resulting in polymers with high sulphur content. Hang S. et al. (2024) describes the development of inverse vulcanisation and its applications in the last decade.

This process has a problem with the production of hydrogen sulphide (H_2S) gas when sulphur is heated in the presence of organic molecules. H_2S is a colourless, highly toxic gas with an odour of rotten egg at concentrations above 0.0047 ppm (HSE, 2009). In the atmosphere, H_2S converts to sulphates leading to sulphurous smog, acid rain, and damage to vegetation.

Lower temperatures, down to 100 °C, were achieved by the development of catalytic inverse vulcanisation in 2019 by Wu et al. with catalysts containing the diethyldithiocarbamate ligands displaying better rate enhancement. The catalysts contain sulphurphilic and oleophilic ends on the DTC ligand. Wu et al. proposed a mechanism as follows: first, the catalyst weakens a S—S bond in a S-ring then the sulphurphilic end cleaves the bond and carries it into the organic phase, where the sulphur chain the catalyst was carrying is added to a C=C double bond attracted by the electrophilic metal ion.

Dodd et al. (2021) found that copper (II) diethyldithiocarbamate has the best rate enhancement ability. However, when factors like cost and processing practicality were considered, Zinc (II) diethyldithiocarbamate ($ZnDEDC_2$) was deemed to be the recommended candidate for industrial purposes hence it was chosen as one of the catalysts in this study. Sodium dimethyldithiocarbamate dihydrate (NaDMDC) was chosen as the other catalyst to be investigated. The rate enhancement ability of this catalyst may be limited as Na^+ is less electronegative than Zn^{2+} and the catalyst only contains one DTC ligand. However, as NaDMDC has shorter carbon chains on the ligand and is smaller than $ZnDEDC_2$, which could possibly be an advantage during the transport step of the reaction mechanism.

Methods

Materials

Sodium dimethyldithiocarbamate dihydrate ($C_3H_6NNaS_2{\cdot}2H_2O$, 97 wt.%, Tokyo Chemical Industry, CAS 128-04-01), Zinc diethyldithiocarbamate ($C_{10}H_{20}N_2S_4Zn$, 99 wt.%, Tokyo Chemical Industry, CAS 14324-55-1), rapeseed oil (CAS 8002-13-09, Oillia Ltd.) and sulphur (S_8, reagent grade, Sigma Aldrich, CAS 7704-34-9) were commercially available and used as received. The waste cooking oil (WCO) was sourced from a restaurant.

Analysis of oils

To compare free fatty acid concentration and acid values, a titration was carried out with 25 ml of rapeseed oil and 2 ml of waste cooking oil, diluted with 23 ml of ethanol, with a 0.1 M standard solution of potassium hydroxide (KOH) and phenolphthalein indicator. Colour change was easily observed in the titration of rapeseed oil. Whereas the WCO formed an opaque emulsion upon stirring, requiring dilution to accurately observe the endpoint. Titration was repeated until concordant results were obtained.

Acid value is the amount of KOH, in milligrams, required to neutralise 1 gram of oil, calculated using equation 1.

$$AV = \frac{56.1056(C_{KOH}V_{KOH})}{1000\rho_{oil}V_{oil}} \quad (1)$$

Where C_{KOH} is the concentration of KOH in mol/dm^3 and V_{KOH} is the volume of KOH required to reach the endpoint in L. 56.1056 g/mol is the molecular mass of KOH. ρ_{oil} is the density of the oil calculated by weighting the mass of 5 ml of oil. V_{oil} is the volume of oil.

^{1}H-NMR spectra of the oils were obtained with a JEOL JNM-ECZ400S/L1 spectrometer, with a field strength of 400 Hz and pulse of 3.225 us, to compare the degree of unsaturation. Sample solutions were prepared with 20 mg of oil in 0.7 ml of deuterated chloroform ($DCCl_3$). NMR data was analysed with the mNova software. The integral of the olefinic hydrogen peak and was normalised by the integral of the glyceride C2 methine hydrogen peak. This is because there is only one hydrogen in this position per triglyceride molecule. Equation 2 was used to calculate the degree of unsaturation.

$$DOU = \frac{I_o}{2I_m} \quad (2)$$

$$DOU = \frac{6.68}{2 \times 1} = 3.34 \quad (3)$$

Where DOU is the degree of unsaturation, I_o is the integral of the olefinic hydrogen peak, and I_m is the integral of the methine hydrogen at the C2 position

of the glyceride. A sample calculation is available as equation 3, showing the calculation process for the degree of unsaturation of waste cooking oil.

Design of experiment

Design of experiment is a statistical tool used to plan experiments to analyse the effects of multiple factors simultaneously on a response variable; vitrification time was chosen to compare the rate of reaction. Vitrification time was defined as the time elapsed between addition of oil to the reaction mixture and the magnetic stirrer stopping due to the viscosity of the mixture. A 3 factor study with 2 levels was carried out to screen the reaction space. Each input factor (temperature, catalyst loading and sulphur to oil ratio) was assigned a high and low value shown in table 1.

Table 1 Experimental conditions used to create design of experiment matrix

Oil type		VO	WCO
Catalyst type		$ZnDEDC_2$	NaDMDC
Factor		Level 1 (–)	Level 2 (+)
1	Catalyst loading (g)	0.01	0.05
2	Temperature (°C)	140	160
3	S : oil ratio (g/g)	1	4

Testing all combinations of factor values resulted in 8 groups of unique experimental conditions shown in table 2. This was repeated for all 4 oil-catalyst pairings, culminating into 32 total experiments covered in this paper.

Table 2 Combination of factors for the experiment groups

Experiment group	Factor 1	Factor 2	Factor 3
1	–	–	–
2	–	–	+
3	–	+	–
4	–	+	+
5	+	–	–
6	+	–	+
7	+	+	–
8	+	+	+

Preparation of polysulphides

The sulphur and catalyst were weighed and loaded into 22 mm diameter glass vials with magnetic stirrers. Three vials were setup as replicates for each set of reaction conditions. These were placed in an aluminium block which was heated via a hotplate and monitored by a temperature probe. Once the temperature set point was reached, the mixture was allowed to stand for 5 minutes to allow the sulphur to become fully molten. The vegetable oil was then added into the mixture using a syringe and stirred at 250 rpm. The total reaction mixture mass excluding the catalyst was 4g. The reaction progressed for 2 h, monitored by a camera positioned directly over the setup, the videos were analysed to calculate the vitrification time. The products were then cured at 110 °C in the hotplate for 12 h.

Unreacted sulphur and oil were removed from the sample via Soxhlet extraction with chloroform as the solvent for 26 h. The samples were then dried in a vacuum oven for 50 h, weighed and stored for analysis.

Samples were named to identify the conditions of its production following the code in table 3, with the characters in the first two positions representing the oil and catalyst types. Separated by a dash, the next three numbers represent the reaction temperature. After the next dash, the sulphur to oil ratio and catalyst loading are represented.

Table 3 Sample code name explanation

Position	Code	Meaning
1	R	Rapeseed oil (VO)
	W	Waste cooking oil (WCO)
2	Z	$ZnDEDC_2$ catalyst
	N	NaDMDC catalyst
3	140	Reaction temperature = 140 °C
	160	Reaction temperature = 160 °C
4	1	S:oil = 1
	4	S:oil = 4
5	01	0.01 g catalyst loading
	05	0.05 g catalyst loading

Characterisation of Polysulphides

Fourier transformed infrared spectroscopy (FTIR) of reactants and polymer samples were conducted with an Agilent Technologies Cary 630 FTIR to compare the bonding present. This was conducted in a scan range of 4000 - 650 cm^{-1} and a resolution of 2 cm^{-1}.

A PANalytical Epsilon3XLE benchtop x-ray fluorescence (XRF) instrument was used to estimate sulphur content in the polysulphides. The samples were milled into fine powder, loaded into x-ray sample holders and then pressed to create a flat surface for consistent thickness (Ichikawa and Nakamura, 2016). Oxygen, carbon and hydrogen are unanalysable by the XRF instrument due to low atomic number however installed software (Omnian) can predict the content of analysable elements such as sulphur by correcting for the effect

of matrix (the oil crosslinker) composition (Rousseau, 2006).

Raman spectroscopy was carried out using a Senterra II by Bruker with a 522 nm laser excitation and 6.25 mW power. The aperture measured 50 × 1000 µm, the exposure time was 500 ms and the resolution was 4 cm^{-1}. The samples were placed onto a microscope slide and labelled, fixed into place using adhesives to prevent shifting during the measurement. The slide was placed inside the machine and the focal point was adjusted to the relevant sample using a joystick for precise control to find sufficiently flat regions for analysis

Thermogravimetric analysis (TGA) was conducted with PerkinElmer TGA 8000, under a nitrogen atmosphere with heating rate of 20 °C/min and temperature range 30 – 850 °C, to compare the thermal stability of the polymers and observe their thermal degradation.

Differential scanning calorimetry (DSC) was carried out with PerkinElmer DSC 8000 and a heating and cooling rate of 20 °C/min and a nitrogen purge rate of 40 ml/min. A heat-cool-heat procedure was used to allow the polymer to undergo a relaxation to erase thermal history (Machado and Sencadas, 2017). The sample was loaded into an aluminium pan with the lid crimped on. An empty pan was used as a reference sample. The glass transition temperature (T_g) was found by using MATLAB to find the inflection point via the gradient function.

Results and Discussion

Analysis of oils

After titration of oils, it was found that WCO had a much greater acid value than rapeseed oil, shown in table 4. This result was expected as the heat of cooking breaks the triglycerides molecules releasing the fatty acids from the glyceride backbone.

Table 4 Acid values of oils

Oil	Acid Value (mg KOH/ g oil)
VO	0.07 ± 0.04
WCO	2.81 ± 0.06

^{1}H-NMR analysis revealed that the VO sample was more unsaturated than the WCO sample. The relative degree of unsaturation can be found in table 5, calculated from the peaks shown in figure 1 at 5.3 ppm and 5.25 ppm of the olefinic hydrogens and the methine hydrogen at C2 of glyceride respectively. The differences in degree of unsaturation could be contributed to different compositions of fatty acids, implying the WCO was not rapeseed oil. As the C=C double bonds is the location where the sulphur chain is added to the organic crosslinker, this result predicts faster rates of inverse vulcanisation in rapeseed oil due to more sites for the addition of sulphur being available.

Table 5 Relative degree of unsaturation of oils

Oil	Relative degree of unsaturation*
VO	3.64
WCO	3.34

*Normalised by the number of triglyceride molecules

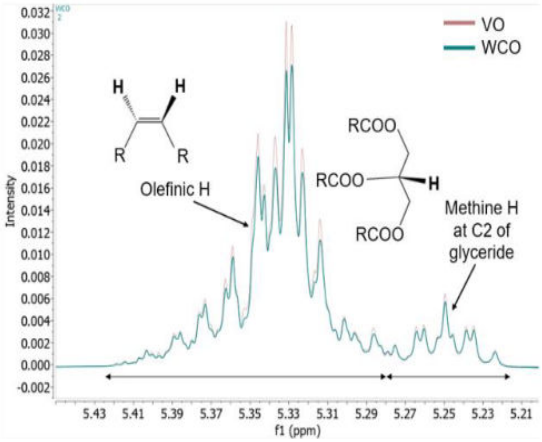


Figure 1 Section of NMR spectrum of VO and WCO showing the olefinic H and C2 glyceride Methine H peaks and the range integrals were calculated

Production of polysulphides

Average vitrification times are summarised in figure 2a, where the variability of this system is visible, attributed to slight differences in heating and stirring in the different positions in the aluminium block. Another source of variability may be caused by the minor differences in shape and weight of the magnetic stirring bars, affecting the amount of resistance required to stop the stirrer. As a result, vitrification was sometimes declared for experiments when the reaction mixture was at different viscosities. This was mitigated by retrieval and reuse of as many stirring bars of the same size as possible. Monitoring of the viscosity would eliminate this issue as the response variable could be changed to the time taken for the reaction mixture to reach a specific viscosity.

Unfortunately, there were many conditions where vitrification was not observed within 2 h, including many experiments conducted at 140 °C. This could be due to this temperature being below the floor temperature of sulphur, at around 159 °C, meaning that the sulphur was initially present as S_8 rings. Therefore, the breaking of these rings becomes entirely dependent on the catalyst. This could explain the lack of vitrification in experiments conducted with the less electronegative metal ion in NaDMDC.

Overall, many of the shortest vitrification times were achieved in experiments catalysed by NaDMDC, including the fastest vitrification for VO and WCO. This could be due to the smaller size of the catalyst facing less mass transfer resistance than $ZnDEDC_2$, speeding up the reaction. There was also a preference for lower S: oil ratio, with a visible loss in rate enhancing ability seen in experiments with a S:oil ratio of 4. This may be due to the lower volume of oil present in higher ratios, and lower transport resistances, so the effect of metal ion electronegativity is greater. Explaining why the experiments with a S:oil ratio of 4 preferred the $ZnDEDC_2$ catalyst.

The optimal catalyst loading of $ZnDEDC_2$ changes between oils. This could be explained by the higher free fatty acid content in WCO, allowing better access of the double bonds present on them to the catalyst so that more of the catalyst can be utilised. Possibly suggesting that in oils with lower free fatty acid content, the $ZnDEDC_2$ catalyst faced difficulties during transport to double bond sites.

Polymers produced with the conditions that resulted in the shortest vitrification time for each oil and catalyst combination were chosen for further analysis. These include the samples: RZ-160-401, RN-160-105, WN-160-105, and WZ-160-405.

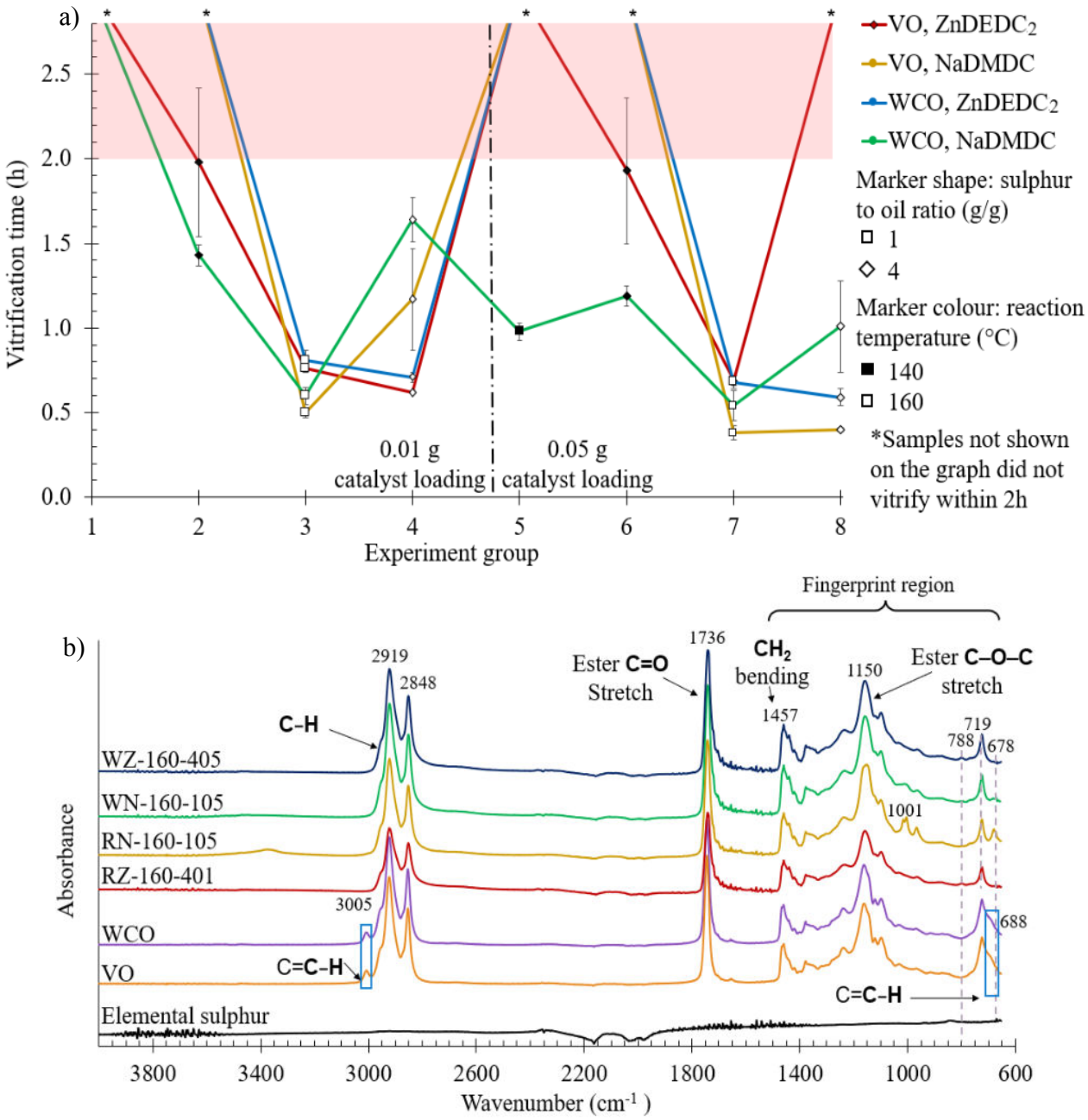


Figure 2 a) Summary of DOE results. b) FTIR spectrum of polysulphides and reactants

Characterisation of Polysulphides

The FTIR spectrum figure 4b, an infrared spectroscopy table from UCLA (2001) was used to identify peaks. Some functional groups can be observed in both the oils and the polysulphides. The stretching of the C=O bond at 1736 cm^{-1} and of the C-O-C group at 1150 cm^{-1} signify the presence of the ester group. This is evidence of triglycerides, and their presence in the polymer product shows that triglycerides did not get completely degraded during the process of inverse vulcanisation.

The peaks at 3005 cm^{-1} and 688 cm^{-1} are likely the result of double bonds as they correspond to a C-H bonded to the C=C group. These peaks are present in the oil samples but not in the polymers. Double bonds were consumed where sulphur chains are added, confirming that inverse vulcanisation had taken place.

New peaks at 788 cm^{-1}, 719 cm^{-1}, and 678 cm^{-1} that appear in the polymers, were not present in the oil samples. They could be the peaks of C-S bonds formed during the addition of the sulphur to the double bond during inverse vulcanisation.

Table 6 Estimated sulphur content in the polysulphide samples obtained from XRF analysis.

Sample	Estimated sulphur content (%)
RZ-160-401	9.067
RN-160-105	28.987
WZ-160-405	37.661
WN-160-105	29.351

The estimated sulphur content in the polysulphides given by XRF is shown in table 6. The samples which were produced using 1:1 sulphur to oil ratio have a lower sulphur content compared to the sample WZ-160-405. One of the samples, RZ-160-401, couldn't be pressed into the sample holder due to the static build up causing the powder to move upwards. The sample didn't cover the entirety of the analysis surface hence the exceptionally low measurement.

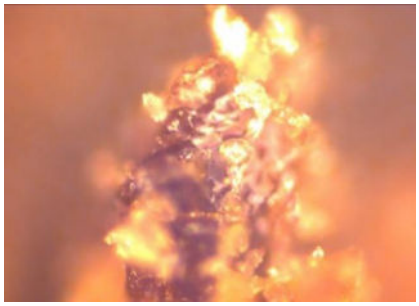


Figure 4 Photo of sample RZ-160-401 taken by Bruker Senterra II Raman Spectrometer

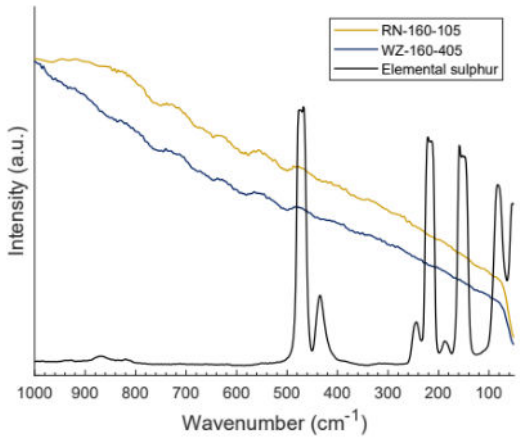


Figure 3 Raman spectrum for elemental sulphur and samples RN-160-105 and WZ-160-405.

A Raman spectrum, shown in figure 3, was collected to identify the nature of the sulphur within the product. Both S-S bonds and C-S bonds are highly Raman active. (Dodd et al., 2023)

The samples exhibit strong fluorescence which is a common issue in polymers produced via inverse vulcanisation (Dodd et al., 2023), partially obscuring the Raman signals. However, some signals are visible through the fluorescent background which can be used to look for unreacted sulphur crystals. Signals between 600 and 800 cm^{-1} suggest the presence of carbon sulphur bonds and the lack of peaks below 250 cm^{-1} indicate there is no elemental sulphur remaining in the sample, further evidence that inverse vulcanisation was successful.

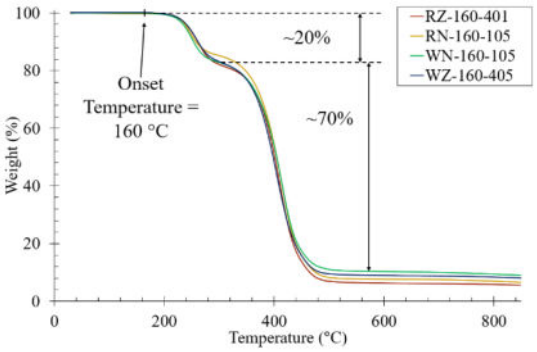


Figure 5 TGA plots over temperature range 30 - 850 °C.

Thermograms of TGA can be found in figure 5. All polymers exhibit two step degradation, implying that the polymers were not homogeneous which can be seen in figure 4, from an initial onset temperature of 160 °C. The first degradation step represents the degradation of the sulphur chains and the second step represents the degradation of the organic part of the polymers.

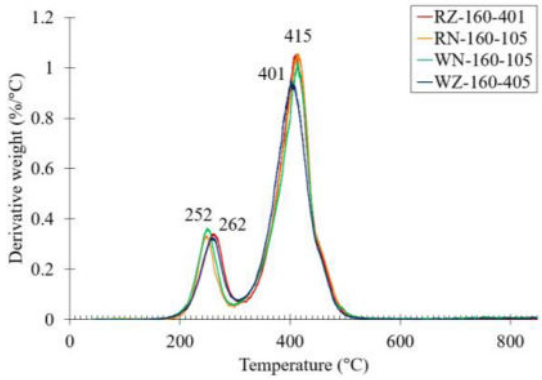


Figure 6 DTA graph showing rate of weight loss from TGA

There are two distinct positions of the first peak in the DTG graph in figure 6, with the polymers produced using a 1:1 sulphur to oil ratio having a lower temperature of maximum rate of weight loss at 252 °C. This implies that the sulphur in these polymers is less thermally stable than the sulphur in polymers produced with a higher sulphur to oil ratio. The optimum degradation temperatures found in the second peak suggest that the organic part of sample WZ-160-405 was less thermally stable compared to the rest, which could be the effect of impurities present in WCO. This can be addressed by pre-processing of the waste oil via centrifugation and microfiltration.

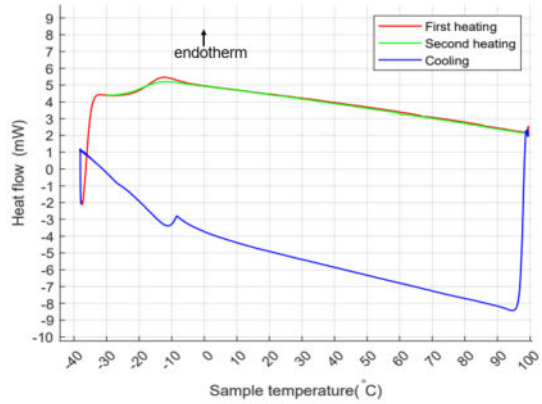


Figure 7 DSC heating and cooling curves for RZ-160-401 over the temperature range -37–100 °C.

The DSC scan for RZ-160-401 is shown in figure 7. A clear glass transition temperature (increase in heat capacity) is observed at -19.2 °C, showing that the polymer has amorphous content which transitions from its glassy state to its rubber state as it is heated past the T_g. This is due to the chains having translational movement in the rubber state compared to just vibrational motion in the glass state (Malmström, 2001).

The T_g is also affected by the curing degree of the polymer. The curing reaction involves the development of the polymer's cross-link network. The more developed the network, the higher the T_g. The relationship between curing time and glass transition temperature has been studied for different polymers by Michel and Ferrier (2020). Following this methodology of varying curing time and temperature and measuring T_g, the optimal curing temperature and duration for the polymer can be potentially quantified.

From the scan, it is not clear whether the polymer is semi-crystalline or fully amorphous. Dodd et al. (2020) suggests that the sulphurous content of the polymer is responsible for the amorphous behaviour. The organic content of the polymer may be crystalline. Since, crystalline polymers can undergo melting transitions while amorphous polymers cannot, the observation of a melting peak would confirm the presence of a crystalline portion in the polymer (Menczel and Prime, 2008). From the TGA thermogram in figure 4a, the onset of the first degradation is 160 °C. Therefore to carry out a heat-cool-heat method whilst preserving the structural integrity of the polymer, a DSC scan from -40 to 160 °C will confirm whether the polymer has a melting point and therefore if it is semi-crystalline or not.

Conclusions

In conclusion, polysulphides were successfully produced using waste cooking and virgin vegetable oils via the process of catalysed inverse vulcanisation, confirmed by FTIR and Raman.

The DOE results allowed for screening of the experimental space, and preferences towards certain conditions were clearly identified. However, the optimisation was not achieved due to the focus on the corners of the variable space with the 2 level study. The next step to find the optimum conditions is to use these trends to select appropriate parameter ranges to carry out a 4 level DOE study. Low temperatures can be excluded due to subpar rates of reaction. This would ensure full data for vitrification time and a 3D response surface could be produced, allowing for the identification of a global optimum.

Utilisation of waste cooking oil in place of rapeseed oil as a crosslinker did not cause significant difference in properties of the product polymer as most differences noted can be attributed to other factors. Firstly, polymers produced with a higher sulphur to oil ratio showed increased thermal stability. The polymer also has amorphous content, given that it undergoes a glass transition.

Sodium dimethyldithiocarbamate dihydrate displayed promising rate enhancement abilities in a narrow set of conditions but showed far less catalytic activity than Zinc diethyldithiocarbamate outside of these conditions, especially in large excesses of sulphur. It was concluded that NaDMDC was not

suitable for inverse vulcanisation. For both catalysts, increasing catalyst loading increased the rate of reaction with the exception of rapeseed oil and zinc-based catalyst combination.

Zinc (II) diethyldithiocarbamate was the better catalyst for the process of catalysed inverse vulcanisation for vegetable oils. $ZnDEDC_2$ had much more consistent rate enhancement ability, achieving vitrification under 2 h even at 140 °C.

Acknowledgements

The authors would like to thank Siyuan Gao and Tianhao Zhang for their assistance with characterisation.

References

BBC News (2017) *'Monster' London fatberg: The smell and other questions* [online] Available at: https://www.bbc.co.uk/news/uk-england-london-41252635 [Accessed 2 December 2024]

Beyza Uçar, Zahra Gholami, Kateřina Svobodová, Ivana Hradecká and Vladimir Hönig (2024) A Comprehensive Study for Determination of Free Fatty Acids in Selected Biological Materials: A review. *Foods*, 13(12), pp. 1891-1891. Doi:https://doi.org/10.3390/foods13121891.

Chung, W.J., Simmonds, A.G., Griebel, J.J., Kim, E.T., Suh, H.S., Shim, I.B., Glass, R.S., Loy, D.A., Theato, P., Sung, Y.E., Char, K., Pyun, J. (2011). Elemental Sulfur as a Reactive Medium for Gold Nanoparticles and Nanocomposite Materials. *Angew. Chem. Int. Ed.,* 50(48), pp 11409-11412. Doi: https://doi.org/10.1002/anie.201104237

Dodd, L. J., Omar, Ö., Wu, X. and Hasell, T. (2021). Investigating the Role and Scope of Catalysts in Inverse Vulcanization. *ACS Catalysis*, 11(8), pp.4441–4455. Doi: https://doi.org/10.1021/acscatal.0c05010.

Dodd, L.J., Lima, C., Costa-Milan, D., Neale, A.R., Saunders, B., Zhang, B., Sarua, A., Goodacre, R., Hardwick, L.J., Kuball, M. and Hasell, T. (2023). Raman analysis of inverse vulcanised polymers. *Polymer Chemistry*, 14. doi:https://doi.org/10.1039/d2py01408d.

Feng, G.D., Ma, Y., Liu, H.H., Zhang, M., Jia, P.Y., Hu, L.H., Liu, C.G., Hu, Y., Zhou, Y.H. (2024). Developing polysulfide castor oil-based polymer coating for controlled-release urea fertilizers. *Progress in Organic Coatings,* 190, 108312. Doi:https://doi.org/10.1016/j.porgcoat.2024.108312

Food and Agriculture Organization of the United Nations (FAO) (2024). *Food Balances.* [Online] Available at: https://faostat.fao.org [Accessed 2 December 2024]

Greene, J.P. (2021). Microstructures of Polymers. *Automotive Plastics and Composites*, pp.27–37. Doi:https://doi.org/10.1016/B978-0-12-818008-2.00009-X..

Health and Safety Executive (2009). *Managing hydrogen sulphide detection offshore.* [Online] Available at: https://www.hse.gov.uk/offshore/infosheets/is6-2009.htm. [Accessed 29 November 2024]

Hoefling, A., Lee, Y.J., Theato, P. (2016) Sulfur-Based Polymer Composites from Vegetable Oils and Elemental Sulfur: A Sustainable Active Material for Li-S Batteries. *Macromol. Chem. Phys.* 218(1), 1600303. Doi:https://doi.org/10.1002/macp.201600303

Ichikawa, S. and Nakamura, T. (2016). Approaches to solid sample preparation based on analytical depth for reliable X-ray fluorescence analysis. *X-Ray Spectrometry*, [online] 45(6), pp.302–307. doi:https://doi.org/10.1002/xrs.2700.

Malmström, E. (2001). *Hyperbranched Macromolecules: Polymerization and Properties*. 2nd ed. [online] pp.3992–3997. Available at: https://www.sciencedirect.com/science/article/pii/B0080431526007026.

Michel, M. and Ferrier, E. (2020). Effect of curing temperature conditions on glass transition temperature . *Construction and Building Materials*, 231(1). Doi:https://doi.org/10.1016/j.conbuildmat.2019.117206..

Rousseau, R.M. (2006). Corrections for matrix effects in X-ray fluorescence analysis—A tutorial. *Spectrochimica Acta Part B,* 61, pp.759–777. doi:http://dx.doi.org/10.1016/j.sab.2006.06.014.

Shen, H., Zheng, B. and Zhang, H. (2024). A Decade Development of Inverse Vulcanisation Towards Green and Sustainable Practices. *Polymer Reviews,* pp.1211-1266. Doi: https://doi.org/10.1080/15583724.2024.2382891

Worthington, M.J.H., Kucera, R., Albuquerque, I.S., Gibson, C.T., Sibley, A., Slattery, A.D., Campbell, J.A., Alboaiji, S.F.K., Muller, K.A., Young, J., Adamson, N., Gascooke, J.R., Jampaiah, D., Bernardes, G.J.L., Chalker, J.M. (2017). Laying Waste to Mercury: Inexpensive Sorbents Made from Sulfur and Recycled Cooking Oils. *Chem. Eur. J.* 23(64), pp. 16219-16230. Doi:https://doi.org/10.1002/chem.201702871

Wu, X.F., Smith, J.A., Petcher, S., Zhang, B.W., Parker, D.J., Griffin, J.M., Hasell, T. (2019). Catalytic inverse vulcanisation. *Nat Commun,* 10, 647. Doi: https://doi.org/10.1038/s41467-019-08430-8

UCLA (2001). *IR table*. [online] Ucla.edu. Available at: https://www.chem.ucla.edu/~bacher/General/30BL/IR/ir.html. [Accessed 11 December 2024]

U.S. Geological Survey. (2024). Sulfur. *Mineral Commodity Summaries*, pp. 172-173. [Online] Available at: https://www.usgs.gov/centers/national-minerals-information-center/sulfur-statistics-and-information [Accessed 20 November 2024]

A Techno-Economic Analysis and Systematic Review of Hydrogen Storage and Transportation Technologies

Kamarul Izzat Bin Nor Azam and Suzulyana Binti Zulkifli
Department of Chemical Engineering, Imperial College London, U.K.

Abstract This study conducts a techno-economic analysis of key hydrogen storage and transportation options. A systematic review of the literature was carried out on hydrogen storage and transportation technologies based on four conditionings: compressed gaseous hydrogen (CGH_2), liquified hydrogen (LH_2), liquid organic hydrogen carriers (LOHC) and ammonia (NH_3). Cost-relevant data were collected for aboveground (AG) tanks, salt caverns, depleted oil and gas reservoirs, aquifers, lined rock caverns, pipelines, ships and trucks. These data were used in economic models to quantify the levelized cost of hydrogen storage (LCOHS) and transportation (LCOHT) for each type of conditioning. A comparison between unharmonized and harmonised costs was conducted, and projections for costs were calculated for 2050. Results show that the mean LCOHS for AG is \$4.23/kg H_2 and \$0.84/kg H_2 for underground, and the mean LCOHT for pipeline is \$2.24/kg H_2, \$2.73/kg H_2 for shipping and \$3.20/kg H_2 for trucks. Minimal changes after harmonisation demonstrate the high degree of standardisation of time, regional and scale-affected variables. The cost projection resulted in a 26%, 36% and 24% cost reduction for salt caverns, H_2 shipping and H_2 trucks, respectively highlighting their potential for future applications. However, the cost reduction for salt caverns seemed to be insignificant as the LCOHS pre-projection was already low at \$0.27/kg H_2, resulting in its cost reduced to only \$0.20/kg H_2.

1. Introduction

The Paris Agreement sets a goal of limiting the temperature increase to 1.5°C above pre-industrial levels which requires greenhouse gas emissions to peak before 2025 at the latest and a decline of 43% by 2030 (UNFCCC, 2015). Hydrogen (H_2) is a vital part of doing so as it has a large gravimetric energy density, approximately 120 kJ/kg, the largest of any known substances (Møller, et al., 2017). Due to its range of use from energy carrier to chemical feedstock to fuel, the demand for H_2 is projected to increase significantly, from 97 Mt in 2023 to 388 Mt by 2050 (IEA, 2024) (IEA, 2023).

While the production of H_2 and its cost are well-researched in academia, it is harder to find a study on the cost of H_2 transport and storage (DESNZ, 2023). Existing literature discusses the technical aspects of H_2 technologies but often lacks a critical analysis of its economics. This gap is important as it is necessary to have an encompassing understanding of the whole H_2 supply chain to facilitate the growth of the H_2 economy. This study aims to address this gap.

The focus of this study is on bulk H_2 storage and transport options in the form of compressed gaseous hydrogen (CGH_2), liquefied hydrogen (LH_2), liquid organic hydrogen carriers (LOHC), and ammonia (NH_3). CGH_2 is widely used for transport and storage, being compressed up to 70 MPa (Du, et al., 2018). LH_2, though denser, requires cryogenic storage at -253°C, making it costly due to liquefaction and boil-off losses (Timmerberg & Kaltschmitt, March, 2019). LOHCs are liquid at ambient conditions and undergo hydrogenation and dehydrogenation, offering easy handling and transport. While NH_3 as a storage option is not explored due to difficulty in purifying the H_2 at the point of use and lack of available cost data, its reliable transport via ships, with a TRL of 9, makes it a viable option for H_2 delivery (Di Molfetta, 2022).

The following H_2 storage options will be analysed: aboveground (AG) tanks for GH_2, LH_2, and LOHCs, as well as underground (UG) storage solutions for GH_2, salt caverns, depleted oil and gas reservoirs, aquifers, and lined rock caverns. H_2 transport will be examined across three primary modes, trucks, pipelines, and ships, considering their application to all of the aforementioned H_2 forms. These storage and transport methods are selected due to their established potential and the availability of relevant data, making them the most viable options for large-scale H_2 distribution (Di Molfetta, 2022). The discussed options are not an exhaustive list of the existing H_2 storage and transport technologies.

This study employs a secondary research methodology approach through systematically reviewing and analysing the existing academic literature on H_2 storage and transport, aiming to consolidate and harmonise the existing data in the form of levelised costs of hydrogen storage (LCOHS) and transport (LCOHT). An economic, model was developed to standardize and compare the cost data from various studies, and future costs for each technology were projected and analysed. To the authors' knowledge, this is the first comprehensive effort to align and compare the LCOHS and LCOHT across studies.

2. Methods

2.1. System boundary

Since H_2 can be transported in various forms, the conditioning and deconditioning of H_2 were taken into account. H_2 production at the early stage and usage at the end were not considered. It was also assumed that in the end, H_2 could be used in both gaseous and liquid form following (Reuß et al., 2017) method.

2.2. Systematic review

A systematic review strategy was implemented to identify relevant cost data for H_2 storage and transportation from the literature. Initial search terms were constructed and ScienceDirect was used as the primary search database. These initial search terms were refined by focussing on topics unique to H_2 storage and transportation. This was achieved by excluding unrelated phrases such as 'production' and 'performance'. Next, the abstracts of the papers from the refined search terms were screened and only relevant articles were filtered through. A 'relevant' paper's abstract:

- 'Appeared to contain information on technology or economics of H_2 storage or transport' or
- 'Appeared to contain indirect information relating to technology or economics of H_2 storage or transport' or
- 'Appeared to contain supporting information that might reasonably be used to contextualise information on H_2 storage or transportation'

A questionnaire was produced to screen the articles satisfying these criteria and their level of relevance was categorised as high, medium or low. Literature with a high level of relevance was reviewed in depth using a cost-focused questionnaire constructed and summary paragraphs were produced for each article. The cost data from this literature were used for the economic analysis as outlined below. In addition, grey literature was consulted to provide contextualisation information on specific parts of the H_2 supply chain where necessary. The search terms and questionnaires are included in the supplementary information (SI 1).

2.3. Economic analysis

2.3.1. Storage

The equations used to calculate LCOHS ($/kg H_2) were extracted from (Pietro Dogliani et al., 2024).

2.3.1.1. Tanks

The LCOHS for tanks was calculated using,

$$\mathrm{LCOHS} = \frac{(\mathrm{a} + \mathrm{OPEX}) \cdot \mathrm{CAPEX}}{\mathrm{Q_{H_2}}} \quad (1)$$

where OPEX is the operational expenditure as a fraction of the capital expenditure (CAPEX) which is in $, a is the amortisation factor and $\mathrm{Q_{H_2}}$ is the annual amount of H_2 stored in kg.

$$\mathrm{a} = \frac{\mathrm{d}}{1 - (\mathrm{d}+1)^{-\mathrm{N}}} \quad (2)$$

a was obtained using equation (2) and it is a function of the discount rate (d) and the economic lifetime (N) in years. $\mathrm{Q_{H_2}}$ was calculated from equation (3)

$$\mathrm{Q_{H_2}} = \mathrm{n_{cycles}} \cdot \mathrm{Q_{storage}} \cdot (1 - \mathrm{m_{losses}}) \quad (3)$$

Where $\mathrm{n_{cycles}}$ is denoted as the number of cycles per year, $\mathrm{Q_{storage}}$ is the storage capacity in kg and $\mathrm{m_{losses}}$ is fractional weight loss.

2.3.1.2. Geological reservoirs

The LCOHS for geological reservoirs were calculated using equation (4).

$$\mathrm{LCOHS} = \frac{(\mathrm{a} + \mathrm{OPEX}) \cdot \mathrm{CAPEX} + \mathrm{a} \cdot \mathrm{C_{gas}}}{\mathrm{Q_{H_2}}} \quad (4)$$

This equation is similar to that for storage tanks but with the addition of the cost of the cushion gas kept in the reservoir ($\mathrm{C_{gas}}$) in $ which was assumed as

$$\mathrm{C_{gas}} = \mathrm{P_{gas}} \cdot \mathrm{Q_{H_2}} \cdot 6 \quad (5)$$

where $\mathrm{P_{gas}}$ is the proportion of cushion gas and $6 refers to the average price of 1 kg of H_2 which is primarily used as a cushion gas (Clean Hydrogen Partnership, 2024).

2.3.2. Transportation

The equations used to calculate LCOHT ($/kg H_2) were extracted from the same article (Pietro Dogliani et al., 2024).

2.3.2.1. Pipelines

The following equation was used to calculate LCOHT for pipelines,

$$\mathrm{LCOHT} = (\mathrm{a} + \mathrm{OPEX}) \cdot \mathrm{CAPEX} \cdot l_s \cdot \frac{\mathrm{LHV_{H_2}}}{\mathrm{Q_{H_2}} \cdot \mathrm{CF} \cdot 8760} \quad (6)$$

where $\mathrm{LHV_{H_2}}$ represents the lower heating value of H_2 in kWh/kg, OPEX is a fraction of CAPEX, CAPEX is in $/km, l_s is the distance between compression stations in km, CF represents the capacity factor of the pipeline and $\mathrm{Q_{H_2}}$ is the pipeline design capacity in kW.

2.3.2.2. Shipping

The LCOHT for shipping was calculated using equation (7) for NH_3 and liquid carriers with negligible boil-off, and equation (8) for LH_2 and liquid carriers with boil-off.

$$\mathrm{LCOHT} = \mathrm{LCOHT_{vessel}} + \mathrm{LCOHT_{fuel}} \quad (7)$$

$$\mathrm{LCOHT} = \mathrm{LCOHT_{vessel}} + \max(\mathrm{LCOHT_{fuel}}, \mathrm{LCOHT_{boiloff}}) \quad (8)$$

The $\mathrm{LCOHT_{vessel}}$ was determined by,

$$\mathrm{LCOHT_{vessel}} = \frac{(\mathrm{a} + \mathrm{OPEX}) \frac{\mathrm{CAPEX_{ship}}}{\mathrm{Q_{ship}}}}{\mathrm{rpa}} \quad (9)$$

where OPEX is denoted as a fraction of CAPEX, Q_{ship} is the H_2 ship capacity in kg/ship and rpa is the number of routes per year which was obtained by,

$$\text{rpa} = \frac{8760}{2 \cdot t_{roundtrip}} = \frac{8760}{2 \cdot \left(\frac{l}{v} + t_{harbour}\right)} \quad (10)$$

where l is the route distance in km, v is the ship velocity in km/h and t_{harbor} is the average time spent in the harbour in hours.

$$LCOHT_{fuel} = \frac{P_{fuel} \cdot E_{fuel} \cdot 2l}{Q_{H_2}} \text{ or } \frac{P_{fuel} \cdot E_{fuel} \cdot t_{roundtrip}}{Q_{H_2}} \quad (11)$$

The $LCOHT_{fuel}$ depends on additional parameters which include the fuel price (P_{fuel}) in \$/L, \$/kg or \$/MJ and fuel consumption (E_{fuel}) in L/km, kg/km or kg/h, or MJ/km. The $LCOHT_{boiloff}$ was calculated by equation (12).

$$LCOHT_{boiloff} = \frac{b}{24} \cdot t_{roundtrip} \cdot LCOHP \quad (12)$$

Here, b denotes the boil-off fraction and LCOHP denotes the levelized cost of hydrogen production in \$/kg.

2.3.2.3. Truck

The LCOHT for trucks were determined using equation (13),

$$LCOHT = LCOHT_{truck} + LCOHT_{fuel} + LCOHT_{driver} + LCOHT_{boiloff} \quad (13)$$

where $LCOHT_{fuel}$ and $LCOHT_{boiloff}$ are similar to the shipping counterpart.

$$LCOHT_{truck} = \frac{CAPEX_{tractor} \cdot a_{tractor} + CAPEX_{tank} \cdot a_{tank} + OPEX \cdot l_n}{Q_{H_2} \cdot \text{rpa}} \quad (14)$$

The $LCOHT_{truck}$ was modified to exclude the trailer chassis-related parameters as the papers reviewed in this study did not include any information on trailer chassis. The CAPEX terms are expressed in \$, the OPEX in \$/km, l_n is the annual distance covered in km and Q_{H_2} is the H_2 truck capacity in kg/truck. The $LCOHT_{driver}$ was obtained by,

$$LCOHT_{driver} = \frac{c_{driver} \cdot t_{op}}{Q_{H_2} \cdot \text{rpa}} \quad (15)$$

where c_{driver} denotes the driver's salary in \$/h and t_{op} denotes the annual hours of operation.

2.3.3. Harmonisation

Harmonisation was implemented to provide an unbiased comparison that could arise due to time, regional and scale differences. The variables harmonised in this study include discount rate, operating hours, lifetime, capacity factor, cushion gas, fuel price and driver wage. Harmonised values were determined by taking the average value of that particular parameter for each technology and conditioning.

2.3.4. Cost escalation

Since this study involved the analysis of historical costs, an escalation model was developed. Firstly, the LCOHS and LCOHT values were converted into USD based on the reference year quoted in the paper using annualised mean exchange rates from the International Monetary Fund (IMF, 2024). Subsequently, the Producer Price Index (PPI) for Transportation and Warehousing Industries, and the PPI for Total Mining Industries were retrieved and used for LCOHT and LCOHS, respectively. These values were then escalated to the chosen comparison year of 2023 from the base year using equation (16).

$$\text{Cost}(2023) = \text{Cost}(\text{base year}) \cdot \frac{\text{PPI}(2023)}{\text{PPI}(\text{base year})} \quad (16)$$

2.3.5. Cost projection

The forecast of cost was carried out by utilising the learning curve framework. This method relates the cost of a technology to its cumulative installed capacity through a learning index which is represented in equation (17) (Schoots et al., 2008),

$$C_{t_2} = C_{t_1} \left(\frac{P_{t_2}}{P_{t_1}}\right)^{-\alpha} \quad (17)$$

where C_{t_1} and C_{t_2} are the cost of the technology at the time t_1 and t_2, respectively and P_{t_1} and P_{t_2} are the cumulative installed capacity at t_1 and t_2, respectively. The learning index, α was calculated using equation (18),

$$LR = 1 - 2^{-\alpha} \quad (18)$$

where LR is the learning rate of the technology.

3. Results & discussion

3.1. Demographics of the systematic review

The initial search terms of this systematic review returned 7557 papers. The refined search terms were able to narrow this down to 676 papers. Abstract screening to eliminate non-relevant papers resulted in 208 papers left to evaluate. Among these, 61 papers were classified as highly relevant, 93 were of medium relevance and 54 had low relevance. From the 61 papers that were reviewed in depth, 41 papers discussed H_2 storage and 33 papers discussed H_2 transportation with an overlap of 13 papers discussing both H_2 storage and transportation. The near balance in the numbers of H_2 storage and transportation papers shows that they are equally significant to consider for H_2 deployment. 25 papers from the 41 papers that discussed H_2 storage focused on only 1 type of technology. In contrast, only 7 papers were particular to a single technology out of the 33 papers for H_2 transportation which demonstrates that technology comparison is given more attention for H_2 transportation compared to

storage. However, an interesting finding from the H_2 storage review is that 11 papers studied the utilisation of H_2 carriers as potential storage options. These carriers include LOHCs, metal hydrides, adsorbents and other liquid carriers such as NH_3. Since these papers defined the carriers as H_2 storage, they lack cost information for the physical storage options that were studied for this paper such as tanks and geological reservoirs thus could not be of use for the economic analysis.

Among the 41 papers for H_2 storage, 22 papers examined tanks, 13 on caverns, 6 on aquifers, 7 on depleted oil or gas fields and 2 on UG pipes. The high number of papers on tanks is due to their common usage and being a rapidly developing technology (Xie et al., 2024). The apparent difference in the number of papers on caverns compared to other UG technology is because it is commercially available and has been used for decades with a Technology Readiness Level (TRL) of 9 (Di Molfetta, 2022). Aquifers and depleted oil and gas fields still face issues regarding insufficient reservoir properties and potential unwanted H_2 reactions (Tarkowski, 2019) that may hinder the investigation of its economic aspects. Only 2 papers covered UG pipes as H_2 storage options. This could be due to the impracticality of using pipes as storage because they require similar high-pressure conditions as caverns but have similar storage capacity as spherical tanks (Papadias and Ahluwalia, 2021).

Of the 33 papers on H_2 transportation, 26 papers discussed pipelines, 15 papers on shipping, 19 papers on trucks and 5 papers on trains. The significantly lower number of research papers on railway H_2 transportation could be because it is similar to trucks but with a more limited route network thus making it impractical to use compared to trucks. This impracticality impedes its research growth as demonstrated previously for H_2 storage using pipes.

Based on the number of papers shown, the economic analysis of H_2 storage technologies focused on tanks and geological reservoirs, while transportation technologies included pipelines, shipping and trucks due to their relative abundance of cost information.

3.2. H_2 storage technology overview & analysis

3.2.1. Underground storage

Underground storage of gasses is not a novel technology with the U.S. Energy Information Administration approximating 379 existing UG storage sites for natural gas in the U.S. with around 3.8 trillion cubic feet of working gas (Michanowicz, et al., 2017). The major advantage of UG storage is the capacity far exceeds AG storage, drastically decreasing the LCOHS. As the H_2 gas stored UG can be pressurised up to 25MPa, compression costs need to be considered. Another critical factor is cushion gas proportion, which maintains storage pressure during H_2 extraction and must be constantly in the formation. While alternative cushion gases, such as CO_2 and CH_4, can be used, most studies reviewed utilize H_2 as the cushion gas unless specified otherwise (Kanaani, et al., 2022). For any omitted data such as cushion gas proportion and discount rate, the average from the rest of the studies will be taken.

3.2.1.1. Salt caverns

Salt caverns are made by leaching out large cavities using injection of fresh water and can be made within salt domes or bedded salt deposits. The surrounding salt provides an impermeable barrier, resulting in an almost leak-proof system with a leakage rate of less than 1%. Caverns are relatively smaller than depleted reservoirs and LRCs but around the same size as aquifers. This smaller size enables higher cycling frequencies, with studies reporting 9–12 cycles per year annually (Van Gerwen, et al., 2019).

In the studies reviewed, most of the CAPEX stems from geologic site characterisations and compressor costs with (Chen, et al., 2023) purported that 72% of the CAPEX comes from site preparation and 17% from compressors. (Lord, et al., 2014) reported a smaller difference with compressor costing at 43% and geological site preparation at 36%. While other studies differ on which costs more, the two factors were consistently cited as the largest contributor to the CAPEX.

Among five studies analysed for the techno-economics of salt caverns, three did not report discount rates (Papadias & Ahluwalia, 2021) (Reuß, et al., 2017) (Reuß, et al., 2019) and two omitted information on the annual cycling frequency (Lord, et al., 2014) (Chen, et al., 2023). Based on these analyses, the average LCOHS for salt caverns was determined to be $0.27/kg.

3.2.1.2. Depleted oil and gas reservoirs

Depleted oil and gas reservoirs are one of the most widely used geological formations for natural gas storage due to their significant capacity, accounting for 75% of the WGC of subsurface storage systems globally (Tarkowski, 2019). Due to existing infrastructure and the maturity of the technology, these reservoirs are easy to develop, operate and maintain. The average reported cushion gas proportion was 50% of the reservoir's total volume. This estimate considers residual gas present in the

reservoir, which can serve as cushion gas (Foh, et al., 1979).

A challenge in assessing the capital costs is the variability in reported costs for gas conditioning. Many studies do not specify the share of capital cost allocated to conditioning, making direct comparisons to grey literature, where conditioning is cited as the most expensive aspect of depleted reservoirs' CAPEX, difficult (Van Gerwen, et al., 2019).

This is contrasted with a study by (Chen, et al., 2023) which attributes over 79% of the CAPEX to cushion gas while compression only accounts for 20% of the capital cost. However, compared to other cushion gasses such as N_2 and natural gas, (Hamidreza Yousefi, et al., 2023) concluded that H_2 would still be the more economical cushion gas in most cases. The average LCOHS calculated for a depleted oil and gas reservoir was $0.68/kg of H2.

3.2.1.3. Aquifers

Aquifers, composed of water-bearing porous rocks such as sandstones, are located thousands of feet underground and offer potential for H_2 storage. Compared to depleted gas reservoirs, aquifers require significantly more cushion gas due to the absence of residual gas to offset the volume needed, the average found is 80% of the total capacity (Beckman, et al., 1995)]. As in depleted reservoirs, (Hamidreza Yousefi, et al., 2023) considered alternative cushion gases such as N_2 and natural gas and they concluded aquifers could benefit economically using the two gasses.

Despite these cost considerations, saline aquifers were reported to have the highest LCOHS compared to other UG options in the same study. In contrast, another study reviewed, (Lord, et al., 2014), identified aquifers as the most economical option even though the rest of the UG storage options remain unchanged. This discrepancy arises from differences in the cost modelling of the compressor in relation to the other options. (Lord, et al., 2014) attributed the lowest capital cost for compressors to aquifers, while (Chen, et al., 2023) applied uniform compressor costs across all storage options, leading to aquifers being the cheapest in the (Lord, et al., 2014).

Only two studies provided sufficient data for a comprehensive economic analysis of aquifers as H_2 storage sites. A harmonisation and cost projection were still performed, as the available data for H_2 storage technologies overall remains limited. Therefore, the relative gap in data for this specific analysis is not substantially larger than for other H_2 storage options. The average LCOHS found for aquifer storage was $1.00/kg.

3.2.1.4. Lined rock cavern

LRC is a novel storage alternative involving the excavation of caverns in hard rock, which are then encased with concrete followed by steel or plastic liners to create an impermeable barrier for gas containment (Sofregaz US Inc, 1999). To date, no industrial-scale LRC storage facilities have been deployed, although pilot projects have demonstrated their feasibility for natural gas storage.

In both (Lord, et al., 2014) and (Papadias & Ahluwalia, 2021), LRC was the more expensive UG storage option. Both papers also cited mining costs as the primary expense, constituting 54% in the former and 72% of the total cost in the latter study. A methodological concern in (Lord, et al., 2014) is the assumption of identical geological characteristics for both salt caverns and LRCs, with differences accounted for only in site development costs. This approach may limit the reliability of cost comparisons.

As with aquifers, only two studies were found that provided sufficient data for detailed economic analysis. The average LCOH for LRCs was found to be $1.40/kg.

3.2.2. AG storage

3.2.2.1. GH2 tanks

GH_2 tanks are one of the most established methods for H_2 storage, with over a century of reliable use. Unlike UG storage methods, H_2 tanks do not require cushion gas and experience minimal to no boil-off losses for GH_2. Two of the papers reviewed did not include discount rates and an approximation had to be made.

(Reuß, et al., 2017) states that the use of GH_2 for storage is not cost-competitive with other storage methods such as salt caverns, LH_2 and LOHC as GH_2 is estimated to be 20% more expensive than these alternatives. This is further reinforced by (Lange, et al., 2024) which states that by 2035, the AG options will have limited market relevance unless H_2 is widely adopted for mobility and not restricted to seasonal electricity storage.

Only three studies provide sufficient techno-economic data to perform detailed analyses, showing a significant gap in the literature where most GH_2 tanks discussed were made for vehicles rather than for bulk storage of hydrogen.

3.2.2.2. LH_2 tanks

Similar to GH_2, LH_2 is recognized as a mature technology requiring cryogenic storage at 20K. Although its gravimetric energy density is 2.7 times greater than gasoline, its volumetric energy density is approximately 3.7 times lower. Applications include NASA's spherical tanks, where up to 230 tonnes of H_2 can be stored.

Reported boil-off losses in studies averaged 0.03% per day across studies, emphasizing the importance of thermal insulation in LH_2 storage. Over time, these losses can reduce the effective capacity of an LH_2 tank by approximately two-thirds.

The majority of the CAPEX for LH_2 storage is attributed to the liquefaction process. This was consistently highlighted in (Parolin, et al., 2022) and (Reuß, et al., 2017), both of which found that liquefaction costs exceeded all other CAPEX components combined. These findings indicate a critical need for further research into improving liquefaction efficiency to reduce overall storage costs.

As with other discussed technologies, some of the (Parolin, et al., 2022) studies lacked discount rates, but some also had WACC which was used as a substitute for discount rates due to its similarity (Fernández, 2011). The LCOHS for LH_2 was calculated to average $3.80/kg.

3.2.2.3. LOHC

LOHC utilize stable, low-toxicity organic molecules to store H_2 at ambient pressure and temperature in liquid form. This eliminates the need for high-pressure or cryogenic tanks, reducing infrastructure costs. However, these savings are offset by the costs of hydrogenation and dehydrogenation processes. Among the reviewed studies, all considered methylcyclohexane (MCH) as a carrier, while only one did not consider dibenzyltoluene (DBT). Other materials, such as N-ethylcarbazole and naphthalene, were included in limited analyses.

Unlike other technologies where the number of data points correlates closely with the number of papers reviewed, (Tsogt, et al., 2024) provided the majority of the data points due to their focus on the techno-economics of all LOHCs discussed. This study identified DBT as the most favourable material when considering factors such as toxicity, price, energy consumption, and material handling. The conclusion aligns with findings from (Jorschick, et al., 2020).

Key cost contributors for LOHC systems include the price of the carrier material and catalysts alongside the conditioning cost. Both (Lange, et al., 2024) and (Tsogt, et al., 2024) conclude that hydrogenation and dehydrogenation temperature are also major contributing factors as it impact the utility requirements (Reuß, et al., 2017). For regions with low H_2 demand and in scenarios where UG storage technologies are impractical, LOHCs have been suggested as a potentially economical storage method.

Despite its potential, the average LCOHS for LOHCs is the highest among the options analyzed, at $7.09/kg. This is attributable to the technology's nascent state, with no existing pilot projects and research primarily focused on discovering new LOHC materials and catalysts.

3.3. H_2 transportation technology overview & analysis

3.3.1. Pipeline

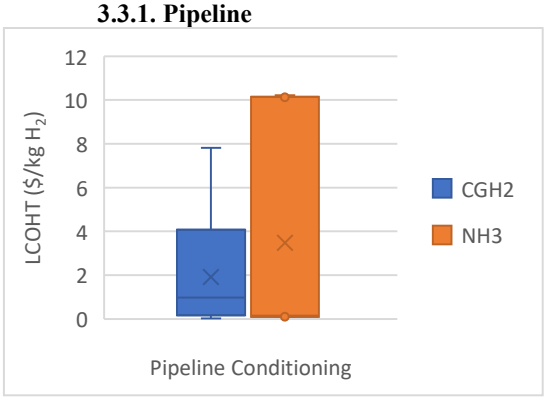


Figure 1. *Cost ranges for pipeline by conditioning*

The mean LCOHT for pipelines with H_2 conditioned as CGH_2 is lower than that for NH_3 conditioning. This could be explained by the conditioning requirements for each type. H_2 compression to produce CGH_2 is regarded as a less energy-intensive process thus less cost demanding (Sehar Tasleem et al., 2024). NH_3 production on the other hand together with its decomposition may demand more energy as these are more complicated processes than simple gas compression.

Figure 1 also shows that the LCOHT data for NH_3 conditioning is more spread out compared to that of CGH_2 which is most probably due to the difference in technology maturity. CGH_2 pipelines have a TRL of 9 (Di Molfetta, 2022) while NH_3 pipelines may still be in the development phase which shows inconsistency in cost data reported.

3.3.2. Shipping

Figure 2 depicts that the mean value of LCOHT for NH_3 is the lowest as it is commercially available with a TRL of 9 (Di Molfetta, 2022), and this trend is followed by LH_2 and LOHC. On top of that, NH_3 shipping does not face boil-off challenges as LH_2 shipping does, which adds additional expenses to mitigate these losses.

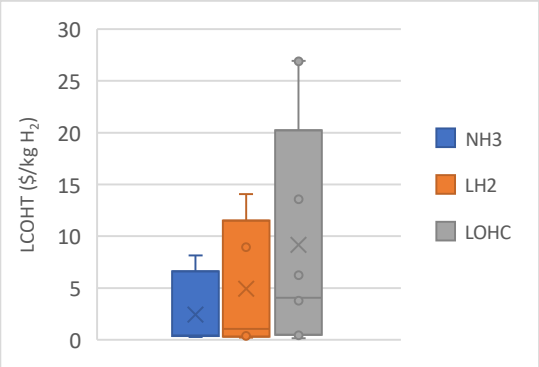


Figure 2. *Cost ranges for shipping by conditioning*

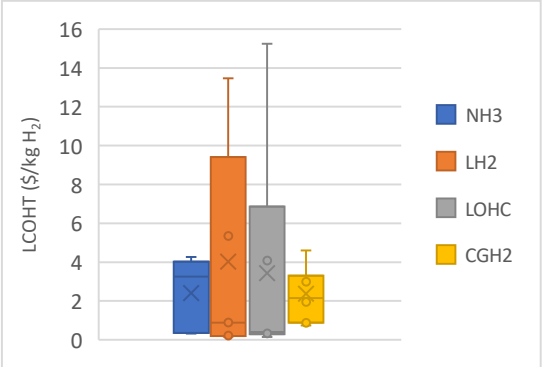


Figure 3. *Cost ranges for trucks by conditioning*

LOHC shipping has the largest IQR and highest mean LCOHT value because of the low H_2 fraction in LOHCs and assumptions made by authors on the ship's fuel consumption. Firstly, the H_2 content in LOHCs ranges from about 6% to 12% (weight basis) (Johnston et al., 2022), and since the calculation of LCOHT is normalised to the weight of H_2 transported, this elevates the cost for transportation of LOHCs. Besides that, (Lee et al., 2022) assumed that the ship's fuel consumption is proportional to maximum loading. Since the maximum loading for LOHC seaborne transport stated in their paper is an order of magnitude larger than NH_3 and LH_2 ships, this accounts for the larger LCOHT for LOHCs which then skews the data distribution leading to a wider IQR.

3.3.3. Truck

The LCOHT for road transportation using CGH_2 tube trailers has the lowest mean value due to the same reasons as before in terms of compression simplicity and technology maturity. Although LH_2 trucks have the highest TRL matching that of CGH_2 trailers, Figure 3 depicts that it has the highest mean LCOHT. Aside from the boil-off issues which was mentioned before, this could be because of the demanding conditions of H_2 liquefaction which require 16kWh/kg H_2 (Ren et al., 2022). This is more energy intensive than H_2 compression which has an energy demand of only 1.1kWh/kg H_2 (GEV, 2021). This increased cost for LH_2 trucks is most probably the foundation of exploring new liquid H_2 carriers for use in road transportation, particularly NH_3 and LOHC. These carriers demonstrate lower mean LCOHT values compared to LH_2 trucks as seen in Figure 3. Since utilisation of H_2 trucks mainly depends on the ability to store H_2 in tanks, NH_3 and LOHC trucks are advantageous because of their mild conditions for storage at ambient temperature and pressure. However, this method has not reached commercial maturity yet, for example, LOHC trucks have a TRL of 6-7 meaning that they are still in the project demonstration stage (Di Molfetta, 2022).

The wide range of LCOHT for LH_2 and LOHC trucks is attributed to the large difference in the assumption of drivers' wages. For instance, (Reuß et al., 2017) assumed the drivers' salary to be 35euro//h while a significantly lower value of 1.67euro/h was used in (Bai et al., 2024)'s study. There is less disparity in the data retrieved for CGH_2 and NH_3 trucks hence the smaller IQR.

3.4. Technology Comparison

3.4.1. H_2 storage

Figure 4 details the aggregated LCOHS of all mentioned H_2 storage technologies. The bubbles in blue hues represent UG storage options, while the red bubbles represent AG storage options. The number in the bubbles represents the number of references used for each technology. However, this value does not directly correlate to the number of data points as a single paper may contribute to multiple technologies. For example, although LOHC has only four papers, it has seven data points. In total, 11 references were used. The range of LCOHS for UG storage is \$0.27/kg to \$1.41/kg and averages at \$0.80/kg/ with the cheapest being salt caverns followed by depleted reservoirs, aquifers and LRC. Salt cavern is the cheapest likely due to its high utilization rate from its frequent cycling which lowers the cost per kg value. For aquifers, the high CAPEX stems from extensive site surveys required to characterize these sites thoroughly, which adds both time and expense to the development process.

LRC has the highest LCOHS among UG storage as it is a relatively new technology with only one pilot study being done. This also parallels other grey and academic literature where LRC was consistently the most expensive option among UG storage (DESNZ, 2023) (Papadias & Ahluwalia, 2021).

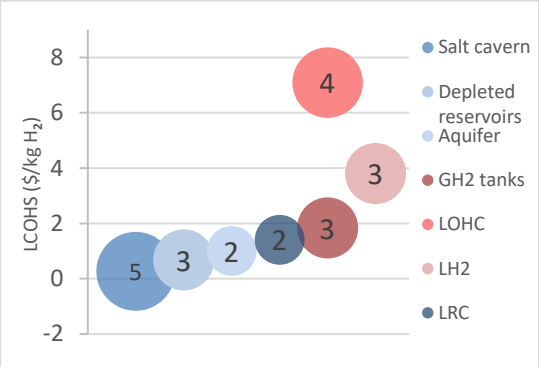


Figure 4. *Mean LCOHS with number of references*

For AG storage, the LCOHS ranges from $1.79/kg to $7.09/kg and averages $4.23/kg. LOHC is the most expensive option and GH_2 is the cheapest. It is noteworthy that the most expensive UG storage still costs less than the most economical AG option. This is concurrent with studies that conclude that AG storage will have lower market relevancy due to how cost-efficient UG storage is (Reuß, et al., 2017) (Lange, et al., 2024).

As seen, the number of papers found is not plentiful. For the two technologies with the least papers, aquifers and LRCs, the small number of papers can be attributed to the novelty of these technologies. Most of the existing literature tends to focus on the technical feasibility of these storage methods rather than an in-depth economic analysis (Kishor Patanwar, et al., 2024) (Liu, et al., 2024). Those that do consider it as a part of a larger H_2 supply chain would not regard it any further than the discussion of storage options due to the lack of reliable cost and location estimates (Reuß, et al., 2019) (Parolin, et al., 2022). This suggests a potential cycle where the limited data available leads to reduced interest in the technology over other more viable options such as salt caverns and depleted reservoirs.

3.4.2. H_2 transportation

At first glance, Figure 5 displays roughly similar mean values for LCOHT with a slightly higher value for H_2 trucks at $2.24/kg H_2 followed by shipping at $2.73/kg H_2 and then pipelines at $3.20/kg H_2. Trucks being more costly than pipelines is consistent with grey literature (Hydrogen Transport and Storage Cost Report, 2023). However, grey literature only considers pipelines and trucks for comparison. This is most probably due to H_2 shipping not being commercially available yet. As stated, LH_2 shipping has a TRL of 5-7 currently and although NH_3 shipping is already a mature technology with a TRL of 9, it is primarily used for the fertiliser industry (Tock et al., 2015) and not as H_2 carriers.

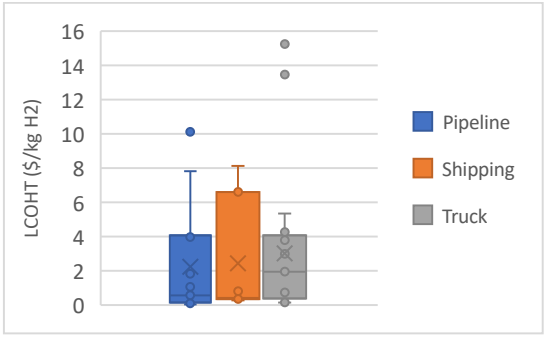


Figure 5. *Cost ranges for all transportation technologies*

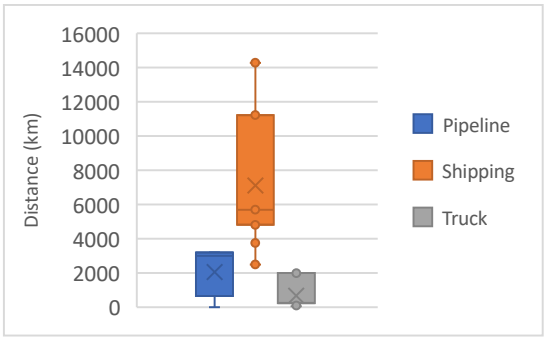


Figure 6. *Distances reported for all transportation technologies*

Figure 6 demonstrates the data points for different technologies based on their distances as reported in their respective papers. It can be deduced that generally, H_2 trucks are used for short-range transportation, pipelines are used for short to medium-range distances and ships are used for long-range H_2 transportation. This explains why the mean LCOHT value from Figure 5 is similar for all types of transportation, that is because each transportation mode has its own advantages at a certain distance. Trucks are utilised for short-distance transport as they have a finite carrying capacity but can benefit from existing road infrastructure. Pipelines have the capacity to transport H_2 at a longer distance due to their continuous operation. However, they are not designed to withstand too high flow rates as this can cause pipe embrittlement which can weaken and degrade the pipelines (Melaina, Antonia and Penev, 2013). Pipelines are also limited by geography hence long-range pipelines for import and export activity are not feasible. Since ships are mainly used for long-range transport, they specialise in international trade of H_2 across sea borderlines. To conclude, the specialised function for each technology at its own distance capacity reflects similar mean values for LCOHT.

In terms of the number of references, Figure 7 shows that CGH_2 pipelines, CGH_2 trucks and LH_2 trucks stand out the most due to having 9, 7 and 5 papers respectively. This number of references is noticeably higher than other technologies. One common feature they share is that they are fully mature and commercialised technologies with a

TRL of 9. Individually, CGH_2 pipelines have been in operation for so long, dating back to 1938, and CGH_2 and LH_2 trucks are the most attractive options for short and sometimes even medium-range distances (Di Molfetta, 2022). The technology maturity together with their utilisation contributes to the larger number of publications compared to other technologies in academic research.

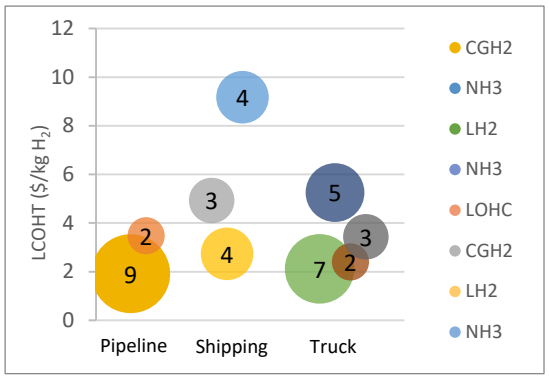


Figure 7. *Mean LCOHT with number of references*

The comparison of different transportation technologies in Figure 5 could be considered fair due to the balance in the number of papers for each technology. However, the investigation of comparing different conditionings within a single technology as in Figure 1, Figure 2, and Figure 3 should be extended before any conclusions can be made due to the uneven number of papers currently available for different types of conditioning. For instance, the comparison of CGH_2 pipelines with NH_3 pipelines as in Figure 1 may require more papers on NH_3 pipelines' costs data as the 2 papers currently available for NH_3 pipelines may be insufficient to draw conclusions for the best conditioning type for H_2 pipelines.

3.5. Harmonisation

The comparison between unharmonised and harmonised values of LCOHS and LCOHT resulted in negligible changes. In many cases, studies did not report, for example, discount rates or cushion gas proportions thus average values derived from existing data were applied during harmonisation. This could explain the minimal differences observed between the harmonised and unharmonised datasets. The small difference after harmonisation might also indicate a high degree of standardization in the methodologies employed across studies which can be caused by how studies tend to build on and refine one another, leading to a convergence in approaches over time. The comparison between unharmonised and harmonised values can be seen in SI 1.

3.6. Cost projections

Due to the scarcity of data regarding installed cumulative capacity for H_2 storage and transportation, the ratio of cumulative capacity for 2050 to 2023 was approximated using the demand for H_2 for each respective year (IEA, 2024). The learning rates for each type of technology were retrieved from the literature and presented in SI 1.

3.6.1. H_2 storage

Figure 8 presents the cost projections for UG H_2 storage technologies from 2023 to 2050. Among the options analysed, salt caverns exhibit the highest percentage decrease in projected costs, with a reduction of 25.9% but this decrease is not significant as the initial LCOHS in salt caverns was already low at $0.27/kg, with the projected reduction lowering the cost to only $0.20/kg. LRC follows with a decrease of 8.29%, from $1.40/kg to $1.29/kg, while depleted reservoirs and aquifers show the least reductions of approximately 5%.

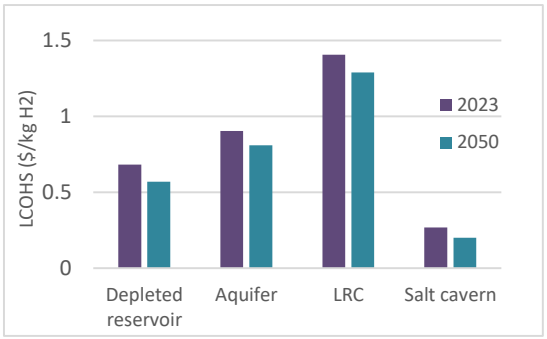


Figure 8. *LCOHS projections for UG storage*

Figure 9 shows the cost projections for AG hydrogen storage from 2023 to 2050. The baseline costs for LH_2, LOHC, and GH_2 are $3.94/kg, $2.80/kg, and $1.93/kg, respectively. The greatest reduction of 6.54% was observed for LOHC due to its higher learning rate of 12%, compared to 2% for both GH_2 and LH_2. This difference is attributed to LOHC is a less mature technology, whereas GH_2 and LH_2 are the most used and established H_2 storage options.

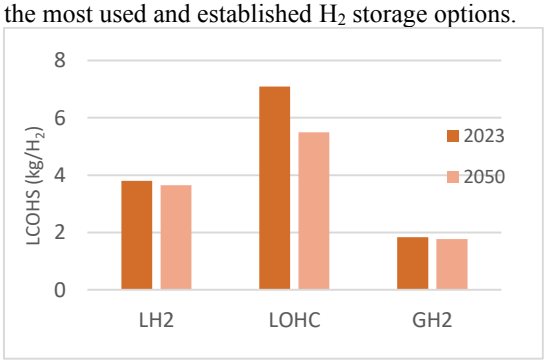


Figure 9. *LCOHS projections for AG storage*

3.6.2. H_2 transportation

Due to the scarcity of data for H_2 trucks' learning rate, it was approximated using the values for fuel cell electric trucks (FCET). The justification for this is that FCET are heavy duty vehicles with H_2 tanks similar to H_2 trucks. Since the LCOHT of H_2 trucks mainly depend on the type of tank to carry the H_2,

they could be considered similar in terms of technology advancements.

The cost projection for pipeline transportation does not show a noticeable change mainly due to its low learning rate at only 3.7%. As discussed previously, H_2 pipelines are already commercially proven with a TRL of 9 thus major improvements from learning-by-doing have already been realised thus reflecting limited opportunities through technology advancement or scale.

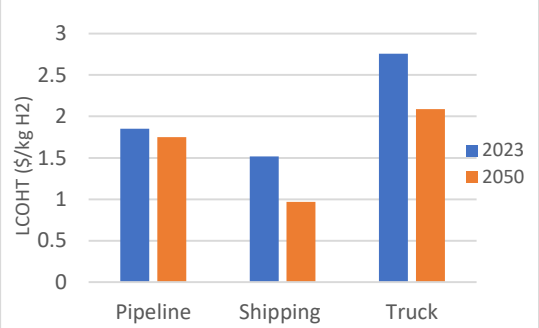


Figure 10. *LCOHT projections*

On the other hand, H_2 shipping and trucks show a dramatic decrease in costs with a reduction of 36% and 24%, respectively. As pointed out earlier for shipping transportation, only NH_3 ships are commercially available. Therefore, the implementation of ships for transporting H_2 may have room for improvement, especially in terms of BOG utilisation as outlined earlier. As for trucks, it was stated that CGH_2 and LH_2 trucks have a maximum TRL of 9 which contradicts the trend for its cost projection. This is because the approximation is made by using the learning rate of FCETs. Nevertheless, this cost reduction and learning rate may still be meaningful because of the technological advancement opportunities that H_2 trucks have such as the improvement of Type 3 and Type 4 trucks for larger H_2 capacity (Xie et al., 2024) and the implementation of H_2 fuel instead of diesel for greener alternatives.

4. Conclusions & Outlook

In this paper, the results of an in-depth systematic review of H_2 storage and transportation technologies are presented, with a particular focus on their economic aspects. The authors reviewed 61 papers from academic literature to gather relevant data for a cost comparison analysis utilising levelised costs and cost projections approach.

Although a targeted search of the literature was implemented, only 16% and 26% of the reviewed papers contained relevant cost data for H_2 storage and transportation, respectively highlighting the challenges in its cost quantification. In general, the values reported in academic literature largely align with those from grey literature, while the overall patterns, such as the relative cost differences between storage options, remain consistent.

UG storage continues to be the most cost-effective option for the large-scale storage of H_2 with an average LCOHS of $0.80/kg. Among the AG storage options, GH_2 represents the least expensive choice but it remains more costly than the costliest UG storage, LRC. For AG storage methods, LCOHS averages at $4.32/kg.

The most cost-effective transportation ultimately depends on the distance and demand. Trucks are mostly used for short distances, pipelines for short to medium-distance and ships for long-range distances. Future studies could examine the change in LCOHT with increasing distance for each technology to observe the transition point of the cheapest technology to utilise. Finally, the cost projections indicate vast cost reductions for H_2 shipping and trucks suggesting innovative improvements could be made which might present a shift in the H_2 supply chain.

The harmonization of data across different storage and transport methods revealed no significant differences. For transport, this has been attributed to the high maturity of the technology across different regions. This also may be an effect of the limited number of studies available, as well as the consistency in the methodologies and input values employed across these studies.

The utilisation of H_2 storage is also a subpart of H_2 transportation as the containment of H_2 is a significant aspect to consider for road and maritime transport. Therefore, understanding the economic requirements for H_2 storage is crucial for transportation as the cost reduction initiatives for H_2 storage would benefit H_2 transportation as well.
Overall, compared to the production costs of blue H_2 and green H_2, which are $3.15/kg H_2 and $4.50/kg H_2 respectively, UG storage adds only a minor expense to the total H_2 cost. However, if AG storage is utilized, it becomes a significant cost factor. Regarding transport, it is relevant compared to the production cost as it averages at $4.00/kg H_2 but this is dependent on distance and demand. Future research should explore various combinations of H_2 storage and transport to identify the most cost-effective supply chain solutions possible.

Supplementary Information

Supplementary information was provided and was sent to Prof. Erich Muller.

References

1. Abdin, Z., Khalilpour, K. & Catchpole, K., 2022. Projecting the levelized cost of large

scale hydrogen storage for stationary applications. *Energy Conversion and Management,* p. 116241.
2. Anon., 2024. *2024 is on track to be hottest year on record as warming temporarily hits 1.5°C.* [Online] Available at: https://public.wmo.int/news/media-centre/2024-track-be-hottest-year-record-warming-temporarily-hits-15degc#:~:text=1.5%C2%B0C-,2024%20is%20on%20track%20to%20be%20hottest%20year%20on%20record,temporarily%20hits%201.5%C2%B0C&text=Baku%2C%20Azerbaijan%20(WMO
3. Bai, F., Zhao, F., Liu, X., Mu, Z., Hao, H. and Liu, Z. (2024). A comparative well-to-wheel analysis of renewable energy pathways for hydrogen and battery electric vehicles. *Journal of Cleaner Production*, [online] 466, pp.142832–142832.
4. Beckman, K., Determeyer, P. & Mowrey, E., 1995. *Beckman, Kenneth L. et al. "Natural gas storage: Historical development and expected evolution. Final report, December 1994-February 1995." (1995).,* Houston, Texas: Gas Research Institute Chicago.
5. Chen, F. et al., 2023. Capacity assessment and cost analysis of geologic storage of hydrogen: A case study in Intermountain-West Region USA. *International Journal of Hydrogen Energy,* pp. 9008-9022.
6. Clean Hydrogen Partnership, 2024. *Hydrogen cost and sales prices.* [Online] Available at: https://h2v.eu/analysis/statistics/financing/hydrogen-cost-and-sales-prices
7. DESNZ, 2023. *Hydrogen Transport and Storage Cost Report,* London: Department for Energy Security & Net Zero.
8. Di Molfetta, R. (2022). *Strategies for regional deployment of hydrogen infrastructure*. [online] Available at: https://upcommons.upc.edu/bitstream/handle/2117/375825/20220405-egi-thesis-report-roberto-di-molfetta.pdf?sequence=1&isAllowed=y [Accessed 28 Oct. 2024].
9. Du, Y. et al., 2018. *Investigation on Standards on Hydrogen Cycle of Composite Tanks for Storage of High Pressure Hydrogen.* Prague, Czech Republic, ASME.
10. Fernández, 2011. *WACC: DEFINITION, MISCONCEPTIONS AND ERRORS,* Barcelona, Spain: IESE Business School.
11. Foh, S., Novil, M., Rockat, E. & Randolph, P., 1979. *UNDERGROUND HYDROGEN STORAGE FINAL REPORT,* Chicago, Illinois: Institute of Gas Technology.
12. GEV (2021). *GEV Scoping Study Delivers Zero Emission Supply Chain for Green Hydrogen*. [online] Available at: https://buyhydrogen.com.au/wp-content/uploads/2021/04/H2-Supply-Chain-Scoping-Study-Issued-GEV-210301.pdf [Accessed 28 Oct. 2024].
13. Gras, C., 2023. *SaltHy: how underground hydrogen storage is supporting the European Hydrogen Backbone.* [Online] Available at: https://hydrogentechworld.com/salthy-how-underground-hydrogen-storage-is-supporting-the-european-hydrogen-backbone
14. Hamidreza Yousefi, S. et al., 2023. Techno-economic analysis of developing an underground hydrogen storage facility in depleted gas field: A Dutch case study. *International Journal of Hydrogen Energy,* pp. 28824-28842.
15. Hu, S., Guo, B., Ding, S., Yang, F., Dang, J., Liu, B., Gu, J., Ma, J. and Ouyang, M. (2022). A comprehensive review of alkaline water electrolysis mathematical modeling. *Applied Energy*, 327, p.120099.
16. Hydrogen Transport and Storage Cost Report. (2023). [online] Department for Energy Security & Net Zero. Available at: https://assets.publishing.service.gov.uk/media/659e600b915e0b00135838a6/hydrogen-transport-and-storage-cost-report.pdf.
17. IEA (2024). *Hydrogen demand – Global Hydrogen Review 2024 – Analysis - IEA*. [online] Available at: https://www.iea.org/reports/global-hydrogen-review-2024/hydrogen-demand [Accessed 18 Nov. 2024].
18. IEA, 2023. *Global hydrogen demand in the Net Zero Scenario, 2022-2050.* [Online] Available at: https://www.iea.org/data-and-statistics/charts/global-hydrogen-demand-in-the-net-zero-scenario-2022-2050
19. IEA, 2024. *Global Hydrogen,* s.l.: International Energy Agency.
20. IMF (2024). *International Monetary Fund* . [online] IMF. Available at: https://www.imf.org/en/Home [Accessed 11 Nov. 2024].
21. Jahanbakhsh, A. et al., 2024. Underground hydrogen storage: A UK perspective. *Renewable and Sustainable Energy Reviews,* p. 114001.
22. Johnston, C., Ali Khan, M.H., Amal, R., Daiyan, R. and MacGill, I. (2022). Shipping the sunshine: An open-source model for costing renewable hydrogen transport from Australia. *International Journal of Hydrogen Energy*.
23. Jorschick, H. et al., 2020. Benzyltoluene/dibenzyltoluene-based mixtures

as suitable liquid organic hydrogen carrier systems for low temperature applications. *International Journal of Hydrogen Energy,* pp. 14897-14906.
24. Kanaani, M., Sedaee, B. & Asadian-Pakfar, M., 2022. Role of Cushion Gas on Underground Hydrogen Storage in Depleted Oil Reservoirs. *Journal of Energy Storage,* p. 103783.
25. Kishor Patanwar, Y., Kim, H.-M., Debasis , D. & Kumar Gujjala, Y., 2024. Underground storage of hydrogen in lined rock caverns: An overview of key components and hydrogen embrittlement challenges. *International Journal of Hydrogen Energy,* pp. 116-133.
26. Lange, J. et al., 2024. Aboveground hydrogen storage – Assessment of the potential market relevance in a Carbon-Neutral European energy system. *Energy Conversion and Management,* p. 118292.
27. Lee, J., Chérif, A., Yoon, H.-J., Seo, S.-B., Bae, J.-E., Shin, H.-J., Lee, C., Kwon, H. and Lee, C. (2022). Large-scale overseas transportation of hydrogen: Comparative techno-economic and environmental investigation. *Renewable & Sustainable Energy Reviews*, 165, pp.112556–112556.
28. Liu, X. et al., 2024. Technical feasibility of converting abandoned calcite mines into lined rock caverns for underground hydrogen storage. *Journal of Energy Storage,* p. 111051.
29. Lord, A. S., Kobos, P. H. & Borns, D. J., 2014. Geologic storage of hydrogen: Scaling up to meet city transportation demands. *International Journal of Hydrogen Energy,* pp. 15570-15582.
30. Melaina, M.W., Antonia, O. and Penev, M. (2013). Blending Hydrogen into Natural Gas Pipeline Networks: A Review of Key Issues.
31. Michanowicz, D. et al., 2017. A national assessment of underground natural gas storage: identifying wells with designs likely vulnerable to a single-point-of-failure. *Environmental Research Letters.*
32. Møller, K. T., Jensen, T. R., Akiba, E. & Li, H.-w., 2017. Hydrogen - A sustainable energy carrier. *Progress in Natural Science: Materials International,* pp. 34-40.
33. Niermann, M., Timmerberg, S., Drünert, S. and Kaltschmitt, M. (2021). Liquid Organic Hydrogen Carriers and alternatives for international transport of renewable hydrogen. *Renewable and Sustainable Energy Reviews*, 135, p.110171.
34. Papadias, D.D. and Ahluwalia, R.K. (2021). Bulk storage of hydrogen. *International Journal of Hydrogen Energy*.
35. Parolin, F., Colbertaldo, P. & Campanari, S., 2022. Development of a multi-modality hydrogen delivery infrastructure: An optimization model for design and operation. *Energy Conversion and Management,* p. 115650.
36. Pietro Dogliani, Ruas, N., Elberry, A.M. and Thakur, J. (2024). Multi-option analytical modeling of levelized costs across various hydrogen supply chain nodes. *International Journal of Hydrogen Energy*, 70, pp.737–755.
37. Raeesi, R. et al., 2024. Hydrogen supply chain and refuelling network design: assessment of alternative scenarios for the long-haul road freight in the UK. *International Journal of Hydrogen Energy,* pp. 667-687.
38. Ren, L., Zhou, S., Peng, T. and Ou, X. (2022). Greenhouse gas life cycle analysis of China's fuel cell medium- and heavy-duty trucks under segmented usage scenarios and vehicle types. *Energy*, 249, p.123628.
39. Reuß, M., Grube, T., Robiniu, M., Preuster, P., Wasserscheid, P. and Stolten, D. (2017). Seasonal storage and alternative carriers: A flexible hydrogen supply chain model. *Applied Energy*, [online] 200, pp.290–302.
40. Reuß, M., Grube, T., Robinius, M. & Stolten, D., 2019. A hydrogen supply chain with spatial resolution: Comparative analysis of infrastructure technologies in Germany. *Applied Energy,* pp. 438-453.
41. Ruffini, E. & Wei, M., 2018. Future costs of fuel cell electric vehicles in California using a learning rate approach. *Energy,* pp. 329-341.
42. Schoots, K., Ferioli, F., Kramer, G. and Vanderzwaan, B. (2008). Learning curves for hydrogen production technology: An assessment of observed cost reductions. *International Journal of Hydrogen Energy*, 33(11), pp.2630–2645.
43. Sehar Tasleem, Bongu, C.S., Krishnan, M.R. and Edreese Housni Alsharaeh (2024). Navigating the hydrogen prospect: A comprehensive review of sustainable source-based production technologies, transport solutions, advanced storage mechanisms, and CCUS integration. *Journal of Energy Chemistry*, [online] 97, pp.166–215.
44. Sofregaz US Inc, 1999. *COMMERCIAL POTENTIAL OF NATURAL GAS,* United States: US Department of Energy (US).
45. Tarkowski, R. (2019). Underground hydrogen storage: Characteristics and prospects. *Renewable and Sustainable Energy Reviews*, 105, pp.86–94.
46. Tarkowski, R., Lankof, L., Luboń, K. & Michalski, J., 2024. Hydrogen storage capacity of salt caverns and deep aquifers versus demand for hydrogen storage: A case study of Poland. *Applied Energy,* p. 122268.
47. Timmerberg, S. & Kaltschmitt, M., March, 2019. Hydrogen from renewables: Supply from North Africa to Central Europe as blend

in existing pipelines – Potentials and costs. *Applied Energy,* pp. 795-809.
48. Tock, L., Maréchal, F. and Perrenoud, M. (2015). Thermo-environomic evaluation of the ammonia production. *The Canadian Journal of Chemical Engineering*, 93(2), pp.356–362.
49. Tsogt, N., Quarme Gbadago, D. & Hwang, S., 2024. Exploring the potential of liquid organic hydrogen carrier (LOHC) system for efficient hydrogen storage and Transport: A Techno-Economic and energy analysis perspective. *Energy Conversion and Management,* p. 117856.
50. UNFCCC, 2015. *The Paris Agreement.* Paris, UNCC.
51. Upstill, G. & Hall, P., 2018. Estimating the learning rate of a technology with multiple variants: The case of carbon storage. *Energy Policy,* pp. 498-505.
52. Van Gerwen, R., Eijgelaar, M. & Bosma, T., 2019. *HYDROGEN IN THE ELECTRICITY VALUE CHAIN,* s.l.: DNV GL.
53. Xie, Z., Jin, Q., Su, G. and Lu, W. (2024). A Review of Hydrogen Storage and Transportation: Progresses and Challenges. *Energies*, [online] 17(16), pp.4070–4070..

Protein-based RNA integrity assay-relationship between mRNA capping ratio and reporter fluorescence

Yujie Mao, Wenqing Ma
Department of Chemical Engineering, Imperial College London, U.K.

Abstract In the realm of industrial vaccine design, the 5' cap of messenger RNA (mRNA) assumes a pivotal role in influencing the efficiency of mRNA vaccines. Here, we describe a fluorescence resonance energy transfer (FRET)--based assay aimed at quantifying the efficiency of mRNA capping across three distinct mRNA lengths. The assay employs pairs of fluorescein isothiocyanate (FITC) dye and yellow fluorescent protein (YFP), facilitating the measurement of fluorescence intensity within the range of 519 nm to 600 nm. A negative correlation was observed between the capping ratio and fluorescence intensity, with shorter mRNA exhibiting heightened sensitivity. Investigations were conducted into additives such as $MgCl_2$ and DTT, which were found to reduce system noise and enhance stability for shorter mRNA; however, limited effects were noted on longer mRNA, possibly due to the intricacies of complex folding structures. These results underscore the assay's promise as a swift and efficient alternative to traditional methodologies for assessing mRNA capping, with significant implications for the quality control of mRNA-based therapeutics. Further enhancements are recommended to amplify sensitivity and reliability, particularly concerning longer mRNA strands.

INTRODUCTION

Messenger RNA (mRNA) serves as a crucial intermediary between encoded DNA and functional proteins. Over the past few decades, the therapeutic use of mRNA has demonstrated considerable advancements. The mRNA vaccines exhibit high efficacy, the ability for swift development, and the potential for economical production as well as safe administration (1). It works by introducing a strand of mRNA that corresponds to a viral protein. Through the translation of this mRNA, cells can synthesise viral proteins, which activate the immune response. The immune system identifies the protein and generates memory cells to safeguard the body against further infections (2). In parallel, researchers exhibit a keen interest in enhancing protein translation, modulating both innate and adaptive immunogenicity, and effectuating improvements in drug delivery (3,4,5).

In this instance, the quality control of mRNA is paramount to the manufacturing process, owing to the inherent instability and susceptibility to mRNA degradation. A key factor for quality control is the integrity of mRNA. In eukaryotic organisms, mRNA is a single-stranded molecule, as shown in Figure 1 (A). At the 5' ends, there exists a structure known as a 7-methylguanylate cap (m^7G), which is composed of a guanine nucleotide and a 5' to 5' triphosphate linkage. At the 3' ends, mRNAs generally feature extended sequences of adenosine residues (polyadenylation) that are not encoded within the DNA sequence, as they are added enzymatically after the transcription process (6). The m7G cap is essential to the biological function of mRNA. It protects the RNA from exonuclease cleavage and serves as an identifier for a cap-binding protein (7). Without the 5' cap, the translation efficiency drops significantly. In the process of RNA manufacture, mRNA is typically synthesised via an *in vitro transcription* (IVT) reaction. The mRNA can be capped by conducting a post-transcriptional reaction or co-transcriptionally introducing a chemical cap analogue into the IVT reaction (8). Thus, assessing the 5' capping is crucial to the quality control of RNA vaccines.

There are a number of analytical methods that evaluate the capping and integrity of mRNA. For instance, fluorescent cap analogues: 5' cap is labelled with a fluorescent marker, which allows for the direct measurement of capped RNA (9). Electrophoresis: RNase H facilitates the cleavage of RNA molecules into smaller fragments, enabling the separation of capped and uncapped RNA based on size differences observed on a polyacrylamide gel (10). qRT-PCR methods: this approach uses a splint to bind near the RNA's 5' end, creating an overhang. An anchor oligonucleotide attaches to the overhang, and uncapped RNA is ligated, adding a detectable sequence for RT-qPCR. Capped RNA cannot be ligated and is not detected (11). Nanopore sequencing: this method offers a platform for measuring the sequence of RNA (12). High-Resolution Liquid Chromatography–Mass Spectrometry(LC/MS): The RNA is cut into small fragments by RNase H and analysed by the Agilent 6545XT AdvanceBio LC/Q-TOF. The entire process takes about 75 minutes (13). Notwithstanding, inevitable drawbacks are associated with the methods enumerated. Some of these methods exhibit low accuracy, while others necessitate intricate sample preparation and extended processing times. Additionally, some of these methods require significant quantities of materials or compromise the integrity of the RNA structure. Consequently, developing a simple and fast assessment approach is paramount for the quality control of RNA vaccines.

Optical biosensor assays emerge as a promising assessment method for mRNA quality control attributable to its sensitivity, non-invasive capabilities, quantitative measurements, and versatile applications. The biosensor described here employs a pair of fluorophores. One fluorophore is an engineered chimeric protein, which consists of a fusion between the murine eIF4E protein and yellow fluorescent protein (YFP). The second fluorophore is a fluorescein isothiocyanate (FITC)-bound oligonucleotide. When the

biosensor is mixed with the RNA sample. The oligonucleotide binds to the mRNA through complementary base pairing, while the eIF4E protein binds to the mRNA cap structure. Förster resonance energy transfer (FRET) may occur when the donor chromophore (FITC) is excited by photonic energy at wavelengths of 499 nm, subsequently emitting light at wavelengths of 518 nm.

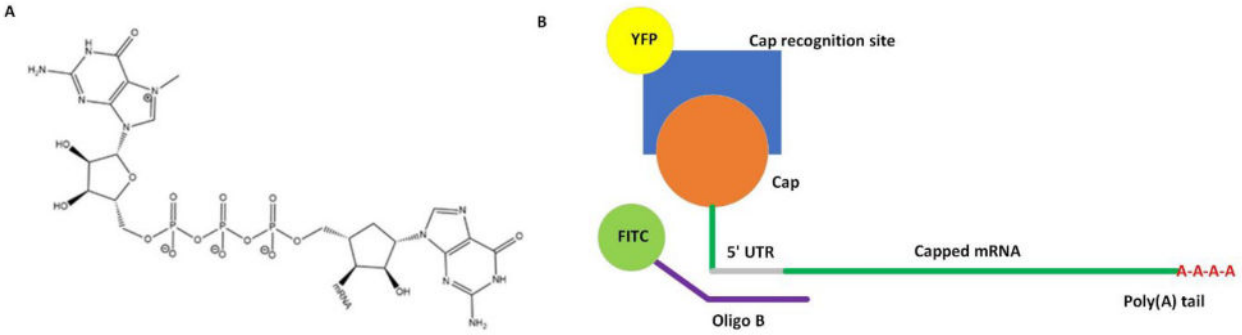


Figure 1 (A) The chemical structure of m^7G. (B) The mechanism of FRET test. A fusion between the eIF4E protein and yellow fluorescent protein (YFP) binding to the cap. FITC-bound oligonucleotide attached to RNA. When the donor (FITC) is excited by photonic energy, the energy may transfer to the acceptor (YFP) via non-nonradiative dipole-dipole coupling

The acceptor (YFP) absorbs light at 516 nm, which is close to the emission wavelength of the donor, facilitating energy transfer through non-radiative dipole-dipole coupling. The efficiency of this energy transfer is inversely proportional to the sixth power of the distance between the donor and the acceptor, which typically occurs at distances ranging from 10 to 100 Å (14). Uncapped RNA lacks a binding site for eIF4E; thus, the occurrence of FRET is improbable. The capping percentage can be quantitatively assessed through the measurement of fluorescence intensity. The assay duration is approximately 20 minutes, which is much shorter compared to the gold standard method (in the best case scenario), which takes 75 minutes. Version 1 takes 210 minutes (another optical biosensor developed in the Polizzi lab (15)). Furthermore, this method can be implemented in either a 96-well or 384-well plate format, providing a high-throughput approach that may enhance process optimisation. The objective of the research is to investigate the correlation between the capping percentage and the fluorescence intensity, which is expected to demonstrate a linear relationship. Additionally, the study will assess the accuracy with various lengths of RNA. The suggested method offers a straightforward, low-tech way to assess the capping status of mRNAs in one step, eliminating the necessity for labelling, amplification, or any physicochemical alteration of the nucleotide chain. Moreover, The adaptability and feasibility of standardisation inherent in this approach render it suitable for a wide array of applications, encompassing comprehensive transcription studies as well as quality control measures at both industrial and end-user levels.

MATERIALS AND METHODS

Chemicals

The chemicals used in this experiment include LB medium, various salts, β-5D-1-thiogalactopyranoside (IPTG), dithiothreitol (DTT), bovine serum albuminates (BSA), and phenylmethylsulfonyl fluoride (PMSF). Additionally, wheat germ tRNA and 2-mercaptoethanol (2-ME) were sourced from Sigma Aldrich (UK). Tween-20 was obtained from Fisher Scientific (UK), and Nitrocefin from Thermo Scientific (UK). Oligonucleotides were supplied by Invitrogen (UK), while restriction enzymes came from New England Biolabs (NEB, UK).

Cell growth

Escherichia coli DH5α (NEB). Cells harbouring the pRSET-T1, pRSET-T2 and pRSET-T3 plasmids were cultured in an LB medium with Ampicillin for 8 hours at 37°C with vigorous shaking for screening. The remaining culture was diluted and grew at 37°C overnight.

DNA extraction

The Qiagen Maxiprep Kit (UK) was utilised for all plasmid purifications, while the DNA Clean & Concentrator Kit (Zymo Research, UK) was used to purify linear DNA fragments.

The bacterial cells were harvested by centrifugation at 6000g for 15 minutes at 4°C and then resuspended in 10 ml of Buffer P1. The suspension was mixed with 10 ml of Buffer P2 and incubated at room temperature for 5 minutes. Next, 10 ml of chilled Buffer P3 was added to lyse the cells, and the lysate was transferred into the barrel of the QIAfilter, where it was incubated at room temperature for 10 minutes. Samples of the lysate were extracted for gel analysis. A HiSpeed Maxi Tip was equilibrated by applying 10 ml of Buffer QBT and allowing the buffer to flow through the column by gravity. The plunger was then inserted into the QIAfilter Maxi to filter the lysate into the HiSpeed Tip. The

clarified lysate was collected. Following this, the Tip was washed with 60 ml of Buffer QC (labelled as "wash"), and the DNA was eluted using 15 ml of Buffer QF (labelled as "elute QF"). DNA precipitation was performed by adding 10.5 ml of isopropanol to the eluate, followed by incubation for 5 minutes at room temperature. The mixture was then transferred to the QIAprecipitator, which was washed with 2 ml of ethanol and dried with air. Finally, the eluted DNA was recovered using 1 ml of Buffer TE (labelled as "elute TE") and labelled T1, T2, and T3 according to their size. The concentration and gel of DNA are shown in Figure 2 (A).

mRNA preparation

Before conducting the *in vitro transcription* reaction, the DNA prepared needs to be linearised by using the restriction enzyme *NotI-HF* (NEB). 10 µg of template DNA (based on concentration shown in Table 1) of T1.T2 and T3 were mixed with 4 µl of rCutSmart (NEB) and 4 µl of *NotI-HF*, topped up to 40 µl with water, respectively. The reaction was incubated at 37°C overnight (the gel analysis of linearisation is shown in Figure 2 (B)). For capped RNA synthesis, an anti-reverse cap
analogue (ARCA) was incorporated using the HiScribe™ T7 ARCA kit (NEB). 1 µg of linearised DNA T1, T2, and T3 were mixed with 2 µl of T7 RNA polymerase Mix, 1 µl DTT and nuclease-free water to a final volume of 20 µl, respectively. The *in vitro reaction* was incubated at 37°C overnight. The same protocol was applied to uncapped RNA, except that the NTP mix was used instead of ARCA. The template DNA was removed by adding 2 µl of DNase I; afterwards, the mixture was incubated at 37°C for 15 minutes. The RNA necessitated Poly(A) tailing before the initiation of FRET. A volume of 20 µl of *in vitro transcription* reaction for both capped and uncapped RNA was combined with 5 µl of Poly(A) Polymerase and Reaction Buffer, then adjusted with nuclease-free water to achieve a final volume of 50 µl, respectively. The reaction was incubated at 37°C for 30 minutes. The RNA Clean & Concentrator kit (Zymo Research) was used to purify the transcribed RNAs. 25 µl LiCl solution was mixed with 50 µl tailing reaction under -20°C overnight. The resulting pellet RNA was rinsed with 500 µl of cold 70% ethanol, dried, and resuspended in 50 µl nuclease-free water for storage purposes. The transcribed RNA had sizes of 255 nt, 1052 nt, and 2700 nt for T1, T2, and T3, respectively. The concentration is shown in Table 1, and the gel analysis is shown in Figure 2 (C).

Protein Production

Two colonies of BL21 cells containing the pET28-elF4E-avi plasmid were inoculated in an LB medium with 50 µl/ml kanamycin at 37°C overnight with vigorous shaking. The initial inoculum cultures were diluted in 100 ml LB into 1 L shake flasks. Cells were shaken and grown at 37°C until OD600 reached 0.4. Samples were collected for SDS-PAGE gel. Cell Pellets were harvested by centrifugation at 3220g at 4°C for 20 minutes. Resuspended with lysis buffer (20 mM Tris-HCl, 6M guanidine hydrochloride, 10% glycerol, 10 mM imidazole, pH 8). The suspension was mixed with 1 mg/ml lysozyme (in isopropanol) and 100 mM PMSP (in Tris pH 8) and incubated for 30 minutes at 4°C. Sonicated the mixture on ice for 8 minutes (10 s ON/10 s cooling bursts, 40% amplitude) with a sonicator. The insoluble fraction was pelletised by centrifugation at 16000g for 30 minutes at 4°C. The fraction was incubated with 1 ml 50% resin (Ni-NTA agarose, Qiagen) for 1 hour at 4°C. After that, the resin was packed into a disposable polypropylene column. Washed the column with 10 ml of wash buffer 1 (20 mM Tris, 20 mM imidazole, 8 M urea, 10% glycerol, pH 8) and 50 ml of wash buffer 2 (20 mM Tris, 20 mM imidazole, 1 M NaCl, 10% glycerol, pH 8). Collected the flow-through for gel analysis. The gel analysis is shown in Figure 2 (C). Finally, Eluted the protein with 15 ml elution buffer in a 50 ml falcon tube (20 mM Tris, 1 M NaCl, 10% glycerol, 250 mM imidazole, pH 8). The eluate was then concentrated using a protein concentrator with a 10 kDa MWCO by centrifuging at 3200g for 15 minutes. Buffer exchange was performed by running the same volume of protein storage buffer (HEPES 50 mM, KCl 100 mM, 10% glycerol, pH 7.4) three times, and the final retentate reached between 0.75 and 1 ml.

FRET assay

The mixture for the FRET assay was composed of 0.45 µM of protein, 0.5 mM of RNA and 0.5 mM of DNA oligo-FITC probe for each sample. The volume of capped RNA and uncapped RNA were calculated as shown below:

$$vol.capped\ RNA = \frac{capping\ ratio \times final\ con \times reation\ vol.}{capped\ RNA\ concentration}$$

$$vol.uncapped\ RNA = \frac{(1 - capping\ ratio) \times final\ con \times reation\ vol.}{uncapped\ RNA\ concentration}$$

The capping ratio for RNA was from 0 to 100%, with an increment of 25%. There is an assumption that the product of the IVT reaction is 100% capped. The calculation of the protein and the DNA oligonucleotide was conducted based on their respective molecular weights and concentrations. A master mix was prepared to mitigate pipetting errors. The protein is added in excess (900:1 protein:mRNA ratio), and oligo is added in 1:1 oligo:mRNA molar ratio. This was labelled master mix 1. Capped and uncapped RNA was subsequently mixed to create master mix 2. The master mix 2 was evenly distributed to 3 or 4 samples, and buffer (HEPES 50 mM, KCl 100 mM, pH 7.4) was added to ensure a 100 µl reaction volume for each sample. The repeat of the samples was intended to analyse the internal reproducibility/noise of the assay.

Finally, the master mix 1 was distributed in the 96-well plate. The same applied to master mix 2. Two mixes

were gently agitated in the well. Each sample exhibiting an identical composition was replicated four times. Additionally, three control groups were established: one comprising protein combined with 100% capped mRNA, another consisting solely of protein, and the final group consisting of oligo combined with mRNA. The assay was conducted four times, and data was collected for analysis purposes. MgCl2 and DTT were used to investigate the experimental noise.

RESULTS AND DISCUSSION

This study aims to evaluate fluorescence patterns across varying capping ratios to characterise a FRET-based RNA capping percentage assay. Additionally, the dynamic range and signal-to-noise ratio are assessed across different RNA lengths. The mRNA samples tested (T1, T2, and T3) consist of 255, 1,052, and 2,700 base pairs, respectively.

To achieve this objective, the study also examines DNA extraction and in vitro transcription (IVT) processes to ensure
the quality and consistency of capped and uncapped RNA for analysis. The relationship between capping ratios, RNA length, and fluorescence intensity is analysed to optimise the assay's performance for accurately determining RNA capping percentages.

DNA Extraction Yield and Gel Analysis

DNA extraction through maxi prep resulted in 1 mL of DNA from each sample (T1, T2, T3). While the anticipated OD600 range for cell culture is between 2 and 3, the recorded OD values for T1, T2, and T3 were 10.08, 9.89, and 8.93, respectively. The DNA concentrations were determined to be 409.8 µg/mL for T1, 427.4 µg/mL for T2, and 625 µg/mL for T3 using Biodrop. The DNA gel shown in Figure 2 (A) illustrates that pure DNA exhibits three distinct bands on the gel: nicked, linear, and supercoiled, arranged from top to bottom. The graph indicates the presence of more than three bands in both the lysate and clarified lysate samples. The lowest band corresponds to the circular and single-stranded plasmid generated during the lysis process. This band is also present in the flow-through and wash samples, which is logically consistent. However, it is absent in the elution sample, indicating that the DNA is pure (16).

IVT Yield Analysis

The gel analysis of linearisation is presented in Figure 2 (B). The image of the gel shows that the DNA has been reduced from three distinct bands to a singular band, indicating that the linearisation process has been executed effectively (16).
The mRNA concentration from the IVT reaction is shown in Table 1.

Table 1 m RNA concentration for the first, second, and third rounds.

	The first round		**The second round**		**The third round:**	
Sample µg/ml	capped	uncapped	Capped	uncapped	Capped	uncapped
T1	148.8	16.8	80.8	113.6	378.4	2411.2
T2	452.8	376	524.8	143.2	412	2442.4
T3	376.2	296.8	547.2	190.4	636	2441.6

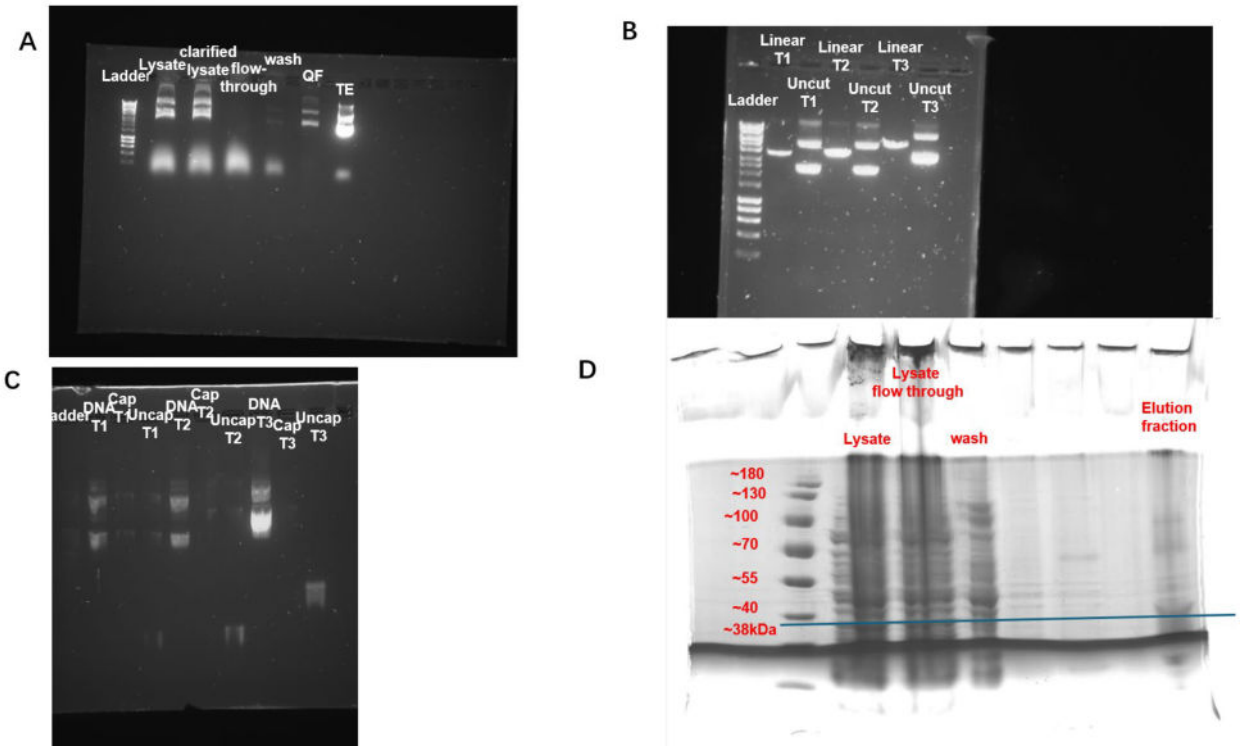


Figure 2 (A) The DNA gel for samples obtained during the purification process. From left to right were Ladder, Lysate, clarified lysate, flow-through, wash, elution (QF), and elution (TE). (B) The DNA gel for linearised and uncut DNA. From left to right were Ladder, Linearised T1, Uncut T1, Linearised T2, Uncut T2, Linearised T3, Uncut T3. The ladder is HyperLadder 1kb. (C) RNA gel for capped and uncapped RNA and DNA. From left to right were Ladder, DNA T1, capped RNA T1, uncapped RNA T1, DNA T2, capped RNA T2, uncapped RNA T2, DNA T2, capped RNA T2, uncapped RNA T2. (D)TheProtein gel

As indicated in the manufacturer's manual. All kit components should be stored at –20°C. Each kit contains sufficient reagents for 20 reactions of 20 µl each. Each standard reaction yields up to 20 µg of capped mRNA from 1 µg control template. Up to 25 µg capped and tailed mRNA can be obtained after poly(A) tailing and purification by LiCl precipitation. The expected yield for IVT is a response to ≥90 µg of RNA, with the hallmark of good yield being a final concentration of ~280 µg/mL. T2 and T3 made good progress, and T1 fell below this threshold. DNA linearisation efficiency (0.025 µg/µL DNA for 7000 µg/mL mRNA) and DNA mRNA yield ratio (~2.8 µg mRNA per µg DNA) confirm the change in T1 performance. Based on rapid LC/MS analysis and capping, the efficiency is approximately 90 %, but between samples, Additional validation is recommended to confirm the foundations. The capping efficiency was estimated to be greater than 90% (17). The faint bands and absence of bands on the agarose gel .2 (C) for RNA suggest that either the detection method was not sensitive enough or insufficient masses of RNA were loaded to be detected. The stain used was SYBR-Safe, which is primarily for RNA detection, which can be improved by the use of an RNA-specific stain, e.g. SYBR Green II (21). This observation suggests that the majority of the mRNA has been successfully capped. However, the exact capping percentage was undetermined. A trend of decrease in yield for shorter mRNA or DNA is observed; this might be caused by shorter mRNA having a less binding site. Therefore, less mRNA is bonded to the column to be restored.

FRET Analysis

Typically, fluorescence absorption peaks occur at 529nm for the dye pair FITC and YFP, but in this experiment, peaks shifted slightly to 523–527 nm. This blue shift can be attributed to variations in buffer pH, ionic strength, or interactions between the mRNA and fluorophores. These factors may have

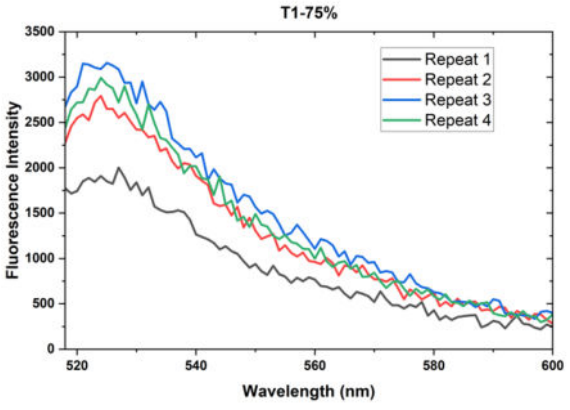


Figure 3 Fluorescence intensity reading for T1, 75% ratio capping, consists of 4 experimental repeats.

influenced the absorption properties of the samples. The FRET intensity was measured at the highest observed intensity for each sample for consistency, and this wavelength has the most efficient photonic energy transfer to the donor molecule. For instance, in T1 at 75% capping, the highest intensity occurred at a wavelength around 525 nm. The average is calculated between the non-anomaly results; for this instance, we take the intensity reading as the average for repeats 2,3 and 4 at the wavelength of 525nm. The highest peak wavelength, 525nm, remained consistent across all mRNA lengths and capping percentages. For each capping ratio (0%, 25%, 50%, 75% and 100%), 3–4 replicates were performed. Any obvious outliers, such as repeat 3, are excluded from the analysis. Any data omitted has been indicated throughout the subsequent analysis. The average FRET intensity was then calculated at the peak wavelength of 525 nm using the remaining replicates.

Control experiment

Typically, fluorescence absorption peaks occur at 529nm for the dye pair FITC and YFP, but in this experiment, peaks shifted slightly to 523–527 nm. This blue shift can be attributed to variations in buffer pH, ionic strength, and interactions between the mRNA and fluorophores. These factors may have influenced the absorption properties of the samples.

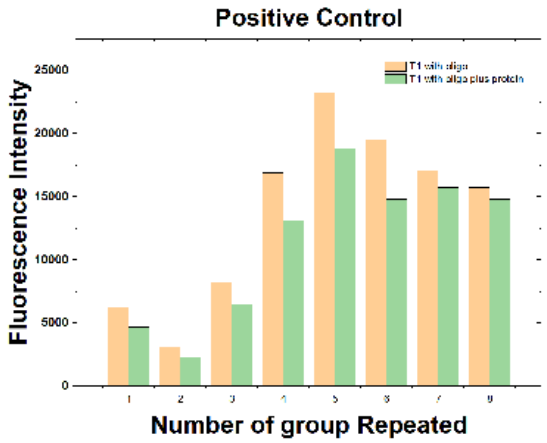


Figure.4 Control experiment: The orange bar is for RNA T1 + oligo probe with attached FITC dye; the green bar is for after adding the YFP-GS-eIF4E protein. All eight samples are identical.

From Figure 4, a decrease in fluorescence intensity is observed for every sample across different oligo concentration ranges. This enables the control. A positive control experiment is conducted to ensure that the fluorescence intensity decreases upon the addition of protein with the acceptor dye. A decrease in fluorescence intensity should be observed when adding the protein to the sample that has capped mRNA and oligo; this decrease in intensity represents effective energy transfer between the dyes. Experiment with capped mRNA with oligo only to act as a baseline measurement. Any fluorescence readings that are higher than this baseline would suggest human error or experiment noise; this proves a reference point for evaluating the accuracy and reliability of the result. The variation in fluorescence intensity is high among each sample; a master mix is introduced to reduce this variation.

Fluorescence result

The table below shows the fluorescence intensity reading across different capping ratios for different mRNA lengths. Figure X shows a capping ratio on the x-axis and fluorescence intensity on the y-axis with a standard deviation error bar on each data point. A negative correlation between the capping ratio and fluorescence intensity is observed. This is expected as a higher capping ratio would result in more binding of the elF4E protein to the 5' end of mRNA. This brings the dye pairs in close proximity, resulting in more energy being transferred to the acceptor dye. The R square values for T1, T2 and T3 are 0.7023, 0.438 and 0.8938. This R square value suggests that T1 and T3 have a stronger linear relationship between capping ratio and fluorescence intensity compared to T2

Table 2 Fluorescence intensity readings and standard deviations for various capping ratios and mRNA lengths.

	Fluorescence intensity			Standard deviation		
Capping Ratio(%)	T1 Avg	T2 Avg	T3 Avg	T1 Avg	T2 Avg	T3 Avg
100	1642.3	2700.6	1310.1	153.52	635.67	365.46
75	2907.6	2556.4	1827.3	207.64	22.28	529.47
50	3505.1	3119.6	1904.6	556.11	332.97	267.36

25	3441.4	2794.5	2637.7	385.06	215.00	218.05
0	3851.9	2380.8	2736.1	747.40	265.45	450.29

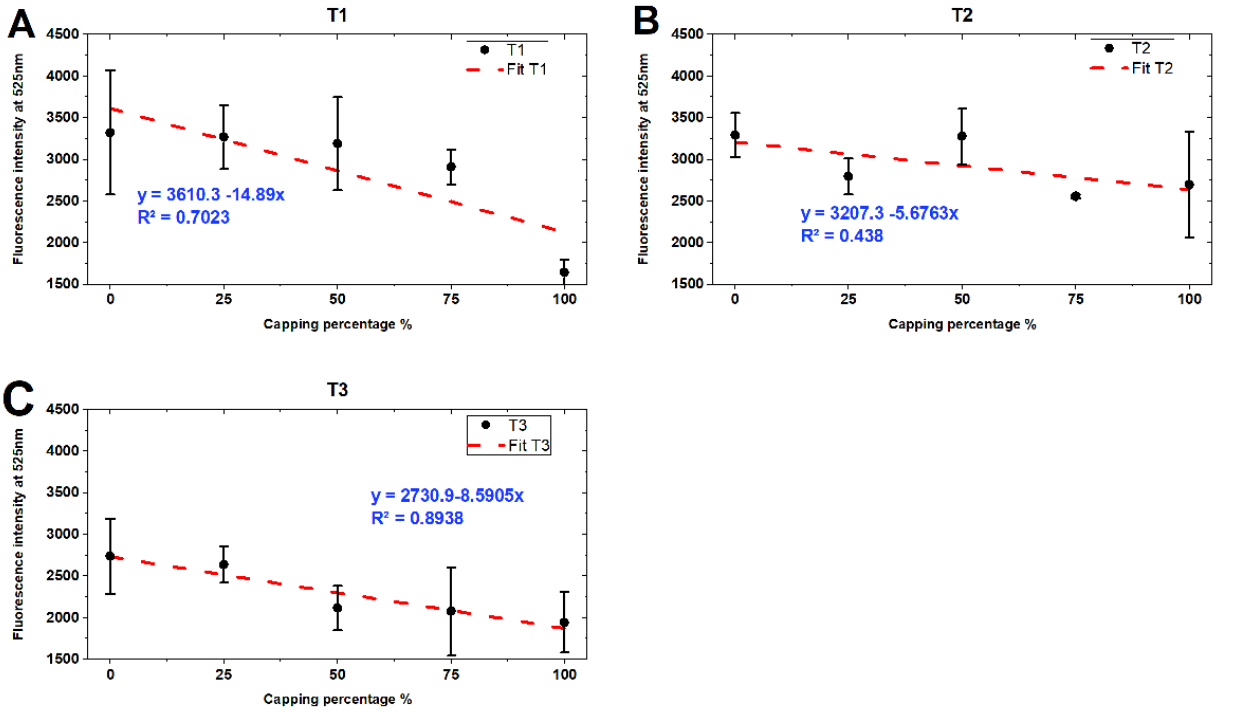


Figure 5 Fluorescence intensity reading for different capping ratios (A) T1, (B) T2, and (C) T3 while other conditions are being kept consistent. A standard deviation for each data point is calculated and demonstrated by the error bar. A regression line is plotted.

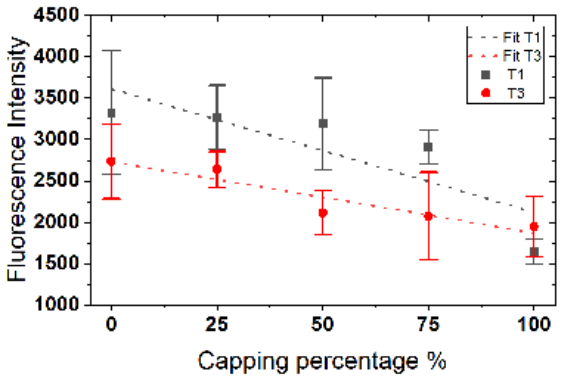


Figure 6 Combined graphs A and C from Figure 5.

T1 and T3 are then put on the same graph for further comparison. Figure 6 shows that T1 always shows higher fluorescence intensity than T3.T1 has a steeper regression line gradient than T3, which indicates a faster decrease in fluorescence intensity. The longer mRNA T3 may form a more stable secondary structure or have a more complex folding compared to shorter mRNA. A steric hindrance effect is created due to mRNA folding; thus, the longer the mRNA, the higher the possibility of a steric hindrance effect to the 5' end of the mRNA; this increases the distance between the mRNA cap and the bounded protein. As a result, the energy transfer between the dye pair could be less efficient compared to a shorter mRNA with less steric hindrance effect. Moreover, T1 has a higher intensity reading than T3 across a capping percentage of 25-75% ratio with no standard deviation error bar being overlapped. This further suggests that the length of mRNA could influence the distance between the dye pair. A statistical T-test is needed to analyse the difference between the two mRNA lengths in order to confirm the significant difference between them.

Signal to noise

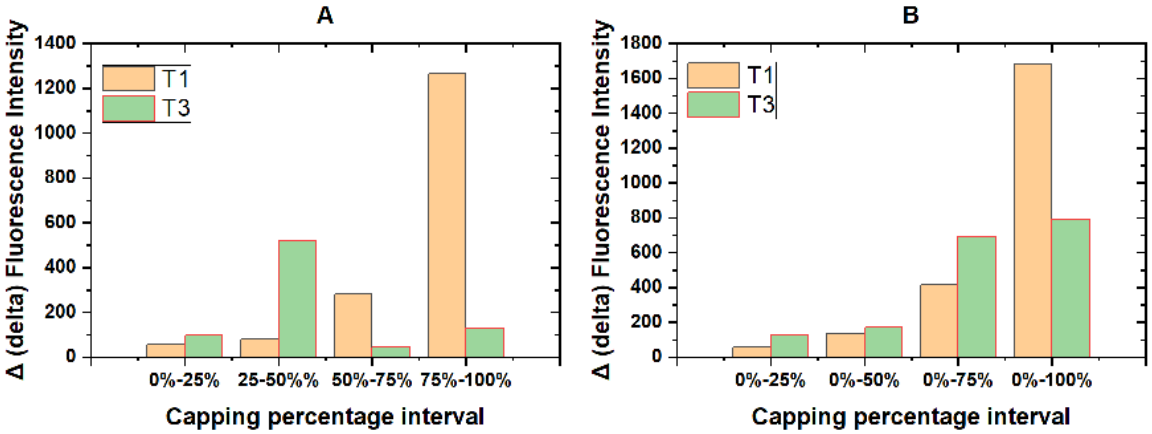


Figure 7 The graphs illustrate the Δ (delta) fluorescence intensity, calculated by subtracting fluorescence values based on the capping percentage intervals shown on the x-axis. The baseline is set at 0% capping, and the delta values represent the changes in fluorescence relative to this reference point. In the A graph, the differences are calculated between consecutive intervals (e.g., 25%-50%), highlighting the sensitivity of T1 and T3 to incremental capping changes. The cumulative graph B shows fluorescence intensity changes from the baseline, 0% capping, to each interval, demonstrating the overall fluorescence behaviour across the full capping range.

The delta difference varies for consecutive delta change graphs A, which could suggest either human error, systematic noise, or a stability threshold that the solution has reached. However, increasing the capping percentage analysed in smaller increments could help to evaluate the potential causes. Cumulatively, T1 exhibits a steeper overall response, with a total fluorescence change of 2209.63, compared to 1907.43 for T3. These results demonstrate that shorter mRNA may be more sensitive to capping ratios due to the dyes being closer together and having less steric hindrance. This paper also indicates that the existence of steric hindrance at the cap can limit the degree of entropy increase during binding, affecting the efficiency and strength of the interaction. Thus, the more the steric hindrance for longer mRNA, the lower the entropy change upon binding and the weaker the interactions between the protein and mRNA

Dynamic value analysis

Table 3 T1 and T3 dynamic range metrics

	T1	T3
0-100%/0-25%	3.783016	6.1514174
biggest/smallest	15.75384	12.324879

The table compares two dynamic range metrics for T1 and T3, offering insights into their fluorescence responses to structural changes. The 0%-100% / 0%-25% dynamic range evaluates total fluorescence change relative to the initial interval. T3 has a value of 6.15, whereas T1 has a value of 3.78. This suggests that T1 is more likely to experience a saturation in the interval of 0-25% compared to T3 due to its lower value. The biggest/smallest dynamic range establishes the difference between the largest and smallest changes. The values for T1 and T3 only have a 25% difference, where both mRNAs have a value that is greater than 10. This indicates that both mRNA have a high signal-to-noise ratio. However, T3 still has a lower value for the biggest/smallest dynamic value, indicating T3 experiences a broader distribution in response.

Unpaired T-test

Table 4. P values for each interval

	T1	T2	T3
0%-25%	0.9042	0.0709	0.5532
0%-50%	0.8274	0.7372	0.0776
0%-75%	0.4116	0.0154	0.0954
0%-100%	0.032	0.3856	0.1436

The T-test analysis evaluates the significance of the difference in fluorescence intensity across intervals relative to the baseline(0% capping ratio). For T1, a significant difference ($p<0.05$) was observed in one interval (0-100%) only. The results showed a high degree of overlap in data over intervals, which is mainly due to the significant standard deviation in T1 readings, as shown by Table X under the florescence section. T2 has a significant difference of $p<0.02$ in the 0-75% interval and with a $p<0.075$ in the 0-25% interval. T2 has the lowest standard deviation among all other lengths, which resulted in higher sensitivity to detect the significant differences. However, the low R^2 value (0.438) indicates poor consistency across the intervals, leading to the observed significance being unreliable and inconsistent. In the case of T3, trends towards significance ($p<0.15$) were observed in all intervals apart from 0-25%. This suggests a potential trend; however, the high level of noise in the system diminutesished the reliability of the trend. Reducing data variation would enhance the reliability of the results

for future analysis. This can be done by reducing the system noise and increasing the sample size to improve statistical power.

Additives and standard deviation

A second round of FRET analysis is carried out, and additives are added to reduce the noise level of the experiment (19). Protocol 2 consists of magnesium ions, which stabilise the binding of elF4E and mRNA by reducing local disorder folding packing of the mRNA backbone (20). DTT is also added; it maintains a reducing environment by preventing the formation of disulfide bonds, thus acting as a stabiliser for protein (18). This ensures the correct conformation of the protein elF4E with the mRNA 5' cap.

The standard deviation decreased upon the addition of the additives by reducing the noise level of the experiment. This reduction in noise level is higher for shorter mRNA molecules. Meanwhile, for longer mRNA, such as T3, there is no significant reduction of the standard deviation. Longer mRNA has a more stable structure and more complex folding; the addition of additives may not be effective enough to overcome the exited interactions between the mRNA and the protein. Moreover, the accessible site of additive attachment may be hindered within the structural folding, thus resulting in less additive binding per unit of mRNA compared to shorter mRNA. Overall, these will result in a less pronounced effect of additives for longer mRNA.

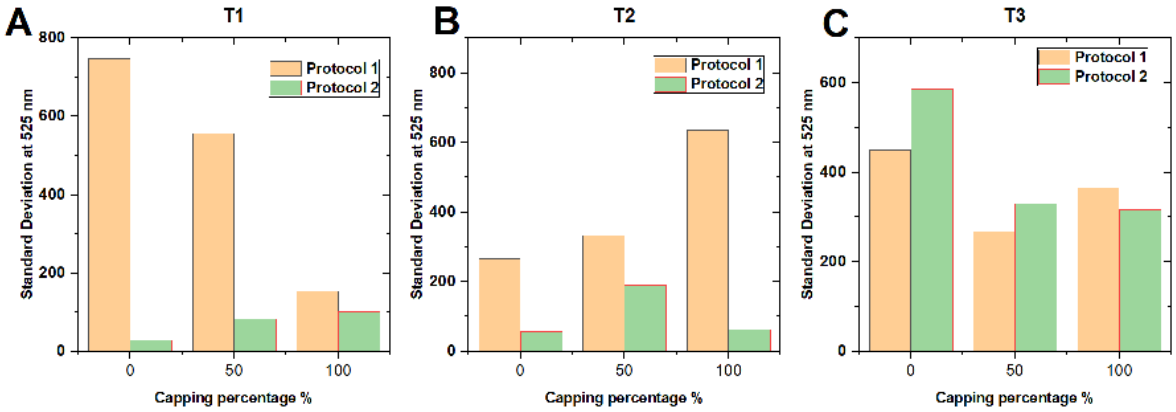


Figure 8 Standard deviation for 3 different capping percentages: 0% ,50% and 100%. The graphs have the y-axis being the standard variation for fluorescence intensity for each protocol and the x-axis being the 3 different capping ratios. Graphs A, B, and C represent the three different mRNA lengths: T1, T2 and T3. The orange bar represents Protocol 1, and the green bar presents Protocol 2, with the addition of additives.

Future recommendation:

Many adjustments could be made to improve the reliability and accuracy of the results. Firstly, use a higher mRNA concentration. This current study used a concentration of 50 nm, whereas previous studies have used 500nm and achieved significance in the result obtained. The choice of 50nm was due to the low yield from in vitro transcription. To address this, finding the optimum transcription and elution temperature for the process can be beneficial. Optimising the oligo: mRNA molar ratio could also lead to an increase in the sensitivity of the FRET analysis. The system's performance has been shown to have a strong sensitivity to the amount of oligo used.

CONCLUSION

This study demonstrates the application of a FRET-based mRNA integrity assay for finding the relationship between mRNA capping ratio and fluorescence intensity across 3 different mRNA lengths. The result suggests a decreasing relationship; as the ratio of capped mRNA increases, the fluorescence intensity decreases. This illustrates the possibility of assessing mRNA capping efficiency using FRET, which is a much faster approach compared to the existing technologies. This study also suggests that steric hindrance may play a crucial role, and complex folding structures in longer mRNA may influence the binding of the protein elF4E and mRNA. The addition of chosen additives reduced the noise but had a minor effect on longer mRNA length. This FRET-based assay provides a rapid, non-resource intensity and easier handling to quantify the mRNA capping ratio. This offers significant potential for integration into the mRNA vaccine manufacturing process and widens the opportunities for faster, simpler mRNA quality control.

ACKNOWLEDGEMENTS

We acknowledge Professor Karen Polizzi and his group for kindly providing the lab. And we want to express our genuine appreciation to Chileab Redwood-Sawyerr for his assistance, time, and wisdom throughout this research project.

REFERENCE

1. Pardi, N., Hogan, M.J., Porter, F.W. and Weissman, D. (2018). mRNA vaccines — a new era in vaccinology. *Nature Reviews Drug Discovery*, [online] 17(4), pp.261–279. Available at: https://www.nature.com/articles/nrd.2017.243.
2. MedlinePlus (2022). What are mRNA vaccines and how do they work? [online] medlineplus.gov. Available at: https://medlineplus.gov/genetics/understanding/therapy/mrnavaccines/.
3. Liang, X., Shen, W., Sun, H., Migawa, M.T., Vickers, T.A. and Crooke, S.T. (2016). Translation efficiency of mRNAs is increased by antisense oligonucleotides targeting upstream open reading frames. Nature Biotechnology, 34(8), pp.875–880. doi:https://doi.org/10.1038/nbt.3589.
4. Lee, Y., Jeong, M., Park, J., Jung, H. and Lee, H. (2023). Immunogenicity of lipid nanoparticles and its impact on the efficacy of mRNA vaccines and therapeutics. Experimental & Molecular Medicine, [online] pp.1–12. doi:https://doi.org/10.1038/s12276-023-01086-x.
5. Paunovska, K., Loughrey, D. and Dahlman, J.E. (2022). Drug delivery systems for RNA therapeutics. Nature Reviews Genetics, 23. doi:https://doi.org/10.1038/s41576-021-00439-4.
6. Rogers, K. (2019). Messenger RNA | genetics. In: Encyclopædia Britannica. [online] Available at: https://www.britannica.com/science/messenger-RNA.
7. Ramanathan, A., Robb, G.B. and Chan, S.-H. (2016). mRNA capping: biological functions and applications. Nucleic Acids Research, [online] 44(16), pp.7511–7526. doi:https://doi.org/10.1093/nar/gkw551.
8. Tu, Y., Das, A., Redwood-Sawyerr, C. and Polizzi, K.M. (2024). Capped or uncapped? Techniques to assess the quality of mRNA molecules. Current Opinion in Systems Biology, [online] 37, p.100503. doi:https://doi.org/10.1016/j.coisb.2023.100503.
9. Marcin Ziemniak, Mariusz Szabelski, Lukaszewicz, M., Nowicka, A., Darzynkiewicz, E., Rhoads, R.E., Wieczorek, Z. and Jacek Jemielity (2013). Synthesis and evaluation of fluorescent cap analogues for mRNA labelling. RSC Advances, [online] 3(43), pp.20943–20943. doi:https://doi.org/10.1039/c3ra42769b.
10. Chan, S.-H., Whipple, J.M., Dai, N., Kelley, T.M., Withers, K., Tzertzinis, G., Corrêa, I.R. and G. Brett Robb (2022). RNase H-based analysis of synthetic mRNA 5′ cap incorporation. RNA, 28(8), pp.1144–1155. doi:https://doi.org/10.1261/rna.079173.122
11. Blewett, N., Coller, J. and Goldstrohm, A. (2011). A quantitative assay for measuring mRNA decapping by splinted ligation reverse transcription polymerase chain reaction: qSL-RT-PCR. RNA, 17(3), pp.535–543. doi:https://doi.org/10.1261/rna.2436411.
12. Wang, Y., Zhao, Y., Bollas, A., Wang, Y. and Au, K.F. (2021). Nanopore sequencing technology, bioinformatics and applications. Nature Biotechnology, [online] 39(11), pp.1348–1365. Available at: https://www.nature.com/articles/s41587-021-01108-x.
13. Liau, B. (n.d.). Rapid Analysis of mRNA 5'-Capping with High-Resolution LC-MS Poster Reprint. [online] Available at: https://storage.googleapis.com/labrulez-bucket-strapi-h3hsga3/po_6545_asms_2021_fp568_en_agilent_c2ea785297/po-6545-asms-2021-fp568-en-agilent.pdf [Accessed 12 Dec. 2024].
14. Wikipedia Contributors (2024). Förster resonance energy transfer. Wikipedia.
15. Moya-Ramírez, I., Bouton, C., Kontoravdi, C. and Polizzi, K. (2020). High resolution biosensor to test the capping level and integrity of mRNAs. Nucleic Acids Research, 48(22), pp.e129–e129. doi:https://doi.org/10.1093/nar/gkaa955.
16. Tirabassi, R. (2017). How to Identify Supercoils, Nicks and Circles in Plasmid Preps - Bitesize Bio. [online] Bitesize Bio. Available at: https://bitesizebio.com/13524/how-to-identify-supercoils-nicks-and-circles-in-plasmid-preps/.
17. Vlatkovic, I., Ludwig, J., Boros, G., Gábor Szabó, Reichert, J., Buff, R., Baiersdörfer, M., Reinholz, J., Azita Josefine Mahiny, Uğur Şahin and Katalin Karikó (2022). Ribozyme Assays to Quantify the Capping Efficiency of In Vitro-Transcribed mRNA. Pharmaceutics, 14(2), pp.328–328. doi:https://doi.org/10.3390/pharmaceutics14020328.
18. www.thermofisher.com. (n.d.). Agarose Gel Electrophoresis of RNA - US. [online] Available at: https://www.thermofisher.com/uk/en/home/references/protocols/nucleic-acid-purification-and-analysis/rna-protocol/agarose-gel-electrophoresis-of-rna.html.
19. New England Biolabs (2016). [online] Neb.com. Available at: https://www.neb.com/en-gb/products/m2080-vaccinia-capping-system [Accessed 12 Dec. 2024].
20. Kraft, J.J., Treder, K., Mariko Sada Peterson and W. Allen Miller (2013). Cation-dependent folding of 3′ cap-independent translation elements facilitates interaction of a 17-nucleotide conserved sequence with eIF4G. Nucleic Acids Research, 41(5), pp.3398–3413. doi:https://doi.org/10.1093/nar/gkt026.
21. Thermofisher.com. (2024). SYBRTM Green II RNA Gel Stain, 10,000X concentrate in DMSO. [online] Available at: https://www.thermofisher.com/order/catalog/product/S7564 [Accessed 12 Dec. 2024].

DEC Production Reactor Design at Commercial Scale

Donggyu Lee
Department of Chemical Engineering, Imperial College London, U.K.

Abstract: This study presents a robust methodology for modeling the non-ideal behaviors of a fixed bed heterogeneous catalyst reactor for the commercial-scale production of Diethyl Carbonate in Aspen Plus software. The model captures the residence time distribution (RTD) by approximating the reactor as a series of continuous stirred-tank reactors (CSTRs). Pressure drop is accuractely estimated by integrating a plug flow reactor (PFR) model using Ergun equation to capture flow resistance. To enhance predictive accuracy, the model incorporates particle size and geometry, accounting for the dual effects on mass transfer and pressure drop. The implementation of the model has been done in Aspen Plus User2 block, utilizing Newton Raphson method on solving non-linear equations in CSTR and Runge-Kutta method on on solving differential equations in PFR and iterative calculations of mixture properties. The results highlight the significance of optimizing catalyst particle characteristics and reactor configuration to balance mass transfer efficiency, pressure drop, and to maximize DEC selectivity.

1 Introduction

The Shell Carbonate Process Development Team has developed a technology for a Diphenyl Carbonate production under conversion of Ethylene Oxide, Carbon Dioxide, and Phenol. This technology offers a significant cost advantage in the polycarbonate market as Diphenyl Carbonate is a monomer for Polycarbonate production, enhancing its competitiveness in the market.

A key step in this process involves the transesterification of Ethylene Carbonate (EC) with Ethanol (EtOH) to produce Diethyl Carbonate (DEC). DEC has a potential to not only serve as an intermediate component in polycarbonate production, but also as a solvent in lithium-on batteries where demand is rising due to the rapid growth of the electric vehicle (EV) market.

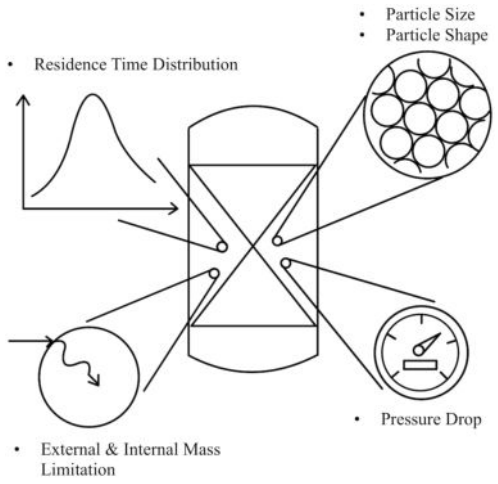


Figure 1.0.1: Outline of the four non-ideal reactor behaviors in commercial scale

The reactor model initially developed by Shell relied on kinetic and equilibrium parameters derived from laboratory and pilot plant scale experiments. However, this model has potential to provide inaccurate prediction when scaled up to commercial operations. Factors such as residence time distribution (RTD), catalyst particle size and geometry, and mass transfer limitation (both internal and external) were not fully captured. Furthermore, the effect of pressure drop in commercial-scale reactors, which can substantially influence reactor performance are not considered.

To overcome these limitations, this study focuses on the development of an advanced reactor model focused on commercial-scale applications. Specifically, the model aims to incorporate non-ideal reactor behaviors through the Aspen Plus User2 block, which allows for the creation of customized reactor models by linking with Excel VBA code. This approach offers greater flexibility compared to Aspen Plus's built-in reactor models, which often fail to represent the complexities of commercial scale processes accurately.

By focusing on developing customized reactor models, this study aims to bridge the gap between model and real-world reactor performance.

2 Background

Various chemical routes for DEC Production have been investigated for DEC production, such as the phosgene route [11], oxidative carbonylation of ethanol [3], the reaction of ethyl nitrite with carbon monoxide [6], and the alcoholysis of urea [2]. However, these methods suffer from the use of toxic gases (phosgene, ethyl nitrite, carbon monoxide) and low product yields and these hinder their usage for industrial applications. In contrast, the transesterification of EC with ethanol offers a safer and more sustainable approach, thereby improving operational safety [7].

Overall main reaction in transesterification is of EC and EtOH to DEC and mono ethylene glycol (MEG) in presence of heterogeneous AlPO catalyst. The main reactions are in two reversible reactions and the equilibrium constants for both are close to thermo-neutral. In this study, we name

the first equilibrium reactions (Equation 2.0.1) as reaction 1 and 2 and the second equilibrium reactions (Equation 2.0.2) as reaction 3 and 4.

$$EC + EtOH \rightleftharpoons HEEC \tag{2.0.1}$$

$$HEEC + EtOH \rightleftharpoons DEC + EG \tag{2.0.2}$$

The reaction rates of the species involved are assumed to be proportional to their stoichiometric coefficients in the balanced reaction equation.

$$r_1 = k_1[EC][EtOH], r_2 = k_2[HEEC] \tag{2.0.3}$$

$$r_3 = k_3[HEEC][EtOH], r_4 = k_4[DEC][EG] \tag{2.0.4}$$

Scaling up chemical reactors from laboratory to industrial scale presents multifaceted challenges due to on transport phenomena, reaction kinetics, and catalyst properties [12]. The transition from small-scale to large-scale operations is not a straightforward linear amplification of reactor dimensions. It requires a comprehensive understanding of how these factors affect performance when the reactor size increases. In particular, catalysts must fulfill several critical requirements. They need to be capable of minimizing pressure drop and exhibit high mechanical, chemical, and thermal stability to withstand commercial scale operation [10].

Despite the importance of catalyst design in commercial scale, significant challenges exists in finding appropriate references from both academic research and industrial practice [9]. Academic researchers often struggle to access detailed information on challenges faced in industry and prepare and handle large scale material quantities and the infrastructure including machinary. Companies, on the other hand, tightly secure their information to maintain a competitive edge in the chemical industry. Nevertheless, catalyst research process need not to be trial-and-error experimentation. By systematically apply chemistry and chemical engineering principles, it is possible to develop reactor models that includes the characteristics of behavior in commercial scale.

In this study we have used Aspen Plus [1] as a software to build the reactor model. Aspen Plus provides a built-in reactor models — RStoic, RYield, REquil, RGibbs, RCSTR, and RPlug, among others — each tailored to specific reaction scenarios. While these built-in models are well-established and user-friendly, they inherently simplify the behaviors present in the reactors that may not capture the influence of complex transport phenomena, catalyst particle size and geometry, or non-ideal fluid dynamics often observed at commercial scale. Such simplications will lead to discrepancies between the model and the actual plant reactor. As a result, this study utilizes the User2 block in Aspen Plus, which enables the integration of custom calculations through VBA code executed in Excel file. This allows the detailed modeling of non-ideal behaviors in the reactor at commercial scale.

3 Methods

3.1 Model Assumptions

A full list of assumptions on transesterification reactions and the reactor are stated in this section.

3.1.1 Reaction Assumptions

1. Kinetic parameters are assumed to be intrinsic. All kinetic parameters - rate constants and activation energies - are derived from experiments conducted under whole particles, which does not guarantee to have intrinsic rate and negligible mass transfer limitations. Experimental results from Shell demonstrated that the reaction rates differ between whole and crushed particles, indicating that the kinetic parameters do not represent intrinsic rates. Despite this, we assume that the kinetic parameters are derived from experiments that had intrinsic rates. Consequently, no scaling or adjustments have been applied to the parameters.

2. Dominance of Main Reactions Main reactions that contribute to DEC formation were only considered. Any potential side reactions are assumed to have negligible impact on overall DEC conversion and selectivity and are thus omitted fro the model. However, they are known to form glycols, which causes AlPO catalyst deactivation, found from Shell Carbonates team.

3. No catalyst deactivation. Catalyst deactivation is not modeled in this study, but catalyst deactivation occur in a real reactor due to glycol byproducts in both main and side reactions.

4. Thermo-neutral reactions. Although the main reaction is slightly exothermic, the net heat release is sufficiently small to be considered negligible [7]. As a result, the reaction is treated as effectively thermo-neutral, and no heat balance equation is required for modeling the reaction.

5. Uncatalyzed thermal reactions are negligible. The rate of uncatalyzed thermal reaction pathway is assumed insignificant compared to the catalyzed reaction rates. This justifies neglecting the Uncatalyzed thermal reaction routes.

3.1.2 Reactor Assumptions

1. Axial Dispersion is negligible. Without axial dispersion, the reactor can be modeled as a plug-flow system without any back-mixing. However, there is no guarantee that the assumption is valid. Back mixing can reduce the distinction between conditions at different reactor lengths and can bring the reactor behavior closer to a CSTR, thus achieving lower conversion per unit reactor length.

2. Isothermal Operation. The reactor is assumed to be isothermal. Including the assumption that the reactions are thermo-neutral, the operating condition of the reactor are such that the temperature gradients along the reactor are negligible.

3. No wall effect. No boundary layer or wall-channeling effects are considered. The reactor walls do not influence the flow pattern or reaction rate distribution, therby having a uniform radial profile. There is no guarantee that the assumption is realistic. Wall effect can create variations in heat transfer, mass transfer and catalyst distributions, which

would alter the performance of the reactor including reaction rates, conversion and yield.

3.2 Ideal Plug Flow Reactor Model

An ideal reactor model for a fixed-bed heterogeneous catalyst reactor operating in continuous flow can be represented by an ideal PFR. The governing equations on molar balances for each component are given as Equation 3.2.1 and total molar balances are given as Equation 3.2.2.

$$\frac{dx_i}{dV_R} = \frac{\rho_{\text{bed}}}{F_{\text{tot}}} \sum_{j=1}^{R} \nu_{i,j}\, r_j \tag{3.2.1}$$

$$\frac{dF_{\text{tot}}}{dV_R} = \rho_{\text{bed}} \sum_{j=1}^{R} \left(\sum_{i=1}^{N} \nu_{i,j}\right) r_j \tag{3.2.2}$$

3.3 Residence Time Distribution (RTD)

To accurately represent non-ideal flow patterns in the reactor, a series of continuous stirred-tack reactors (CSTRs) can capture deviations from ideal plug flow and ideal CSTR on mixing and residence time distribution [4]. The ideal CSTR has governing equation on molar balance of each component as Equation 3.3.1 and total molar flowrate as Equation 3.3.2

$$\frac{dx_i}{dt} = F_{\text{tot}}^{\text{in}}\, x_{i,\text{in}} - F_{\text{tot}}\, x_i + \rho_{\text{bed}} \frac{V_{\text{R}}}{N_{\text{CSTR}}} \sum_{j=1}^{R} \nu_{i,j}\, r_j \tag{3.3.1}$$

$$\frac{dF_{\text{tot}}}{dt} = F_{\text{tot}}^{\text{in}} - F_{\text{tot}} + \rho_{\text{bed}} \frac{V_{\text{R}}}{N_{\text{CSTR}}} \sum_{j=1}^{R} \left(\sum_{i=1}^{N} \nu_{i,j}\right) r_j \tag{3.3.2}$$

From these equations, an equation on RTD for a system comprising N CSTRs in series can be derived [5], which is expressed as:

$$E(t) = \frac{t^{N-1}}{(N-1)!} \left(\frac{N}{\tau}\right)^{N} e^{-Nt/\tau} \tag{3.3.3}$$

To match the RTD of the reactor model with the RTD of the actual reactor, we determine the optimal number of CSTRs (N)by fitting the RTD Equation 3.3.3 to the measured RTD and the mean residence time (τ) of the actual reactor. To incorporate the RTD into the reactor model, the total outlet molar flowrate and mole fractions from the last CSTR in the series are used to represent the overall output of the reactor. An example of fitting RTD of the CSTRs in series to the dataset is shown in Section 5.1 based on data of residence time distribution in Fogler's book [4].

3.4 Internal Mass Transfer Limitations

To quantify internal mass transfer limitations, the Thiele modulus (ϕ) and effectiveness factor (η) were employed. Using the principles of molar flux based on Fick's law of diffusion combined with a first-order reaction rate inside a catalyst particle, formulas for Thiele modulus and effectiveness factor are derived [4] as shown in Equation 3.4.1 and 3.4.2.

$$\phi = \frac{V_p}{A_p} \sqrt{\frac{k_{\text{cat}}\, \rho_{\text{bed}}}{D_e}} \tag{3.4.1}$$

$$\eta = \frac{3}{\phi^2} (\phi \coth \phi - 1) \tag{3.4.2}$$

However, reactions r_1, r_2, r_3 are not in first order, as their rate constant equations are involved with two components. To simplify the analysis, these reactions were approximated as pseudo-first-order by lumping the concentration of one component into the rate constant ($k_{\text{cat}} \cdot C_j$).

The effective diffusion coefficient (D_e) is derived from diffusion coefficient (D_{AB}) using Equation 3.4.4 and D_{AB} is calculated from Wilke-Chang Equation 3.4.3, an emperical correlations for liquid diffusivity of a component [8].

$$D_{AB} = \frac{7.4 \times 10^{-8} (\phi_a M_B)^{1/2} T}{\eta_B V_A^{0.6}} \tag{3.4.3}$$

$$D_{e_A} = D_{AB} \left(\frac{\epsilon}{\tau_{\text{Tor}}}\right) \tag{3.4.4}$$

where A is a solute and B is a solvent. In this study, we have assumed that the solvent is ethanol, which the association factor (ϕ_a) equals to 1. Also, the solute has been chosen to be the components that its concentration is not lumped into the rate constant in pseudo-first-order rate equation.

3.5 External Mass Transfer Limitations

To incorporate the effects of external mass transfer, the global effectiveness factor (η_G) was used to combine both internal and external limitations. The global effectiveness factor accounts for both diffusion constraints and is defined as:

$$\frac{1}{\eta_G} = \frac{1}{\eta} + \frac{\phi^2}{Bi} \tag{3.5.1}$$

where effectiveness factor (η) is from Equation 3.4.2 and the Boit number (Bi) represents a and it is expressed as:

$$Bi = \frac{k_{m_c}\, r_p}{D_e} \tag{3.5.2}$$

The external kinetic mass tranfer (k_{m_c}) cannot be easily measured and it needs to be derived from Sherwood number (Sh) in Equation 3.5.3 and the Sherwood correlation for spherical particles in Equation 3.5.4.

$$Sh = \frac{k_{m_c}\, d_p}{D_{AB}} \tag{3.5.3}$$

$$Bi = \frac{k_{m_c}\, r_p}{D_e} \tag{3.5.4}$$

Hence, the intrinsic reaction rate can be adjusted by multiplying it with the global effectiveness factor (η_G), accounting for both internal and external mass transfer limitations.

3.6 Pressure Drop

The pressure drop across a packed bed reactor was calculated using the Ergun equation, which accounts for both viscous and inertial contributions to fluid flow resistance through a porous medium. The differential form of the Ergun equation is expressed as:

$$\frac{-dP}{dh} = 150\left(\frac{\mu q}{d_p^2}\right)\frac{(1-\epsilon)^2}{\epsilon^3} + 1.75\left(\frac{\rho q^2}{d_p}\right)\frac{(1-\epsilon)}{\epsilon^3} \tag{3.6.1}$$

The Ergun equation applies specifically to packed bed reactors where fluid flows through a fixed bed of particles. It cannot be applied to continuous stirred-tank reactors (CSTRs), which assume uniform mixing and do not consider flow through a packed bed. Therefore, pressure drop calculations were performed for the plug flow reactor (PFR) model.

The pressure drop in commercial-scale reactors significantly impacts the annual operating cost due to the energy required for pumping. In this study annual operating cost has been calculated using work required to overcome the pressure drop. The work done is calculated with Equation 3.6.2 and the corresponding annual operating cost is calculated with Equation 3.6.3.

$$W = \Delta P \cdot Q \tag{3.6.2}$$

$$\text{Annual operating cost} = W \cdot C \cdot t_{\text{work}} \tag{3.6.3}$$

where total hours of work in a year (t_{work}) is assumed as 8150 h/yr and cost of electricity is assumed to be 0.09 USD$/kWh$.

3.7 Particle Size and Geometry

The range of particle sizes considered in this study spans from laboratory scale to beyond commercial scale, covering diameters from 10 micrometers (10 µm) to 10 centimeters (10 cm). This comprehensive range enables the evaluation of mass transfer effects and pressure drop across different scales of operation.

Three primary particle geometries were considered in this study: **spheres, cylinders, and slabs**. Equivalent diameter of a particle was used in comparison with particle geometry. Equivalent diameter is a measure used to compare particles of different shapes by expressing them in terms of the diameter of a sphere with the same volume as the particle of interest. This allows particles to have same volume of catalyst, but different surface area.

Although complex geometries such as miniliths, wagon wheels, and trilobes are commonly used for catalyst pellets [10], they were excluded due to their significant deviation from spherical shapes. These deviations make the standard correlations used in the Sherwood equation for mass transfer (Equation 3.5.4) and the Ergun equation for pressure drop (Equation 3.6.1) inapplicable. In contrast, cylinders and slabs exhibit geometric characteristics that allow them to be reasonably approximated using the same formulas as spheres, facilitating their inclusion in this analysis.

3.8 Model Development

The reactor model developed in this study integrates two primary reactor configurations: a series of CSTRs and a PFR. Both models incorporate the global effectiveness factor (η_G) in their rate equations to account for internal and external mass transfer limitations and they are dependent to particle size and geometry. The reactor model's outlet conditions are derived from the final CSTR in the series, ensuring that residence time distribution (RTD) effects are captured. In parallel, pressure drop calculations are obtained from the PFR model using the Ergun equation (Equation 3.6.1). This combined approach effectively integrates non-ideal reactor behaviors.

The reactor model was implemented using Aspen Plus V14. As the software does not natively support simultaneous CSTR and PFR calculations within a single block, a customized model was developed using the User2 block feature in Aspen Plus. This feature allows the block to interface with an Excel file, which both inputs from Aspen are sent to excel to conduct core reactor calculations and the corresponding output is sent back to Aspen. For the CSTRs in series, the model solves a set of non-linear algebraic equations derived from steady-state differential equations using the Newton-Raphson method. In contrast, the PFR model involves solving differential equations along the reactor length, which is performed using the fourth-order Runge-Kutta (RK4) method.

A comprehensive flowchart illustrating the entire reactor model calculation process is provided in Figure 5.3.1.

3.9 Aspen User2 Block Development

Table 3.9.1: Reactor and Catalyst Input Parameters

Parameter	Unit	Value
A_1	$m^6/kmol.kg(AlPO).s$	8.58 E-05
A_2	$m^6/kmol.kg(AlPO).s$	1.93 E-03
A_3	$m^6/kmol.kg(AlPO).s$	4.20
A_4	$m^6/kmol.kg(AlPO).s$	6.91
$E_{a,1}$	kJ/mol	20.0
$E_{a,2}$	kJ/mol	20.0
$E_{a,3}$	kJ/mol	56
$E_{a,4}$	kJ/mol	5.70 E+01
ρ_{AlPO}	kg/m^3	603
ϵ		0.35
τ_{tor}		5
V_R	m^3	62.8319
D_R	m	2
L_R	m	20
N		
d_P	m	
Geometry$_{cat}$		

Aspen User2 block has been implemented within the Aspen Plus software by establishing a connection to an Excel file. This excel file contains VBA code that defines the reactor models and the numerical methods explained in Section 3.8. The User2 block is designed to receive specific input parameters, which must be defined in prior.

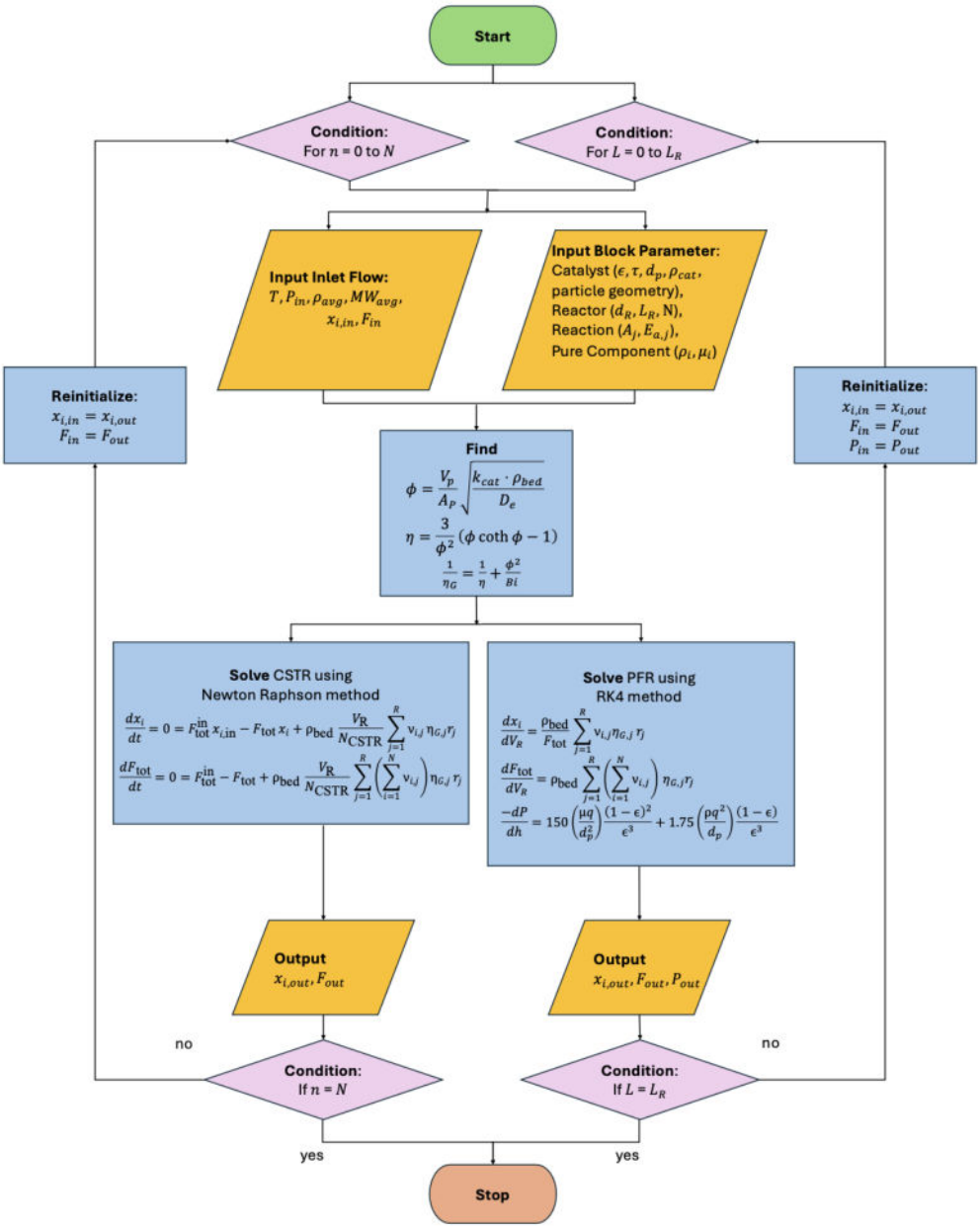


Figure 3.8.1: Flowchart for whole

Aspen User2 Block has been implemented in Aspen Plus software by connecting to Excel file. The Excel file has VBA code for reactor models and numerical methods explained in section 3.8. The User2 block has been set to get values for certain parameters, which needs to be inserted before running the block.

A full list of input parameters required are listed in table 3.9.1. Reaction kinetic parameters and catalyst properties are provided from Shell. Reactor kinetics parameters are derived from experiments and the catalyst parameters are referred from pilot plant design both conducted from Shell. Reactor parameters -V_R, D_R, L_R- are arbitrary set in this study. Viscosity and molar density of pure components are the input parameters for the model as they are required to calculate mixture properties, which they are directly measured from Aspen Plus. Finally, there are input parameters - N, d_P, Catalyst Particle Geometry - that are not constant, but are free independent variables.

Table 3.9.2: Reactor and Catalyst Parameters

Parameter	Unit	Value
T	$°C$	140
P	bar	10
F_{in}	$kmol/hr$	380
$x_{EC,in}$		0.167
$x_{EtOH,in}$		0.833

For catalyst particle geometry, three geometries - sphere, cylinder, slab - can only be chosen as explained in section 3.7.

The inlet temperature, pressure and mole fractions of components are set as in Table 3.9.2. EC/EtOH mole fraction is coherent with experiments conducted from Shell Carbonate team and the study from (Qiu et al., 2012)[7]

where the optimal ratio was in 1:4.5. However, the total molar flow rate at commercial scale was not known, which had to be set with arbitrary value.

4 Model Validation

4.1 Model Accuracy Validation

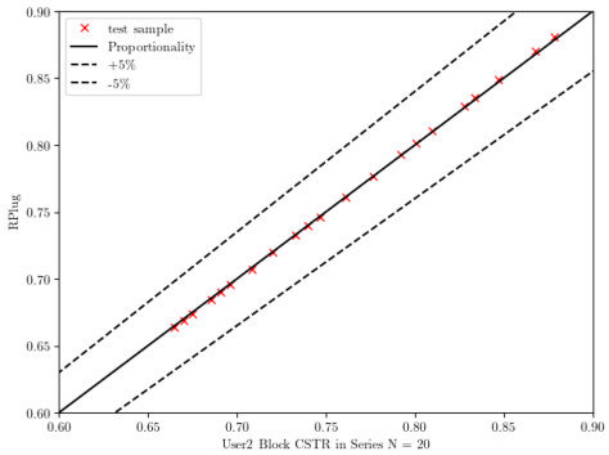


Figure 4.1.1: Parity Plot on Comparing a Set of Results on DEC Selectivity From a Aspen User2 Block against Aspen RPlug Block

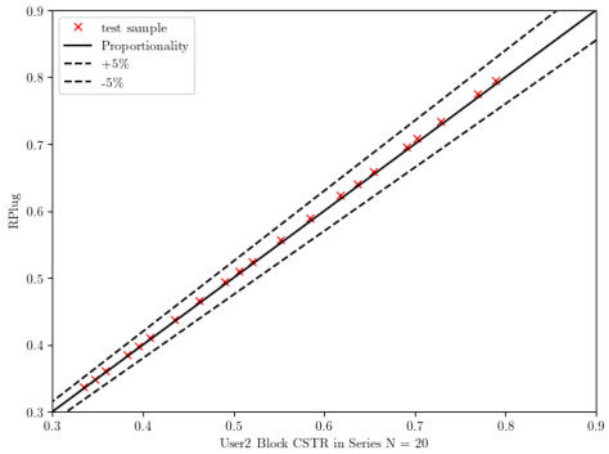


Figure 4.1.2: Parity Plot on Comparing a Set of Results on EC Conversion From a Aspen User2 Block against Aspen RPlug Block

A comparison of the Aspen User2 block, configured as 20 CSTRs in series, with the Aspen RPlug model was performed to evaluate the resulting DEC selectivity and EC conversion at the outlet. A total of 20 samples were generated for both approaches, with each sample featuring a randomly selected inlet flowrate and EC-to-EtOH ratio. The results are presented as parity plots in Figure 4.1.1 and Figure 4.1.2. Both the DEC selectivity and EC conversion obtained from the Aspen User2 block closely matched by the Aspen RPlug model, as indicated by the data points aligning closely with the proportionality line and falling within 5% error margins. These findings confirm that the Aspen User2 block can reliably reproduce the performance predicted by the Aspen RPlug model.

To further validate the behavior of the User model, a sensitivity analysis was conducted by varying the number of CSTRs in series and observing the changes in outlet mole fractions, shown in Figure 4.1.3. At the start and the last of the graphs, results on Aspen Plus block RCSTR and RPlug blocks are added for additional comparison between Aspen User2 block and Aspen Plus pre-built blocks.

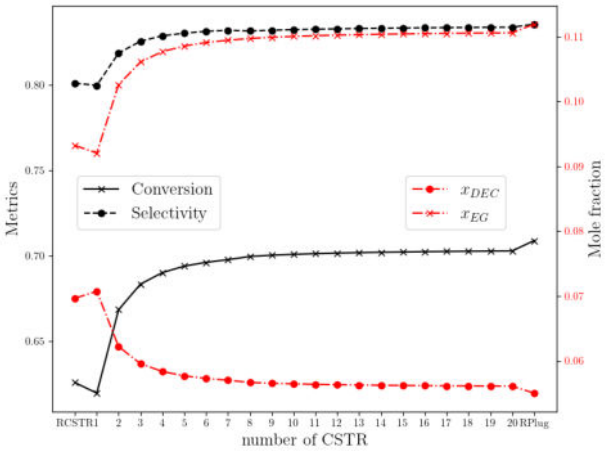


Figure 4.1.3: Conversion, Selectivity, DEC and EC Mole Fraction in increasing number of CSTRs in Series

The analysis revealed that increasing the number of CSTRs led to a decrease in EC mole fractions, while DEC mole fractions, selectivity, and EC conversion increased. The results confirm that, the reactor behavior increasingly approximates that of a PFR, which achieves the highest possible EC conversion for a given reactor length with the uniform mixing. These findings validate that the Aspen User block accurately captures the expected reactor dynamics and provides reliable results when modeling both CSTR and PFR configurations.

5 Results and Discussions

5.1 CSTRs in Series and Residence Time Distribution

To model the residence time distribution (RTD) accurately, two configurations of the User2 model were implemented: one with 4 CSTRs in series and another with 2 CSTRs in series. The RTD data obtained from these models highlight the effect of increasing the number of CSTRs on the shape of the distribution. Specifically, as the number of CSTRs increases, the distribution shift occurs toward the longer residence times, the variation in distribution decreases, and the peak of the distribution is higher. This trend reflects a transition from complete mixing toward plug flow behavior.

By selecting an optimal number of CSTRs, the model can closely approximate the actual RTD observed in the reactor. This approach ensures that non-ideal mixing characteristics are captured, thereby enhancing the predictive accuracy of the reactor model.

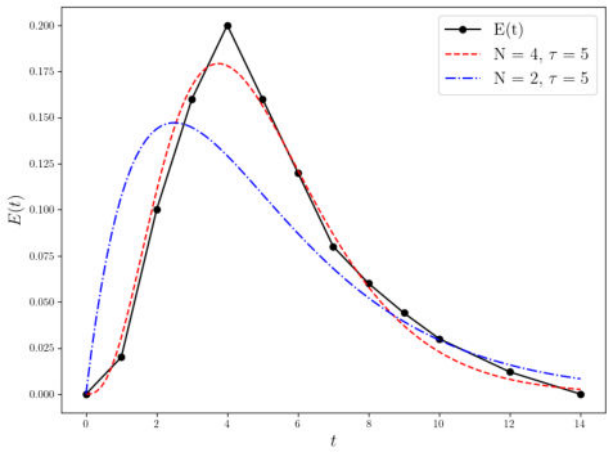


Figure 5.1.1: 3 different residence time distribution in experiments from 1) Fogler's book [4] 2) User2 block with 2 CSTRs and 3) User 2 block with 4 CSTRs

5.2 Global Effectiveness Factor and Particle Diameter

The relationship between the global effectiveness factor (η_G) and the diameter of the particles was analyzed to understand the impact of mass transfer limitations on reaction performance. The results, shown in Figure 5.2.1 indicate that as the particle diameter increases, the global effectiveness factor decreases in Aspen User2 block, while the global Effectivness factor for Aspen RCSTR stays constant at 1. This trend reflects the increasing significance of external and internal mass transfer limitations for larger particles.

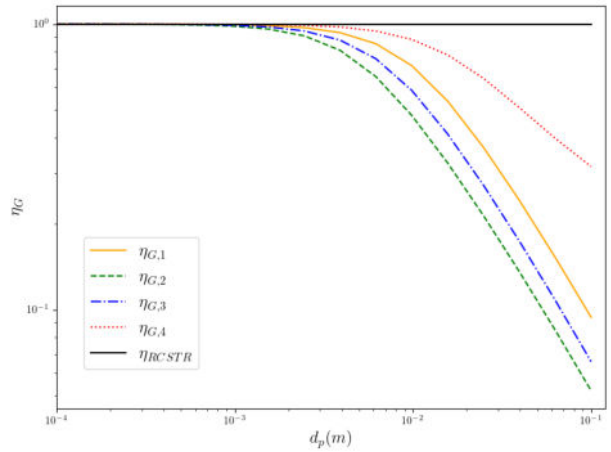


Figure 5.2.1: The relationship of the global effectiveness factor of four reactions on catalyst particle diameter, shown on a log-log scale

A notable observation is that the global effectiveness factor for reaction 4 decreases at a slower rate compared to the other reactions. This will cause the negative impact on the production of DEC and EG and their selectivity, as reactions 1 and 3 that produces EC and EtOH to HEEC and HEEC to DEC and EG will have relatively smaller rates compared to the reaction 4 as shown in Figure 5.3.2.

5.3 Global Effectiveness Factor and Particle Geometry

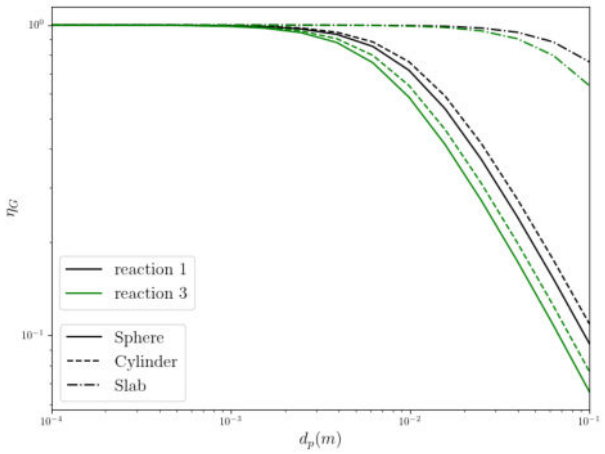


Figure 5.3.1: The relationship of the global effectiveness factor on catalyst particle diameter and geometry, shown on a log-log scale

The effect of particle geometry on the global effectiveness factor (η_G) was analyzed based on particles with equivalent volume diameters for reactions r_1, r_3. The results demonstrate that the global effectiveness factor increases in the following order of particle geometries: sphere $<$ cylinder $<$ slab.

This trend is primarily due to the differences in the surface area-to-volume ratio (V_p/A_p) for each geometry. Spherical particles exhibit the smallest V_p/A_p ratio, while cylindrical particles have a ratio approximately 1.15 times greater than spheres, and slabs exhibit a ratio approximately 11.5 times greater given cylinder has length ($L_P = 2 \cdot d_P$) and slab have lengths in ratio of $1 : 2 : 4$.

Since the Thiele modulus (ϕ) is proportional to V_P/A_P ratio, particles with higher surface area-to-volume ratios (e.g., slabs) exhibit higher global effectiveness factors due to reduced internal mass transfer limitations. In contrast, spheres and cylinders experience greater internal mass transfer resistance due to lower surface area-to-volume ratios, leading to lower effectiveness factors.

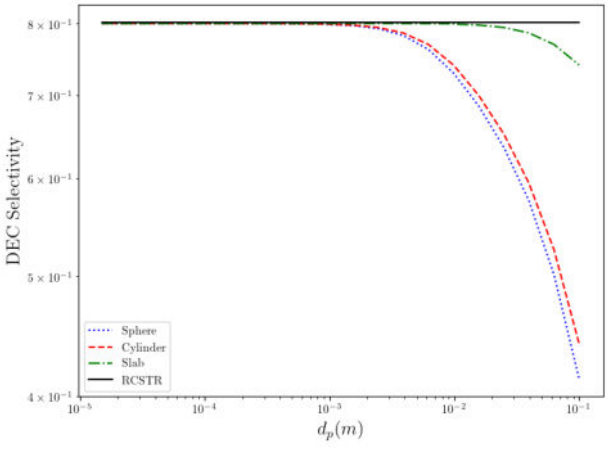


Figure 5.3.2: The relationship of the DEC selectivity on catalyst particle diameter and geometry, shown on a log-log scale

Figure 5.3.2 illustrates how both internal and external mass transfer limitations affect DEC selectivity. As particle size increases and particle geometry shifts towards a smaller V_P/A_P ratio, DEC selectivity declines. This reduction can be attributed to a bigger decrease in global effectiveness factor for reactions 1 and 3 (shown in Figure 5.2.1), causing their overall rates to slow under mass transfer limitations. In contrast, reaction 4 experiences a comparatively smaller decrease in rate, leading to disproportionately higher consumption of DEC and EG - the target products. This results presents a significant challenge at the commercial scale, as it effectively diminishes the yield of DEC and EG. To overcome this issue, additional modifications must be implemented in designing commercial scale reactors to reduce both internal and external mass transfer limitations.

For scale-up applications, these findings highlight that particle geometry with higher V_P/A_P ratio like slab will be favored. However, they are typically not suitable for commercial-scale processes due to mechanical constraints [10] where particles may lack the necessary structural strength to withstand the high fluid velocities and pressures encountered. Therefore a balance must be set between maximizing the effectiveness factor and ensuring mechanical integrity, while selecting particle geometry for scale-up.

5.4 Pressure Drop and Particle Diameter

The effect of particle dimaeter on the pressure drop in a reactor was analyzed to assess the feasibility of using small particles on a commercial scale. Figure 5.4.1 demonstrates that the pressure drop exponentially increases with decreasing particle diameter, making the pressure drop significantly higher at the laboratory-scale particle diameter in a commercial-scale reactor.

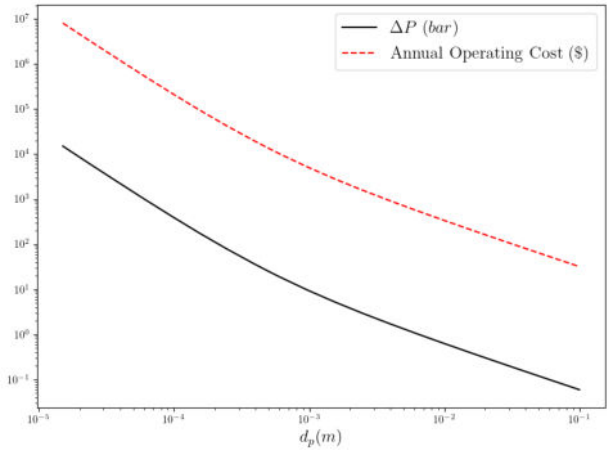


Figure 5.4.1: The relationship of Pressure Drop and Annual Operating Cost on catalyst particle diameter, shown in log-log scale

The underlying mechanism can be explained by the Ergun equation 3.6.1, which describes pressure drop in packed beds as a combination of viscus and inertial contributions. Both terms in the Ergun equation are inversely proportional to the square of the particle diameter for the viscosity term and inversely proportional to the particle diameter for inertial term. As the diameter of the particles decreases, the V_P/A_P ratio increases, causing the fluid to flow through narrower flow channels. This results in higher frctional resistance to flow. Additionally, the fluid must navigate through increasingly tortuous paths, leading to more collisions and directional changes. These effects generate a greater energy loss, contributing to the overall pressure drop.

The relationship between pressure drop ad particle size has a direct impact on the annual operating cost of the reactor. Higher pressure drop necessitate greating pumping power to maintain target flow rates. At extremely small particle diameters, this increased power requirement can escalate annual operating costs to magnitudes approaching 10^7 USD per year, making the use of such particles economically unsustainable for commercial-scale operations.

6 Conclusion

This study has developed a reactor model using Aspen User2 block connected to Excel file for Diethyl Carbonate (DEC) production that accounts for a range of non-ideal behaviors of a reactor at commercial scale, including residence time distribution, internal and external mass transfer limitation, particle size and geometry and pressure drop.

Using a series of continuously stirred tank reactors (CSTRs), the observed RTD can be represented by adjusting the number of reactors to best match the RTD of CSTR in series to the experimentally measured RTD.

The reactor model in Aspen User2 block effectively incorporated both mass trasnfer limitation (internal and external) in the rate of reaction and pressure drop. The analysis of the model revealed that optimizing the catalyst particle diameter is critical in DEC production operation. While smaller particles can enhance reaction kinetics due to improved mass transfer, they also increase pressure drop and the corresponding operating cost. This trade-off highlights the importance of selecting an optimal particle size to balance improved mass transfer against the increasing pressure drop and operating costs.

A key observation from the study is that reaction 4 is less affected by mass transfer limitations compared to other reactions when particle size decreases. Consequently, the DEC selectivity reduces at commercial scale catalyst particle size because the reaction 4 rate that uses DEC and EG as reactants is less hindered by mass transfer limitations compared to the other reactions. This highlights the need for strategies to mitigate mass transfer limitations in commercial operations, where the use of larger catalyst particles seems unavoidable.

Finally, the Aspen User2 block can incorporate the effect of three particle geometries - sphere, cylinder and slab - and this study demonstrated that the catalyst particle geometries with higher surface area-to-volume ratios, such as slabs, help reduce the impact of mass transfer limitations and improve DEC selectivity.

7 Future Work

This study established a practical methodology for modelling the non-ideal behavior of reactors for DEC produc-

tion at a commerical scale. However, several areas for improvement and extensions were identified, as discussed below.

The current work did not address the optimal reactor design required to maximize the yields of these desired products. Further research should focus on identifying the optimal combination of reactor and catalyst particle design to enhance DEC production.

Several limitations were identified in the current modeling approach. The model have neglected side reactions in DEC production and heat released from the reactions. Additionally, the kinetic parameters used were assumed to represent intrinsic reaction rates, but in reality, they were influenced by internal mass transfer limitation. Removing these assumptions will enhance the model's accuracy in predicting DEC production.

The study also relied on arbitarily chosen parameters, particulary, the inlet flowrate and porosity, which significantly affect results on pressure drop. For instance, if the flowrate is low, the pressure drop will be negligible even at pilot plant scale catalyst particle size. Therefore, future work should incorporate actual parameters to provide a more realistic and robust analysis.

The calculation method in model requires further refinement to handle non-linearities in mixture properties, such as viscosity and density, for accurate performance predictions. One potential improvements involves integrating Aspen Plus calculations for mixture properties.

Finally, Strength of catalyst particles under commercial scale must also be incorporated. While slab-shaped particles exhibit the highest reaction rates due to their superior surface area-to-volume ratio, they lack the mechanical strength needed to withstand high flow rates and pressures encountered in industrial reactors [10]. Experimental data on the mechanical strength of AlPO catalyst particles with various geometries are necessary to identify shapes that provide an optimal balance of surface area and durability. Since such data were unavailable in this study, future research should focus on experiments to evaluate catalyst strength.

8 Acknowledgements

The authors would like to acknowledge the help and support provided by Javier.Moreno, David Pinilla, and Fred van der Steen from Energy Transition Campus Amsterdam at Shell. Their foundational work on DEC production technology and generous knowledge sharing provided the basis in developing the model presented in this study. I would also like to acknowledge the help and support provided by Dr Lu Ai from Chemical Engineering department at Imperial College London on providing initial guidance on building Aspen Plus model.

References

[1] AspenPlus. URL: http://www.aspentech.com.

[2] X. Zhai L. Zhou D. Wang, B. Yang. Synthesis of diethyl carbonate by catalytic alcoholysis of urea. *Fuel Processing Technology*, 88(9):807–812, 2007. doi:doi.org/10.1016/j.fuproc.2007.04.003.

[3] Brian C. Dunn, Catherine Guenneau, Steven A. Hilton, Jorg Pahnke, Edward M. Eyring, Jacek Dworzanski, Henk L. C. Meuzelaar, J. Z. Hu, Mark S. Solum, and Ronald J. Pugmire. Production of diethyl carbonate from ethanol and carbon monoxide over a heterogeneous catalyst. *Energy Fuels*, 16(1):177–181, 2002. doi:doi.org/10.1021/ef0101816.

[4] H. SCOTT FOGLER. *ELEMENTS OF CHEMICAL REACTION ENGINEERING Fifth Edition*. 2002.

[5] LearnChemE. Residence Time Distribution Screencasts. Accessed: 2024-06-07. URL: https://learncheme.com/quiz-yourself/interactive-self-study-modules/residence-time-distribution/residence-time-distribution-screencasts/.

[6] X. Ma M. Fan, P. Zhang. Study on wacker-type catalysts for catalytic synthesis of diethyl carbonate from ethyl nitrite route. *Fuel*, 86(5-6):902–905, 2007. doi:doi.org/10.1016/j.fuel.2006.09.005.

[7] X. Jiang P. Qiu, L. Wang and B. Yang. Synthesis of diethyl carbonate by the combined process of transesterification with distillation. *Energy Fuels*, 26(2):1254–1258, 2012. doi:doi.org/10.1021/ef201732w.

[8] T.K. Sherwood R.C. Reid, J.M. Prausnitz. *The Properties of Gases and Liquids*. McGraw-Hill, fifth edition, 1977.

[9] N.L. Michelsa S. Mitchell and J. Pérez-Ramírez*a. From powder to technical body: the undervalued science of catalyst scale up. *Chem. Soc. Rev.*, 42:6094–6112, 2013. doi:doi.org/10.1039/C3CS60076A.

[10] S.T. Sie and R. Krishna. Process development and scale up: Ii. catalyst design strategy. *Reviews in Chemical Engineering*, 14(3):159–202, 1998. doi:doi.org/10.1515/REVCE.1998.14.3.159.

[11] X. Ma J. Gong W. Wang, S. Wang. Crystal structures, acid–base properties, and reactivities of cexzr1xo2 catalysts. *Catalysis Today*, 148(3–4):323–328, 2009. doi:doi.org/10.1016/j.cattod.2009.07.090.

[12] M. He Y. Tian W. Wang Y. Xu, B. Lu. *Fundamentals of Reactor Design and Scale-Up. In: Diameter-Transformed Fluidized Bed. Particle Technology Series*. Springer, Cham, 2020.

The effect of bentonite content on the equilibrium and kinetic CO_2 adsorption properties of zeolite 13X shaped adsorbents

Aylar Gylyjova and Francesca Rowley

Department of Chemical Engineering, Imperial College London, UK

Abstract

To be utilised effectively for post-combustion carbon capture processes, shaped zeolites need to be manufactured with a binder to increase mechanical strength. This study investigates the impact of bentonite as a binder on the CO_2 adsorption performance of zeolite 13X beads, specifically the final equilibrium uptake and kinetic behaviour. Beads with varying bentonite content (0–100 wt%) were produced via extrusion-spheronisation and the textural and morphological properties of these beads were established using N_2 physisorption at 77 K and mercury intrusion porosimetry (MIP). These methods provided an insight into pore size distributions, pore volumes and characteristics of the beads. In addition, we carried out CO_2 physisorption at 303 K to evaluate equilibrium uptake, and dynamic sorption experiments using thermogravimetric analysis at 303 K between relative pressure, p/p_0 (CO_2) from 0.03 to 0.06. Alongside our experimental work, we incorporated multi-component behaviour into a computational model to simulate equilibrium and kinetic performance by solving coupled diffusion-adsorption equations. Our findings revealed that bentonite content increased the number of larger pores, decreasing the number of micropores and thus surface area of zeolite 13X. In turn, bentonite content decreased the amount of CO_2 adsorbed at equilibrium in a non-linear way, indicating that at higher wt%, bentonite has a pore-blocking effect. The CO_2 kinetic experiments revealed that bentonite may shift the diffusion mechanism in the beads to be more micropore controlled, although this analysis was largely qualitative. These conclusions provide valuable information into the role of bentonite in modifying the CO_2 adsorption properties of zeolite 13X beads. Thus, offering important guidance for the design of optimised zeolite 13X-based adsorbents for post-combustion CO_2 capture.

Keywords: *adsorption, CO_2 separation, beads, zeolite 13X, bentonite, equilibrium, kinetics*

1 Introduction

As atmospheric levels of anthropogenic greenhouse gases (GHG) are increasing at unprecedented rates every year, the need for climate change mitigation technologies is more apparent than ever. Since the Paris agreement was signed in 2015, states around the world have been striving to significantly reduce the global temperature increase caused by the GHG surplus [1]. The most prevalent GHG is carbon dioxide (CO_2), which makes up for 76% of total GHG emissions [2] and subsequently has become the primary focus for tackling climate change. There are various ways to mitigate carbon emissions, one of them being the removal of CO_2 from syngas in post-combustion carbon capture [3].

A promising technique within this field, which has arisen in the last decade, is adsorption [4]. Here, materials provide a solid surface for CO_2 molecules to adhere to, separating these molecules from the bulk gas [5]. One reason adsorption is well regarded is due to its high energy efficiency. For example, in regeneration alone, adsorption methods consume less energy than alternative methods [6]. Over the last decade, adsorbents such as activated carbon, metal-organic frameworks (MOFs) and zeolites have been circulating in literature [6, 7, 8].

Zeolites are highly microporous, crystalline structures that perform as molecular sieves and are formed by TO_4 tetrahedrons, where T is predominantly Si or Al [9]. They have been thoroughly studied for their CO_2 adsorption capabilities [10]. A barrier to the widespread application of zeolites for use in post-combustion carbon capture is their hygroscopic nature, as water molecules present in syngas may block adsorption sites from CO_2 molecules [11]. Regardless, zeolites have been found to outperform other adsorbents in this field [12]. It is important to note that adsorption occurs in the micropores of zeolites, which act as adsorption sites for CO_2 [13]. A notable zeolite for this application is zeolite 13X, which accommodates for the larger CO_2 molecules better than smaller pored zeolites due to it's unique faujasite structure. Zeolite 13X has outperformed other zeolites in CO_2 uptake [14], and therefore has been chosen for this study.

Typically, adsorbents are synthesised in powder form as they provide application flexibility and homogeneity [15]. However, problems can arise when using equipment such as packed beds. Introducing powder to a gaseous system causes dustiness and handling issues, as well as inducing larger pressure drops across the bed. Large pressure drops should be avoided as they affect flow distribution and the overall performance of a separator [16, 17]. To combat this, powders can be shaped to create dense structures through techniques such as extrusion-spheronisation. Common shapes include extrudes, beads, pellets and granules [16, 18, 19]. Beads are a suitable option for use in packed beds as they provide the homogeneity that powders offer without the undesirable

pressure drop.

To avoid the risk of bead crumbling in high gas flow conditions, binders can be incorporated to provide mechanical strength [20]. Promising binders for zeolite 13X include kaolin, silica and bentonite [21, 22, 23]. Abundant clay binders such as bentonite provide low-cost malleability to extrusion pastes, allowing for easy post-treatment processes such as spheronisation to occur [24]. Although binders provide strength and plasticity, at high amounts they can inhibit the adsorption capacity of the zeolite by reducing the adsorbent surface area [22, 25]. Therefore, an optimal ratio needs to be struck.

In this paper, we study quantitively and qualitatively the impact of bentonite on the CO_2 adsorption properties, namely equilibrium and kinetic performance, of zeolite 13X. We initially designed and created blended beads using an extruder-spheroniser. To investigate the textural and morphological properties of these beads, we utilised the analytical techniques of N_2 physisorption at 77 K and mercury intrusion porosimetry (MIP). This provided us with insight into how bentonite content affects the porous structure of zeolite 13X, in turn helping us understand how and why it affects the adsorption properties. We obtained equilibrium adsorption data volumetrically and kinetic adsorption data gravimetrically. A key portion of our study included the incorporation of multi-component behaviour into a computational model to help us analyse adsorption behaviour. The results across our experiments served as inputs to our model, allowing us to apply well-studied diffusion-adsorption relationships to our experimental outcomes to gain insight into the effect of bentonite content on the diffusion mechanism of zeolite 13X. Overall, our study serves as a valuable resource for the preliminary design of zeolite 13X-based adsorbents containing bentonite.

2 Experimental Framework

2.1 Materials

Zeolite 13X powder (CAS# 63231-69-6) and bentonite powder (CAS# 1302-78-9) were provided by Sigma-Aldrich, and binderless zeolite 13X beads were provided by CWK. We used N_2 (zero grade, 99.9995 %), He (CP grade, 99.9990 %) and CO_2 (research grade, 99.999 %) gases in this study. All chemicals and materials were used in their original form unless stated otherwise.

2.2 Preparation of zeolite 13X beads

To develop a thorough understanding of the effect of bentonite content on the adsorption performance of zeolite 13X, we chose equally spaced weight ratios of bentonite over a 0-100 % range. We used commercially-obtained binderless zeolite 13X beads for comparison, as the lack of stability from the absence of a binder made them difficult to produce experimentally. In this study, zeolite 13X beads with nominal binder contents of 0 wt%, 20 wt%, 33 wt%, 50 wt%, 67 wt% and 100 wt% are referred to as Z100-B0, Z80-B20, Z67-B33, Z50-B50, Z33-B67 and Z0-B100 respectively.

To create the beads we used a Caleva Multi Lab Scientific (MLS) machine - a mixing, extruding and spheronising device. The bentonite and zeolite 13X powders were mixed in the specified ratios, where we determined the correct appearance of the granulated mixture by roughly adding 3.5-4 mL of de-ionised water dropwise. We extruded the mixture in an axial screw extruder at 100 RPM and finally spheronised the extrudes, both until completion. If the material was too elastic for spheronisation, it was put in a dryer beforehand. We measured the final average bead radius across each sample, R_p, using the Image Processing Toolbox in MATLAB R2020a and kept all samples in Parafilm sealed containers.

2.3 Materials characterisation

2.3.1 Textural and morphological properties

We determined the pore size distributions of our samples with N_2 physisorption at 77 K using a Micromeritics 3Flex Adsorption Analyser and a Quantachrome Autosorb iQ. Prior to measurements, we degassed samples ex-situ at 623 K for a minimum of 12 h. For samples with higher zeolite 13X content, we introduced a temperature ramp: 1 h at 363 K, 1 h at 473 K then 12 h at 623 K. We measured the N_2 isotherms in the pressure range of 0.003 mbar to 0.99 bar and used the non-local density functional theory (NLDFT) from the equipment's software to obtain the pore-size distribution of the different samples. The final pore size distributions were normalised for $\frac{dV}{dD}$ ($cm^3 \cdot g^{-1} \cdot nm^{-1}$) to determine the average pore size of each sample. At low (p/p_0)s during the N_2 physisorption measurement, the N_2 forms a monolayer on the internal pore walls which gives the monolayer capacity of the bead. With this value, we applied the Brunauer–Emmett–Teller (BET) method to determine the specific surface area of the bead [26].

We performed mercury intrusion porosimetry (MIP) using a Micromeritics AutoPore IV 9500 porosimeter in the range of 0.036 to 2275 bar. Prior to these measurements, we degassed samples ex-situ following the same protocol as for N_2 physisorption at 77 K. The results were analysed using the AutoPore IV 9500 v1.09 software.

An important pre-requisite in adsorption behaviour is characterisation of textural properties, which can be determined by combining data from N_2 physisorption at 77 K and MIP. We determined the pore size distribution, total pore volume, porosity, tortuosity, and bulk density. To evaluate theoretical gas diffusivity, we used the parallel-pore model proposed by Wang and Smith [27, 28]. This framework uses the pore size distribution to evaluate the contributions of molecular and Knudsen diffusion, which occur simultaneously in different pores, to determine the theoretical gas diffusivity.

2.3.2 CO_2 adsorption equilibrium

CO_2 adsorption isotherms were collected volumetrically on a Quantachrome Autosorb iQ at 303 K in the pressure range of 1.3×10^{-9} to 1.0 p/p_0. Prior to measurements, samples were degassed ex-situ following the same protocol as for N_2 physisorption at 77 K. We analysed only the adsorption branch and fitted its points to a Dual Site Langmuir (DSL) adsorption equilibrium model [29], given by:

$$q^*_{CO_2} = \frac{q_{s1} b_1 p}{1 + b_1 p} + \frac{q_{s2} b_2 p}{1 + b_2 p} \tag{1}$$

$$b_i = b_{0i} \exp\left(\frac{\Delta U_{bi}}{RT}\right) \tag{2}$$

Where $q^*_{CO_2}$ is the equilibrium CO_2 adsorption capacity ($mol \cdot kg^{-1}$), q_{si} is the saturation capacity of a site ($mol \cdot kg^{-1}$), b_i and b_{0i} are site adsorption coefficients (bar^{-1}), p is the adsorption pressure (bar), ΔU_{bi} is the heat of adsorption ($J \cdot mol^{-1}$), R is the ideal gas constant (8.314 $J \cdot mol^{-1} \cdot K^{-1}$) and T is the adsorption temperature (K).

2.3.3 CO_2 adsorption kinetics

We carried out dynamic adsorption experiments on a Netzsch TG 209F1 Libra thermogravimetric analyser (TGA) at 303 K and CO_2 relative pressure (p/p_0) in the range 0.03 to 0.06. This temperature was selected to align with the industry standard for solid sorbent capture, as it represents the optimal temperature for maximizing adsorption capacity [30]. This relative pressure range was chosen to replicate the concentration of CO_2 used in post-combustion carbon capture, which is typically 4-14% [31]. To improve the accuracy of capturing the kinetics of our samples, a narrow range of 3% within this interval was chosen, considering the known fast kinetics of zeolite 13X. To ensure we used the correct kinetic regime, we conducted preliminary experiments to test the CO_2 gas flowrate, testing flowrates of $F = 100$, 200 and 350 $mL \cdot min^{-1}$. All measurements were repeated twice and corrected for buoyancy and drag effects using an empty crucible.

Prior to measurements, samples were degassed ex-situ at 623 K for a minimum of 12 h following the same protocol as for N_2 physisorption at 77 K. Prior to CO_2 adsorption, samples were degassed in-situ at 623 K for 1.5 h and then held at 303 K for 0.5 h under purge and protective gas flow, for which helium was used. Upon the introduction of CO_2, samples were held at p/p_0 (CO_2) = 0.03 for 1.7 h.

The CO_2 adsorption was measured by the weight increase upon switching from p/p_0 (CO_2) = 0.03 to 0.06, which was held for a following 1.7 h. The final kinetic curves analysed display the fractional uptake of CO_2, normalised to this region, as a ratio of uptake at time t to uptake when equilibrium was reached. This was achieved when the mass of the sample became unvaried, seen over the final 0.5 h of the experiment, over which an average mass was calculated as the final uptake. We modelled the rate of uptake of CO_2 in the adsorbent with the linear driving force (LDF) model, given by:

$$\frac{dq_{CO_2}}{dt} = k_{LDF}\left(q^*_{CO_2} - q_{CO_2}\right) \tag{3}$$

Where $q^*_{CO_2}$ is the equilibrium CO_2 adsorption capacity obtained from a suitable adsorption isotherm ($mol \cdot kg^{-1}$) and k_{LDF} (s^{-1}) is a lumped kinetic rate constant that describes the resistance to mass transfer from the gas to the solid phase.

3 Computational Framework

A key part of this study was to develop a computational framework to model the adsorption equilibrium and kinetic performance of our samples to provide us with further insight into our experimental results. To address this, we incorporated multicomponent behaviour into a script previously developed by Prof. Petit's group in MATLAB R2020a.

3.1 CO_2 adsorption equilibrium

To generate CO_2 adsorption isotherms, a simple weighted average approach was used, where the contribution of zeolite 13X and bentonite to the total equilibrium loading of a bead was determined based on their respective wt% in the sample, seen as follows:

$$\bar{q}^*_{CO_2,T} = \rho_{ads} \cdot [(wt\%_Z \cdot \bar{q}^*_{CO_2,Z100-B0}) + (wt\%_B \cdot \bar{q}^*_{CO_2,Z0-B100})] \tag{4}$$

Where $\bar{q}^*_{CO_2,T}$ is the total equilibrium CO_2 adsorption capacity of a bead ($mol \cdot m^3$), ρ_{ads} is the adsorbent skeletal density ($kg \cdot m^3$), $wt\%_i$ is the wt% of zeolite 13X (Z) or bentonite (B) in the bead, and $\bar{q}^*_{CO_2,i}$ ($mol \cdot kg^{-1}$) is the equilibrium CO_2 adsorption capacity of either Z100-B0 or Z0-B100.

The data for $\bar{q}^*_{CO_2,i}$ for Z100-B0 and Z0-B100 was obtained volumetrically, as outlined in Section 2.3.2, and served as an input to generate the theoretical CO_2 isotherms at 303 K for the remaining combined beads. This allowed us to qualitatively compare the expected impact of bentonite on the adsorption of CO_2 with the true impact seen from experimental data.

3.2 CO_2 adsorption kinetics

To estimate CO_2 adsorption kinetics, we created a multi-scale mass transfer model to simulate the mass transfer of CO_2 across a bead with time. Zeolites contain both macroporous and microporous structures, thus requiring distinct diffusion mechanisms to accurately describe the behaviour in each region [32]. First, we defined intra-crystalline diffusion as follows:

$$\frac{\delta \bar{q}_{CO_2}}{\delta t} = \frac{15}{t_{mic}}\left(\bar{q}^*_{CO_2,T} - \bar{q}_{CO_2}(R,t)\right) \tag{5}$$

Where t_{mic} is the micropore diffusion time constant (s) and $\bar{q}_{CO_2}$ ($mol \cdot m^3$) is the local average adsorbed amount as a function of the variables (R,t), where R is a location along the radius of a bead (m) and t is time (s).

Next, we defined macropore diffusion as follows:

$$\frac{\delta c}{\delta t} = \frac{1}{R^2}\frac{\delta}{\delta R}\left(R^2 D^e_{mac}\frac{\delta c}{\delta R}\right) - \frac{(1-\epsilon_p)}{\epsilon_p}\frac{\delta \bar{q}_{CO_2}}{\delta t} \tag{6}$$

$$D^e_{mac} = \frac{\epsilon_p D_p}{\tau} \tag{7}$$

Where c is the concentration of CO_2 in the macropores of a bead ($mol \cdot m^{-3}$), D^e_{mac} is effective macropore diffusivity ($m^2 \cdot s^{-1}$), ϵ_p is the porosity of a bead (-), D_p is the theoretical gas diffusivity ($m^2 \cdot s^{-1}$), and τ is the tortuosity of a bead (-).

To predict the behaviour of the system over time, Equations 5 and 6 were solved as a system of ordinary differential equations using ode15s, subject to the following boundary conditions:

$$\frac{\delta c}{\delta R}(R = 0, t) = 0 \tag{8}$$

$$c(R = R_p, t) = c_0 \tag{9}$$

And the following initial conditions:

$$c(R, t = 0) = c_0 \tag{10}$$

$$\bar{q}_{CO_2}(R, t = 0) = \bar{q}^*_{CO_2,T} \tag{11}$$

To generate CO_2 adsorption kinetic curves, the following data was inputted into our model. The data for ρ_{ads}, ϵ_p, D_p and τ was obtained for each bead from MIP, as outlined in Section 2.3.1. Bead composition and R_p were obtained as described in Section 2.2. To match the data obtained gravimetrically, as outlined in Section 2.3.3, the experiment conditions were specified as $T = 303\ K$, $y_0 = 0.03$ and $y_Feed = 0.06$, the latter describing the initial and feed mole fractions of CO_2 respectively. To obtain the same kinetic curve showing fractional uptake against t, we calculated the average loading across the bead as shown below, converted the units to ($mol{\cdot}kg^{-1}$) and finally computed fractional uptake.

$$q_{avg} = \frac{3}{R^3} \cdot \int_0^{R_p} \bar{q}_{CO_2}(R, t) \cdot R^2 \, dR \tag{12}$$

The only unknown parameter remaining was t_{mic}. We determined $\frac{1}{t_{mic}}$ values by manually adjusting this input to achieve the best fit between the model-predicted kinetic curves and those obtained experimentally, as outlined in Section 2.3.3. Given the equipment and time restraints of our study, we deemed this to be a sufficient methodology to provide preliminary insights to kinetic performance. Finally, the macropore diffusion time constant can be calculated as follows:

$$t_{mac} = \frac{R_p^2}{D^e_{mac}} \tag{13}$$

Then, t_{mac}/t_{mic} can be evaluated to determine the presence of a driving diffusion mechanism within our samples.

4 Results and Discussion

4.1 Bead shaping

Images of the created beads are shown in Figure 1. Visually, as bentonite content increases, the beads appear darker. During bead creation, moisture content is an important factor to consider. If too little granulation liquid is added, material will be lost as dust during spheronisation and beads will lack plasticity. If too much is added, agglomeration of beads in spheronisation can occur, producing non-uniform batches [33]. Additionally, not all the beads are uniform in size as some dumbbell shapes can be seen, for example, in Z80-B20. When latter experiments were carried out, we chose the most spherical and consistent beads. The average bead radius for each sample can be seen in Table A1.

4.2 Textural and morphological properties

The structure of porous materials can be divided into three regions: micropores ($< 2\,nm$), mesopores ($2\,nm < x < 50\,nm$), and macropores ($> 50\,nm$) [34]. For bead characterisation, we used N_2 physisorption at 77 K to determine information about the porous structure i.e. pore size distribution in the micro- and mesopore range, total pore volume, micropore volume and BET surface area. N_2 physisorption isotherms at 77 K can be observed in Figure 2. Z100-B0 exhibits a type 1A isotherm which is characteristic of microporous materials [35], and Z50-B50 exhibits a type H4 hysteresis loop which has been seen in mesoporous, aggregated zeolite crystals [34]. Across all the samples, increases in uptake in the region p/p_0 (N_2) > 0.95 suggest the presence of larger pores [36]. The Z80-B20 sample shows significant deviation from the expected trends, with uptake declining at higher relative pressures. An anticipated N_2 adsorption isotherm was recorded by Amir et. al for Z80-B20 suggesting that our results are inaccurate for this sample and a repeat should be performed [22]. Finally, Z0-B100 displays a type H3 hysteresis loop which is common in certain clays and when the pore structure is macroporous [34].

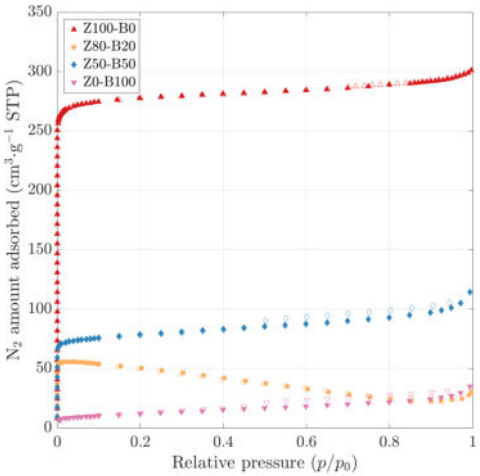


Figure 2: N_2 physisorption isotherms at 77 K for Z100-B20, Z80-B20, Z50-B50 and Z0-B100.

Pore volume data was also obtained from N_2 physisorption at 77 K and is displayed in Table 1. Experimentally, two key observations can be seen: as bentonite content increases, both the BET surface area and total pore volume decrease. When there is less zeolite 13X present in the beads, there are relatively less micropores, meaning there is relatively less surface area [9]. For our Z80-B20, Z67-B33 and Z50-B50 samples, we also determined the theoretical BET surface

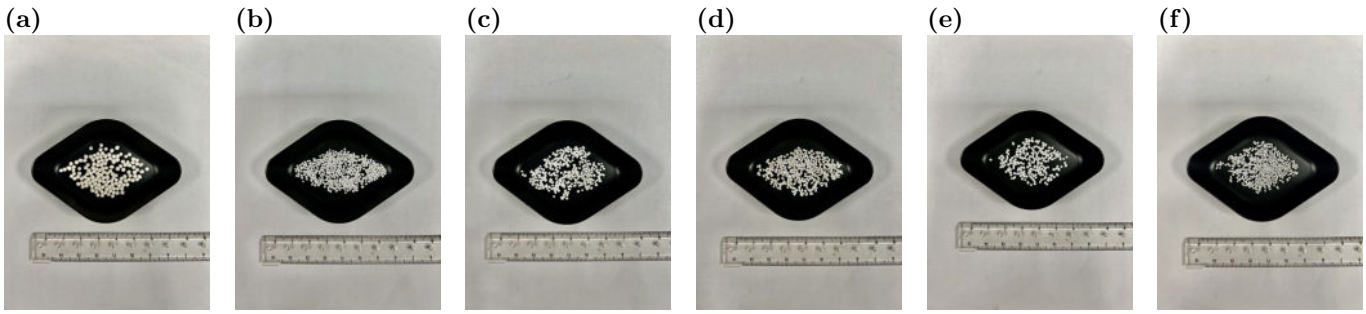


Figure 1: Optical images of commercial sample: (a) Z100-B0 and manually created beads: (b) Z80-B20; (c) Z67-B33; (d) Z50-B50; (e) Z33-B67; and (f) Z0-B100.

area and total volume of pores using a weighted average of Z100-B0 and Z0-B100. The idea that the BET surface area and total pore volume of a blended material often exhibits a linear relationship with the proportions of its components is widely observed in studies of porous materials [37]. This approach is deemed valid when components retain their individual characteristics in the mixture. For Z50-B50, the difference between the theoretical and experimental values for BET surface area and total pore volume could be attributable to the addition of bentonite not having this proportional effect. This may be indicative of pore blockage, where bentonite may partially coat micropores present in zeolite 13X [22]. Given the deviated results of the Z80-B20 isotherm, the pore volume data was likely affected by the same issue so may not be reliable. Amir et. al obtained an experimental BET surface area value of 551 ($m^2 \cdot g^{-1}$) which is still less than our theoretical prediction [22]. This similarly supports the theory of pore blockage. The volume of micropores in Z100-B0 is 86 % of the total volume of pores, aligning with the type 1A isotherm observed previously in Figure 2. It is interesting to note that Z0-B100 has a negligible micropore volume, meaning that negligible amounts of CO_2 adsorption will occur in this sample.

To obtain the full pore size distributions, the data from N_2 physisorption at 77 K and MIP was combined. The results of these can be seen in Figures 3, A1 and A2. The overall trend is best outlined by Figure 3(c), which shows the cumulative pore volumes from stitched N_2 physisorption and MIP data. We observe that as bentonite content increases, the number of larger pores also increases. In Figure 3(a), the normalised micro- and mesopore range pore size distributions can be seen for Z100-B0, Z50-B50 and Z0-B100. Z80-B20 was removed from the plot as it displayed differential pore volumes with values of zero across the micro- and mesopore range, which was caused by the aforementioned reasons. A Z100-B0 peak appears in the small micropore range, which has also been observed in literature where the zeolite 13X micropores are < 0.7 nm [38]. As bentonite content increases, this peak shifts into the mesopore region which could be due to the creation of inter-particle voids. These arise as bentonite agglomerates onto the microporous structure of zeolite 13X leading to the formation of hierarchical porosity [22, 39]. When the pore diameter is increased into the macropore region, seen in Figure 3(b), the peaks for all samples suggest that bentonite also creates inter-particle voids in the macropore region. However, this cannot be the case for Z100-B0 due to the absence of zeolite 13X. As previously mentioned, in powder form zeolite 13X is purely microporous, but when shaped into beads, we observe macropores. Compacting powdered zeolite 13X during extrusion-spheronisation may rearrange the microporous framework, leaving inter-particle voids in the macropore region as well [40]. The size of the macropores is dependent on the level of compression used during extrusion i.e. greater compression would cause smaller macropores. Therefore, the presence of meso- and macropores in the blended samples could be due to a combination of increased bentonite content and the shaping process of the beads.

MIP also allowed for exploration into the textural properties of the beads. Out of tortuosity, bulk density and porosity, porosity gives the greatest insight into the characteristics of a porous material [41]. As observed in Figure

Table 1: Textural, morphological and pore volume properties of Z100-B0, Z80-B20, Z67-B33, Z50-B50, Z0-B100, determined by N_2 physisorption at 77 K, MIP, or a combination of both. Exp. and Theor. refer to the experimentally obtained values and theoretical predictions respectively. Theoretical values were obtained using a weighted average of Z0-B100 and Z100-B0.

Sample	BET ($m^2 \cdot g^{-1}$)		V_{total} ($cm^3 \cdot g^{-1}$)		V_{micro} ($cm^3 \cdot g^{-1}$)	τ (-)	ρ ($g \cdot mL^{-1}$)	ε_p (-)	D_p ($\times 10^{-5} m^2 \cdot s^{-1}$)
	Exp.	Theor.	Exp.	Theor.					
Z100-B0	1023	-	0.462	-	0.40	1.9	1.024	0.29	7.33
Z80-B20	202	827	0.038	0.379	0.11	1.9	1.122	0.32	9.51
Z67-B33	-	699	-	0.324	-	1.9	1.245	0.28	9.03
Z50-B50	292	532	0.177	0.254	0.10	1.9	1.347	0.27	5.64
Z0-B100	41	-	0.045	-	nil	1.8	1.317	0.40	4.76

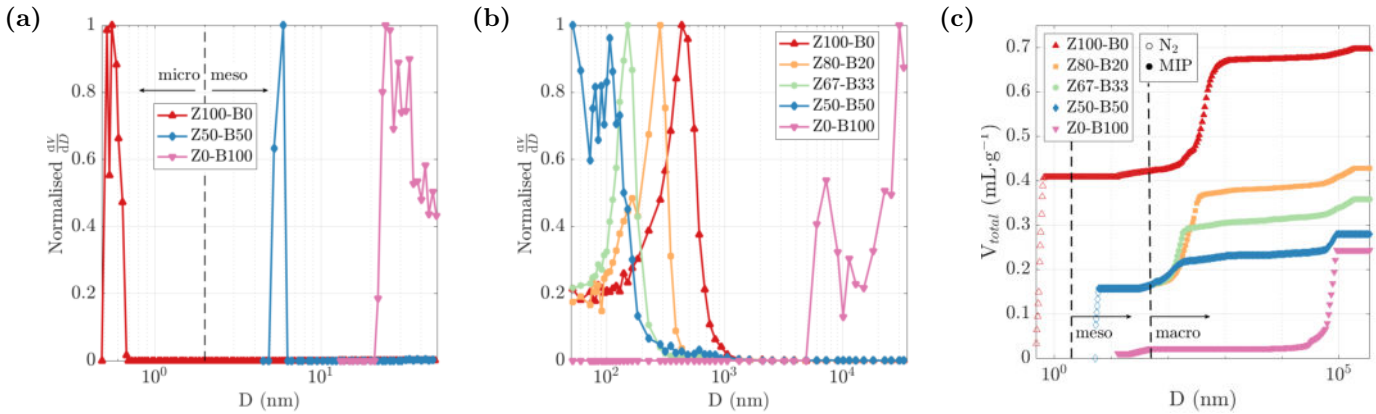


Figure 3: Pore size distributions (PSD) of the samples: (a) normalised differential PSD in the micro- and mesopore range for Z100-B0, Z50-B50 and Z0-B100 obtained from N_2 physisorption at 77 K; (b) normalised differential PSD in the macropore range for Z100-B0, Z80-B20, Z67-B33, Z50-B50 and Z0-B100 obtained from MIP; and (c) cumulative PSD for Z100-B0, Z80-B20, Z67-B33, Z50-B50 and Z0-B100 obtained from N_2 physisorption at 77 K and MIP combined.

3(b), the average size of the macropores in the blended beads decreases with increasing bentonite, excluding Z0-B100. This may indicate that bentonite fills the larger pores, causing pore blockage. This is in line with the decreasing D_p values shown in Table 1. This suggests that the rate of macropore diffusion decreases due to the additional resistances to mass transfer from pore blockage.

We did not obtain the full data set for N_2 physisorption at 77 K due to time constraints, so to increase the reliability of our results, we would need to obtain credible data for Z80-B20, Z67-B33 and Z33-B67. We also would need to establish MIP data for Z33-B67.

4.3 CO_2 adsorption equilibrium

CO_2 adsorption equilibrium isotherms at 303 K are shown in Figure 4. The associated DSL model parameters can be seen in Table A2. The DSL model (solid line) exhibits excellent agreement with our experimental data (markers). Additionally, the values for final CO_2 uptake at 303 K and 1 bar for Z100-B0 and Z0-B100 were both in line with literature [23, 42]. Given the alignment, we used these isotherms as inputs to our model for subsequent analysis on the equilibrium and kinetic performance of our samples.

Generally, as the content of bentonite increases, the CO_2 uptake decreases. This is in line with the pore size distributions seen in Figure 3 and the trend in BET surface area seen in Table 1. Bentonite content reduces the number of micropores and thus internal surface area, limiting the number of adsorption sites available for CO_2 molecules. It is interesting to note that this reduction is not linear. At low wt% of bentonite, the measured and model-predicted isotherms are very similar, as seen in Z80-B20 and Z67-B33. This indicates that within these samples, bentonite does not interact or interfere with the uptake of zeolite 13X. Contrastingly, the measured uptake of Z50-B50 is considerably less than the theoretical prediction. This reinforces the idea of pore blockage presented in Section 4.2. As such, bentonite may not only reduce the number of micropores, but it also may inhibit the accessibility of CO_2 molecules to the remaining micropores. This trend could be confirmed by completing an additional measurement on Z67-B33 and repeating experiments.

4.4 CO_2 adsorption kinetics

Prior to conducting kinetic experiments, preliminary tests were performed to determine the correct CO_2 gas flowrate to use. Since bead size influences adsorption kinetics [43], we carried out two sets of flowrate tests. One for the beads we created, and another for the commercially obtained Z100-B0 beads due to the noticeable disparity in R_p. The results of these preliminary experiments can be seen in Figures 5 and Figures A4 to A17. For kinetic experiments, it is important to perform experiments in the correct kinetic regime, where gas flowrate is sufficiently high to ensure that total uptake and rate of uptake of CO_2 remain unvaried with further increases in flowrate [44]. This is to ensure that any observations in kinetic behaviour can be solely attributed to differences in sample properties rather than external mass transfer limitations.

Considering Z50-B50, Figure 5(a) shows that CO_2 uptake is relatively consistent across the three flowrates whilst the

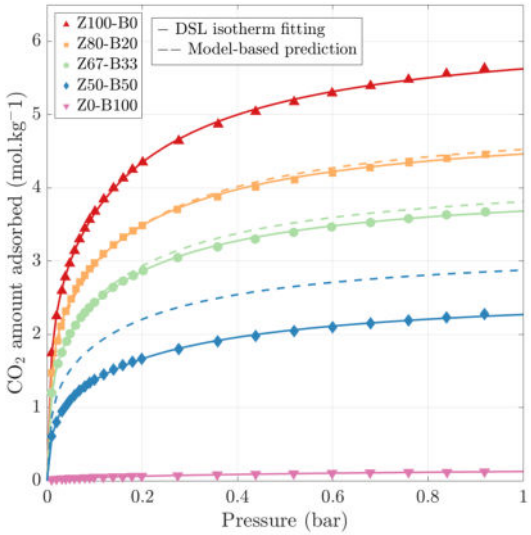


Figure 4: CO_2 adsorption isotherms at 303 K with fitted adsorption isotherms (solid line, DSL model), and model-based predicted adsorption isotherms (dashed line) for Z100-B0, Z80-B20, Z67-B33, Z50-B50, and Z0-B100.

kinetics still vary, as seen by the increasing k_{LDF} values. This suggests that at F = 350 mL·min^{-1}, the system has reached equilibrium but has not yet achieved the ideal kinetic regime. However, the decreasing slope between increasing flowrates signals that the system is approaching this regime, where further increases in flowrate will likely result in smaller increases in k_{LDF}. Considering Z100-B0, Figure 5(b) shows that both CO_2 uptake and k_{LDF} are still increasing, suggesting that the system has neither reached equilibrium, nor the ideal kinetic regime. With this sample, it is difficult to conclude which flowrate would result in the correct kinetic regime. Despite this, we proceeded with F = 350 mL·min^{-1} to perform kinetic experiments on all our samples as our equipment was constrained by this flowrate. We acknowledged that the results of our kinetic experiments may be subject to both sample properties and external mass transfer limitations.

The results of our CO_2 adsorption kinetic studies obtained gravimetrically and by fitting the LDF model as described in Section 2.3.3 for each sample are presented in Figures 6 and Figures A18 to A25. Figure 6(a) displays the normalised fractional uptake curves and Figure 6(b) displays how k_{LDF} varies with bentonite content, allowing us to qualitatively and quantitatively assess the impact of bentonite content on kinetic performance. Figure 6(a) includes the repeat of each sample that qualitatively had the best LDF fit, and excludes the result for Z0-B100. For this sample, the equipment could not accurately record the CO_2 uptake as it was too low, producing a noisy result. This experiment can be seen in Figure A24. Our kinetic experiments were carried out between p/p_0 (CO_2) = 0.03 to 0.06 at 303 K. The CO_2 uptake between these values, measured both volumetrically and gravimetrically can be seen in Figure A3.

Although there was doubt from our preliminary experiments that these tests were recorded under the correct kinetic regimes, the systematic variation of the kinetic curves with each sample clearly show that kinetic behaviour is sample dependent, with k_{LDF} decreasing as the wt% of bentonite increases. This is in line with the trend seen in

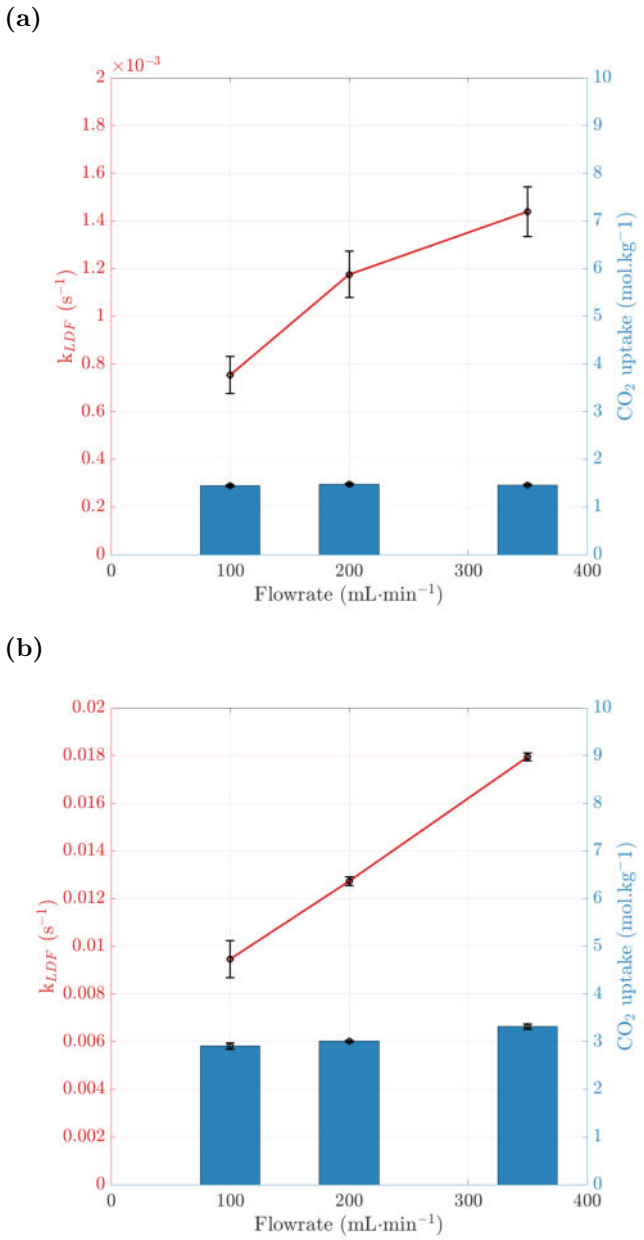


Figure 5: Quantitative analysis of the flowrate tests carried out gravimetrically at F = 100, 200 and 350 mL·min^{-1} between p/p_0 (CO_2) = 0.03 to 0.06 at 303 K for: (a) Z50-B50 and (b) Z100-B0. In red: k_{LDF} obtained from the LDF fit; blue: CO_2 uptake at p/p_0 (CO_2) = 0.06.

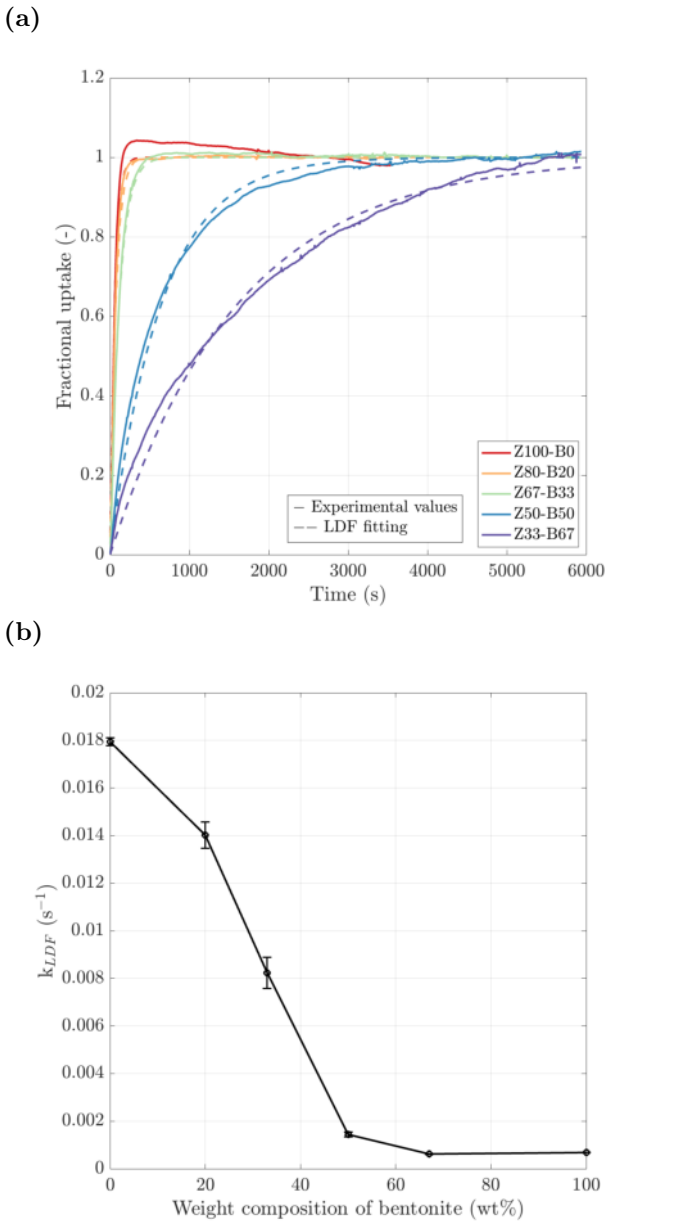


Figure 6: Quantitative and qualitative analysis of the kinetic tests carried out gravimetrically at F = 350 mL·min^{-1} between p/p_0 (CO_2) = 0.03 to 0.06 at 303 K for Z100-B0, Z80-B20, Z67-B33, Z50-B50, Z33-B67: (a) Experimental normalised adsorption curves (solid line) with fitted adsorption curves (dashed line, LDF model); (b) Variation of k_{LDF} obtained from the LDF fit with bentonite content.

Section 4.3. An intrinsic assumption of the DSL model is that the rate of adsorption is directly proportional to the number of adsorption sites. Bentonite content reduces the number of micropores, thus reducing the number of adsorption sites and in turn resulting in a slower rate of CO_2 uptake. The reduction in k_{LDF} is not linear, as seen with the equilibrium uptake, reinforcing the idea that bentonite causes pore blockage once again.

The impact of bentonite on the kinetics of zeolite 13X was further evaluated by analysing the fit of each sample to the LDF model, as seen in Figure 6(a). Due to the fast kinetics of zeolite 13X, the kinetic tests of Z100-B0 and Z80-B20 required modification to the programme specified in Section 2.3.3. To improve the accuracy of the CO_2 adsorption measurements, both samples were held at p/p_0 (CO_2) = 0.03 for 2.3 h and p/p_0 (CO_2) = 0.06 for 1 h as opposed to 1.7 h at each p/p_0. Whilst this adjustment corrected the response of Z80-B20, the rapid kinetics of Z100-B0 caused an overshoot above a fractional uptake of one, highlighting that there were limitations in our experimental setup for such a fast process. Regardless, the LDF model (dotted line) exhibits reasonably good agreement with our experimental data (solid line) for Z80-B20 and Z67-B33. However, as bentonite content increases, the agreement reduces, as seen in Z50-B50 and Z33-B67. Wang, J.Y. et al. [43] describes common "poor fits" between kinetic models and gravimetric data, which allowed us to qualitatively assess the LDF fit. In our study, Z50-B50 displays a Type 1 "poor fit", where the curves intersect at one point and Z33-B67 displays a Type 2 "poor fit", where the curves intersect twice. Both of these indicate the presence of two time constants in the process, possibly reflecting the presence of distinct diffusion mechanisms in

Table 2: Macropore and micropore diffusion time constants for Z100-B0, Z80-B20, Z67-B33, Z50-B50, and Z0-B100 obtained using our computational model.

Sample	D^e_{mac} ($\times 10^{-5}$ m^2·s^{-1})	$\frac{1}{t_{\text{mic}}}$ ($\times 10^{-3}$ s^{-1})	t_{mac} ($\times 10^{-2}$ s)	t_{mic} ($\times 10^{3}$ s)	$t_{\text{mac}}/t_{\text{mic}}$
Z100-B0	1.13	1.34 ± 0.04	10.70	0.75	1.43×10^{-4}
Z80-B20	1.62	1.04 ± 0.09	1.71	0.96	1.78×10^{-5}
Z67-B33	1.31	0.59 ± 0.08	2.91	1.96	3.38×10^{-5}
Z50-B50	0.78	0.09 ± 0.01	3.36	10.40	3.22×10^{-6}
Z0-B100	1.07	–	1.96	–	–

the macropore and micropore regions. However, a more likely explanation for the presence of multiple time constants in the process is the use of a suboptimal flowrate. As previously discussed, the selection of $F = 350$ mL·min^{-1} was likely not representative of the correct kinetic regime. From the DSL model, the rate of adsorption is also directly proportional to the bulk concentration of CO_2 molecules. Thus, if the flowrate is not sufficiently high enough, the rate of adsorption may be restricted by the number of CO_2 molecules, affecting the kinetics of our samples alongside any internal diffusion effects. This is a common source of error when using gravimetric equipment to capture the kinetics of fast processes [43].

To gain further insight into the internal diffusion mechanisms within our samples, we analysed our experimental data using our computational model. The macropore and micropore diffusion time constants obtained for each sample can be seen in Table 2. In shaped zeolite 13X adsorbents, the diffusion mechanism is typically macropore-controlled [45]. In a macropore controlled system, one would expect $t_{\text{mac}}/t_{\text{mic}} > 1$. However, our values show that $t_{\text{mac}}/t_{\text{mic}} < 1$ for all samples, which is particularly surprising for Z100-B0. This contradiction to literature suggests that our numerical data is unreliable, likely due to the following compounded errors: (i) our kinetic experiments were carried out in the incorrect kinetic regime, using equipment that was unsuitable for the fast kinetics of zeolite 13X, and (ii) the manual parameter selection required to extract $\frac{1}{t_{\text{mic}}}$ introduced bias and limited reproducibility. Despite this, our values for $\frac{1}{t_{\text{mic}}}$ and D_p for Z100-B0 and Z80-B20 were in the same order of magnitude and showed trends consistent with those reported in literature [22]. This gave us sufficient confidence to draw a qualitative assessment from our data, which shows that $t_{\text{mac}}/t_{\text{mic}}$ decreases with increasing bentonite content, indicating that the system becomes more micropore controlled. In Section 4.2, we established that increasing bentonite content reduces the rate of macropore diffusion. Therefore, the decreasing $t_{\text{mac}}/t_{\text{mic}}$ values may imply that bentonite content additionally reduces the rate of micropore diffusion. The reduction in micropore volume and BET surface area, alongside the potential pore blocking effect outlined in Section 4.2 provides insight into this trend, suggesting that bentonite introduces additional resistances to mass transfer within the crystalline structure of zeolite 13X.

To allow conclusions to be drawn with greater certainty, future work should aim to minimise external mass transfer effects to ensure more reliable kinetic measurements. Alternate equipment could be used that is better suited to capturing the kinetics of fast processes, such as the zero length column (ZLC) technique [46].

5 Conclusion

In this study, we created five batches of zeolite 13X beads blended with 20, 33, 50, 67 and 100 wt% of bentonite using an extruder-spheroniser. We investigated the difference in the porous structure of these samples, alongside binderless, commercially obtained zeolite 13X, and the subsequent effects on CO_2 adsorption properties for application in post-combustion carbon capture. Bentonite content increases the number of larger pores, seen by the shift in the pore size distributions toward the larger meso- and macropore ranges, by agglomerating onto the microporous framework of zeolite 13X. These findings seem to directly impact the equilibrium uptake and kinetics of zeolite 13X, providing an explanation for key observations in our experimental data. With regards to CO_2 adsorption equilibrium, reduced equilibrium CO_2 uptake values are observed with increased amounts of bentonite, attributable to the fewer number of micropores. Not only this, but bentonite reduces equilibrium uptake in a non-linear way. The theoretical isotherms produced by our computational model brought us to conclude that bentonite may cause pore blockage as higher wt%, seen by the increasing deviation of the experimental data from our theoretical isotherms. Finally, from our kinetic studies, we found that bentonite content may alter the diffusion mechanism of zeolite 13X, shifting it towards being more micropore controlled. The impacted diffusion pathways can again be associated with the altered structure that bentonite creates in the beads, causing a decrease in average pore size and reducing the number of micropores. This can be seen with the increasing deviation of our kinetic adsorption curves from the LDF model with increasing bentonite content, indicating the presence of two time constants. We supported this by exploring the coupled diffusion-adsorption mechanism with our computational model. Overall, this study provides a preliminary qualitative assessment of the impact of bentonite on the porous structure and thus CO_2 adsorption properties of zeolite 13X. A primary limitation of our study was the use of unsuitable equipment for kinetic experiments that was not fit for the fast kinetics of zeolite 13X. This work would be benefited by repeating kinetic experiments using alternate equipment, such as ZLC. Additionally, we would benefit from obtaining complete data sets and repeats for all our experiments, especially for N_2 physisorption at 77 K.

6 Acknowledgements

Acknowledgements go to Prof. Camille Petit, Mr. Killian Gmyrek and the entirety of the Petit Group for their continued assistance with this research. We would also like to thank Ms Patricia Carry and Mr Kaho Cheung for their

support in using the analytical services equipment.

References

[1] United Nations. *Paris Agreement*. 2015. URL: https://unfccc.int/sites/default/files/english_paris_agreement.pdf.

[2] United States Environmental Protection Agency. *Inventory of U.S. Greenhouse Gas Emissions and Sinks*. 2024. URL: https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks.

[3] D. Baskaran et al. "An overview of technologies for Capturing, Storing, and utilizing carbon Dioxide: Technology Readiness, large-scale Demonstration, and cost". In: *Chemical Engineering Journal* (2024), pp. 151998–151998. DOI: https://www.sciencedirect.com/science/article/pii/S1385894724034855.

[4] A. Das et al. "Advancements in adsorption based carbon dioxide capture technologies- A comprehensive review". In: *Heliyon* 9.12 (2023), e22341. DOI: 10.1016/j.heliyon.2023.e22341. URL: https://www.sciencedirect.com/science/article/pii/S2405844023095491X1.

[5] S. Sjostrom and H. Krutka. "Evaluation of solid sorbents as a retrofit technology for CO_2 capture". In: *Fuel* 89.6 (2010), pp. 1298–1306. DOI: 10.1016/j.fuel.2009.11.019. URL: https://www.sciencedirect.com/science/article/pii/S0016236109005286.

[6] M. Kong et al. "A review on development of post-combustion CO_2 capture technologies: Performance of carbon-based, zeolites and MOFs adsorbents". In: *Fuel* 371 (2024), p. 132103. URL: https://www.sciencedirect.com/science/article/pii/S0016236124012511.

[7] Z. Hu et al. "CO_2 Capture in Metal-Organic Framework Adsorbents: An Engineering Perspective". In: *Advanced Sustainable Systems* 3.1 (2018), p. 1800080. DOI: 10.1002/adsu.201800080. URL: https://doi.org/10.1002/adsu.201800080.

[8] N. Hedin et al. "Adsorbents for the post-combustion capture of CO_2 using rapid temperature swing or vacuum swing adsorption". In: *Applied Energy* 104 (2013), pp. 418–433. DOI: 10.1016/j.apenergy.2012.11.034. URL: https://www.sciencedirect.com/science/article/pii/S0306261912008276.

[9] A. Maghfirah et al. "A review on the green synthesis of hierarchically porous zeolite". In: *Materials Today Chemistry* 17 (2020), p. 100348. DOI: 10.1016/j.mtchem.2020.100348. URL: https://www.sciencedirect.com/science/article/pii/S2468519420301087.

[10] D.G. Boer, J. Langerak, and P.P. Pescarmona. "Zeolites as Selective Adsorbents for CO_2 Separation". In: *ACS Applied Energy Materials* 6.5 (2023), pp. 2634–2656. DOI: 10.1021/acsaem.2c03605. URL: https://pubs.acs.org/doi/10.1021/acsaem.2c03605.

[11] S.G. Hill and D. Seddon. "The hygroscopic nature of H-ZSM—5". In: *Zeolites* 5(3) (1985), pp. 173–178. DOI: https://doi.org/10.1016/0144-2449(85)90026-0.

[12] Z. Zhang et al. "Adsorption of CO_2 on Zeolite 13X and Activated Carbon with Higher Surface Area". In: *Separation Science and Technology* 45.5 (2010), pp. 710–719. DOI: 10.1080/01496390903571192. URL: https://www.tandfonline.com/doi/full/10.1080/01496390903571192.

[13] S. Wang and Y. Peng. "Natural zeolites as effective adsorbents in water and wastewater treatment". In: *Chemical Engineering Journal* 156.1 (2010), pp. 11–24. DOI: 10.1016/j.cej.2009.10.029. URL: https://www.sciencedirect.com/science/article/pii/S1385894709007219#:~:text=Natural%20zeolites%20are%20abundant%20and,capacity%20for%20organics%20and%20anions..

[14] A. Arefi Pour et al. "Performance evaluation of clinoptilolite and 13X zeolites in CO_2 separation from CO_2/CH_4 mixture". In: *Journal of Natural Gas Science and Engineering* 26 (2015), pp. 1246–1253. DOI: 10.1016/j.jngse.2015.08.033. URL: https://www.sciencedirect.com/science/article/pii/S1875510015300986.

[15] G. I. Berezin and A. V. Kiselev. "Adsorbate-adsorbate association on a homogenous surface of a nonspecific adsorbent". In: *Journal of Colloid and Interface Science* 38.1 (2004), pp. 227–233. DOI: 10.1016/0021-9797(72)90238-X. URL: https://www.sciencedirect.com/science/article/abs/pii/002197977290238X#:~:text=Various%20methods%20for%20calculating%20the%20degree%20of,association%20of%20the%20molecules%20parallel%20to%20the.

[16] M. Nedoma et al. "The effect of adsorbent shaping on the equilibrium and kinetic CO_2 adsorption properties of ZIF-8". In: *Microporous and Mesoporous Materials* 380 (2024), p. 113303. URL: https://www.sciencedirect.com/science/article/pii/S1387181124003251?via%3Dihub.

[17] Z. Vajglová et al. "Effect of Binders on the Physicochemical and Catalytic Properties of Extrudate-Shaped Beta Zeolite Catalysts for Cyclization of Citronellal". In: *Organic Process Research & Development* 23.11 (2019), pp. 2456–2463. DOI: 10.1021/acs.oprd.9b00346. URL: https://pubs.acs.org/doi/full/10.1021/acs.oprd.9b00346#.

[18] D.-A. Chisăliță, J. Boon, and L. Lücking. "Adsorbent shaping as enabler for intensified pressure swing adsorption (PSA): A critical review". In: *Separation and Purification Technology* 353 (2025), p. 128466. DOI: 10.1016/j.seppur.2024.128466. URL: https://www.sciencedirect.com/science/article/pii/S1383586624022056.

[19] F.A.A. Mamaghani, A. Salem, and S. Salem. "A novel technique for fabrication of rod-like shape zeolite LTA and hydroxysodalite by extrusion of bentonite powder: Effects of technical factors on structural characteristics". In: *Advanced Powder Technology* 34.8 (2023), p. 104085. DOI: 10.1016/j.apt.2023.104085. URL: https://www.sciencedirect.com/science/article/pii/S0921883123001486.

[20] Z. Asgar Pour et al. "A Comparative Review of Binder-Containing Extrusion and Alternative Shaping Techniques for Structuring of Zeolites into Different Geometrical Bodies". In: *Catalysts* 13.4 (2023), p. 656. DOI: 10.3390/catal13040656. URL: https://www.mdpi.com/2073-4344/13/4/656.

[21] Z. Zhao et al. "Adsorption of carbon dioxide on alkali-modified zeolite 13X adsorbents". In: *International Journal of Greenhouse Gas Control* 1.3 (2007), pp. 355–359. DOI: 10.1016/s1750-5836(07)00072-2. URL: https://www.sciencedirect.com/science/article/pii/S1750583607000722.

[22] A. Najafi et al. "Effect of binder on CO_2, CH_4, and N_2 adsorption behavior, structural properties, and diffusion coefficients on extruded zeolite 13X". In: *Chemosphere* 324 (2023), pp. 138275–138275. URL: https://www.sciencedirect.com/science/article/pii/S0045653523005428.

[23] C. Chen, D.-W. Park, and W.-S. Ahn. "CO_2 capture using zeolite 13X prepared from bentonite". In: *Applied Surface Science* 292 (2014), pp. 63–67. DOI: 10.1016/j.apsusc.2013.11.064. URL: https://doi.org/10.1016/j.apsusc.2013.11.064.

[24] H. Liu, B. Xie, and Y. Qin. "Effect of Bentonite on the Pelleting Properties of Iron Concentrate". In: *Journal of Chemistry* 2017 (2017), e7639326. DOI: 10.1155/2017/7639326. URL: https://www.researchgate.net/publication/317805616_Effect_of_Bentonite_on_the_Pelleting_Properties_of_Iron_Concentrate.

[25] S.R. Jeong et al. "Adsorption characteristics and separation behavior of binder and binderless zeolite LiX Pellets: O_2, N_2, and CO_2". In: *Separation and Purification Technology* 355 (2024), p. 129702. DOI: 10.1016/j.seppur.2024.129702. URL: https://www.sciencedirect.com/science/article/pii/S1383586624034415.

[26] S. Brunauer, P. H. Emmett, and E. Teller. "Adsorption of Gases in Multimolecular Layers". In: *Journal of the American Chemical Society* 60(2) (1938), pp. 309–319. URL: https://pubs.acs.org/doi/pdf/10.1021/ja01269a023.

[27] C. Wang and J.M. Smith. "Tortuosity factors for diffusion in catalyst pellets". In: *AIChE Journal* 29.1 (1983), pp. 132–136. DOI: 10.1002/aic.690290118.

[28] Jörg Kärger, D.M. Ruthven, and P. Doros. "Diffusion Mechanisms". In: *Diffusion in Zeolites and Other Microporous Solids*. 2012, pp. 85–110. DOI: 10.1002/9783527651276.ch4. URL: https://onlinelibrary.wiley.com/doi/10.1002/9783527651276.ch4.

[29] Imperial College London. *Isotherm Fitting Tool*. https://github.com/ImperialCollegeLondon/IsothermFittingTool. Accessed: 2024-12-04. 2024.

[30] O. Akeeb et al. "Post-combustion CO2 capture via a variety of temperature ranges and material adsorption process: A review". In: *Journal of Environmental Management* 315 (2022), p. 115026. DOI: 10.1016/j.jenvman.2022.115026.

[31] A. Olajire. "CO_2 capture and separation technologies for end-of-pipe applications – A review." In: *Journal of Environmental Management* 91.5 (2010), pp. 1261–1268. DOI: 10.1016/j.jenvman.2010.01.001. URL: https://www.sciencedirect.com/science/article/pii/S0360544210000848.

[32] K. Kamiuto, A. Goubaru, and Ermalina. "Diffusion coefficients of carbon dioxide within type 13X zeolite particles". In: *Chemical Engineering Communications* 193.5 (2006), pp. 628–638. DOI: 10.1080/00986440500193970.

[33] R. Dreu et al. "Physicochemical properties of granulating liquids and their influence on microcrystalline cellulose pellets obtained by extrusion-spheronisation technology". In: *International Journal of Pharmaceutics* 291(1-2) (2004), pp. 99–111. URL: https://www.sciencedirect.com/science/article/abs/pii/S0378517304006581.

[34] M. Thommes et al. "Physisorption of gases, with special reference to the evaluation of surface area and pore size distribution (IUPAC Technical Report)". In: *Pure and Applied Chemistry* 87(9-10) (2015), pp. 1051–1069. URL: https://www.3p-instruments.com/wp-content/uploads/2017/04/2015-IUPAC-Technical-Report.pdf.

[35] Y. Sun et al. "Adsorption, separation and regeneration of cation-exchanged X zeolites for LNG purification: Li+, K+, Mg2+ and Ca2+." In: *Microporous and Mesoporous Materials* 340 (2022), pp. 112032–112032. URL: https://www.sciencedirect.com/science/article/abs/pii/S138718112200350X.

[36] C. Chen et al. "Carbon dioxide capture using amine-impregnated HMS having textural mesoporosity." In: *Chemical Engineering Journal* 161(1-2) (2010), pp. 46–52. URL: https://www.sciencedirect.com/science/article/abs/pii/S1385894710003451.

[37] B. Freitas et al. "Robust, flexible, freestanding and high surface area activated carbon and multi-walled carbon nanotubes composite material with outstanding electrode properties for aqueous-based supercapacitors". In: *Materials Advances* 2.13 (2021), pp. 4264–4276. DOI: 10.1039/d0ma00783h.

[38] T. Zhao et al. "Microporous carbon coated zeolite particles for efficient carbon capture from wet flue gas". In: *Separation and Purification Technology* 317 (2023), p. 123762. DOI: 10.1016/j.seppur.2023.123762. URL: https://www.sciencedirect.com/science/article/pii/S1383586623006706.

[39] C. Trayford and S. van Rijt. "In situ modified mesoporous silica nanoparticles: synthesis, properties and theranostic applications". In: *Biomaterials Science* (2024). DOI: 10.1039/d4bm00094c. URL: https://pubs.rsc.org/en/content/articlehtml/2024/bm/d4bm00094c.

[40] W. Schwieger et al. "Hierarchy concepts: classification and preparation strategies for zeolite containing materials with hierarchical porosity". In: *Chemical Society Reviews* 45.12 (2016), pp. 3353–3376. DOI: 10.1039/c5cs00599j. URL: https://pubs.rsc.org/en/content/articlelanding/2016/cs/c5cs00599j/unauth.

[41] F. A. L. Dullien. *Porous Media: Fluid Transport and Pore Structure*. Elsevier BV, 1979. DOI: https://doi.org/10.1016/b978-0-12-223650-1.x5001-3.

[42] Elham Khoramzadeh, Masoud Mofarahi, and Chang-Ha Lee. "Equilibrium Adsorption Study of CO2 and N2 on Synthesized Zeolites 13X, 4A, 5A, and Beta". In: *Journal of Chemical Engineering Data* 64.12 (2019), pp. 5648–5664. DOI: 10.1021/acs.jced.9b00690.

[43] J. Y. Wang, E. Mangano, and S. Brandani. "A review of common practices in gravimetric and volumetric adsorption kinetic experiments." In: *Adsorption* 27.5 (2020), pp. 295–318. DOI: 10.1007/s10450-020-00276-7.

[44] D. O. Ojwang, J. Grins, and G. Svensson. "The adsorption kinetics of CO2 on copper hexacyanoferrate studied by thermogravimetric analysis." In: *Microporous and Mesoporous Materials* Volume Number.Issue Number (2018), Page Range. DOI: 10.1016/j.micromeso.2018.06.019. URL: https://doi.org/10.1016/j.micromeso.2018.06.019.

[45] X. Hu, E. Mangano, and D. Friedrich. "Diffusion mechanism of CO2 in 13X zeolite beads." In: *Adsorption* 19.4 (2013), pp. 701–714. DOI: 10.1007/s10450-013-9554-z. URL: https://doi.org/10.1007/s10450-013-9554-z.

[46] S. Brandani and E. Mangano. "The zero length column technique to measure adsorption equilibrium and kinetics: lessons learnt from 30 years of experience". In: *Adsorption* 27 (2021), pp. 319–351. DOI: 10.1007/s10450-020-00273-w.

Techno-Economic Assessment of Oxyfuel-Based Carbon Capture and Storage (CCS) in the Cement Industry

Xiang Hong Tan and Hamad Al Khalifa
Department of Chemical Engineering, Imperial College London, U.K.

Abstract
Oxyfuel-based Carbon Capture and Storage (CCS) is an emerging technology that significantly reduces CO_2 emissions by capturing CO_2 at high concentrations for transport and storage. This study presents a techno-economic assessment of integrating oxyfuel CCS into cement production processes using technical key performance indicators (KPIs). This research found that oxyfuel-based CCS has the potential to capture up to 86% of carbon dioxide generated, reducing net CO_2 emission to 0.11 tonne CO_2 per tonne of cement produced with lower thermal duty (-26%) but higher electrical duty (+100%). The cost of CO_2 avoided (CAC) was found to be €76.4, which is highly sensitive to equipment purchase costs and discount rates, whereas labour and fuel costs had minimal impact. This study underlines the potential of oxyfuel CCS to decarbonise the cement industry, identifies critical cost drivers and presents avenues to improve the performance of oxyfuel-based CCS.

Keywords – Oxyfuel combustion, Techno-economic assessment, Cost of CO_2 avoided (CAC)

1 Introduction

Cement production is essential for construction and infrastructure development, with worldwide demand persistently increasing. The global cement industry is projected to grow from $423 billion in 2024 to $592 billion by 2032 (Fortune Business Insights, 2024). Consequently, the industry accounts for approximately 8% of global carbon dioxide (CO_2) emissions (World Economic Forum , 2024). International institutions, such as the International Energy Agency (IEA) and the Cement Sustainability Initiative (CSI), have outlined measures to mitigate these emissions. This includes decreasing the clinker-to-cement ratio, utilising alternative fuels with lower carbon content, and employing carbon capture and storage (CCS) technology (Gardarsdottir et al., 2019).

Oxyfuel combustion is a carbon capture and storage (CCS) technique that substitutes air with pure oxygen in the kiln. This removes nitrogen (N_2) from the combustion process, yielding flue gas predominantly consisting of CO_2 and water vapour (H_2O). The elevated CO_2 concentration facilitates capture and decreases emissions (IEAGHG, 2013). In addition, oxyfuel combustion enhances combustion efficiency but elevates flame temperatures beyond 3500°C, potentially damaging kiln components. To alleviate this, a segment of the CO_2-rich flue gas is reintroduced into the kiln to lower the temperature to controllable levels (D.A. Granados et al., 2015). Additional CCS techniques consist of amine scrubbing and calcium looping. Amine scrubbing works by chemically absorbing CO_2 from flue gases using a solvent like monoethanolamine (MEA), succeeded by a heat-induced regeneration process to liberate the CO_2. Calcium looping uses the reversible reaction between calcium oxide (CaO) and CO_2 to reduce emissions, and uses high-temperature heat to effectively capture the carbon (IEAGHG, 2013). These methods demonstrate continuous endeavours to decarbonise cement manufacturing.

The primary aim of the report was to simulate a cost-effective full oxyfuel combustion cement plant. This was compared to a baseline cement plant without carbon capture that was also simulated utilising Aspen Plus V11. Each configuration produces 1 million tonnes of clinker per annum. Analysis consisted of investigating ways of optimising CO_2 capture, while achieving total conversion of calcium carbonate. Moreover, a technical and economic assessment of the oxyfuel plant was carried out to determine its suitability as a cost-effective carbon capture technology. As well as that, a sensitivity analysis on the cost of CO_2 avoided (CAC) was also conducted.

2 Background

The process of cement production begins with the extraction and crushing of raw materials, primarily limestone ($CaCO_3$) and clay, which are proportioned and finely milled into a raw meal. This is preheated in a succession of cyclones, where partial calcination takes place, before being introduced into a rotary kiln. In the kiln, temperatures reach up to 1450°C, facilitating a series of chemical reactions necessary for clinker formation (Civil Today, 2023). One of the primary reactions in cement production is the calcination of calcium carbonate, occurring at around 900°C. This reaction generates calcium oxide (CaO) and emits carbon dioxide (Eqn. (1)), making up approximately 60% of the CO_2 emissions in the cement manufacturing process.

$$CaCO_3\ (s) \rightarrow CaO\ (s) + CO_2 \quad (1)$$

Alongside the calcination of calcium carbonate, carbon dioxide is emitted during the synthesis of magnesium oxide from magnesium carbonate (Eqn. (2)),. This reaction adds to CO_2 emissions throughout the cement production process.

$$MgCO_3 \rightarrow MgO + CO_2 \quad (2)$$

After calcination, the synthesis of clinker compounds commences. This consists of multiple reactions occurring at approximately 1450°C, which are exothermic and influence the physical qualities and performance of cement.

The initial reaction involves the synthesis of dicalcium silicate (belite), resulting from the interaction between calcium oxide and silica (SiO_2) (Eqn. (3)). This

reaction yields belite, a phase that enhances the long-term strength of cement.

$$2CaO + SiO_2 \rightarrow Ca_2SiO_4 \quad (3)$$

As the temperature increases, tricalcium silicate (alite) is produced when belite interacts with an extra molecule of calcium oxide (Eqn. (4)). Alite contributes to the initial strength progression of cement.

$$3CaO + SiO_2 \rightarrow Ca_3SiO_5 \quad (4)$$

The interaction between calcium oxide and alumina (Al_2O_3) produces tricalcium aluminate (Eqn. (5)). This chemical affects the setting time of cement by hydrating with water to produce ettringite.

$$3CaO + Al_2O_3 \rightarrow Ca_3Al_2O_6 \quad (5)$$

Finally, the interaction between calcium oxide, alumina, and iron oxide produces tetracalcium aluminoferrite (Eqn. (6)), which contributes to the durability of cement and its characteristic grey colour.

$$4CaO + Al_2O_3 + Fe_2O_3 \rightarrow Ca_4Al_2Fe_2O_{10} \quad (6)$$

These processes combined contribute close to 40% of the CO_2 emissions in cement manufacture. The remaining emissions stem from the combustion of fossil fuels needed to attain the elevated temperatures essential for clinker production. Furthermore, indirect emissions arise from electricity usage and the transfer of raw materials and completed goods (Environment et al., 2018).

Oxyfuel combustion has been the focus of a variety of research and studies in its development for implementation in cement production. Among carbon capture technologies, its cost of CO_2 avoided (CAC) ranges from €30 to €70 per tonne of CO_2 avoided. Comparatively, this is significantly lower compared to those of post-combustion capture technologies, which generally range between €50–€100 for every tonne of CO_2 (Leeson et al., 2017). When compared in terms of configurations, full oxyfuel systems tend to be more economically feasible than their partial ones; the CAC for both configurations are €42.4 and €55.4 per tonne of CO_2, respectively (Quevedo et al., 2022). Although oxyfuel combustion is associated with high efficiency, it involves substantial energy penalties related to the required air separation and CO_2 purification steps. These energy demands can increase thermal energy consumption by up to 40–50% compared to conventional operations. However, continuing improvements in design and optimisation of the system are increasingly mitigating these penalties, which improves the viability of carbon capture and storage (CCS) technologies in industrial contexts (Czakiert et al., 2022).

Global agreements, such as the Paris Agreement and the Glasgow Climate Pact, have set clear targets for limiting temperature increases to within 1.5°C and achieving net-zero emissions by 2050. The Intergovernmental Panel on Climate Change (IPCC) has emphasised the critical role of carbon capture and storage (CCS) in meeting these goals by capturing emissions, that are otherwise difficult to mitigate.

In addition to oxyfuel combustion , other innovative concepts for carbon capture are attracting attention. One of which is chemical looping combustion (CLC). CLC uses oxygen carriers, such as metal oxides, to facilitate combustion without the need for air separation, thereby enabling the direct capture of CO_2 during the process. This technology achieves high CO_2 capture efficiencies while lowering the energy demands compared to conventional carbon capture methods (Czakiert et al., 2022). Furthermore, advancements in reactor designs and the development of oxygen-transport materials have improved the scalability and economic viability of CLC for industrial applications. Another emerging technology is Direct Air Capture (DAC), which captures CO_2 directly from the atmosphere. Driver et al. (2022) highlighted the potential of integrating DAC facilities into cement production as an effective strategy to achieve net-zero emissions. While DAC systems face challenges related to high energy input and associated costs, ongoing advancements in sorbent materials and process optimisation show promise for improving the feasibility of this approach. Other, under-investigation options for cement industry decarbonisation include hydrogen fuel and biofuels. Hydrogen fuel can offer the potential for emissions reductions by replacing traditional fossil fuels, while biofuels combined with coal are considered renewable options with a smaller carbon footprint (Roussanaly et al., 2021).

3 Methodology

3.1 Research Design

This study employed a computational simulation approach to evaluate the technical and economic performance of oxyfuel-based CCS in cement production. Additionally, key parameters such as coal flow rate, oxygen purity and flue gas recirculation ratio were varied to study their effect on CO_2 purity, conversion and kiln temperatures.

3.2 Process simulation

Simulations were conducted using Aspen Plus V11 to model clinker production, CO_2 compression and purification. The IDEAL property method was selected due to the absence of vapor-liquid equilibrium and the low-pressure nature of the process (John, 2020). The plant was assumed to operate 24 hours per day and 330 days per year. Coal and ash were simulated as non-conventional components with their composition shown in Table 2. Input materials (CEMCAP, 2016) and fuels compositions (Shahabuddin & Bhattacharya, 2021) are provided in Tables 1 & 2.

3.3 Modelling baseline cement plant

3.3.1 Raw Meal & Clinker Mills

Mills, modelled as jaw crushers, were used to grind both raw meal and the mixture of clinker, gypsum ($CaSO_4$) and granulated blast furnace slag (GBFS) to improve calcination and clinker formation. The Bond work index

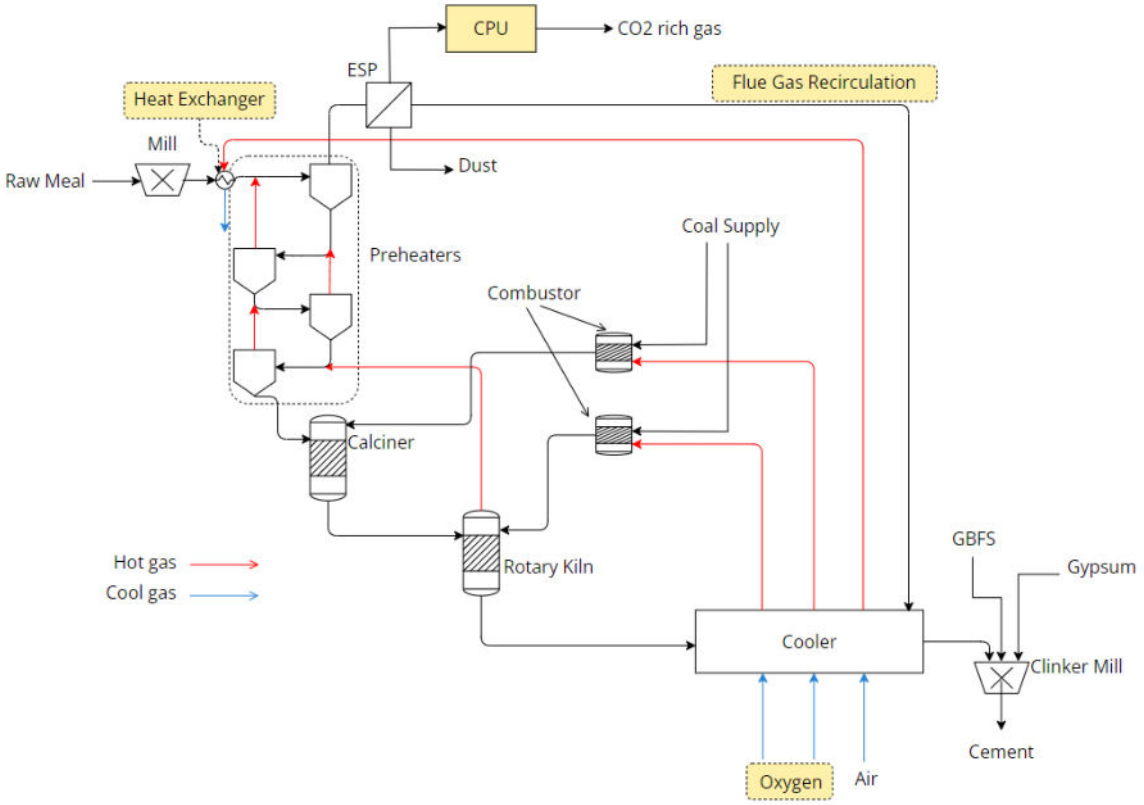


Figure 1 Process flowsheet for oxyfuel combustion cement plant

Table 1 Assumed composition of input materials

Material	Mass Composition (wt%)								
	CaCO3	SiO2	Al2O3	Fe2O3	MgCO3	CaO	MgO	CaSO4	H2O
Raw Meal	78.2	13.9	3.3	2.0	1.5	-	-	-	1.0
Gypsum	-	-	-	-	-	-	-	100.0	-
GBFS	-	43.1	15.0	-	-	34.3	7.6	-	-

Table 2 Assumed composition for coal, oxygen and ambient air

Material	HHV	Proximate (wt%)				Ultimate (wt%)					Sulfanal (wt%)		
	(MJ/kg)	Moisture	VM	FC	Ash	C	H	N	S	O	Pyritic	Sulfate	Organic
Coal	27	3.1	30.6	57.8	11.6	74.2	4.6	1.5	0.6	7.6	0.3	0.1	0.3
		Molar Composition (mol%)						N2		O2			
Oxygen	-							1.0		99.0			
Ambient Air	-							79.0		21.0			

was set at 20 kWh/ton, and mechanical efficiency was assumed to be 0.615 (Atmaca & Kanoglu, 2012) .

3.3.2 Preheaters & Calciner

Cyclone preheaters were used to heat raw materials before entering the rotary kiln, improving thermal efficiency and reducing fuel consumption. Four cyclone preheaters were modelled as a series of *RGibbs* reactors (for equilibrium-based reaction modelling) and *SSplit* solid separators. The primary reaction modelled in the preheaters was the decomposition of limestone (Eqn. (7-8)), with approximately 30% calcination completed in the preheaters. Heat losses were 19kJ/$kg_{clinker}$ for preheaters and 95.4 kJ/$kg_{clinker}$ for the calciner. Mechanical efficiencies were 96% (first stage), 86% (second and third stages) and 76% (final stage).

At 500-650°C,

$$MgCO_3 \rightarrow MgO + CO_2 \; \Delta Hr = +118 kJmol^{-1} \quad (7)$$

At 900 to 1050°C,

$$CaCO_3 \rightarrow CaO + CO_2 \quad \Delta Hr = +174 kJmol^{-1} \quad (8)$$

3.3.3 Combustor

Decomposition and combustion of fuel were modelled using *RYield* and *RGibbs* reactors. Air supply was connected to the combustor. Hot gaseous products supplied heat to the calciner and rotary kiln.

3.3.3 Rotary kiln

The rotary kiln was used to convert calcined raw materials into clinker. *RStoic* reactors were used to model the process as reaction extents were predefined based on empirical or literature data to achieve the

target composition of clinker (Driver et al., 2022). Clinker compounds simulated include primarily dicalcium silicate (Ca_2SiO_4, C_2S, belite), tricalcium silicate (Ca_3SiO_5, C_3S, alite) and aluminate ($Ca_3Al_2O_6$, C_3A) (Eqn. (9-11)).

Key reactions:
At 650-900°C,

$$2CaO + SiO_2 \rightarrow C_2S \quad \Delta Hr = -126 kJmol^{-1} \qquad (9)$$

At 700-900°C,

$$3CaO + Al_2O_3 \rightarrow C_3A \quad \Delta Hr = -7.3 kJmol^{-1} \qquad (10)$$

At 1300-1425°C,

$$3CaO + SiO_2 \rightarrow C_3S \quad \Delta Hr = -113 kJmol^{-1} \qquad (11)$$

3.3.4 Clinker cooler

The clinker cooler reduced the temperature of clinker from ~1600°C to 113°C (CEMCAP, 2018). It was modelled using three counterflow heat exchangers in Aspen Plus, where cold ambient air absorbed heat from the hot clinker. First and second cold stream outlet temperatures were set at 1137°C and 1000°C respectively.

3.3.5 Electrostatic Solids Precipitator (ESP)

An *ESP* was used to filter dust from the flue gas exiting the preheaters. The efficiency of the *ESP* was assumed to be 98%.

3.3.6 Cement formulation

Cooled clinker was mixed with gypsum and GBFS to form cement, which was then grinded in the clinker mill. A clinker-to-cement ratio of 0.72 was used. The cement composition is detailed in Table 3.

3.4 Modelling oxyfuel combustion

3.4.1 Air Separation Unit (ASU)

The air separation unit (ASU) was not explicitly modelled in this study but assumed to produce 99% purity oxygen for oxyfuel combustion.

3.4.2 Flue Gas Recirculation

A portion of the dust-filtered CO_2-rich flue gas was recycled into the combustors through the oxygen supply streams to recover heat.

3.4.3 Heat Exchanger

The hot stream of ambient air from the clinker cooler was connected to the heat exchanger to heat raw meal before it entered the preheaters.

3.4.4 Compression and Purification Unit (CPU)

The carbon dioxide rich flue gas was compressed and passed through 2 absorber towers with water as solvent to remove sulphur oxides and nitrogen oxides as modelled by Meunier et al., 2014. A dehydration unit using triethylene glycol (TEG) was modelled to remove water from the gas. TEG was selected for its high water absorption capacity, low vapour pressure, and regeneration capability (G. Speight, 2017). The water-free CO_2-rich gas was assumed to be liquefied using an ammonia refrigeration process ammonia (Oi et al., 2016). This step was not modelled in the study.

3.5 Key performance indicators (KPIs)

Key performance indicators (KPIs) were used to assess the technical and economic performance of baseline and oxyfuel cement plants.

3.5.1 Energy & Environmental indicators

Conversion measured the calcination efficiency of limestone ($CaCO_3$) into lime (CaO) and CO_2, using inlet ($\dot{m}_{CaCO_{3,0}}$)& outlet ($\dot{m}_{CaCO_{3,n}}$) flowrates (Eqn. (12)). Specific Thermal Duty quantified the energy required for clinker production, calculated by multiplying the coal supplied ($\dot{m}_{coal}$) with higher heating value (HHV_{coal}), then dividing by mass of clinker produced ($\dot{m}_{clinker}$) (Eqn. (13)). Specific Electrical Duty was calculated by summing the total power needed (*P*) and dividing it by the mass of clinker produced ($\dot{m}_{clinker}$) (Eqn. (14)). Specific CO_2 generation evaluated the amount of CO_2 emitted by the process ($\dot{m}_{CO_2,emitted,process}$) and electricity generation ($\dot{m}_{CO_2,emitted,electricity}$) per tonne of cement, assuming 0.2 $kgCO_2/kWh$ produced (Eqn. (15)). The Capture Rate reflected the effectiveness of the CCS system in capturing CO_2, expressed as the fraction of captured CO_2 ($\dot{m}_{CO_2,captured}$) relative to total generated CO_2 ($\dot{m}_{CO_2,generated}$) ((Eqn. (17)).

$$\textbf{Conversion} = \left(\frac{\dot{m}_{CaCO_{3,0}} - \dot{m}_{CaCO_{3,n}}}{\dot{m}_{CaCO_{3,0}}}\right) * 100\% \qquad (12)$$

$$\textbf{Specific Thermal Duty} = \frac{\sum(\dot{m}_{coal} * HHV_{coal})}{\dot{m}_{clinker}} [GJ/t_{clinker}] \qquad (13)$$

$$\textbf{Specific Electrical Duty} = \frac{\sum P_i}{\dot{m}_{clinker}} [kWh/t_{cement}] \qquad (14)$$

$$\textbf{Specific } \boldsymbol{CO_2} \textbf{ generation, } \boldsymbol{e} = \frac{\left(\dot{m}_{CO_2,emitted,process} + \dot{m}_{CO_2,emitted,electricity}\right)}{\dot{m}_{cement}} [t_{CO_2}/t_{cement}] \qquad (15)$$

$$\textbf{Specific } \boldsymbol{CO_2} \textbf{ avoided} = e_{base} - e_{oxyfuel} \qquad (16)$$

$$\textbf{Capture Rate} = \left(\frac{\dot{m}_{CO_2,captured}}{\dot{m}_{CO_2,generated}}\right) * 100\% \qquad (17)$$

3.5.2 Economic indicators

Capital (CAPEX) and operating (OPEX) costs were estimated using a factorial method developed by Coulson & Richadson (Sinnot, 2005). Purchase Cost of Equipment were estimated using either six-tenths rule or interpolation from correlations in 1987 by Garrett, adjusted with $CEPCI_{1987}$ = 320 & $CEPCI_{2024}$ = 798.8. The six-tenths rule is particularly useful for cases where the cost and capacity of a similar reference unit are known, allowing for the estimation of the cost of the target unit of interest. The equation used is as follows:

$$C_2 = C_1 \times \left(\frac{S_2}{S_1}\right)^{0.6} \qquad (18)$$

C_2 represents the estimated cost of the unit of interest, C_1 is the cost of the reference unit, S_2 and S_1 are the capacities of the target and reference units respectively. Data from the IEA report *Deployment of CCS in the Cement Industry* (2013) and Gardarsdottir et al. (2019) were used to obtain reference values.

Annualised CAPEX (Eqn. (19)), Production Cost (Eqn. (20)) and Cost of CO_2 avoided (Eqn. (21)) were then

calculated. Discount rate (i) and project lifetime (t) used were 10% and 25 years.

$$\textbf{Annualised CAPEX} = Total\ CAPEX * \left(\frac{i(1+i^t)}{(1+i^t)-1}\right)[€M] \quad (19)$$

$$\textbf{Production Cost, PC} = \frac{Annualised\ CAPEX+OPEX}{\dot{m}_{cement,annual}}\ [€/t_{cement}] \quad (20)$$

$$\textbf{Cost of } \boldsymbol{CO_2} \textbf{ avoided, CAC} = \frac{PC_{oxyfuel}-PC_{base}}{e_{base}-e_{oxyfuel}}[€/t_{CO_2}] \quad (21)$$

4 Results & Discussion

4.1 Effect of fuel on oxyfuel combustion

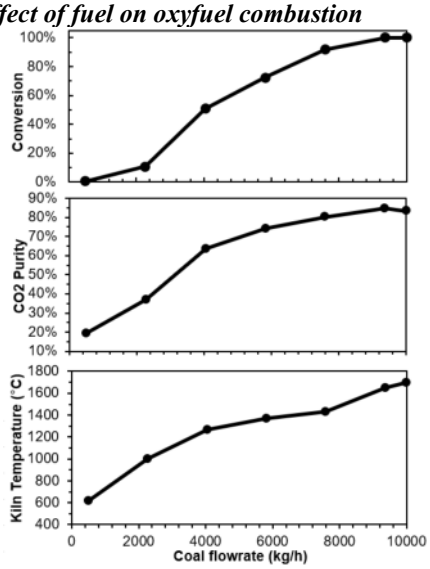


Figure 2 Effect of coal flowrate for calciner on conversion, CO_2 purity and kiln temperature.

Conversion increases sharply at low coal flow rates, reaching 100% at 10,000 kg/h. Increased coal flow supplies more thermal energy, enhancing endothermic reactions like limestone decomposition and clinker formation. CO_2 purity increases steadily with coal flow rate, peaking at approximately 0.8 around 10,000 kg/h, and then declines slightly as coal flow continues to increase. This indicates the presence of incomplete combustion. However, the changing coal flowrate for rotary kiln did not seem to have as significant effect as the calciner.

4.2 Effect of oxygen purity

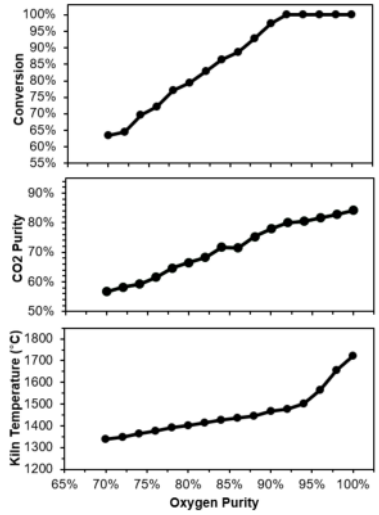


Figure 3 Effect of oxygen purity on conversion, CO_2 purity and kiln temperature.

Conversion increases steadily with oxygen purity, reaching 100% as oxygen purity approaches 92.5%. Higher oxygen purity enhances the combustion process by providing more oxygen for complete combustion, minimising the presence of unburned carbon and carbon monoxide. This leads to a steady improvement in the conversion efficiency.

CO_2 purity increases linearly with oxygen purity, rising from approximately 60% at lower purities to about 85% at 100% of oxygen purity. Increased oxygen purity reduces the presence of inert gases like nitrogen gas in the system, resulting in higher CO_2 concentration in the flue gas. This directly improves CO_2 purity, highlighting the advantage of using oxyfuel combustion.

Kiln temperature also increases with oxygen purity, rising sharply beyond 0.95 oxygen purity and reaching approximately 1750°C at full oxygen purity. The sharper increase near 100% oxygen purity indicates that higher combustion efficiencies are achieved under pure oxygen conditions, resulting in maximum thermal energy utilisation. However, the significant increase in temperature can have structural damage to refractory linings of the kiln thus flue gas was recirculated back to the rotary kiln through the combustors.

4.3 Effect of flue gas recirculation ratio

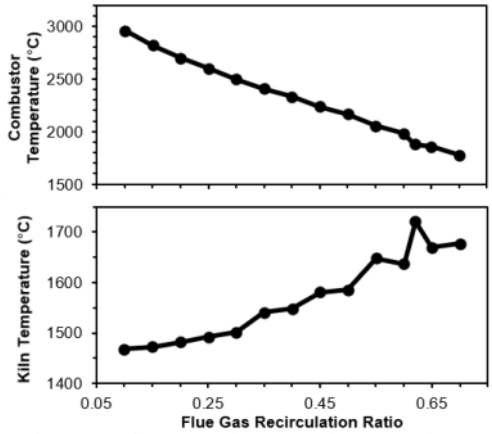


Figure 4 Effect of flue gas recirculation ratio on combustor temperature and kiln temperature.

Combustor temperature decreases steadily as the flue gas recirculation ratio increases, dropping from approximately 3000°C at 0.1 ratio to about 1800°C at 0.7 ratio. Flue gas recirculation introduces cooler exhaust gases back into the combustion chamber, reducing the peak combustion temperature. This reduction in temperature can help prevent overheating and material degradation in the combustor but may also reduce the overall thermal efficiency of the process.

In contrast to the combustor temperature trend, kiln temperature increases as the flue gas recirculation ratio increases, peaking around 0.62. The recirculated flue gas minimises heat loss through exhaust and helps balance energy distribution in the kiln.

Table 3 Simulated compositions of output streams

Material	Mass Composition (wt%)									
	SiO2	**Al2O3**	**Fe2O3**	**CaO**	**MgO**	**CaSO4**	**C3S**	**C2S**	**C3A**	**Ash**
Clinker	-	1.82%	3.11%	4.49%	1.12%	-	71.94%	7.75%	8.80%	0.98%
Cement	8.10%	4.20%	2.37%	9.87%	2.28%	4.99%	54.82%	5.90%	6.70%	0.75%
Material	**Mole Composition (wt%)**									
	CO2	**N2**	**O2**	**H2O**	**CO**	**SO2**	**H2**			
Flue gas	83.1	0.5	0.5	13.8	1.4	0.1	0.5			
Liquefied gas	97.0	1.0	1.0	46ppm	190ppm	289ppm	1. 0			

4.4 Technical & Economic analyses

Overall, the performance parameters for both the baseline and oxyfuel scenarios align well with literature values reported for cement plants. The baseline simulated thermal duty of approximately 3.5 GJ/tclinker and electrical demand of about 82 kWh/tcement are within the ranges of 2.9–6.7 GJ/tclinker and 90–110 kWh/tclinker documented in prior studies (IEA, 2018; Jankovic et al., 2004; (Ecofys et al., 2009). These values suggest that the modelling assumptions and process conditions reflect a representative base configuration. Similarly, the baseline net CO_2 emissions of around 0.6 tCO_2/tcement align with the established range of 0.5–1.0 tCO_2/tcement reported in several investigations of conventional cement kilns (WBCSD/CSI ,2009; García-Gusano et al., 2015; Huntzinger and Eatmon, 2009)[4–6]. The CAPEX of €170M for baseline cement plant is within the range of €150 to €200 million (IEA, 2018).

Under oxyfuel combustion conditions, the specific thermal duty declines notably—from 3.5 GJ/tclinker to 2.6 GJ/tclinker—highlighting the potential energy efficiency gains arising from oxyfuel operation. While the specific electrical duty nearly doubles from 82 to 162 kWh/tcement, largely due to the energy requirements of the Air Separation Unit (ASU) and Compression and Purification Unit (CPU), this increase is consistent with projections for integrated oxyfuel capture systems in the cement industry. These technologies inherently demand greater electrical input, but the trade-off is a more favourable environment for carbon capture.

The capture rate reaches approximately 86%, reducing net CO_2 emissions to about 0.1 tCO_2/t_cement and translating into an 82% avoidance rate.

Economically, while the CAPEX and OPEX rise under oxyfuel conditions, resulting in a cost of CO_2 avoided (CAC) of about €76.4/tCO_2, this figure remains competitive within the current landscape of carbon capture solutions which is discussed next.

Table 4 Unit prices and annual material and utilities cost

Input	Unit Price Cost	Unit	Annual Cost (€M/y)	
Materials			**Baseline**	**Oxyfuel**
Raw Meal	4.8	€/t	7.5	7.5
GBFS	2.4	€/t	0.8	0.8
Gypsum	7.2	€/t	0.5	0.5
Coal	78	€/t	10.2	7.5
Electricity	0.1	€/kWh	9.2	180.4
Total	0		28.1	196.7

Table 5 Purchase cost of equipment for baseline and oxyfuel configuration

Equipment	Purchase Cost of Equipment (€M)		Costing Method
	Baseline	**Oxyfuel**	
Cement Production			
Pneumatic Conveyor	4.5	4.5	Six tenths
Preheaters	0.22	0.22	Six tenths
Raw Meal Mill	4.30	4.30	Six tenths
Clinker Mill	1.52	1.52	Six tenths
Calciner	5.74	5.74	Six tenths
Rotary Kiln	17.08	17.08	Six tenths
Clinker Cooler, fans	0.80	0.80	Six tenths
Electrostatic Precipitator	0.80	0.80	1st principle
CPU			
Compressor	-	5.00	1st principle
Cooler 1	-	0.70	Six tenths
Cooler 2	-	0.77	Six tenths
Flash 1	-	0.18	1st principle
Flash 2	-	0.18	1st principle
DESOX Absorber	-	0.05	1st principle
DEHY Absorber	-	0.02	1st principle
Air Separation Unit	-	12.84	Six tenths
Cryogenic Unit	-	20.11	Six tenths
Total Purchase Cost	34.95	74.81	

Table 6 Electrical duty of equipment for baseline and oxyfuel configuration

Component/Process	Baseline (kWh/tcem)	Oxyfuel (kWh/tcem)
Raw Meal Mill	46.4	46.4
Clinker Mill	25.2	25.2
Conveyor	3.3	3.4
Raw mill fan	7.2	7.2
ASU	0	26.3
CPU	0	53.2
Total electricity	**82.1**	**161.7**

Table 7 Calculation of CAPEX and OPEX

Cost type	Cost (€M)	
	Baseline	Oxyfuel
Purchase Cost of Equipment	35.0	74.8
Physical Plant Cost	110.1	235.6
Fixed Capital Cost	154.2	329.9
Working Capital Cost	15.4	33.0
CAPEX	**169.6**	**362.9**
Materials	19.2	16.5
Utilities	9.2	18.0
Maintenance	11.0	23.6
Miscellaneous	1.1	2.4
Labour	10.8	13.0
Other	7.7	16.5
Fixed Operating Cost	29.5	53.0
Variable Operating Cost	29.4	36.8
Direct Production Cost	59.0	89.9
Indirect Production Cost	5.9	9.0
OPEX	**64.9**	**98.8**

4.4.1 Comparison with other technologies

The effect of CO_2 capture rate on cost of CO_2 avoided (CAC) is highlighted in Figure 6. There is a clear trend in the data: higher CO_2 capture rates are generally associated with lower costs of avoided carbon, stressing the importance of efficiency in advanced CCS technologies. For example, the oxyfuel and calcium looping configuration achieves the lowest CAC at €30.8/tCO_2 , with the highest capture rate amongst all technologies at 94%.

As well as that, calcium looping (integrated) reaches its lowest cost of €61.98/tCO_2 at its highest capture rate of 93%. Oxyfuel combustion (Oxy) also follows this trend with its lowest cost of €55.8/tCO_2 at a capture rate of 90%. The oxyfuel system simulated in this report achieves a cost of €76.41/tCO_2 for a capture rate of 86%, showing a good balance in the cost-performance profile. In contrast, amine-based post-combustion techniques show much higher costs: €88.3–119/tCO_2 while offering less efficient capture rates of 60–77%.These results underline the benefits in terms of optimal economic and environmental.

A comparison of CAPEX and OPEX was also carried out against different technologies, as seen in Figure 6.

With a CAPEX of €363 million, oxyfuel is slightly more expensive than MEA Scrubbing (€347 million) but significantly cheaper than the Chilled Ammonia Process (€468 million) and Calcium Looping (€537–589 million). Its OPEX of €99 million is notably lower than MEA (€105 million) and Chilled Ammonia (€196 million), owing to the direct production of concentrated CO_2 during combustion, which eliminates the need for energy-intensive solvent regeneration or cooling systems.

Table 8 Comparison of the performances of the baseline and oxyfuel combustion cement

Parameter	Unit	Base Line	Oxyfuel
Technical Parameters			
Clinker Production	Mtclinker/y	1.0	1.0
Cement Production	Mtcement/y	1.4	1.4
Specific Thermal Duty	GJ/ tclinker	3.5	2.6
Specific Electrical Duty	kWh/tcement	82	162
Environmental Parameters			
Indirect CO2 generation (@ 0.2 tCO2/MWh)	tCO2/tcement	0.0	0.0
Direct CO2 generation	tCO2/tcement	0.6	0.5
Specific CO2 generation	tCO2/tcement	0.6	0.6
Capture rate	%	-	86
Net CO2 emission	tCO2/tcement	0.6	0.1
CO2 captured	tCO2/tcement	-	0.5
Specific CO2 avoided	tCO2/tcement	-	0.5
Avoidance rate	%	-	82.0
Annual captured CO2	MtCO2/year	-	0.6
Economic Parameters			
Capital Cost, CAPEX	€M	170	363
Annualised CAPEX	€M	17.6	37.7
Operating Cost, OPEX	€M	64.9	98.8
Production Cost	€/tcement	58.8	97.4
Cost of CO2 avoided	€/tCO2	-	76.4

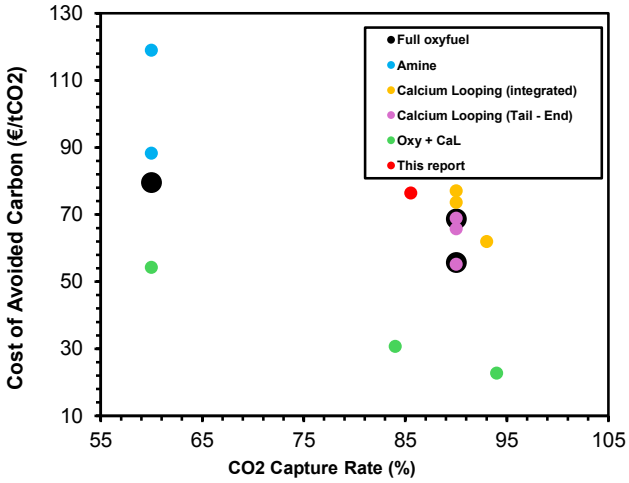


Figure 5 CAC and CO_2 capture rate of different CCS technologies [(Leeson et al., 2017), Gardarsdottir et al., 2019)]

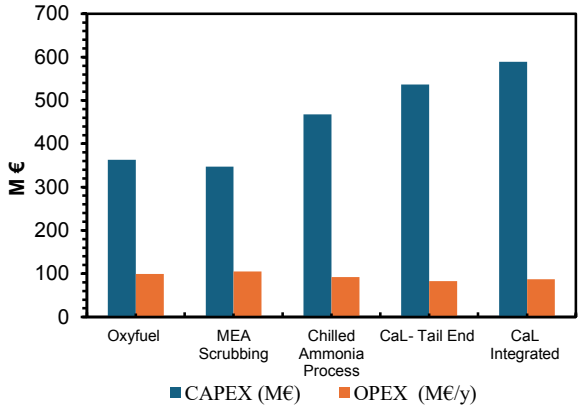


Figure 6 Comparison of CAPEX and OPEX of different CCS technologies

While Calcium Looping offers the lowest OPEX (€83–87 million), its high CAPEX makes it less feasible for immediate adoption. Oxyfuel's compatibility with high-temperature cement kilns (~1450°C) further reduces integration costs and operational complexity, making it a practical option for retrofitting or new installations. Overall, with a competitive CO_2 capture rate (86%) and low CAC, CAPEX and OPEX values, oxyfuel combustion provides a cost-effective approach compared to other technologies to implement into cement plants to achieve significant emissions reductions.

Table 6 Comparison of specific thermal and electrical duty for different CCS technologies

Technology	Specific Thermal Energy Required (GJ/ tclinker)	Specific Electrical Energy Required (kWh/ tcement)	Capture Rate (%)
Simulated Oxyfuel	2.6	162	86
MEA Scrubbing	3.6	230	90
SPEI Adsorption	3.6	220	90

A comparison of specific thermal and electrical energy requirements of MEA scrubbing and silica-alkoxylated polyethyleneimine (SPEI) adsorbent-based processes to the simulated oxyfuel combustion is shown in the Table 7. The simulated oxyfuel combustion system exhibits a specific thermal energy requirement of 2.6 GJ/tclinker, lower than both MEA and SPEI systems (3.6 GJ/tclinker). This corresponds to an essential gain since it aligns with the aim of the industry to lessen the clinker production energy intensity. Furthermore, the simulated oxyfuel system requires 162 kWh/tcement of specific electrical energy, which is significantly lower than the 230 kWh/tcement for MEA scrubbing and the 220 kWh/tcement for SPEI adsoprtion.

According to the IEA's Net Zero Scenario (NZE), the thermal energy intensity of clinker production is expected to decrease to less than 3.4 GJ/ton clinker by 2030, while electrical intensity is projected to fall below 90 kWh/tcement. The simulated oxyfuel combustion system already lies below this target in terms of thermal energy use, indicating good accordance with future cement industry energy efficiency targets. However, its electrical energy use still lies significantly above the NZE target, so potential improvement exists in this regard.

4.5 Sensitivity analyses

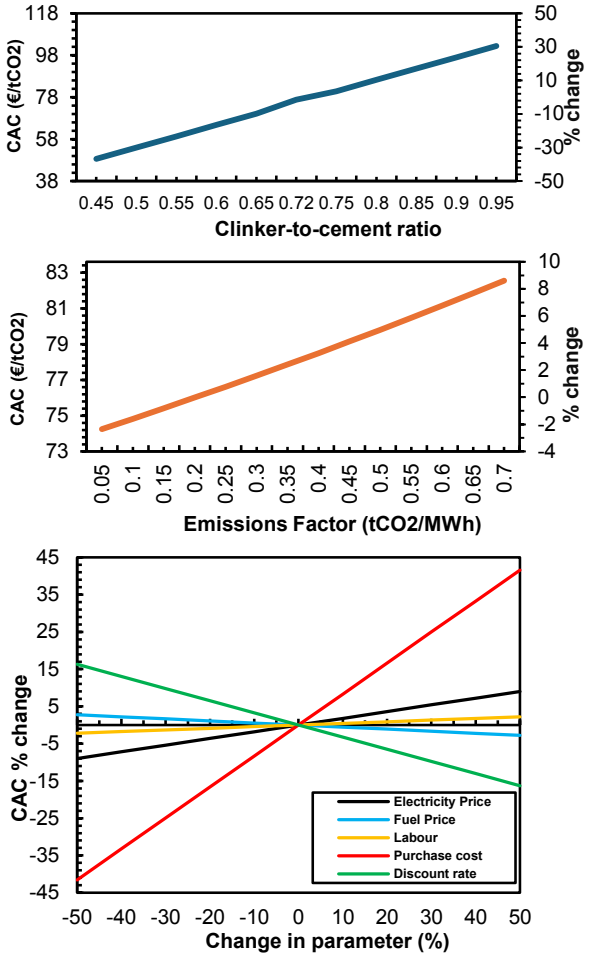


Figure 7 Sensitivity analyses

Sensitivity analyses were conducted to demonstrate how variations in both material and operational parameters influence the cost of CO_2 avoided (CAC) (shown in Figure 8). As the clinker-to-cement ratio decreases from 0.72 to 0.5, the CAC decreases significantly by 30%. This indicates that incorporating supplementary cementitious materials such as fly ash, silica fumes, GBFS and gypsum reduces both production costs and also emissions, which was also shown in a research done by Yang et al. As the emissions factor of the electricity supply increases from 0.2 to 0.4, the CAC also rises by around 4%, this shows that a dirtier electricity mix enhances indirect emissions and increases overall capture costs. The affordability of CO_2 capture system improves when powered by low-carbon electricity.

The purchase cost of equipment has the most significant impact on CAC, with a 10% increase in equipment costs resulting in an 8% rise in CAC. Conversely, labour cost has the least impact, as a 50% increase leads to only a 2% rise in CAC. Interestingly, an increase in fuel cost slightly decreases the CAC, though the effect is minimal. The discount rate also has a notable influence, with a 1% increase in value reducing the CAC by approximately 3%. Additionally, a 50% increase in electricity price results in a 9% increase in CAC, reflecting its moderate but expected impact.

5 Conclusion & Outlook

This research assessed the implementation of oxyfuel combustion technology within the cement sector to mitigate carbon dioxide emissions. Results showed the application of oxyfuel combustion facilitated the capture of up to 86% of CO_2 emissions, thereby reducing net emissions to 0.11 tCO_2 per tonne of cement—an 82% reduction relative to the baseline scenario. In addition, the thermal energy requirement declined to 2.6 GJ per tonne of clinker (representing a 26% improvement compared to the reference case), although electrical energy demand rose by 97% to 162 kWh per tonne of cement. This elevated electricity consumption was primarily associated with the energy-intensive Air Separation Unit (ASU) and the Compression and Purification Unit (CPU).

The estimated cost of CO_2 avoided (CAC) was €76.4 per tonne of CO_2, demonstrating oxyfuel combustion as a competitive CCS option. Further cost reductions may be achieved through economies of scale, the incorporation of renewable energy sources and process optimisations. The model can be improved by making more accurate technical assumptions such as simulating efficiency losses in all parts of the processes, simulating air leaks and using different property methods for different processes. Future research should focus on adopting cheaper alternative fuels, optimising heat recovery system and investigating the effect of clinker-to-cement ratio on cement quality.

Acknowledgements

The authors would like to extend their gratitude to Tumisang Tsolope (Imperial College London) for his continuous support and guidance throughout this research project.

References

Atmaca, A. and Kanoglu, M. (2012) Reducing energy consumption of a raw mill in cement industry, Energy. Available at: https://www.sciencedirect.com/science/article/pii/S0360544212002630 (Accessed: 12 December 2024).

CEMCAP (2016) *Design and performance of CEMCAP cement plant with MEA post combustion capture*. Available at: https://www.sintef.no/globalassets/sintef-energi/cemcap/d4.2-design-and-performance-of-cemcap-cement-plant-with-mea-post-combustion-capture_rev1~1.pdf (Accessed: 12 December 2024).

CEMCAP (2018) *CEMCAP comparative techno-economic analysis of CO2 capture in cement plants*. Available at: https://www.sintef.no/globalassets/sintef-energi/cemcap/d4.2-design-and-performance-of-cemcap-cement-plant-with-mea-post-combustion-capture_rev1~1.pdf (Accessed: 12 December 2024).

Czakiert, T. et al. (2022) Chemical Looping Combustion: A brief overview, MDPI. Available at: https://doi.org/10.3390/en15041563 (Accessed: 29 November 2024.

Driver , J. et al. (2022) Simulation of direct separation technology for carbon capture and storage in the cement industry, Chemical Engineering Journal. Available at: https://www.sciencedirect.com/science/article/pii/S1385894722032089?ref=pdf_download&fr=RR-2&rr=8ea469721a5ebd89 (Accessed: 29 November 2024).

Ecofys , Fraunhofer Institute for Systems and Innovation Research and Öko-Institut (2009) *Methodology for the free allocation of emission allowances in the EU ETS post 2012*. Available at: https://climate.ec.europa.eu/system/files/2016-11/bm_study-cement_en.pdf (Accessed: 12 December 2024).

Environment, U. *et al.* (2018) *Eco-efficient cements: Potential economically viable solutions for a low-CO2 cement-based Materials Industry*, *Cement and Concrete Research*. Available at: https://www.sciencedirect.com/science/article/abs/pii/S0008884618301480?via%3Dihub (Accessed: 11 December 2024).

Fortune Business Insights, 2024. Cement Market Size, Share & Industry Analysis by Type (Ordinary Portland Cement, Portland Pozzolana Cement, White Cement, Others), By End-Use (Residential, Non-residential, Infrastructure), and Regional Forecast, 2023-2030. Available at: https://www.fortunebusinessinsights.com/industry-reports/cement-market-101825 [Accessed 20 November 2024].

G. Speight, J. (2017) *Glycol dehydration process*, *Glycol Dehydration Process - an overview | ScienceDirect Topics*. Available at: https://www.sciencedirect.com/topics/engineering/glycol-dehydration-process#:~:text=Glycols%20are%20commonly%20used%20to,typical%20acid%20gas%20removal%20process. (Accessed: 11 December 2024).

García-Gusano, D., Cabal, H., Lechón, Y., & Izquierdo, L. (2015). Life Cycle Assessment of applying CO_2 post-combustion capture to the Spanish cement production. *Journal of Cleaner Production,* 104, 328–338. [Accessed 20 November 2024].

Gardarsdottir, S.O., De Lena, E., Romano, M., Roussanaly, S., Voldsund, M., Pérez-Calvo, J.F., Berstad, D., Fu, C., Anantharaman, R., Sutter, D., Gazzani, M., Mazzotti, M. & Cinti, G., 2019. Comparison of technologies for CO_2 capture from cement production—Part 2: Cost analysis. MDPI Energies, 12(3), p.542. Available at: https://www.mdpi.com/1996-1073/12/3/542 [Accessed 20 November 2024].

Garrett, D.E. (1989) *Chemical Engineering Economics*, *SpringerLink*. Available at: https://link.springer.com/book/10.1007/978-94-011-6544-0 (Accessed: 11 December 2024).

Granados, D.A., Chejne, F. and Mejia, J.M. (2015) *Oxy-fuel combustion as an alternative for increasing lime production in Rotary Kilns*, *Applied Energy*. Available at: https://www.sciencedirect.com/science/article/pii/S0306261915009162?via%3Dihub (Accessed: 12 December 2024).

Hasan, S.T. *Syeda Tahsin Hasan*, *Civil Engineering*. Available at: https://civiltoday.com/civil-engineering-materials/cement/106-cement-manufacturing-process (Accessed: 11 December 2024).

Huntzinger, D. N., & Eatmon, T. D. (2009). A life-cycle assessment of Portland cement manufacturing: comparing the traditional process with alternative technologies. *Journal of Cleaner Production,* 17(7), 668–675. (Accessed: 11 December 2024).
IEA *Cement*, *IEA*. Available at: https://www.iea.org/energy-system/industry/cement?utm_source (Accessed: 11 December 2024).
IEA (2018) *Technology roadmap - low-carbon transition in the cement ...* Available at: https://iea.blob.core.windows.net/assets/cbaa3da1-fd61-4c2a-8719-31538f59b54f/TechnologyRoadmapLowCarbonTransitionintheCementIndustry.pdf (Accessed: 12 December 2024).
IEAGHG, 2013. Deployment of CCS in the Cement Industry. International Energy Agency Greenhouse Gas R&D Programme. Available at: https://ieaghg.org/publications/deployment-of-ccs-in-the-cement-industry [Accessed 20 November 2024].
Jankovic, A., Valery, W. and Davis, E. (2004) *Cement grinding optimisation*, *Minerals Engineering*. Available at: https://www.sciencedirect.com/science/article/pii/S0892687504002043 (Accessed: 12 December 2024).
John, J.P. (2020) *Parametric studies of cement production processes - john - 2020 - journal of energy - wiley online library*. Available at: https://onlinelibrary.wiley.com/doi/10.1155/2020/4289043 (Accessed: 12 December 2024).
Leeson, D. *et al.* (2017) *A techno-economic analysis and systematic review of Carbon Capture and storage (CCS) applied to the iron and steel, cement, oil refining and pulp and paper industries, as well as other high purity sources*, *International Journal of Greenhouse Gas Control*. Available at: https://www.sciencedirect.com/science/article/pii/S1750583617302890X (Accessed: 29 November 2024).
Meunier, N. *et al.* (2014) *CO2 capture in cement production and re-use: First step for the optimization of the overall process*, *Energy Procedia*. Available at: https://www.sciencedirect.com/science/article/pii/S1876610214025004 (Accessed: 11 December 2024).
Oi, L.E. *et al.* (2016) *Simulation and cost comparison of CO2 liquefaction*, *Energy Procedia*. Available at: https://www.sciencedirect.com/science/article/pii/S1876610216000539?via%3Dihub (Accessed: 11 December 2024).
Quevedo, S. *et al.* (2022a) *Techno-Economic Assessment of hybrid CO2 capture processes in cement manufacturing based on partial Oxyfuel & Post-Combustion Technologies*, *SSRN*. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4286127 (Accessed: 10 December 2024).
Roussanaly, S. et al. (2021) Towards improved cost evaluation of carbon capture and storage from industry, International Journal of Greenhouse Gas Control. Available at: https://www.sciencedirect.com/science/article/pii/S1750583621000153 (Accessed: 29 November 2024).
Shahabuddin, M. and Bhattacharya, S. (2021) *Process modelling for the production of hydrogen-rich gas from gasification of coal using oxygen, CO2 and Steam reactants*, *International Journal of Hydrogen Energy*. Available at: https://www.sciencedirect.com/science/article/pii/S0360319921016852 (Accessed: 12 December 2024).
Sinnot, R. (2005) *Chemical Engineering Design*, *Elsevier Shop*. Available at: https://shop.elsevier.com/books/chemical-engineering-design/sinnott/978-0-08-049255-1 (Accessed: 12 December 2024).
WBCSD/CSI (2009). *Cement Technology Roadmap 2009: Carbon Emissions Reductions up to 2050.* World Business Council for Sustainable Development and International Energy Agency. (Accessed: 12 December 2024).
World Economic Forum, 2024. Cement production: A major contributor to global CO_2 emissions and sustainable concrete solutions. Available at: https://www.weforum.org/stories/2024/09/cement-production-sustainable-concrete-co2-emissions/ [Accessed 20 November 2024].
Yang, K.-H. *et al.* (2014) *Effect of supplementary cementitious materials on reduction of CO2 emissions from concrete*, *Journal of Cleaner Production*. Available at: https://www.sciencedirect.com/science/article/pii/S095965261400239X?via%3Dihub (Accessed: 12 December 2024).

Reinforcement Learning for Constrained Continuous Stirred-Tank Reactors Systems

Shunhua Jiao Tachun Chen

December 12, 2024

Abstract

Reinforcement learning (RL) faces challenges in real-world process systems due to the critical need to handle state constraints, which conventional RL frameworks do not explicitly address. This research aimed to address this limitation by integrating state-dependent control constraints into RL frameworks. The proposed method was validated on a Continuous Stirred-Tank Reactor system using the PC-Gym environment, evaluating the performance of two RL algorithms, Deep Q-Network (DQN) and Proximal Policy Optimisation (PPO), in both constrained and unconstrained settings. A significant future direction involves extending the framework to stochastic environments to incorporate uncertainties in state transitions and constraints. This study highlights the potential of constrained RL as a practical solution for optimising dynamic and real-world chemical processes.

Keywords: Constrained Reinforcement Learning, Process Control Optimisation, Deep Q-Network, Proximal Policy Optimisation

1 Introduction

Sequential decision-making plays a critical role in process systems engineering (PSE), with applications ranging from scheduling and supply chain management to process control and optimisation. These tasks require a series of optimal decisions to ensure system efficiency. However, the inherent uncertainties in real-world PSE systems pose significant challenges to traditional methods like stochastic programming, which relies on predefined probability distributions and approximations. While effective in some cases, stochastic programming often becomes computationally intractable for complex systems, motivating the exploration of reinforcement learning (RL) as a viable alternative (J. Shin et al. (2019)[14]).

RL has emerged as a promising solution for addressing uncertainty in PSE by leveraging data to learn optimal control policies (J. W. Kim et al. (2020) [8]). Its effectiveness has been demonstrated in several applications, including set-point tracking and control optimisation problems (Lee and Lee (2005) [9]; Spielberg et al. (2019) [10]). Typically, RL formulates these problems within the framework of Markov Decision Processes (MDPs), which provide a structured approach to sequential decision-making (Puterman, M.L. (2014) [11]). However, a significant limitation of this framework is its inability to explicitly handle state constraints, which are common and normally necessary in real-world settings.

Common strategies to address this include penalty functions and Lagrangian multipliers, but these methods often lead to convergence instability and require extensive tuning to align with RL architectures (H.Yu et al (2021)[12]). Alternative methods, like domain restrictions and control masking, offer partial solutions but are typically limited to discrete control settings (A. Kanervisto et al. (2020) [15]).

This work aims to extend the capability of RL by addressing state constraints within continuous control environments. Inspired by the observation that MDPs can manage control constraints in a state-dependent manner, we built RL framework upon the decomposition feasibility method by M. Mowbray et al. (2024) [6], abstracting state constraints into state-dependent control constraints. The proposed method is initially developed in a deterministic setting for clarity and validation, with the ultimate goal of extending it to probabilistic environments to incorporate uncertainty effectively. This approach represents a significant step toward robust RL-based solutions for real-world PSE challenges.

This report is organized as follows: Section 2 reviews reinforcement learning and constrained RL, highlighting key frameworks and challenges. Section 3 outlines the methodology, including state constraint transformation and RL implementation. Section 4 presents a case study on a Continuous Stirred-Tank Reactor (CSTR), with results discussed in Section 5. Section 6 concludes with key findings and future directions.

2 Background

2.1 Reinforcement Learning

RL involves an agent interacting with an environment E to determine optimal actions $\mathbf{u}_t \in \mathcal{U}$ based on ob-

served system states $\mathbf{x}_t \in \mathcal{X}$. At each discrete time step $t \in \mathbb{Z}$, the agent selects actions based on its policy $\pi(\mathbf{u}_t|\mathbf{x}_t)$, aiming to maximize an expected cumulative return $\mathbb{E}_\pi[G_t]$ from a discount factor γ and numerical reward $R_{t+1} \in \mathbb{R}$ received from environment E after taking control $\mathbf{u}_t$:

$$\mathbb{E}_\pi[G_t] = \mathbb{E}_\pi\left[\sum_{t=1}^{T} \gamma^{t-1} R_{t+1}(\mathbf{x}_t, \mathbf{u}_t)\right] \tag{1}$$

The interaction between the RL agent and the environment is assumed to follow a Markov Decision Process (MDP), defined as a five-tuple framework $< \mathcal{X}, \mathcal{U}, \mathcal{P}, \mathcal{R}, \gamma >$, where the transition probability of the environment is given by:

$$p(\mathbf{x}_{t+1}|\mathbf{x}_t, \mathbf{u}_t) \approx \mathbf{f}(\mathbf{x}_t, \mathbf{u}_t, \mathbf{w}_t) \tag{2}$$

with $\mathbf{f}(\cdot)$ approximating a nonlinear state space model, $\mathbf{f}(\cdot)$, and $W \sim p(\mathbf{w})$ representing general uncertain parameters. In this study, a deterministic setting was assumed:

$$p(\mathbf{x}_{t+1}|\mathbf{x}_t, \mathbf{u}_t) \approx \mathbf{f}(\mathbf{x}_t, \mathbf{u}_t) \tag{3}$$

2.2 Constrained Reinforcement Learning

Constrained Reinforcement Learning (CRL) has gained significant attention from the research community, as real-world dynamics often demand fast and safe control mechanisms. Previous optimisation routines in CRL required either the challenging Lagrange method for constrained Markov Decision Processes (CMDPs) or mathematical simulations for probabilistic state constraints within policy optimisation. In the context of our problem setting, we purposed a solution method for the nominal optimal control problem:

$$\pi(\cdot) := \begin{cases} \max_{\pi(\cdot)} G_t \\ \text{s.t.} \\ \mathbf{x}_{t+1} = \mathbf{f}(\mathbf{x}_t, \mathbf{u}_t) \\ \mathbf{u}_t = \pi(\mathbf{x}_t) \\ \mathbf{u}_t \in \mathbb{U} \\ \mathbf{x}_t \in \overline{\mathbb{X}_t} \\ \overline{\mathbb{X}_t} := \{\mathbf{x}_t \in \mathbb{X} \mid \mathbf{g}_t(\mathbf{x}_t) \leq 0\} \\ \forall t \in \{0, ..., T-1\} \end{cases} \tag{4}$$

Here, $\mathbf{g}_t(x_t)$ represents the state constraints imposed at time t. This, however, highlights the challenges in CRL. Despite numerous efforts in previous works to develop more reliable state constraint approximations, these overall process often exhibit inherently non-trackable method, particularly in state constraint identification or in optimising the corresponding Lagrange function (E. Pan et al.[1], P. Petsagkourakis et al.[5]).

2.3 Constraints Identification

M. Mowbray et al. (2024) [6] presented a sampling-based decomposition approach to the feasible space via a directed acyclic graph (DAG). Instead of directly solving a large problem defined on control trajectories, $\mathbb{U}_\tau \subseteq \prod_{t=0}^{T-1} \mathcal{U}$ space, a time-wise decomposition is applied, and local subproblems are solved individually to exploit problem structure. Each pair of state and control at time t is denoted by a node in the graph and the subproblems are denoted $\mathbb{XU}_t$. Nodes are connected in a way that follows the MDP. Throughout the method, the DAG is updated, through a policy of backward propagation according to a precedence ordering on the time grid, i.e. $t \prec t+1$, with nodes subproblems solved incrementally, enforcing additional constraints on the original $\mathcal{X} \times \mathcal{U}$ space. Figure 1 visualizes the DAG, with each circle representing a node in the graph.

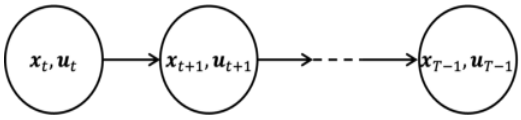


Figure 1: Example of DAG

This decomposition technique enables us to transform state constraints, which cannot be handled directly, into surrogate constraints that can be applied explicitly on control input variables $< \mathbf{x}_t, \mathbf{u}_t, t >$. In other words, it abstracts the state constraint feasibility into state dependent control constraints enforced at time t:

$$\begin{aligned} \mathbb{XU}_t &= \left\{ (\mathbf{x}_t, \mathbf{u}_t) \in \mathcal{X} \times \mathcal{U} \;\middle|\; \begin{array}{c} \mathbf{x}_t \in \overline{\mathbb{X}}_t \\ \exists (\mathbf{x}_{t+1}, \mathbf{u}_{t+1}) \in \mathbb{XU}_{t+1} : \\ \mathbf{x}_{t+1} = \mathbf{f}(\mathbf{x}_t, \mathbf{u}_t) \end{array} \right\} \\ &\approx \{(\mathbf{x}_t, \mathbf{u}_t) \in \mathcal{X} \times \mathcal{U} | 0 \geq \overline{\mathbf{g}}_t(\mathbf{x}_t, \mathbf{u}_t)\}. \end{aligned} \tag{5}$$

In this notation, $\overline{\mathbf{g}}_t(\mathbf{x}_t, \mathbf{u}_t)$ represents a scalar function parameterising the set of safe states and controls at time t. In this work, this is identified as a neural network binary classifier.

3 Methodology

By taking advantage of the constraint transformation, state constraints can be easily imposed on any RL framework. To explore the potential and feasibility of this novel CRL framework, both Deep Q-network [3] (action-value function) and Proximal Policy Optimisation [4] (actor-critic method) were experimented with, as they are both easy to implement and widely known among researchers, while providing completely different routines in terms of agent training.

Proximal Policy Optimisation (PPO) [4] is an actor-critic method that utilizes neural network architectures for both its policy (actor) and value function (critic). The agent's objective is to learn an optimal policy through an approximated policy gradient descent. This approximation is necessary due to the complexity of computing the exact gradient and the inherent bias in the advantage function. Therefore, a

generalized advantage function [13] is employed:

$$\hat{A}_t = \sum_{l=0}^{T-t-1} (\gamma\lambda)^l \delta_{t+l}$$
$$\text{where} \quad \delta_t = r_t + \gamma V(s_{t+1}) - V(s_t) \tag{6}$$

Where λ controls the balance between bias and variance of the approximation, $V(\cdot)$ is the value function (critic) of the PPO.

To facilitate exploitation, PPO samples actions from a normal distribution centered around the mean action.

$$\pi_\phi(u_t|x_t) \sim \mathcal{N}(\mu_\phi(x_t), \sigma) \tag{7}$$

Ultimately, PPO aims to maximize a clipped surrogate objective, which enhances stability by limiting the magnitude of policy updates while ensuring convergence to an optimal policy.

$$L^{CLIP}(\theta) = \mathbb{E}_t[\min(r_t(\theta)A_t^n, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon))]$$
$$\text{where} \quad r_t(\theta) = \frac{\pi_\theta}{\pi_{\theta,old}} \tag{8}$$

Deep Q-network (DQN) [3] is a value function method that aims to learn an optimal Q-value function. By selecting the action with the highest Q-value in each state, DQN implicitly learns the optimal policy. To stabilize training and improve convergence, DQN uses a target network. This network is a copy of the behavior network that is updated partially. By using a more stable target, the Temporal difference (TD) updates become less noisy, leading to more efficient learning.

TD updates are applied to the behavior network to minimize the TD error. By minimizing this loss, the behavior network learns to predict accurate Q-values, which in turn leads to better decision-making

$$y = r + \gamma \max_{u'} Q_\theta^{\text{target}}(x', u') \tag{9}$$

$$\mathbb{E}_{(x,u,r,x')\sim\text{replay buffer}}\left[\left(Q_\theta^{\text{behaviour}}(x,u) - y\right)^2\right] \tag{10}$$

To avoid over-exploitation, DQN employs an epsilon-greedy exploration strategy, where a random action is chosen with probability ϵ.

3.1 Constrained PPO

Vanilla PPO (J. Schulman et al. [4]) was adapted with minor changes to the logic of action selection. An additional constrained optimisation problem is solved at each action selection stage (the filter in Figure 2). The objective is to minimize the Euclidean distance between the safe control $\bar{\mathbf{u}}_t$ and the unconstrained control $\mathbf{u}_t$ generated by PPO's actor network, thereby determining the best safe control with the smallest distance using a multi-start approach. Algorithm 1 details the instruction.

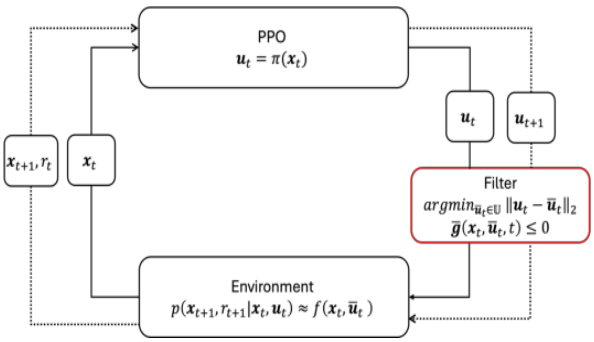


Figure 2: PPO process scheme

Algorithm 1 PPO's Action Selection with Constraint Handling

Input: state $\mathbf{x}_t$, horizon t
Output: best safe control $\bar{\mathbf{u}}_t^*$
Sampling a list of initial guesses for $\mathbf{u}_{init} \in \mathcal{U}$.
$d^* \leftarrow \infty$.
$\bar{\mathbf{u}}_t^* \leftarrow$ None.
$\mathbf{u}_t \leftarrow \text{actor}(\mathbf{x}_t)$
for $\bar{\mathbf{u}}_t$ in $\mathbf{u}_{init}$ **do**
  Minimize$_{\bar{\mathbf{u}}_t \in \mathcal{U}} \|\mathbf{u}_t - \bar{\mathbf{u}}_t\|_2$
  subject to: $\bar{\mathbf{g}}(\mathbf{x}_t, \bar{\mathbf{u}}_t, t) \leq 0$
  $d_t \leftarrow \|\mathbf{u}_t - \bar{\mathbf{u}}_t\|_2$
  if $d_t \leq d^*$ **then**
    $d^* \leftarrow d_t$
    $\bar{\mathbf{u}}_t^* \leftarrow \bar{\mathbf{u}}_t$
  end if
end for
return $\bar{\mathbf{u}}_t^*$

3.2 Constrained DQN

The general DQN framework (V. Mnih et al. [3]) requires the discretization of the continuous action space to avoid expensive computation and to identify a global optimum within the discrete action set through enumeration. However, discretized actions have inherent limitations in achieving fine control, particularly when the discrete set fails to encompass all meaningful decisions. This limitation is especially critical in the context of safe control, where precise adjustments are often necessary to ensure compliance with safety constraints. Therefore, DQN was modified to handle continuous control for this task. To select the best safe control, $\bar{\mathbf{u}}_t^*$, with the highest Q-value, a multi-start approach is employed. This involves starting from multiple initial control points and iteratively improving them using gradient-based optimisation.

Unlike PPO, which uses an actor network to directly determine the control action for a given state at time t, DQN relies on a Q-value function $Q : f(\mathbf{u}_t, \mathbf{x}_t)$ to make decisions. This allows us to incorporate the surrogate constraint function directly into the greedy action selection process of the DQN policy. As outlined in Algorithm 2, the constrained DQN selects only safe controls when maximizing the Q function. A similar approach can be applied to the exploration phase of

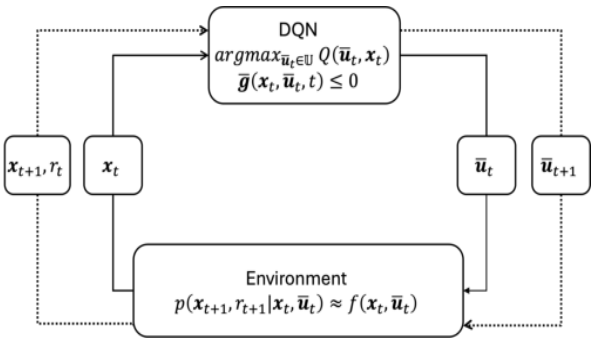


Figure 3: DQN process scheme

the epsilon-greedy method. Instead of sampling random controls from the entire designed control space, only those that satisfy the constraints are considered. Sampling can be done recursively until a feasible solution is found (see Algorithm 3).

Algorithm 2 DQN's Greedy Action Selection with Constraint Handling

Input: state $\mathbf{x}_t$, horizon t
Output: safe control $\bar{\mathbf{u}}_t$
Sampling a list of initial guesses for $\mathbf{u}_{init} \in \mathcal{U}$.
$Q^* \leftarrow -\infty$.
$\bar{\mathbf{u}}_t^* \leftarrow$ None.
for $\bar{\mathbf{u}}_t$ in $\mathbf{u}_{init}$ **do**
    $\underset{\bar{\mathbf{u}}_t \in \mathcal{U}}{\textbf{Maximize}} \quad Q(\mathbf{x}_t, \bar{\mathbf{u}}_t)$
    subject to: $\bar{\mathbf{g}}(\mathbf{x}_t, \bar{\mathbf{u}}_t, t) \leq 0$
    $Q_t \leftarrow Q(\mathbf{x}_t, \bar{\mathbf{u}}_t)$
    if $Q_t \geq Q^*$ **then**
        $Q^* \leftarrow Q_t$
        $\bar{\mathbf{u}}_t^* \leftarrow \bar{\mathbf{u}}_t$
    end if
end for
return $\bar{\mathbf{u}}_t^*$

The exploration part of the epsilon-greedy policy involves sampling from a uniform distribution within the designed control space. However, without any restrictions, a random control sampled from this distribution may lead to state-constraint violations in subsequent time steps. To address this, control samples are continuously drawn from the distribution until a sample satisfying the state constraints is obtained.

Algorithm 3 DQN's Random Sampling with Constraint Handling

Input: state $\mathbf{x}_t$, horizon t
Output: safe control $\bar{\mathbf{u}}_t$
$\mathbf{u}_t \sim$UniformSampling(-1,1) //For normalized control
while $\bar{\mathbf{g}}(\mathbf{x}_t, \mathbf{u}_t, t) \geq 0$ **do**
    $\mathbf{u}_t \sim$UniformSampling(-1,1)
end while
$\bar{\mathbf{u}}_t \leftarrow \mathbf{u}_t$
return $\bar{\mathbf{u}}_t$

4 Case study

4.1 PC-Gym

PC-Gym (Bloor et al., 2024 [2]) is an open-source environment that simulates complex chemical systems, such as Continuous Stirred-Tank Reactors (CSTRs), multistage extraction columns, and crystallization processes. These environments incorporate nonlinear dynamics, disturbance, and constraints, making them highly realistic for industrial process control research.

The CSTR system (APMonitor [7]) models a first-order irreversible reaction, with cooling provided by water flow. Two nonlinear differential equations govern its dynamics:

$$\frac{dC_A}{dt} = \frac{q}{V}(C_{Af} - C_A) - kC_A e^{-\frac{E_A}{RT}} \tag{11}$$

$$\frac{dT}{dt} = \frac{q}{V}(T_f - T) - \frac{\Delta H_R}{\rho C_p} kC_A e^{-\frac{E_A}{RT}} + \frac{UA}{\rho C_p V}(T_c - T) \tag{12}$$

Here, C_A is the concentration of reactant A, and T is the reactor temperature, which together form the state vector $x = [C_A, T]^\top \in \mathbb{R}^2$. The action variable, $u = T_c$, represents the cooling water temperature. Observations in this environment include the state values and their setpoints (if defined), structured as an array of shape (`(1, 2 + N_SP)`), where N_{SP} is the number of setpoints. For example, with setpoints for both states, the observation is `[CA, T, CA_Setpoint, T_Setpoint]`. The action space is defined as a continuous range $[290, 302]$, corresponding to a jacket temperature in Kelvin. The reward function computes the squared error between the state variables and their setpoints, scaled by a factor (`r_scale`), and sums the results into a single value. This setup enables training effective control policies for maintaining reactor stability and optimising performance.

4.2 Environment Setup

In this study, a customized simulation environment was configured using the PC-Gym framework to evaluate control performance for a CSTR system. A setpoint change is introduced at $t = 5$, adjusting the setpoint of `CA` from 0.80 to 0.85. The objective is to train an optimal policy for a RL agent that maximizes the episodic reward when interacting with the environment. For this relatively simple problem, such as setpoint tracking, the cost function to be minimized can be defined as the sum of the squared distances between the state of interest and its setpoint over a trajectory:

$$\text{Cost} = \sum_{i=0}^{T-1} r_{scale}(s_t - s_t^{SP})^2 \tag{13}$$

Table 1 summarizes the configuration of the unconstrained environment used within the PC-Gym framework.

By rolling out and visualizing the state and control trajectories from the model (Section 5.1), a box

state constraint for the CSTR temperature (T) was determined. Subsequently, stage-1 constraint identification was performed, making the state-dependent control constraints (Section 2.3) available for stage-2 RL agent training.

Table 2 lists the additional parameters required to implement constrained dynamics control within the PC-Gym framework. Notably, our method directly handles state constraints using a 3-tuple input $\langle \mathbf{x}_t, \mathbf{u}_t, t \rangle$. This approach eliminates the need for penalty-based methods to address constraint violations. Instead, we can explicitly solve the constrained optimisation problem at each time step using a constrained nonlinear local solver like `COBYLA` with multi-restart or any suitable constrained global optimisation solver.

Parameter	Value
Time Step	`nsteps = 12`
Simulation Time	`T = 26`
Environment Model	`cstr_ode`
Reward Scaling	`r_scale = {'Ca': 5e2}`
Setpoints	`Ca` transitions from `0.80` (initial 5 steps) to `0.85` (next 7 steps)
Observation Space	`[CA, T, CA_Setpoint]`: low `[0.75, 310, 0.80]`, high `[0.90, 350, 0.85]`
Action Space	`T_c` in `[295, 305]` K
Initial Conditions	`[0.80, 329, 0.80]`
Normalization	Actions and states are normalized
Noise	Disabled
Integration Method	`casadi`

Table 1: Environment Parameters for the unconstrained CSTR

Parameter	Value
constraint	`{'T':[320,330]}`
constraint type	`{'T':['<=','>=']}`
constraint violation penalty	`False`
done on constraint violation	`False`

Table 2: Environment Parameters for the constrained CSTR

5 Results & Analysis

5.1 Unconstrained RL

Initially, unconstrained RL agents (PPO and DQN) were trained to assess the feasibility of the current environment configuration. As shown in Figure 4, the bottom subplot illustrates the control trajectory (cooling jacket temperature `Tc`), while the two upper subplots depict the state variables. The first subplot shows the state variable (concentration of `Ca`) where a setpoint

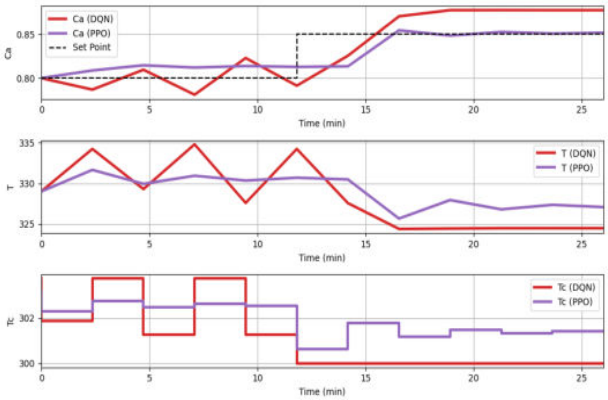


Figure 4: Unconstrained RL trajectories

change was applied from 0.8 to 0.85. The dotted line in the first subplot represents the setpoint at each time step, while the solid lines show the model's predictions. Notably, the second state variable, (temperature `T`), remained within a narrow range of 325 to 335 for both PPO and DQN.

To further challenge the RL agents, we introduced a box constraint on the temperature state variable `T`, limiting its range to [320, 330]. This constrained environment forced the agents to learn policies that not only achieved the desired control objective but also adhered to the imposed constraint. By imposing this constraint, we aimed to investigate how the agents would adapt their behavior and how well they could maintain state satisfaction while optimising the control objective.

5.2 Constrained RL

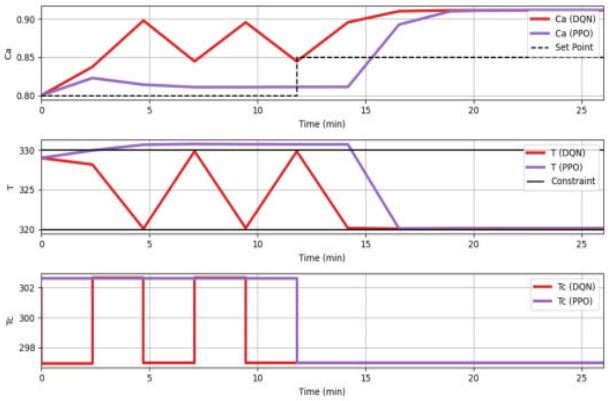


Figure 5: Constrained RL trajectories

After applying the constraint on state variable `T`, both RL agents were retrained, and their control trajectories are shown on the Figure 5. It was observed that both agents' predictions for the first state variable `Ca` deviated significantly from the setpoints. This suggests that the constraint may have been too restrictive, hindering the agents' ability to track the setpoints effectively. This is further evidenced by the trajectory of the second state variable, T, where the models frequently reached the boundary of the constraint.

Agent	Cost	Constraints violation
PPO	-5.88	5/12
DQN	-14.90	0/12

Table 3: Performance metrics for RL agents

Table 3 summarizes the agents' performance metrics in Figure 5. While PPO exhibits a slightly lower overall cost, it violates the constraint nearly half the time. This suggests that current PPO's decision-making process may not be well-suited to handling the imposed constraints, rather than a limitation in the constraint identification process itself. The primary reason for the observed constraint violations can be attributed to the exploration strategy defined defined in Eq 7. The PPO takes samples control from a normal distribution centered around the safe control $\bar{\mathbf{u}}_t$. However, there is no guarantee that this sampling process will yield control that satisfy the state constraint. When an unsafe control is obtained, the calculated gradient information can be misleading, potentially leading the PPO agent astray. This can ultimately result in the learning of a wrong policy that frequently violates the constraints.

In contrast, DQN demonstrates the effectiveness of the constraint identification process, exhibiting no constraint violations throughout the entire trajectory.

6 Conclusion and Future Work

In this research, we proposed a novel approach to apply reinforcement learning algorithms to constrained environments by leveraging the feasible space decomposition method developed by M. Mowbray [6]. This method enables the decomposition of a large joint state-control trajectory space problem $\prod_{t=0}^{T-1} \mathcal{U}_t \times \mathcal{V}_t$, into a sequence of local subproblems on time-dependent state-control spaces $\mathcal{U}_t \times \mathcal{V}_t$. The feasibility of this approach is tested by a case study involving setpoint tracking in a CSTR system within the PC-gym framework [2]. Both DQN and PPO were experimented with, and two naive ways, which are applicable to many other RL algorithms, of incorporating this state-dependent control constraint were investigated to assess the compatibility of action-value-based RL and actor-critic methods with this new approach. The success of DQN in making safe control in the constrained case study demonstrates the feasibility of the work, while further research is needed to adapt PPO to make safe control.

Despite its promise, current constrained PPO framework does not guarantee the feasibility of sampled controls. When an unsafe control is sampled, it can lead to the calculation of problematic gradient information, which hinders the learning process. A potential solution involves recursively sampling controls around a safe control until a feasible (safe) exploratory control is obtained. Future work will focus on exploring unbiased sampling strategies to further enhance the performance.

Another limitation of the current constrained PPO approach is its reliance on a biased policy gradient approximation. The filtering mechanism applied to the policy's output can be reinterpreted as an additional layer within the neural network architecture. To ensure accurate gradient information, future work will focus on identifying the exact form of this reparameterization and incorporating it as a regularization term in the loss function.

Code & Data Availability

The code used in this research is available online for public access. Specific implementations for the employed algorithms are provided in the following repositories:

`PPO` https://github.com/TACHUNC/PPO

`DQN` https://github.com/TACHUNC/DQN

Acknowledgement

The authors would like to acknowledge Dr. Max Mowbray for his consistent assistance, supervision, and insightful suggestions throughout the research process. The authors would also like to thank Prof. Benoit Chachuat, Prof. Christos Markides, and Ph.D. candidate Max Bloor for their valuable suggestions.

References

[1] E. Pan et al., "Constrained model-free reinforcement learning for process optimization," vol. 154, pp. 107462–107462, Nov. 2021.

[2] M. Bloor et al., "PC-Gym: Benchmark Environments For Process Control Problems," arXiv.org, 2024.

[3] V. Mnih et al., "Human-level Control through Deep Reinforcement Learning," Nature, vol. 518, no. 7540, pp. 529–533, Feb. 2015.

[4] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal Policy Optimization Algorithms," arXiv.org, 2017.

[5] P. Petsagkourakis et al, "Chance constrained policy optimization for process control and optimization," Journal of Process Control, vol. 111, pp. 35–45, Mar. 2022.

[6] M. Mowbray et al, "A Decomposition Approach to Characterizing Feasibility in Acyclic Multi-Unit Processes," IFAC-PapersOnLine, vol. 58, no. 14, pp. 216–221, Sep. 2024.

[7] APMonitor, "pdc/CSTR_Control.ipynb at master · APMonitor/pdc," *GitHub*, 2019.

[8] H. Yoo, B. Kim, J. W. Kim, and J. H. Lee, "Reinforcement learning based optimal control of batch processes using Monte-Carlo deep deterministic policy gradient with phase segmentation," Computers & Chemical Engineering, vol. 144, p. 107133, Jan. 2021.

[9] J. M. Lee and J. H. Lee, "Approximate dynamic programming-based approaches for input–output data-driven control of nonlinear processes," Automatica, vol. 41, no. 7, pp. 1281–1288, Jul. 2005.

[10] S. Spielberg, A. Tulsyan, N. P. Lawrence, P. D. Loewen, and R. Bhushan Gopaluni, "Toward self-driving processes: A deep reinforcement learning approach to control," AIChE Journal, vol. 65, no. 10, Jun. 2019.

[11] L. M. Puterman, Markov decision processes : discrete stochastic dynamic programming. New York Etc.: John Wiley & Sons, 2005.

[12] H. Yoo, V. M. Zavala, and J. H. Lee, "A Dynamic Penalty Function Approach for Constraint-Handling in Reinforcement Learning," IFAC-PapersOnLine, vol. 54, no. 3, pp. 487–491, Jan. 2021.

[13] J. Schulman, P. Moritz, S. Levine, M. Jordan, and P. Abbeel, "High-Dimensional Continuous Control Using Generalized Advantage Estimation," arXiv:1506.02438 [cs], Oct. 2018.

[14] J. Shin, T. A. Badgwell, K.-H. Liu, and J. H. Lee, "Reinforcement Learning – Overview of recent progress and implications for process control," Computers & Chemical Engineering, vol. 127, pp. 282–294, Aug. 2019.

[15] A. Kanervisto, C. Scheller, and V. Hautamäki, "Action Space Shaping in Deep Reinforcement Learning," arXiv.org, 2020.

Assessing different methodologies for characterisation of carbon persistence in biochar

Karim Alami Sounni
Department of Chemical Engineering Imperial College London, U.K.
Shell Campus The Hague, Netherlands

Abstract
Biochar use for Carbon Dioxide Removal (CDR) is recognise by both the IPCC and the EU. It could play a key role in climate change mitigation and meeting Paris agreement goals, if applied on a large scale. However, to confidently do so more insight on the persistence of carbon in biochar is needed. Petrological methodologies were considered to assess biochar persistence. Random Reflectance (R_o) measurements and the newly proposed inertinite benchmark for permanence were investigated. This benchmark corresponds to R_o above 2% and was derived from the R_o of inertinite, a maceral stable in soil over geological timescale. R_o measurements were compared to H/C_{org}, a persistence proxy already in use. It was found Ro and H/C_{org} are linearly correlated and R_o should be used as a complementary tool whenever H/C_{org} falls in a grey area. the use of Ro as a proxy also suggests current estimates are too conservative. Moreover, to meet current standards for permanence the mean Ro of sample needs to be above 2.3%, challenging the inertinite benchmark. Additionally, Raman spectroscopy parameter relevant to the study of macerals were tested as potential carbon stability proxies. The 3 following parameters were identified as preliminary/complementary proxies: Raman Band Separation (RBS), the width ratio and the saddle point intensity (S_i).

Key words Carbon Dioxide Removal, Biochar, Random Reflectance, H/C, Raman spectroscopy

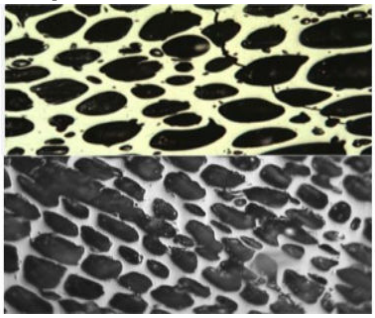
1. Introduction

Despite heated talks and negotiations at the Baku Cop29, many scientists insist limiting temperature increase below 1.5°C is unfeasible. 2024 is set to be the hottest year on record according to 3 of the main temperature monitoring research group, increasing the feeling of urgency toward climate change (Milman 2024). Consequently, the current benchmarks for climate change mitigations, built around science-based targets, seem to have missed the mark. The focus on carbon neutrality and "achieving net zero by 2050" is misleading as it neglects other greenhouse gas (GHG) and global inequalities in terms of responsibilities in the climate crisis (Resinger 2024). Net zero by 2050 is only consistent with the Paris agreement' goals if followed by a long period of net negative CO_2 emissions. (IPCC 2022) This is especially true as current national contributions exceed the requirements of the Paris agreement leading to an overshoot in temperature after 2030. Negative emissions appear to be the only efficient mitigation measure to avoid a prolonged overshoot, with disastrous implications for the environment (Fuss 2018) To achieve said negative emissions, Carbon Dioxide Removal (CDR) solutions can be used as they are able to remove CO_2 from the atmosphere. To match IPCC scenarios expectations CDR technology would require significant amounts of research and investment. The current portfolio of CDR solutions recognised by the IPCC includes Afforestation, Reforestation, Soil Carbon Sequestration, Biochar, Bion-Energy with Carbon Capture and Storage (BECCS), Direct Air Carbon Capture and Storage (DACCS), Enhanced Rock Weathering and Peatland Restoration. This paper focuses on biochar as it gained interest when the EU recognised it as a CDR method for its voluntary carbon market (European Commission 2024).

2. Background

2.1. Biochar

Figure 2.1.1 (a) fusinite taken from (SIU 2024) (b) biochar sample under microscope



Biochar is a carbon rich solid produced by pyrolyzing (heating up in anaerobic conditions) biomass. Biochar production is influenced by a large array of factors such as the biomass feedstock (herbaceous, soft wood and hard wood), the pyrolysis type (artisanal, industrial) and pyrolysis conditions (temperature, heating rate, heat transfer, residence time). It is used a soil amendment and yields several benefits for agriculture. For example, it improves water and nutrients retention in the soil

and has the potential to reduce greenhouse gas (N_2O,CH_4) emissions (Sanchez 2019). Moreover, producing biochar as a waste management technique for manure and crops residues significantly reduces methane emissions from agriculture (Sanchez 2019). To achieve CDR, biochar needs to be mixed with a C-sink matrix (building materials or soil), the combination of biochar and a C-sink matrix constitutes a C-sink. C-sink are the product of (i) an active removal of CO2 from the atmosphere, (ii) the transformation of the removed carbon into a storable form and (iii) its verifiable storage outside the atmosphere. In the case of biochar, the growth of the biomass removes the CO2 through photosynthesis, the pyrolysis transforms the removed carbon, and storage occurs with application to a matrix. The only option to form permanent C-sink is to apply biochar to soil, this is incentivised by the agricultural benefits mentioned earlier. (Schmidt 2024).

CDR use of biochar is reminiscent of the organic pathway of carbon dioxide removal via geological processes, known as "maceralization". Maceralization is responsible for the storage of 15 million gigatons of carbon within sedimentary rocks (Kroeger 2011) as it transfers carbon from the biosphere to the geosphere. Macerals are the organic counterparts of minerals; they are formed from the remains of biomass which went through several stages of carbon maturation. As the biomass undergoes gradual burial in the ground it is subjected to bacterial degradation as well as thermal degradation when burial temperature increases. Macerals are progressively enriched in carbon and gain a higher degree of aromaticity, synonymous with carbon stability. Inertinites are mature macerals which are stable over geological timeframe, they are the benchmark for carbon permanence. (Sanei 2024) One could think of pyrolysis as an artificial maceralization process. This comparison is particularly striking when considering the formation of fusinites. Indeed, fusinites are a subclass of inertinite whose maceralization started with "pyrolyzation" from ancient wildfires (Sanei 2024). However, this parallel alone between fusinite and biochar is not enough to systematically guarantee carbon persistence within biochar.

2.2. Carbon persistence in biochar

To generate carbon credits biochar projects must guarantee persistence of removed carbon above a given threshold. For reference the European Union require guarantee for persistence above 200 years (European Commission 2024) whereas Puro Earth uses 100 years as a threshold (Puro Earth 2022). Biochar projects are usually ex ante and highly additional, meaning they are financed based on predictive results and rely on incentives such as carbon credits to be financially viable (MSCI 2024). Thus, it is critical to effectively quantify and predict the amount of removed carbon which would fulfil the fixed persistence requirements. There is an ongoing debate in the scientific community regarding the best proxies and methodologies to assess this critical metric.

2.2.1. H/C_{org}

It is widely recognised stability of carbon increases with the degree of aromaticity. Thus, common practices involve establishing a correlation between a proxy for aromaticity and empirical persistence estimates. H/C_{org} is the prevalent proxy for aromaticity, aromaticity increases as H/C_{org} decreases, and they are highly correlated (Wang 2013). To be qualified as biochar samples need to have H/C_{org} below 0.7 (Schmidt 2024). On the other hand, persistence estimates are obtained through bacterial incubation experiments (Wolf 2021). Understandably, incubation experiments cannot run over 100 years, the longest one ran over for 8.5 years (Kuzyakov 2014). Carbon persistence estimates for longer timescale are given by regression and projection of best fit curves. Consequently, the results are mainly influenced by labile carbon outliers which are prone to early-stage degradation. A large share of carbons is considered to stay in storage for more than 100 years according to those experiments. The work of Wolf et al 2021 provides the following equation to determine stored carbon in biochar (C_{stored}):

$$C_{stored} = M_{bc} \times F_c \times F_p^{TH,TS} \times \frac{44}{12} \qquad (1)$$

M_{bc} is the dry mass of biochar, F_c is the organic carbon fraction of the biochar, $F_p^{TH,TS}$ is the permanence factor of carbon in biochar for a given soil temperature and time horizon, and the $\frac{44}{12}$ factor is the ratio of carbon dioxide and carbon molar mass (converts carbon stored in its equivalent carbon dioxide stored). $F_p^{TH,TS}$ is the most important term of this equation as it introduces the notion of carbon stability. It is expressed as a percentage and given by a linear relationship:

$$F_p^{TH,TS} = c + m \times H/C_{org} \qquad (2)$$

H/C_{org} here is used in molar basis and the linear regression coefficient were estimated by Wolf et al for given time horizon (usually 100 years) and soil temperature. The biochar C-sink methodology uses H/C_{org} to introduce discontinuity in their C_{stored} assessment. Samples with H/C_{org} below 0.4 are considered to contain 75% Permanent Aromatic Fraction (virtually stable forever) and 25% Semi Persistent Carbon. In this case C_{stored} is given by the following decay functions:

$$C_{stored}(y) = \frac{M_{bc}}{1000} \times F_c \times (750 + 45e^{-0.5232y} + 205e^{-0.00996y}) \quad (3)$$

Samples with H/C_{org} above 0.4 are treated as fully SPC and C_{stored} is given by the following decay function:

$$C_{stored}(y) = M_{bc} \times F_c \times (0.1787e^{-0.5337y} + 0.8237e^{-0.00997y}) \quad (4)$$

This introduces the idea of different carbon pools with different decay rates in soil, there is a non-normal distribution of carbon stability within biochar (Wang 2016). This is coherent with best fit regression to incubation experiments are multi pool exponential (Azzi 2024). An effective description of carbon persistence within biochar includes the distinction of different carbon pools with their respective decay rates.

2.2.2. Random Reflectance (R_o%)

Random reflectance R_o% is an established parameters in petrology used in the study of organic carbon. It corresponds to the ratio of reflected incident light. Similarly to H/C_{org} it is an indirect proxy for aromaticity, it is influenced by the degree of aromaticity and size of aromatic rings, R_o increases with aromaticity (Carr 1990). Over the last century $R_o$has been measured over a wide range of sedimentary organic matter. Building from those observation, R_o became an effective characterisation method for maceral types and their maturity, a key parameter for coal ranking and hydrocarbon exploitation. This is of particular interest as maceral types have known carbon stability. Stability ranges of naturally occurring macerals are not subject to debate; they are based on geological observations of sometimes millions of years old sediments. 3 maceral groups are relevant in the comparison with biochar: Liptinite with $0.2 < R_o < 1.2\%$, Semi-inertinite with $1.2 < R_o < 2\%$ and Inertinite with $R_o > 2\%$ (Sanei 2024). Liptinite and semi-inertinite are still considered as reactive and would be assimilate to SPC, liptinite is by far the most labile fraction. Inertinite is chemically inert and has the highest degree of carbonisation. It is by far the most relevant maceral type in the discussion around carbon stability because it is a reference in term of carbon permanence in the soil. As discussed previously, biochar is produced by "artificial" maceralization and share commonalities with macerals. Biochar can be characterised by its R_o leading to estimation of their carbon stability, mirroring maceral types. This is the idea proposed by Hamed Sanei in his 2024 paper, it is crystallised by his proposal of the inertinite benchmark IBR_O%. The $R_o\% > 2\%$ indicate the organic carbon's permanence. In their research Sinai et al use Ro frequency histogram to determine the fraction of different pools of carbon within the biochar leading to an accurate description of its potential for CDR use.

2.2.3. Raman spectroscopy

Most publications on carbon material include at least one Raman spectrum (Ferrari 2007) it is imposing itself as a backbone of research in the field. It is a non-destructive characterisation method, fast with high resolution and the deployable at lab scale or industry scale. It easily identifies different type of carbon structure and indicates their graphiticity. Graphiticity expresses the extent of graphitic like characteristic such as lattice structure and carbon hybridisation. Raman spectroscopy use is already a standard practice of nano tubes as well as amorphous and diamond-like carbons (Ferrari 2006). Characterisation of biochar for CDR purposes through Raman spectroscopy appears to be a promising pathway. This would build on efforts from the petrogeological community to use Raman spectroscopy as a geothermometer for organic matter maturity, once again a key parameter in coal and hydrocarbon exploitation. Additionally, studies have established correlations between Raman spectroscopy parameters and equivalent random reflectance (Henry 2019). The concepts of graphiticity and aromaticity both tie to how structured and stable (chemically and thermally) carbons are. Thus, this further confirms the relevance of Raman spectroscopy to assess carbon persistence within biochar.

Raman spectra of carbon system only show a limited number of prominent features, regardless of the material. This makes spectral interpretation challenging as the spectra appear to be deceptively

simple. However, they allow to distinguish a range of material from amorphous carbon to metallic nanotubes. The most prominent features are the D (disordered) and G (graphite) bands occurring respectively around 1360 cm^{-1} and 1560 cm^{-1} (Ferrari 2007). The D band is associated with structural defects and amorphous organic matter. It is caused by the breathing motion of sp^2 carbon in aromatic rings G band is associated with ideal graphitic structure, it's the result of bond stretching of all sp^2 carbon either in rings or chains (Bassez 1985). The bands are in the first order region of the spectra ($1000 - 1800\ cm^{-1}$). The bands in the second order region (2400-3500 cm-1) of the spectra are only defined for highly mature and graphitic organic matter (Henry 2019). Therefore, parameters linked to the D and G peaks are more suitable to the study of biochar.

3. Aim

This paper aims to compare existing proxies and investigate proposed benchmarks/thresholds for carbon permanence in biochar. As discussed in the previous section the initial reasoning behind established methodologies diverges. They are either based on incubation experiments or geological observations combined with petrological knowhow. Thus, there is a need to check for convergence of given methodologies to understand when they are more suitable. Moreover, it is critical to investigate the validity of permanence standards as they are bound to become the basis of policies for biochar CDR projects. Indeed, for biochar to be deployed at the necessary scale standardised procedure for decision-making are required.

Furthermore, the use of Raman spectroscopy to study biochar aims at finding new benchmarks and proxies for permanence and stability. Therefore, several Raman spectroscopy parameters will be investigated.

4.Methods

4.1 Sample Inventory

A total of 26 samples were studied displaying a range of feedstock origin and pyrolysis temperature. Sample can be divided into 3 categories: hard wood, soft wood and herbaceous. Pyrolysis temperature range was likewise divided into 3 categories: low temperature ($< 500°C$), medium temperature ($500 - 600°C$) and high temperature ($> 600°C$). The set includes 9 low temperature samples, 9 medium temperature samples and 8 high temperature samples. This is of particular interest as Hamed Sanei paper sample set contains more than 75% of high temperature samples, mostly from gasification processes. Thus, this paper fills an existing gap in the data for the study of R_o for biochar characterisation. Samples are either commercial biochar or provided by research group in Spain and Mexico. It is important to note that sample 11, tomato straw produced at 400°C has a H/C_{org} of 0.80 and does not qualify as biochar. H/C_{org} were either measured by the biochar producer or in the SHELL ETCA labs.

Table 4.1.1. list of studied samples with feedstock origin, pyrolysis temperature (PT) and H/C_{org}

#	FEEDSTOCK	PT [°C]	H/C_{org}
1	Stika spruce	700	0.27
2	Soft wood	450	0.46
3	Untreated wood TCR	n/a	0.50
4	Wood TCR method	n/a	0.54
5	Hard wood	450	0.45
6	Untreated wood	700	0.30
7	Rice husk	600-800	0.37
8	Grape stalks	400	0.57
9	Grape stalks	600	0.26
10	Giant reed straw	400	0.54
11	Tomato straw	400	0.80
12	Orange tree pruning	600	0.25
13	Agave bagasse	600	0.44
14	Corn stover	600	0.35
15	Corn stover	600	0.25
16	Aleppo pine	450	0.50
17	Holm oak	450	0.47
18	Stone pine	500	0.43
19	Aleppo pine (with leaf)	500	0.41
20	Aleppo pine	500	0.35
21	Holm oak	500	0.43
22	Arbutus unedo	500	0.45
23	Erica arborea	500	0.44
24	Quercus suber	500	0.40
25	Aleppo pine	500	0.40
26	Chestnut tree	500	0.46

4.2. Random Reflectance Measurements

Figure 4.2.1. frequency histogram and normal distribution for single population biochar

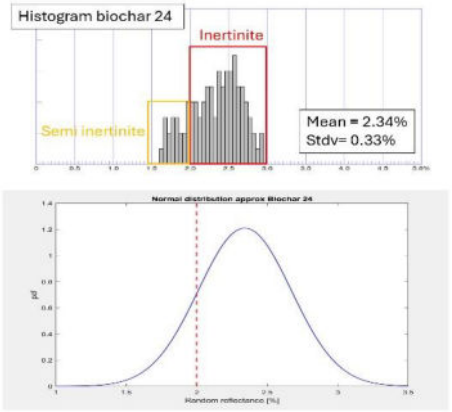
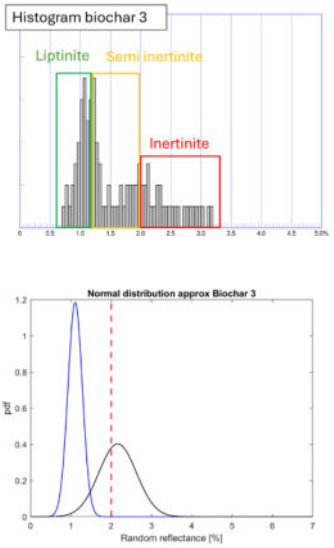


Figure 4.2.2. frequency histogram and normal distribution for 2 populations biochar



To guarantee efficient measurement of R_o in biochar samples need to be prepared thoroughly. Grains were crushed to smaller particles if too coarse and embedded in cold setting epoxy resin blocks. The blocks are shaped in cubes of side 1 inch and their upper surface was grinded and polished to expose the cross section of the biochar sample. This process aims at creating the most level and polished cross section area for each sample according to petrological methodology (ref iso 7404).

The measurements were carried out in incident white light using a Leica microscope integrated with the Discus-Fossil system. A 50× oil immersion objective was selected. The system was calibrated with a standard of know reflectance $R_O = 0.904$.

$R_O$measures the incident light reflected by the surface of an organic carbon material (ISO). For the measurement to be valid it needs to be taken on a flat surface in focus of the microscope. To guarantee statistical significance of the measurements 500 points were measured for each sample. Operator should cover the entirety of the cross-sectional area by doing a series of horizontal traces across the epoxy cube, starting from the bottom left of the cube. Suitable fragments with flat area should be identified and put into focus. However, no more than 5 points must be measured per view to avoid introducing biases linked to the presence of different morphotypes within biochar. It is imperative that the epoxy cube is not slanted as it would limit the number of measurable flat surfaces. A mean R_O and a standard derivation are obtained from the 500 manual measurements. Those parameters are used to approximate the biochar population by using a normal distribution. The PAC fraction is obtained by calculating the probability of having an $R_O > 2\%$ in accordance with the inertinite benchmark. Samples 3,7,10 and 16 display 2 separate pics in their R_O frequency histogram and are modelled by using 2 different normal distributions. Samples can be separated on 3 distinct categories depending on their PAC estimate: high persistence (PAC > 75%), medium persistence (PAC between 75% and 35%) and low persistence (PAC < 25%).

4.3 Raman Spectroscopy

Figure 4.3.1 Raman spectroscopy parameters taken from (Henry D.G 2019)

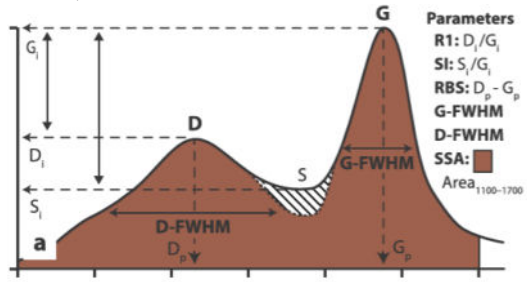
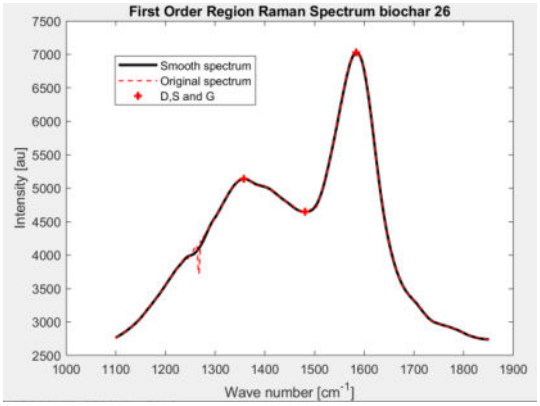


Figure 4.3.2 processed spectrum biochar 26



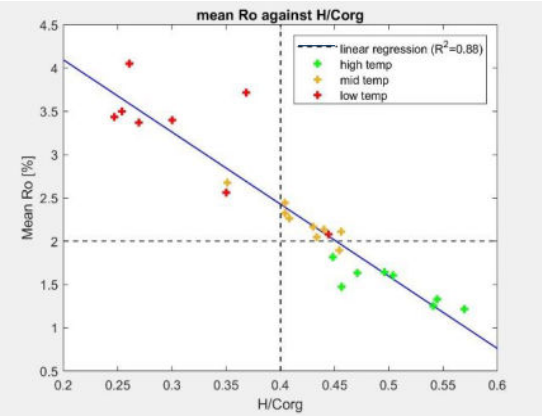
Raman Spectroscopy measurements were carried out on samples 2, 4, 5, 8, 9, 11, 12, 13, 18, 19, 20, 22, 24 and 26. Samples with 2 populations were avoided as they are not distinguishable by the eye and could have produced outliers. The samples were selected to cover a wide range of H/C_{org} in the hopes of covering extremes in terms of carbon persistence Sample 11 was still selected even if it does not fit the technical definition of biochar Samples were prepared by drop casting a solution of water, ethanol and grinded biochar in suspension onto a glass microscope slide. This allows to have a flat and homogenous surface to facilitate measurement. Measurements were carried out on a Senterra 2 Raman microscope with the 50× objective. The following parameters were selected: laser wavelength of 532 nm, power of 2.5 mW, aperture of

50×1000 μm, integration time of 8 seconds. The use of higher power or integration time leads to the sample burning. For the results to be statistically significant, more than 100 points were measured for each sample. Initially, 5 by 5 mapping was used but it resulted in too many outliers as the covered surface was not totally in focus. 4 by 4 mapping proved more reliable and was used throughout the experiments. The resulting 100 plus spectra per sample were exported to Excel and clear outliers were manually deleted. The remaining spectra were averaged and the primary region isolated as the secondary region bands were unexploitable. The smooth function in MATLAB allowed to dampen the noise from the spectra. The peaks and saddle point were identified by using the findpeaks function in MATLAB. The following parameters from figure 4.3.1 were calculated: peaks and saddle intensities position (D_i, G_i, S_i), the peaks positions (D_p, G_p), the Raman Band Separation (RBS), the ratio of Raman band height (R1), the full width at half maximum of both peaks (D_{FWHM} and G_{FWHM}) and the Saddle Index (SI). Additionally, the width ratio (D_{FWHM}/G_{FWHM}) was calculated.

5. Results and Discussion

5.1. Mean Random Reflectance

Figure 5.1.1. mean R_o against H/C_{org}



The relationship between measured mean R_o and measured H/C_{org} can be described by the following linear regression:

$$R_o = -8.34 \times (H/C_{org}) + 5.76 \qquad (5)$$

The high determination coefficient ($R^2 = 0.88$) indicates the goodness of fit. R_o decreases as H/C_{org} increases. In figure 5.1.1., the vertical dotted line indicates the $H/C_{org} < 0.4$ threshold (ref ithaka) and the horizontal dotted line indicates the inertinite benchmark $R_o > 2\%$ (hamed). At the H/C_{org} threshold the corresponding $R_o$is 2.42%. The H/C_{org} value corresponding to the inertinite benchmark is 0.45.

The linear relationship between R_o and H/C_{org} is most likely due to their dependence on the degree of aromaticity, as explained in previous sections. Since they are both proxies for aromaticity. Due to this relationship their simultaneous use seems redundant. However, the discrepancy between both permanence standards creates an added layer of complexity. Indeed, some points meet the inertinite benchmark while violating the H/C_{org} threshold. Thus, the H/C_{org} threshold appears to be too conservative. There is a grey area for biochar samples with H/C_{org} between 0.4 and 0.45. In this grey area additional R_o measurements are necessary to assess persistence.

5.2 PAC Fraction Estimate

Figure 5.2.1 Box plot of PAC fraction against temperature

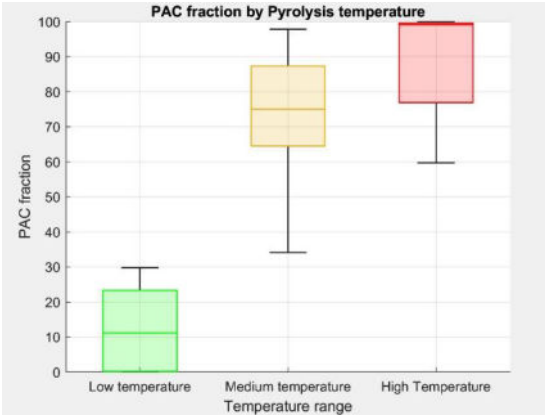


Figure 5.2.2. PAC fraction estimate against H/C_{org}

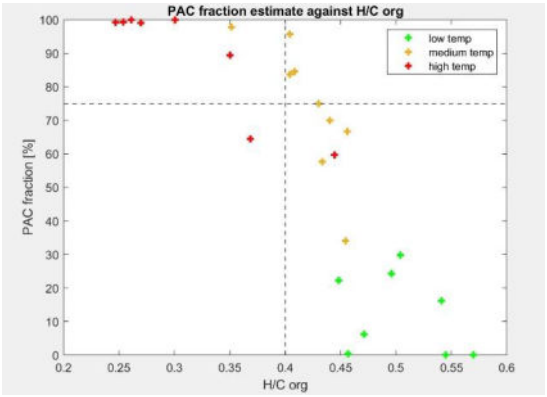


Figure 5.2.1 shows low temperature samples have a median PAC of 11% and all have PAC under 30%. They have no overlap with other temperature categories. Medium temperature samples have a median PAC fraction of 75% and range from 35% PAC fraction to 98% PAC fraction. High temperature samples have a median PAC fraction of 99% and range from 100% PAC fraction to 60% PAC fraction.

Thus, low temperature samples can be characterized systematically as low persistence and are not suitable for permanent CDR. Medium temperature and high temperature boxes overlap and both include medium persistence and high persistence samples. For those 2 categories pyrolysis temperature cannot be used to describe carbon persistence in a systematic manner.

On the other hand, figure 5.2.2 shows the PAC fraction decreases as H/C_{org} increases. Except for sample 7, PAC fraction is close to 100% for samples with H/C_{org} below 0.4. Samples with H/C_{org} between 0.4 and 0.45 display a wide range of PAC fraction from 98% to 35%. All samples with H/C_{org} above 0.45 are both low temperature and low persistence.

Those observations confirm the H/C_{org} threshold is too conservative and potentially leads to missed opportunities in term of carbon crediting. Samples below the threshold display PAC fraction higher than the proposed 75%. Additionally, samples with between 0.4 and 0.45 H/C_{org} form a grey area and contain non negligeable amounts of permanent carbon. This reinforces the requirement of further characterisation using R_o expressed in section 5.1. Sample 7 position as an outlier can be explained by the presence of 2 distinct population within the biochar.

Figure 5.2.3.PAC fraction estimate against R_o

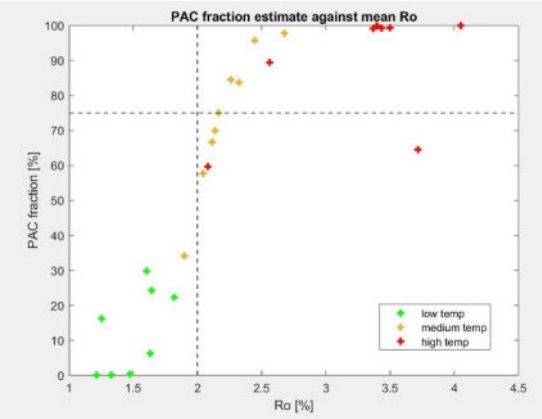


Figure 5.2.3 shows the PAC fraction estimate increases with random reflectance. Except for sample 7, all tested biochar with R_o above 2.3% classify as high persistence. Samples with R_o between 2.3% and the inertinite benchmark have PAC fraction ranging from 75% to 58%. Samples below the benchmark are all low persistence or close to low persistence.

At first, those results challenge the validity of the inertinite benchmark as it is not synonymous with high persistence. However, this arises from the lack of distinction between a single $R_o$measurement and the mean $R_o$of the sample. If a measured point has an R_o of 2% it still qualifies as an inertinite point whereas the sample requires a higher mean R_o to meet the inertinite standard. For example, if a single population sample had a 2% mean R_o it would only have a 50% PAC fraction, regardless of standard deviation. Consequently, the use of R_o above 2.3% a benchmark for high permanence is more suitable.

5.3. Raman Spectroscopy

Out of the parameters introduced in section 4.3, the Raman Band Separation (RBS), the width ratio and the saddle point intensity (S_i) were the most promising proxies for carbon persistence. Their potential as a proxy was established through 2 criteria: (i) their ability to identify different persistence categories and (ii) corelation to other proxies or PAC fraction estimate.

Figure 5.3.1 RBS against R_o

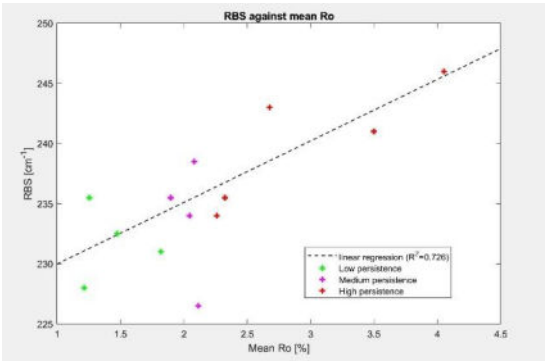


Figure 5.3.2 box chart of width ratio against persistence

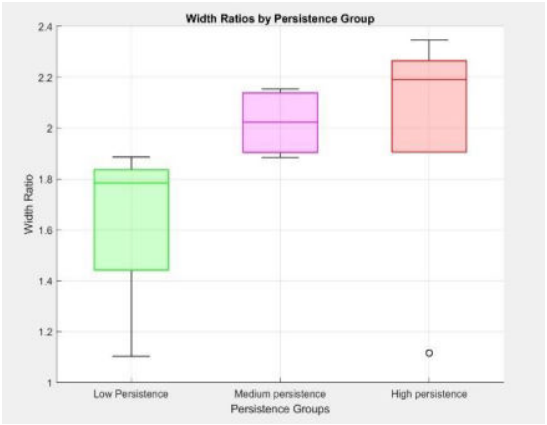


Figure 5.3.1 shows RBS increases with the mean R_o. The 2 extremum points are sample 11 with an RBS of 213 cm^{-1} for a mean $R_o$of 1.13% and Sample 9 with an RBS of 246 for a mean R_o of 4.05%. The relationship between RBS and mean R_o can be described by the following linear regression:

$$RBS = 5.23 \times R_o + 223.91 \quad (6)$$

Sample 11 was not considered for this linear regression. Although being promising as a fast to measure R_o estimator, RBS fails to identify different

persistence categories. Sample 4, 22 and 26 all have RBS of 235.5 cm^{-1} but all belong to different persistence categories. On its own it would not be a reliable carbon stability proxy.

Figure 5.3.2 focuses on the width ratio as a proxy. It shows median width increases with carbon persistence. Low persistence samples have a median width ratio of 1.785 with values ranging from 1.102 to 1.880. Medium persistence samples have a median width ratio of 2.023 with values ranging from 1.884 to2.138. High persistence samples have a median width ratio of 2.191 with values ranging from 1.906 to 2.346. Sample 19 is a clear outlier with high persistence and a width ratio of 1.116.

Width ratio can only be used to identify low persistence sample. Indeed, there is too much of an overlap between medium persistence and high persistence value ranges to make an accurate distinction. As a proposed guideline, samples with width ratio below 1.9 should be classified as low persistence. This a conservative benchmark as it would classify sample 19 as a low persistence biochar. Similarly to RBS on its own the width ratio would not be a reliable proxy for carbon stability.

Figure 5.3.3 box plot of S_i against persistence

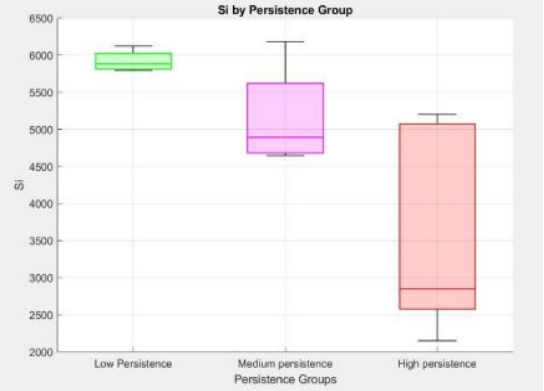


Figure 5.3.3 focuses on saddle intensity S_i as a proxy. It shows median S_i decreases as carbon persistence increases. This corresponds to a more pronounced "valley" between the D peak and G peak in stable biochar samples. Low persistence samples have a median S_i of 5880 au with values ranging from 5800 to 6120 au. Medium persistence samples have a median S_i of 4890 with values ranging from 4650 to 5620 au. High persistence samples have a median S_i of 2850 au with values ranging from 5200 to 2600 au.

It is important to note S_i trends were not studied in macerals (ref George). However, S_i is the most promising proxy as it would allow a quick approximation of the persistence category. The following conservative guidelines could be put in place: (i) samples with S_i above 5700 au are labelled as low persistence, (ii) sample with S_i between 5700 and 4600 au are labelled as medium persistence and (iii) samples with S_i below 4600 au are labelled as high persistence. Those are conservative guidelines which fail to describe carbon stability as accurately and reliably as R_o. Thus, S_i should only be used for preliminary estimation and in combination with other proxies.

Figure 5.3.4 3d plot of persistence proxies including S_i

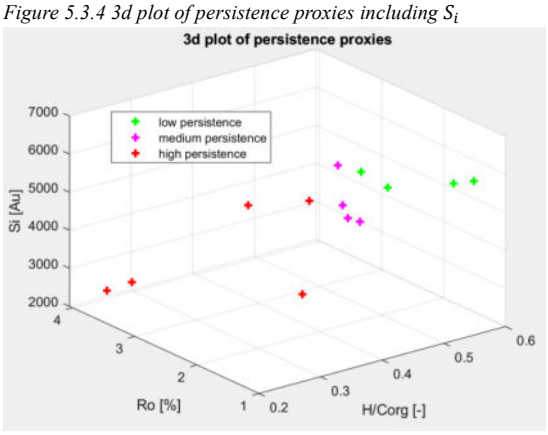
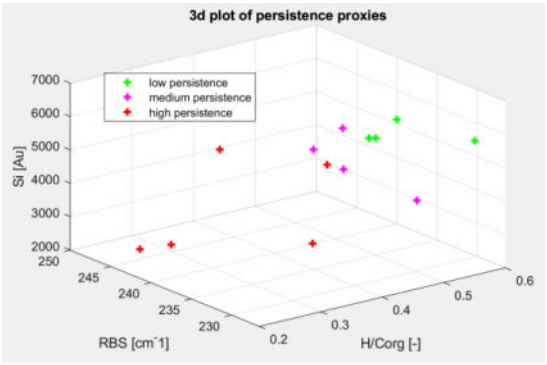


Figure 5.3.5 3d plot of persistence proxies including RBS and S_i



Figures 5.3.4 and 5.3.5 demonstrate the relevance of Raman spectroscopy parameters as complementary proxies. Their combined use with H/C_{org} and R_o allows to identify clusters, corresponding to each persistence categories. The clusters for high persistence and low persistence correspond to large and distinct regions of the 3d plots. In comparison, the medium persistence cluster is less distinct and encompasses a smaller region of the 3d plots. This is due to the arbitrary definition of persistence bounds for each group. The medium persistence group effectively acts as a buffer zone. Additionally, figure 5.3.5 highlights the effectiveness of clustering without R_o. Thus, this method bypasses the uncertainties of more widely available proxies. This enables a more reliable and accessible assessment of biochar CDR capacities. Moreover, further samples could be classified by identifying their nearest cluster and its corresponding persistence category.

6. Conclusion

CDR technologies are bound to develop significantly due to their key roles in climate change mitigation. Biochar is a part of the CDR portfolio and has received considerable interest following its recognition by the IPCC and the EU. However, for biochar use to reach the expected scale, it is imperative to propose methodologies to characterise its CDR capacities in a systematic and standardised manner. In existing methodologies 2 proxies for carbon persistence biochar were identified H/C_{org} and R_o. They either stem from natural science or petrological knowledge. The study and comparison of those proxies allowed to establish the following guidelines:

1. Samples with Pyrolysis temperature (PT) below 450°C should be classified as low persistence.
2. Samples with H/C_{org} above 0.45 should be classified as low persistence.
3. Samples with H/C_{org} below 0.4 should be classified as high persistence.
4. Samples with H/C_{org} between 0.45 and 0.4 fall in a grey area and require additional characterisation, ideally with R_o.
5. R_o must be used on samples with high likelihood of irregularities (Artisanal biochar)
6. Biochar samples with mean R_o above 2.3% should be classified as high persistence. This challenges the inertinite benchmark.
7. R_o should be considered as the most precise characterisation method as it identifies different carbon pools within biochar.

Furthermore, Raman spectroscopy was investigated as a potential characterisation method for carbon persistence in biochar. This also stems from efforts in the petrological community to describe organic matter through their Raman spectra. The 3 following Raman parameters were identified as promising proxies:

1. Raman Band Separation (RBS), corresponding to the difference of the D band and G band positions. It can be used as a rough estimator for R_o. It is unable to identify persistence categories due to non-negligible overlaps.
2. Width ratio, corresponding to the ratio of the Full Width at Half Maximum (FWHM) of the D and G band. It can be used to identify low persistence biochar. As conservative guideline, samples with width ratio below 1.9 should be classified as low persistence.
3. S_i, corresponding to intensity of the saddle point between the D band and G band. It is the most promising proxy from Raman spectroscopy. It allows to approximatively identify the different persistence categories. The following conservative guidelines could be applied: (i) samples with S_i above 5700 au should be classified as low persistence, (ii) sample with S_i between 5700 and 4600 au should be classified as medium persistence and (iii) samples with S_i below 4600 au should be classified as high persistence

Raman spectroscopy parameters are not robust enough to be used as standalone proxies. Indeed, they lack the required precision and should be considered as preliminary/complementary proxies. The simultaneous use of various characterisation methodologies is recommended as it allows to perform clustering techniques. This is strongly recommended, since it enables enable precise persistence estimates without relying on R_o.

To increase the understanding of carbon persistence in biochar other characterisation methods must be considered. For example, pyrolysis-GC/MS and Hydrogen Pyrolysis could be investigated. Moreover, the clustering methods offers the possibility to apply machine learning techniques to biochar persistence classification. To handle the simultaneous use of a large array of proxies, methods such as Principal Component Analysis (PCA) or K nearest neighbour (KNN) could be considered in further work.

7. Acknowledgements

Acknowledgements go to Dr. Marta Camps-Arbestain, Dr. George Siavalas and Prof. Chris Tighe for their guidance and support. Acknowledgement also go to Dr. Niamh Hartley, Dr. Kalpana Balaskandan and Mrs. Patricia Carry for their help with Raman spectroscopy.

8. References

Azzi, E.S et al. 2024. Modelling biochar long-term carbon storage in soil with harmonized analysis of decomposition data.
Geoderma. Volume 441

Beny-Bassez, C et al. 1985. Characterization of carbonaceous materials by correlated electron and optical microscopy and Raman micro spectroscopy. Scan. Electron Micros. 1, 119–132.

Carr, A.D et al. 1990. The relationship between aromaticity, vitrinite reflectance and maceral composition of coals: Implications for the use of vitrinite reflectance as a maturation parameter.

Organic Geochemistry, Volume 16. Issues 1–3. pp. 313-323.
European Commission. 2024. provisional agreement on the Carbon Removals and Carbon Farming (CRCF) Regulation. [Online] Available at : https://climate.ec.europa.eu/eu-action/carbon-removals-and-carbon-farming_en
Ferrari, A.C et al. 2006. Raman spectroscopy in carbons : From nanotubes to diamond. Philos. Trans. R. Soc. Ser. A 362, 2267–2565
Ferrari, A.C. 2007. Raman spectroscopy of graphene and graphite: Disorder, electron–phonon coupling, doping and nonadiabatic effects. Solid State Communications. Vol 143. Issues 1–2. pp 47-57,
Fuss, S et al. 2018. Negative emissions—Part 2: Costs, potentials and side effects. *Environ. Res. Lett.* Volume 13
Henry, D.G et al. 2019. Raman spectroscopy as a tool to determine the thermal maturity of organic matter : Application to sedimentary, metamorphic and structural geology. Earth-Science Reviews. Volume 198.
IPCC. Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Shukla, P. R., et al. (Eds.). (Cambridge University Press, Cambridge, UK and New York, NY, 2022).
Kroeger, K.F et al. 2011, Atmospheric methane from organic carbon mobilization in sedimentary basins — The sleeping giant? Earth-Science Reviews, Volume 107, Issues 3–4,
Kuzyakov, Y et al., 2014. Biochar stability in soil: decomposition during eight years and transformation as assessed by compound-specific 14 analysis. Soil Biol. Biochem. 70, 229–236.
Milman, O. 2024. World's 1.5C climate target 'deader than a doornail', experts say. The Guardian. [Online] Available at https://www.theguardian.com/environment/2024/nov/18/climate-crisis-world-temperature-target
MSCI Carbon Market. 2024. Outlook for the Global Biochar Market [Online]. Available at https://www.msci.com/www/blog-posts/outlook-for-the-global-biochar/04633228838
Puro earth. 2022. Biochar Methodlogy version 3 [Online]
Available at https://puro.earth/biochar
Reisinger, A et al. 2024. Science-based targets miss the mark. Communications Earth & Environment. Volume 5. Article 383
Sanchez Garcia, A et al 2019. Biochars from Mediterranean Agroindustry Residues: Physicochemical Properties Relevant for C Sequestration and Soil Water Retention. ACS Sustainable Chem. Eng.7, 4724−4733
Sanei, H et al. 2024. Assessing biochar's permanence: An inertinite benchmark. International Journal of Coal Geology, Volume 281
Schmidt, H.P et al. 2024. Global Biochar C-Sink Standard 3.1. the Ithaka Institute for Carbon Strategies
Southern Illinois University, 2024. Coal Macerals. Crelling's Petrographic Atlas of Coals and Carbons
Wang, T et al. 2013. Predicting C aromaticity of biochars based on their elemental composition, Organic Geochemistry. Volume 62. pp 1-6
Wang, J et al., 2016. Biochar stability in soil: meta-analysis of decomposition and priming effects. GCB Bioenergy 8, 512–523.
Wolf, D et al. 2021. Greenhouse Gas Inventory Model for Biochar Additions to Soil. Vol 55. Issue 22. pp 14795-147805

Energy System Modelling to Decarbonise Zambia's Transport Sector

Elise Cerullo and Yaroslav Zakharov
Department of Chemical Engineering, Imperial College London, U.K.

Abstract Developing a feasible pathway to decarbonise Zambia's transport sector is crucial for meeting emission targets outlined in Zambia's Nationally Determined Contributions in line with the Paris Agreement. As the impacts of climate change in Zambia intensify, with droughts stunting the nation's GDP growth, implementing climate-compatible initiatives for decarbonisation has become essential. In this paper, the practicality and implications of proposed climate policies were explored in OSeMOSYS. Findings showed that the optimal pathway for transport decarbonisation involves incorporating a biofuel blending mandate in conjunction with making dedicated efforts to electrify freight and light passenger vehicles. This resulted in a 41% reduction of 2022 emission levels by 2050, and a total additional capital investment cost of $5.67 billion relative to business-as-usual. The cost of supporting the blending mandate was found to be significantly lower by switching to domestically produced biofuels. However, the development of the industry would also be negatively affected by droughts. Results from a sensitivity analysis emphasise Zambia's vulnerability to climate change, and the need to diversify the energy mix to reduce the reliance on hydropower. Although achieving net zero emissions by 2050 was theoretically possible in the model through intense electrification, the predicted decreases in rainfall showed this was infeasible. Therefore, climate adaptation is vital for Zambia's economic and environmental prosperity.

Keywords: Zambia, transport sector, carbon emissions, optimisation, OSeMOSYS

1. Introduction

Zambia is a landlocked country located in Sub-Saharan Africa (SSA) with one of the fastest-growing populations globally, it is anticipated that by 2050 the population will triple. The country is currently experiencing rapid economic growth as it is the second largest copper producer in Africa.[1] With the copper demand increasing exponentially as global trends towards "net zero" promote accelerated uptake of electric vehicles, the economy is projected to continue its economic progress.[2]

Transportation is an essential driver for economic growth, as it influences business productivity, labour market functionality and geographical distribution of economic activity. However, domestic and international transport are leading contributors with a share of up to 20% of global greenhouse gas emissions.[3] Currently it is projected that global transport emissions will increase by over 60% by 2050.[4] To meet the targeted temperature warming limit of well below 2 degrees compared to pre-industrial levels, as highlighted by the 2015 Paris Agreement, urgent action is required to decarbonise the transport sector. Since the adoption of the UN's 2030 Agenda for Sustainable Development,[5] there has been increased international momentum for transformation in the transport sector. Many of the Sustainable Development Goals (SDGs) are directly linked to transportation, revitalising the drive to implement change.[6]

Consistent with global trends, Zambia's transport sector is a major contributor to carbon emissions, accounting for 27.6% of the country's total emissions.[7] Challenges in infrastructure and rising vehicle emissions have created a complex landscape for decarbonisation. Zambia's transport sector alone represents 38% of the country's total consumption of petroleum products and has an energy intensity 14% higher than the global average. According to Zambia's Nationally Determined Contributions (NDCs), commitments have been made to reduce greenhouse gas emissions by 25% or 47% with substantial international support from 2010 levels by 2030.[8] These commitments highlight the pressing need for proactive measures to help Zambia meet its climate targets.[9]

Zambia's Integrated Resource Plan (IRP) further emphasises the government's focus to develop its transport sector to minimise carbon emissions in alignment with the Paris Agreement. To achieve these proposed targets, international climate finance is crucial. As the impacts of climate change intensify, the average annual climate finance cashflows need to increase exponentially to account for both climate mitigation and adaptation. In the upcoming years, global requirements for climate finance are increasing to over $1 trillion per year to mitigate the worst impacts of climate change. Current business-as-usual investments are insufficient to maintain the global temperature rise to well below 2 degrees. Since transportation is one of the highest emitting sectors, it currently receives 29% of mitigation finance. However, developing countries receive only a small fraction of the necessary funding, with just 3% of global climate finance allocated to them.[10] Discussions during the COP29 in Azerbaijan have led to agreements of providing developing countries with $300 billion per year by 2035, which is a fraction of the required climate financing of over $1 trillion.[11]

Zambia is one of the most vulnerable countries to climate change due to its inadequate infrastructure and substantial rural population.[12] Over the past three decades, climate-related extreme weather events, such as floods and droughts, have cost the nation 0.4% of its economic growth, with their severity and frequency continuing to rise.[13] These weather occurrences have significantly affected the agricultural sector, which relies heavily on rainwater, jeopardizing the country's food security and exacerbating poverty levels.[14]

These challenges highlight the growing pressure on Zambia's transport sector to evolve. The 2024 Zambian drought, the worst in over 2 decades, has contributed to increased poverty levels, crop losses and food insecurity. Zambia requires urgent assistance to finance climate adaptation and mitigation initiatives. In this study, the impact of proposed climate policies was investigated to identify the optimal decarbonisation pathway for Zambia's policymakers while balancing economic prosperity, food security and energy accessibility. Scenarios investigated included business-as-usual (BAU), achieving Zambia's NDC emission reduction targets, implementing a biofuel blending mandate, and reaching "net zero". Modelling was conducted to determine the optimal policies for achieving these emission reduction goals.

2. Background

Zambia's transport sector heavily relies on fossil fuels, with a huge prevalence of diesel and gasoline vehicles. As of 2022, electric passenger and freight vehicles did not exist in the country. Infrastructure challenges across Zambia for road transport electrification is a limiting factor for achieving "net zero" in the transport sector.

Existing research by Moono *et al.* (2012) demonstrates potential for biofuel policy implementation in Zambia, with predicted emission reductions of 10% if blending occurs in accordance with the 2008 National Energy Policy.[15] Research by Hartley *et al.* (2018, pp. 213-232) explored Zambia's significant potential for domestic bioethanol production, particularly for export to South Africa. Their findings suggest that developing this industry could boost Zambia's economic growth while safeguarding food security.[16] While local biodiesel production is unlikely to profitably support a biodiesel mandate in Zambia, Drabik *et al.* (2016) concluded that expanding the domestic soybean supply could make such an initiative "welfare beneficial".[17]

Due to Zambia's low subsidised cost of electricity and the high efficiency of electric vehicle (EV) engines, Collett *et al.* (2021) has illustrated that there are projected cost savings of over $0.15/km in Zambia associated with switching to EVs from internal combustion engine (ICE) vehicles. However, the key barrier is the high capital cost for EVs, thus implementing governmental subsidies could be a strategy to reduce the burden.[18]

The largest issue with implementing electrification in Zambia, is the hydro-based electricity system which is highly sensitive to weather events. McPherson *et al.* (2018) highlights the shortfalls in the hydropower-based Zambian power sector at the consequence of population, economic growth and climate change.[19]

To optimise electrification and ensure the energy balance remains stable, focus on decarbonising freight is crucial. Aryanpur and Rogan (2024) highlight the critical need for "heavy truck decarbonization", noting that while trucks only account for 9% of global vehicle fleet, they contribute a staggering 39% of the life-cycle emissions from road transport. Since there is a strong correlation between economic growth and freight, the activity of freight is modelled to double by 2050. Thus, freight electrification is pivotal to prevent "carbon lock-in effects".[20]

ICE buses are another key contributor to emissions, accounting for 25% of global black carbon output. Predictions expect that the global percentage share will grow to 50% by 2030. Ayetor *et al.* (2021) suggests that African countries need to focus on the electrification of buses, as ICE buses emit 329 times more carbon dioxide than electric buses. Africa's high renewable potential provides opportunity for electrification if harnessed.[21]

Another key area to focus on for electrification is two and three-wheelers (2&3Ws), due to their rising prevalence in SSA countries. The United Nations Environment Program (UNEP) has listed 2&3Ws as the leading transportation to decarbonise in developing countries due to their rapid growth in the transport mix. UNEP is currently providing financial support for electric 2&3W uptake in nine African developing countries, presenting potential opportunities for Zambia to benefit in the future.[22]

These papers offer key insights into effective strategies for decarbonising transportation across SSA, which can be applied to Zambia. This study aimed to evaluate proposed policies to decarbonise the Zambian transport sector while improving the accuracy of the transport sector model within Zambia's broader energy system model.

3. Methods

3.1 Overview - OSeMOSYS

OSeMOSYS (Open-Source Energy Modelling System) is an optimisation system used for energy system planning and sustainable policy development through open-source modelling. It uses linear programming to compute a minimisation of the system's total cost over a model period.[23] The objective function was bounded with constraints related to capacity, activity and emissions. Input

data for the model was collected from online open data resources or estimated based on historical trends and assumptions in line with existing literature.

3.2 Model Implementation

The transport sector model was developed in tandem with the existing power sector model created by the Climate Compatible Growth research group.

Appendix A provides a simplified reference energy system created for the transport sector, displaying the connection between fuel, electricity generation and vehicle technology in Zambia. Passenger transport included cars, buses, 2-wheelers (2W) and 3-wheelers (3W), whereas, freight transport combined both light and heavy goods vehicles. In this study, cars and buses were referred to as light (LPVs) and heavy passenger vehicles (HPVs) respectively. Railway and domestic aviation were incorporated in the model to satisfy the energy balance, but the associated costs, emissions and energy consumption were not investigated. Demand projections for these technologies were kept constant across different scenarios. Water-based transport was not included in the model as data was not readily available. Furthermore, international aviation from Zambia was not included in the model as only strictly domestic transport was considered.

3.3 Data Inputs

Since the most up-to-date International Energy Agency (IEA) energy balance for Zambia was from 2022, the model's base year was set to 2022. The demand and the activity of a technology were expressed in billion-passenger-kilometres (bpkm) for passenger vehicles, and billion-tonne-kilometres (btkm) for freight vehicles. The capacity of each technology in the model was expressed as 1000 vehicles.

Using the sources outlined in Appendix B, fuel efficiency, average occupancy and average annual mileage of a typical vehicle for each technology was assumed. Multiplying the former two values yielded the input activity ratios. This ratio represented the amount of a particular fuel in petajoules (PJ) that was used to produce 1 bpkm of activity. Ratios were assumed to be constant over time on the basis that as Zambia developed, improvements in fuel efficiency of private transportation technologies would be offset by decreasing average occupancy. For ease of implementation, the same assumption was made for HPVs and freight vehicles. Thus, these ratios were the basis of each technology's emission factor. Vehicle emissions were assumed to only include CO_2 tailpipe emissions. NO_x and THC emissions were not considered in this model.

Capital costs of each technology were assumed to be the upfront cost of purchase in Zambia, incorporating manufacturing and transportation costs. The costs of diesel and gasoline-fuelled vehicles were assumed to remain constant, while the costs of electric and hybrid vehicles were assumed to decrease over time due to the reduction in the price of batteries and the improvement in the manufacturing process.[24] Appendix B provides the estimated capital and fixed costs for the base year, rate of change and associated assumptions. Based on findings from Grube *et al.* (2021), hybrid vehicles were projected to have a zero-cost difference to ICE vehicles by 2030.[25] Thus, their costs in the model were assumed to decrease by 2.5% annually in line with the suggested trends. Furthermore, South Africa has a rapidly growing market for hybrid vehicles, which reinforced this assumption.[26]

Fixed costs were assumed to include the cost of vehicle maintenance, repair and insurance, and were estimated based on prices in similar countries and average proportions of capital costs, outlined in Appendix B. These proportions were assumed to remain constant for all modelled scenarios, balancing any potential increase in insurance and labour cost with the potential reduction in price of parts due to improving availability in Zambia.

3.4 Fuel Types

Vehicle fuels used in freight and passenger transportation included electricity, diesel, and gasoline. Hybrid vehicles were assumed to be powered by electricity 40% of the time, and the remainder by gasoline. The cost of imported fuels was assumed to vary identically in relation to each other, and the projected change was based on historic market trends. In the blending scenarios, biofuels were blended into gasoline and diesel fuels. Flex-fuel vehicles were not considered in the model as their widespread adoption would be unrealistic and would conflict with existing plans to integrate biofuels into conventional fuels.[27] Since green hydrogen is produced through the electrolysis of water and Zambia is currently experiencing severe droughts,[28] hydrogen powered vehicles were not considered in this model. Furthermore, their high capital costs would provide high barriers for market entry in Zambia.[29]

Freight in Zambia was assumed to be 100% diesel powered, based on studies in Ireland.[30] HPVs were also assumed to be 100% diesel powered, based on data found in the UK.[31] Since data on the fuel mix of LPVs in Zambia was not readily available, it was estimated based on the IEA energy balance, outlined in Appendix B. All 2&3Ws were assumed to run on gasoline.[23]

3.5 Scenario Development

3.5.1 Business-as-usual

A business-as-usual (BAU) scenario was developed, which projected future transport demand and corresponding emissions assuming no climate policy changes.

Demand projections for each transportation type were predicted based on historical data trends. For LPVs, growth was projected using predicted population and vehicle ownership growth rates. Arndt *et al.* (2013) highlights the strong correlation between population growth, automobile use and motor vehicle fleet.[32] Regarding HPVs, progression was driven by population growth as studies in rural America had illustrated that public transportation use increases with higher populations.[33] 2&3Ws had the same driver for demand as HPVs, as studies in SSA countries showed evidence for correlations between uptake and population growth rates.[34] Freight demand was projected with GDP growth rates as there is a close relationship between freight demand and economic development.[35]

3.5.2 Biofuel Blending

Zambia's 2019 National Energy Policy outlined the government's plan to begin blending bioethanol and biodiesel into fossil fuels in 2020 - a plan that is yet to be implemented.[27] Therefore, a scenario was explored to model the introduction of the mandate to blend biofuels into gasoline and diesel fuel across Zambia. In line with Nigeria's blending mandate policy, the model assumed an immediate switch from pure diesel to 20% biodiesel fuel (B20) at all fuel filling stations.[36] Similarly, the replacement of pure gasoline with 10% bioethanol fuel (E10), was also assumed to take place in 2023, with a step increase to 20% bioethanol fuel (E20) in 2035. These fuels remain compatible with conventional engines while lowering CO_2 emissions. The change in fuel efficiency of each technology associated with the blend mandate was assumed to be negligible, as these blends are said to insignificantly impact fuel consumption.[37]

To model the most likely implementation of the biofuel blending mandate, all bioethanol and biodiesel were assumed to be imported.[38] It was assumed that these biofuels produced zero vehicle carbon emissions, as the CO_2 produced from their combustion is excluded from national greenhouse gas emissions inventories.[39] This assumption was based on the premise that growing the biomass feedstocks used for biofuel production would offset the produced CO_2.[39] Moreover, a further scenario was explored to incorporate Zambia's promising potential in domestic biofuel production and examine the economic impact of the potential switch from imported to locally produced biofuels.

3.5.3 NDC Targets

Zambia's NDC targets are to reduce carbon emissions by 25% or 47% with substantial international support compared to 2010 levels by 2030. To reduce the emissions, constraints on the minimum capacity investment for electric vehicles were introduced.

3.5.4 Net Zero

To limit global warming by well under 2 degrees in correspondence with the Paris Agreement, a "net zero" scenario was developed. This scenario focused on electrification as the most practical method to decarbonise the transport sector is through the introduction of EVs.[40] Furthermore, achieving "net zero" in the current model was only possible using solely electric vehicles, as they were the only zero emission vehicles.

4. Results

4.1 Overview

Findings illustrated that electrification efforts in alignment with Zambia's NDCs are not feasible by 2030. The NDC targets to reduce emissions by 25% or 47% compared to 2010 levels required emission reductions to 712 kilotons or 1008 kilotons of CO_2 by 2030, which was not possible. Hence, a refined version of the NDC targets was explored: to achieve a reduction in carbon emissions of 25/47% by 2050 relative to 2022 levels. This scenario was titled "Modified NDC Targets", and it showed that the proposed carbon emission reductions were possible with sufficient policy. Henceforth, "25% reduction" and "47% reduction" results refer to these respective targets.

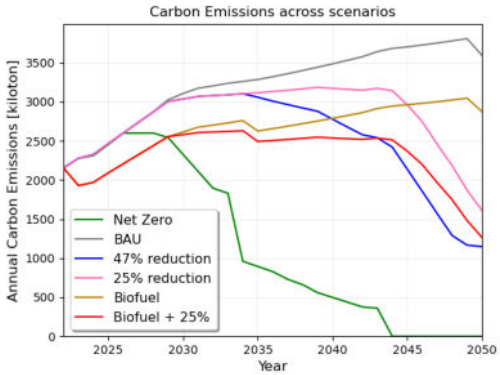


Figure 1: Projected Annual Carbon Emissions across different scenarios from 2022 to 2050.

Emission trajectories seen in *Figure 1* illustrate that the "Net Zero" scenario could reduce emissions to zero before 2050. For the 25% and 47% reduction scenarios, the overall emissions relative to 2022 were reduced. However, the emission reductions occurred at a much slower rate compared to the net zero scenario.

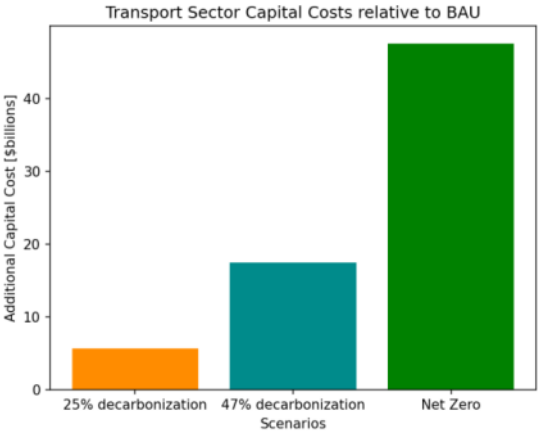


Figure 2: Additional Transport Sector Capital Costs Relative to BAU under Different Decarbonisation Scenarios

As seen in *Figure 2*, the scenarios with a higher extent of carbon emission reductions had higher additional capital costs compared to BAU. "Net zero" had the highest additional capital expenditure of $47.6 billion compared to BAU.

4.2 Biofuel blending

As expected, the biofuel blending mandate achieved a 20% reduction in CO_2 emissions by 2050 compared to the BAU scenario. Moreover, total CO_2 emissions were 17.9% lower across the whole modelling period. The transport sector capital costs and the activity projections remained identical to the BAU scenario, hence the biofuel blending scenario was not included in *Figure 1* and *Figure 2*, respectively. In this scenario, the annual variable cost of all imported fuels, including gasoline, diesel, bioethanol and biodiesel, was 4.8% higher than in the BAU scenario. However, exploring the switch to locally produced biofuels resulted in lower fuel variable costs than in BAU, as shown by *Figure 3*.

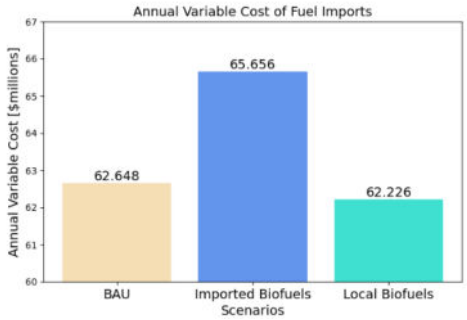


Figure 3: Annual Variable Cost of Fuel Imports under Different Scenarios

4.3 Modified NDC Targets

Constraining the minimum capacity investment in EVs resulted in the achievement of modified NDC targets. To achieve a 25% reduction in carbon emissions, a minimum annual investment of $20 million for electric LPVs was introduced in 2030. This constraint underwent a step increase to $50 million by 2040. Freight electrification began with a constraint of $25 million annual minimum capacity investment introduced in 2045, gradually increased to $50 million in 2049.

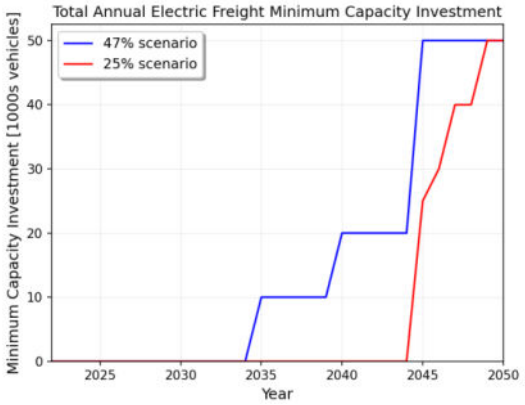


Figure 4: Total Annual Electric Freight Minimum Capacity Investment for Modified NDC Scenarios

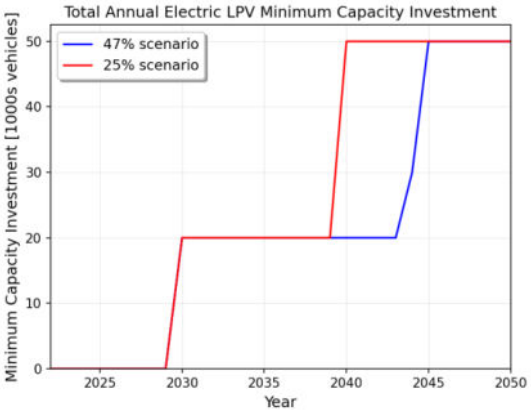


Figure 5: Total Annual Electric LPV Minimum Capacity Investment for Modified NDC Scenarios

For the 47% reduction scenario, the minimum annual capacity investment for electric LPVs was introduced in a similar manner as the 25% reduction scenario. However, freight electrification was introduced earlier in 2035 with a constraint of $10 million invested annually into electric freight. This constraint gradually increased to $50 million by 2045. These constraints are demonstrated by *Figure 4* and *Figure 5* for freight and LPVs, respectively.

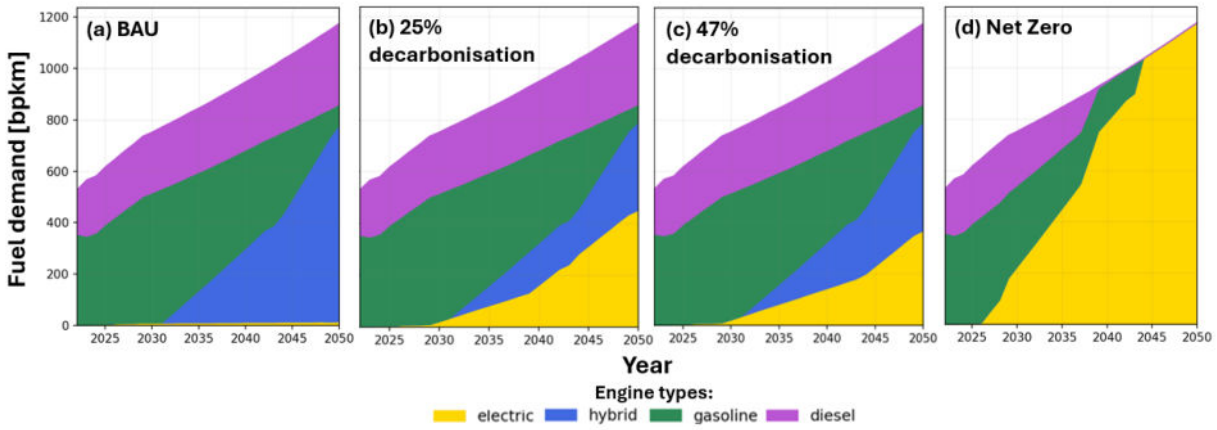


Figure 6: Passenger Transport Activity across different scenarios from 2022 to 2050.

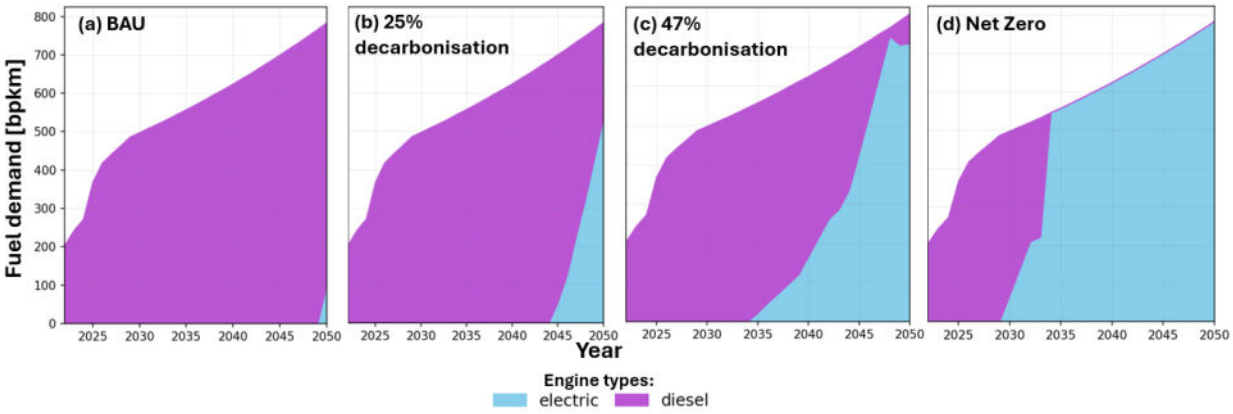


Figure 7: Freight Transport Activity across different scenarios from 2022 to 2050.

Figure 6 highlights the change of the passenger transport activity mix over the modelling period for each scenario. In the BAU scenario, the activity mix transitions from 100% fossil fuels to 65% hybrid in 2050. There was a small introduction of electric vehicles from 2023, but over the demand period the electric fuel share did not exceed 1.3%.

In the 25% reduction scenario, intensive LPV electrification from 2030 and freight electrification from 2045 was the optimal solution to achieving the modified NDC target. On the other hand, in the 47% reduction scenario the optimal solution was obtained through intensive freight electrification from 2035 and gradual LPV electrification from 2030. As shown by *Figure 6* and *Figure 7*, the 25% reduction scenario had a stronger focus on passenger transport electrification whereas the 47% reduction scenario had a stronger emphasis on freight electrification.

Compared to BAU, the additional investment for the 25% reduction scenario was $5.669 billion between 2023 and 2050. For the 47% reduction scenario, the additional investment required was $17.5 billion compared to BAU.

4.4 Combined scenario

Noticing the immediate emissions reduction caused by the biofuel blending mandate, and the positive impacts of a long-term electrification strategy, it was decided to combine the constraints from the biofuel blending mandate and the 25% reduction scenarios. This resulted in an emission reduction of 41% by 2050 compared to 2022 levels with an additional capital expenditure of $5.670 billion, which was only $1 million higher than for the 25% reduction scenario alone.

4.6 Net Zero

Net zero emissions across the transport sector could be achieved by 2050. However, it would only be possible if investment into electric vehicles began in 2025. The ideal transport mix to reach net zero emissions was fully electrified. However, this scenario was highly cost intensive with an additional investment requirement of $47.6 billion.

5. Discussion

Biofuels are growing in popularity across Africa, with policies being implemented across Kenya, South Africa and more.[41] A biofuel blending mandate appeared to be a promising policy for the

Zambian government to enact. As shown in the results, the mandate only increased the cost of fuel imports by less than 5% compared to the BAU scenario. Thus, there was no impact on the total capital expenditure and activity mix projections due to technology investment decisions being primarily driven by vehicle capital costs. By contrast, the 20% reduction of emissions in 2050 and the 17.9% reduction in total emissions up to that year make the biofuel blending mandate a worthwhile commitment. This policy would also make Zambia less vulnerable to the volatility in the fossil fuels' market and improve the energy mix of the transport sector.

A scenario was explored investigating a switch to locally produced biofuels, assuming that Zambia's bioethanol and biodiesel capacity would be able to accommodate the supply necessary for the mandate. The project by the Food and Agriculture Organisation of the United Nations found that while there was sufficient cassava and sugarcane molasses to provide a sustainable supply of bioethanol, it concluded that neither soybeans nor sunflower seeds would be able to produce the biodiesel required for the mandate proposed in this study, necessitating a combination of locally produced and imported biodiesel.[38] Nevertheless, this scenario showed the potential for a lower variable cost of vehicle fuel in relation to the BAU scenario, highlighting the appeal of developing the domestic biofuel production industry. Undoubtedly, this would require significant upfront investment, but this strategy would carry economy-wide benefits. For example, the existing sugar mills would be able to utilise sugarcane waste from bioethanol production to generate electricity, selling excess supply to the national grid. It was estimated in Hartley *et al.* (2018, pp. 213-232) that the bioethanol production industry alone could provide new employment opportunities for up to 260,000 people and would increase both rural and urban household welfare.[16] Therefore, on top of implementing the biofuel blending mandate, it is advisable for policymakers to explore the development of domestic biofuel production.

As mentioned in the previous section, achieving Zambia's NDC targets in the current model proved impossible. The required reductions in CO_2 emissions by 2030 were not even possible with the most extreme Net Zero scenario. It was shown that the annual carbon emission level would be 2327 kilotons in 2030, which is 8% higher than the 2022 value. Thus, this highlighted that NDC targets must be adjusted for the reduction to be feasible.

To reach the modified NDC emission reduction targets, intensive 2&3W electrification was explored in line with plans outlined by the UNEP to electrify 90% of 2&3Ws by 2030 in developing countries. Literature had illustrated that moving towards high electrification rates for 2&3W transport was crucial to meet climate goals in developing countries, accounting for up to 5-10% of total emissions for some nations.[42] Given Zambia's low adoption of 2&3Ws, which account for less than 10% of the total vehicles, there was minimal change in emissions during modelling. Thus, other electrification scenarios were explored to achieve the emission targets.

Another policy investigated to meet NDC goals was the mandate of a carbon tax. A value of $100/ton of CO2 was tested as it represented the most extreme tax currently implemented in the European Union.[43] However, the carbon tax had no impact on the optimal solution as the same fuel mix was selected by the model across the scenarios and the emissions stayed constant. Thus, this was not an effective policy to explore in the context of Zambia's transport sector.

In Zambia's Energy Efficiency Strategy Action Plan (2022),[10] one of the strategic objectives was to increase public transport use. The target involved reducing private transportation by 5% and incentivising public transportation. Since HPVs were the only form of public transport used in the model, the constraint was implemented through reducing bpkm demand for LPVs by 5% and increasing bpkm demand for HPVs by 5%. However, the results highlighted that the policy was not highly impactful as emissions reduced by 50 kilotons in 2050 compared to BAU, which is just 1%.

Results from BAU projections showed that most emissions from Zambia's transport sector were derived from freight or LPVs. Policies related to incentives for electrification of these transportation forms were considered to achieve the modified NDC targets. To achieve a 25% reduction in emissions, a stronger focus on LPV decarbonization is recommended. Conversely, for a 47% reduction, prioritising freight electrification proved more effective, as it was more carbon intensive. In the BAU scenario, 49% of emissions were derived from freight, while 43% originated from LPVs. LPVs had a lower decarbonization cost of approximately $10,000, making them the focus for the 25% reduction scenario. Freight, however, faced a higher electrification cost difference of around $100,000 per vehicle, hence why it was targeted for electrification in the 47% reduction scenario.

By incorporating key insights from both biofuel blending and the modified NDC target of 25% reduction scenarios, a model was developed using constraints derived from each. Emissions were reduced by 41% with a capital cost of only $1 million extra compared to the 25% reduction scenario. This was a significant improvement to

either of those scenarios in isolation. Moreover, a staggering 65% reduction compared to BAU levels was achieved in 2050, showing the positive impact of implementing the biofuel blending mandate in conjunction with electrifying freight and LPVs.

During the exploration of the net zero scenario, only intensive electrification was able to reduce emissions to zero by 2050. Although electrification of the transport sector in Zambia has proven to provide promising emission reductions, there are currently large power shortages in the country. Kariba is a dam located in Zambia which provides 85% of the nation's electricity generation capacity.[8] Due to the extreme droughts, which are worsened by climate change, the dam has insufficient water to run its turbines.[44] These power shortages are predicted to worsen amid more frequent climate-related weather events driven by climate change. Since the IRP has predicted a reduction of rainfall of between 7.5% and 15% by 2050, a sensitivity analysis was conducted on the Net Zero scenario to assess the robustness of the results.

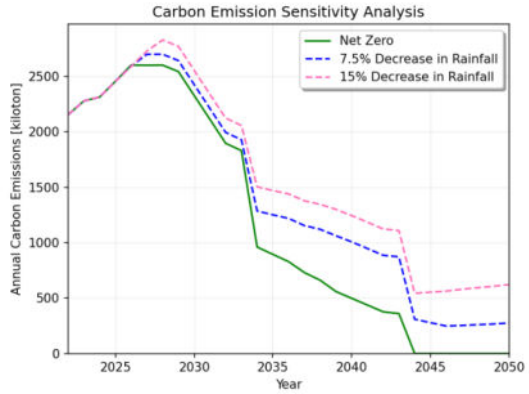


Figure 8: Sensitivity Analysis on the Impact of Rainfall Reductions on achieving Net Zero

The sensitivity analysis illustrated the reliance of Zambia's power sector on climate-related weather events. Achieving net zero emissions across the transport sector was shown to be infeasible if the projected decrease in rainfall occurs, as seen in *Figure 8*. The results illustrated a concerning reliance of decarbonisation on stable power supply, assuming there would be insufficient adaptation to climate events. Thus, diversifying country-wide power generation would be another key area for the Zambian government to explore to negate the reliance on climate-vulnerable technologies such as hydropower.

Due to the volatility of energy markets, cost of electricity generation is subject to variability. To represent this, a sensitivity analysis was also conducted on the economic potential. Using the approach described by Hofbauer (2017, p.66), electricity prices were changed by $\pm 30\%$.[45] Emissions from all scenarios stayed constant across the price changes, demonstrating that the optimal solution for the transport mix was not impacted by electricity costs.

6. Conclusions

This paper aimed to develop a representative model of Zambia's transport sector and project the associated costs and CO_2 emission reductions from various climate strategies. To provide insights for Zambian policy development, the study aimed to assess the feasibility of these policies by incorporating the context of Zambia's ongoing challenges.

Although achieving Zambia's NDC goals of 25% or 47% reduction in emissions by 2030 compared to 2010 levels proved infeasible, there are other strategies the government can implement across policymaking. Findings from the study show that implementing the biofuel blending mandate and in conjunction with dedicated efforts to electrify freight and LPVs is the most cost-effective pathway to reduce emissions in the transport sector. This was shown to be significantly more impactful than either of those strategies in isolation, achieving a 41% reduction in emissions from 2022 by 2050.

The potential worsening of power shortages caused by projected decreases in rainfall and droughts showed that achieving "net zero" emissions by 2050 could be infeasible if Zambia's power sector is not diversified. The significant increase in capital costs of $47.6 billion compared to the BAU scenario strengthen this prediction, given that climate cashflows from developed countries are not meeting requirements for adequate climate adaptation and mitigation.

It was also found that switching to domestically produced biofuels could significantly reduce the cost of supporting the blending mandate. However, the increasing severity of droughts would have a substantial impact on the developing industry, hence climate adaptation is vital for Zambia to prosper, and future studies could investigate solutions to dampen this impact.

Additional policies investigated included carbon tax, an intense electrification of 2&3W, and incentives to increase public transport use. However, in isolation, they all proved to have little-to-no effect on the transport sector emissions.

Future pathways for research could include incorporating other technologies of low emission transmission transport, such as 100% biofuel-powered vehicles. Furthermore, looking at the entire life cycle emissions associated with each transportation type could provide valuable insights into the decarbonisation of the transport sector.

7. Acknowledgements

We would like to express our gratitude to Dr. Ariane Millot for her extensive support throughout this project.

8. References

[1] BBC (2023) *Zambia Country Profile*. Available at:https://www.bbc.co.uk/news/world-africa-14112449 (Accessed: 12 December 2024)

[2] Mills, R. (2023) *The global copper market is entering an age of extremely large deficits*. Available at: *https://www.mining.com/the-global-copper-market-is-entering-an-age-of-extremely-large-deficits/* (Accessed: 12 December 2024)

[3] Greener Vision (n.d.) *Transport and the economy: The relationship between transport and the economy*. Available at: *https://greener-vision.com/guidance-tool/relationship-between-transport-economy/* (Accessed: 12 December 2024)

[4] Scronce, E. (2024) *World Bank Group Transport Overview*. Available at: *https://www.worldbank.org/en/topic/transport/overview* (Accessed: 12 December 2024)

[5] United Nations Climate Change (n.d.) *Key aspects of the Paris Agreement*. Available at: *https://unfccc.int/most-requested/key-aspects-of-the-paris-agreement* (Accessed: 12 December 2024)

[6] Jaramillo, P., *et al.* (2022) *IPCC Sixth Assessment Report.* Available at: https://www.ipcc.ch./report/ar6/wg3/chapter/chapter-10/ (Accessed: 12 December 2024)

[7] International Energy Agency (2022) *Zambia Emissions.* Available at: https://www.iea.org/countries/zambia/emissions (Accessed: 12 December 2024)

[8] Ministry of Energy (2023) *Integrated Resource Plan*. Available at: https://www.moe.gov.zm/irp/?wpdmpro=summary-report (Accessed: 12 December 2024)

[9] Kapala *et al.* (2022) *Zambia Energy Efficiency Strategy and Action Plan 2022* Available at: https://www.moe.gov.zm/wp-content/uploads/2022/08/Zambia-Energy-Efficiency-Strategy-and-Action-Plan-2022.pdf (Accessed: 12 December 2024)

[10] Buchner *et al.* (2023) *Global Landscape of Climate Finance 2023*. Available at: https://www.climatepolicyinitiative.org/publication/global-landscape-of-climate-finance-2023/ (Accessed: 12 December 2024)

[11] Valdre and Barbara (2024) *Climate finance, carbon markets and more: 4 key takeaways from COP29*. Available at: *https://www.weforum.org/stories/2024/11/cop29-4-key-takeaways/*(Accessed: 12 December 2024)

[12] Winthrop *et al.* (2018) *Zambia Country Climate Risk Assessment Report*. (Accessed: 12 December 2024)

[13] World Bank Group (2021) *Zambia*. Available at: https://climateknowledgeportal.worldbank.org/country/zambia/vulnerability (Accessed: 12 December 2024)

[14] Climate Links (2016) *Climate Risk Profile: Zambia.* Available at: https://www.climatelinks.org/resources/climate-risk-profile-zambia (Accessed: 12 December 2024)

[15] Moono, J. (2012) *Illustrative scenarios of biofuel policies for transport sector in a developing country: case study of Zambia*. Availabe at: *https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A548124&dswid=-7709* (Accessed: 12 December 2024)

[16] Hartley *et al.* (2018, pp. 213-232) *Economy-wide implications of biofuel production in Zambia*. Available at: https://doi.org/10.1080/0376835X.2018.1485552 (Accessed: 12 December 2024)

[17] Drabik *et al.* (2016) *Producing biodiesel from soybeans in Zambia: An economic analysis*. Available at: https://doi.org/10.1016/j.foodpol.2016.01.001 (Accessed: 12 December 2024)

[18] Collett *et al.* (2021) *Can electric vehicles be good for Sub-Saharan Africa?* Available at: *https://www.sciencedirect.com/science/article/pii/S2211467X21001085* (Accessed: 12 December 2024)

[19] McPherson, M. (2018) *Planning for variable renewable energy and electric vehicle integration under varying degrees of decentralization: A case study in Lusaka, Zambia.* Available at: *https://www.sciencedirect.com/science/article/pii/S0360544218304821?via%3Dihub* (Accessed: 12 December 2024)

[20] Aryanpur and Rogan (2024) *Decarbonising road freight transport: The role of zero-emission trucks and intangible costs*. Available at: *https://www.nature.com/articles/s41598-024-52682-4* (Accessed: 12 December 2024)

[21] Ayetor *et al.* (2021) *Impact of replacing ICE bus fleet with electric bus fleet in Africa: A lifetime assessment.* Available at: *https://www.sciencedirect.com/science/article/pii/S0360544221001018* (Accessed: 12 December 2024)

[22] Ayetor *et al.* (2023) *Feasibility of electric two and three-wheelers in Africa.* Available at: *https://www.sciencedirect.com/science/article/pii/S2773153723000427* (Accessed: 12 December 2024)

[23] KTH-dESA (2018) *Introduction to OSeMOSYS.* Available at: https://osemosys.readthedocs.io/en/latest/manual/lntroduction.html (Accessed: 12 December 2024)

[24] Goldman Sachs (2024) *Electric vehicle battery prices are expected to fall almost 50% by 2026.* Available at: https://www.goldmansachs.com/insights/articles/electric-vehicle-battery-prices-are-expected-to-fall-almost-50-percent-by-2025 (Accessed: 12 December 2024)

[25] Grube *et al.* (2021) *Passenger car cost development through 2050.* Available at: https://doi.org/10.1016/j.trd.2021.103110 (Accessed: 12 December 2024)

[26] The Outlier (2024) *SA is embracing hybrids, but all-electric and plug-in sales lag*. Available at:

https://theoutlier.co.za/economy/2024-07-15/88376/electric-vehicle-car-sales-south-africa (Accessed: 12 December 2024)
[27] Kaunda, T. (2019) *National Energy Policy (2019)* (Accessed: 12 December 2024)
[28] Chugh and Taibi (2021) *What is green hydrogen and why do we need it? An expert explains.* Available at: *https://www.weforum.org/stories/2021/12/what-is-green-hydrogen-expert-explains-benefits/* (Accessed: 12 December 2024)
[29] Arnold, N. (2013) *Hydrogen fuel cell cars: everything you need to know*. Available at: *https://www.bmw.com/en/innovation/how-hydrogen-fuel-cell-cars-work.html* (Accessed: 12 December 2024)
[30] Aryanpur and Rogan (2024) *Decarbonising Road transport: The role of zero-emission trucks and intangible costs*. Available at: *https://www.nature.com/articles/s41598-024-52682-4* (Accessed: 12 December 2024)
[31] Jesson, N. (2019) *Road fuel consumption and the UK motor fleet.* Available at: *https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/812622/Road_fuel_consumption_and_the_UK_motor_vehicle_fleet.pdf* (Accessed: 12 December 2024)
[32] Arndt, W.H. (2013) *Transport in Megacities - development of sustainable transportation systems*. Available at: *https://www.researchgate.net/publication/273996052_Transport_in_Megacities_-_development_of_sustainable_transportation_systems* (Accessed: 12 December 2024)
[33] Zhang and Xu (2022) *The bus is arriving: Population growth and public transportation ridership in rural America*. Available at: *https://www.sciencedirect.com/science/article/pii/S0743016722002182#:~:text=Despite%20this%20stereotype%2C%20population%20growth%20does%20increase%20public,public%20transportation%20does%20not%20affect%20the%20commuting%20time* (Accessed: 12 December 2024)
[34] Alimo, P. (2022) *Investigating the increasing demand and formal regulation of motorcycle taxis in Ghana.* Available at: *https://www.sciencedirect.com/science/article/pii/S0966692322001211?casa_token=502n3QF-LC0AAAAA:jvycXgkTmky3BSyaBD-maCZAWFUqeZcnaMduD-deLGvpyxVeL7klWUOJKZWRWDeBcAQUUyQXVjU* (Accessed: 12 December 2024)
[35] Wang *et al.* (2021) *The relationship between freight transport and economic development: A case study of China.* Available at: *https://www.sciencedirect.com/science/article/pii/S0739885920300834#abs0015* (Accessed: 12 December 2024)
[36] International Energy Agency (2021) *Biofuels blending mandate*. Available at: https://www.iea.org/policies/5696-biofuels-blending-mandate (Accessed: 12 December 2024)
[37] U.S. Energy Information Administration (2024) *How much ethanol is in gasoline, and how does it affect fuel economy?* Available at: https://www.eia.gov/tools/faqs/faq.php?id=27&t=10#:~:text=In%20general%2C%20vehicle%20fuel%20economy,does%20not%20contain%20fuel%20ethanol (Accessed: 12 December 2024)
[38] Food and Agriculture Organization of the United Nations & Ministry of Energy of Zambia (2020) *Sustainable Bioenergy Potential in Zambia* (Accessed: 12 December 2024)
[39] U.S. Energy Information Administration (2022) *Biofuels explained.* Available at: https://www.eia.gov/energyexplained/biofuels/biofuels-and-the-environment.php (Accessed: 12 December 2024)
[40] Patil *et al.* (2024) *Sustainable Decarbonization of Road Transport: Policies, Current Status, and Challenges of Electric Vehicles*. Available at: https://doi.org/10.3390/su16188058 (Accessed: 12 December 2024)
[41] SMEBluePages (2023) *Biofuel Industry Statistics in Africa.* Available at: https://smebluepages.com/biofuel-industry-statistics-in-africa/ Accessed: 12 December 2024)
[42] Gupta *et al.* (2023) *The real global EV buzz comes on two wheels.* Available at: *https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/the-real-global-ev-buzz-comes-on-two-wheels* Accessed: 12 December 2024)
[43] Mengden, A. (2023) *Carbon Taxes in Europe, 2023.* Available at: https://taxfoundation.org/data/all/eu/carbon-taxes-in-europe-2023/ Accessed: 12 December 2024)
[44] Zimba and Adebayo (2024) *Ongoing African drought has plunged Zambia into daily blacouts as hydroelectric dam unable to run.* Available at: https://www.euronews.com/green/2024/10/14/ongoing-african-drought-has-plunged-zambia-into-daily-blackouts-as-hydroelectric-dam-unabl (Accessed: 12 December 2024)
[45] Hofbauer, L. (2017) *Shaping India's Energy Future: Energy Demand and Transport Sector Pathways.* Available at: https://www.sev-bayern.de/wp-content/uploads/2019/01/Masters_Thesis_Hofbauer.pdf (Accessed: 12 December 2024)

Optimisation of Electrochemical Bleaching for Lignocellulosic Biomass

Hannah Coy and Amy McLoughlin
Department of Chemical Engineering
Imperial College London

Abstract

The bleaching of lignocellulosic biomass is a crucial step in the production of high-purity cellulose for industrial applications, particularly in the paper and textile industries. Traditional bleaching methods, such as chlorine gas and chlorine dioxide, produce dioxins. Dioxins are harmful compounds that have been shown to cause cancer and have negative environmental impacts such as bioaccumulation. Electrochemical bleaching is a promising development that can produce bleaching agents such as hypochlorite, chlorite and chlorate that have fewer negative societal impacts. This study focuses on optimising the conditions of NaCl electrolysis, with the aim to minimise the production of hypochlorite and maximise the concentration of chlorite and chlorate ions present. Hypochlorite concentrations were decreased when temperatures were elevated, and the time of electrolysis was increased. This highlights an initial step into developing a bleaching process that is more sustainable.

Keywords: Electrochemical Bleaching, Hypochlorite, Chlorite, Box Behnken Design, Optimisation

1. Introduction

The bleaching of lignocellulosic biomass improves the purity of cellulose, which significantly increases its value for industrial applications, particularly in the paper and textile sectors(1). Pulp and paper production relies mainly on cellulosic fibres derived from wood, although other plant materials are also used. In addition to paper, cellulose from wood pulp is used in manufacturing various goods, including rayon, cellulose acetate, and cellulose esters, which are essential for products like textiles and packaging films(2).

Bleaching agents primarily target lignin, a polymer responsible for the brown colouration of plant materials(3). These agents break down lignin, resulting in the decolourisation of the pulp. The removal of lignin is crucial to purify cellulose, the main structural component of plant fibres. An effective bleaching process focuses on maximising lignin removal while preserving the integrity of cellulose which contributes to the strength of paper through hydrogen bonding between its fibres(2).

The bleaching of biomass can be done by many methods. Traditionally, chlorine gas was used to decolourise and whiten the pulp(4). However, it was discovered during the mid-1980s that dioxins were present in pulp mill effluents(5). Dioxins are highly toxic and carcinogenic compounds that pose severe risk to aquatic ecosystems and humans. They are formed during reactions between lignin and chlorine containing compounds(2).

Following this discovery, Elemental Chlorine-Free (ECF) bleaching became the dominant technology for

bleaching biomass, accounting for 87% of the global bleached kraft pulp capacity(4). This process uses chlorine dioxide, a highly reactive and unstable compounds, to bleach the pulp. While ECF significantly reduces the production of dioxins compared to chlorine gas bleaching, these compounds can still be generated, raising environmental concerns(5).

Electrochemical bleaching is a promising alternative, leveraging in situ generation of hypochlorite, chlorite and chlorate – oxidising agents capable of bleaching lignocellulosic biomass. Sodium hypochlorite has been widely used due to its strong bleaching efficiency. However, it has many drawbacks as a bleaching agent for pulp. Firstly, it is a non-selective bleaching agent, meaning it attacks not only the lignin but also the cellulose in the pulp(6). Furthermore, its instability and potential to form dioxins under certain conditions present challenges for transportation, storage, and environmental safety(7). Chlorite and chlorate, being less reactive than hypochlorite, may offer a safer alternative as they are less likely to produce dioxins.

2. Background

2.1. The Process

The formation of the hypochlorite (ClO^-), chlorite ($ClO_2{}^-$) and chlorate ($ClO_3{}^-$) ions is of a result of a series of oxidation reactions starting from chloride ions, which occur at the positive anode. Corresponding reduction reactions happen at the site of the negative cathode. This is due to flow of the electrons in the system, negative ions are attracted to the anode which also accept the electrons released from the oxidation reactions. The series of half reactions describing the system are shown in Equations 1- 5 (8).

$$2\,Cl^- \longrightarrow Cl_2 + 2\,e^- \tag{1}$$

$$Cl_2 + H_2O \longrightarrow HClO + Cl^- + H^+ \tag{2}$$

$$HClO \longrightarrow ClO^- + H^+ \tag{3}$$

$$ClO^- + 2\,OH^- \longrightarrow ClO_2{}^- + H_2O + 2\,e^- \tag{4}$$

$$ClO_2{}^- + 2\,OH^- \longrightarrow ClO_3{}^- + H_2O + 2\,e^- \tag{5}$$

2.2. Aims of Project

The primary aim of this study is to address the environmental concerns associated with traditional bleaching methods by minimising the production of dioxins. This will be achieved by optimising the conditions for electrochemical bleaching to favour the formation of chlorite and chlorate over hypochlorite, as hypochlorite is more likely to generate dioxins because of its higher reactivity with lignin.

An additional objective is to improve the efficiency of the bleaching process, by reducing the need for multiple bleaching steps traditionally required. This improvement is expected to be achieved by directly oxidising chloride ions to higher oxidation states.

To achieve these objectives, the effects of the initial sodium chloride concentration, temperature, and electrolysis time on the concentrations of the bleaching agents were investigated. The parameters chosen are shown in Table 1.

Table 1: Parameter Ranges for Electrochemical Bleaching.

Parameter	Range
Initial NaCl Concentration (M)	0.5, 1, 2
Temperature (°C)	25, 50, 75
Electrolysis Time (hours)	0.5, 1, 2, 4, 8, 24

The selected NaCl starting concentration range was chosen to determine whether lower or higher NaCl concentrations minimise hypochlorite production. It is hypothesised that higher NaCl concentrations may lead to increased hypochlorite formation due to the greater availability of chloride ions for oxidation.

The broad temperature range was selected to investigate previous findings from literature suggesting that lower temperatures favour the formation of hypochlorite, while higher temperatures promote the production of chlorite and chlorate(9).

The electrolysis was ran for a long period of time to allow sufficient time for chloride ions to oxidise to higher oxidation states. Literature indicates that hypochlorite oxidises to chlorite relatively slowly, while chlorite reacts to form chlorate at a much faster rate(10).

3. Materials and Methods

3.1. Calibration Curves for Regression Models

Standard solutions of $NaClO_2$ and $Ca(ClO)_2$, with respective chlorite and hypochlorite concentrations of 12 mM, were prepared by dissolving the associated solids in DI water. To achieve solutions at set pHs of 8, 9 and 10, solutions of HCl and NaOH were used to acidify or basify the solutions respectively. To make the set of calibration curves Ultraviolet–visible (UV-Vis) spectroscopy analysis (Shimadzu, UV-2600) was used to find a relationship between absorbance and concentration of the active anions, for the wavelength range 400-200 nm. This was completed on a dilution series, for all standards, at concentrations 12 mM, 6 mM, 3 mM 1.5 mM, 0.75 mM and 0.375 mM.

3.2. Electrochemical Set Up and Analysis

To carry out the electrolysis of NaCl, standard solutions for the chosen concentration variables were made (0.5 M, 1 M, 2 M). The electrolytic cell was placed in a bath of glycerine oil, which was used to maintain the temperature of the experiment at the desired values (25°C, 50°, 75°C). The mesh RuO_2 coated titanium anode (16 cm^2), titanium cathode (40.50 cm^2) and silver reference electrode were all connected to the power supply and submerged into the NaCl solution, shown in Figure 1. The cathode was only submerged to half of the depth of the anode to ensure the correct current density of 49.66 mA/cm^2 was achieved for the experimental runs. The electrode spacing was 5.5 cm, as this has been previously optimised, like the other constant parameters(11). Each electrolysis experiment was run for 24 hours and tested at various time intervals (0.5, 1, 2, 4, 6, 8, 24 hours). For each time point a 500 µL sample of the electrolyte was taken and diluted with DI water by factor of 10. UV-Vis analysis was then carried out and the pH of the samples tested. The absorbance values from the UV-Vis were used to estimate the concentrations of the chlorite and hypochlorite ions in the electrolyte.

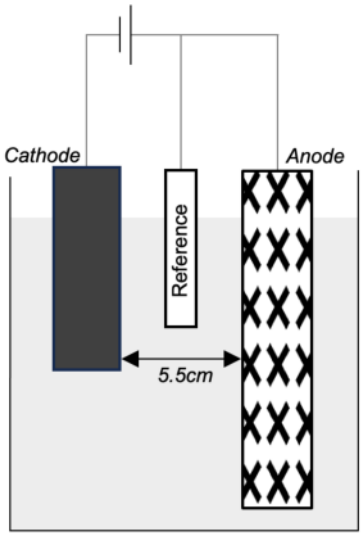


Figure 1: Schematic of Electrolysis Set Up.

3.3. Box Behnken Analysis

To investigate the effects of sodium chloride concentration, temperature and electrolysis time on the production of hypochlorite, Box Behnken Design was implemented for analysis using the software Minitab. This design is a type of response surface methodology (RSM) used for optimising processes involving three or more factors.

The three factors were tested at a low, medium and high level. The design featured 15 experimental runs,

including three replicates at the centre point to estimate experimental error and ensure model reliability. Each run was conducted in an undivided electrochemical cell, with the response variable being the concentration of hypochlorite produced, which was quantified using the linear regression model.

4. Results and Discussion

4.1. Regression Analysis

For both the chlorite and hypochlorite ions it was demonstrated that there is a linear relationship between concentration of the ions and absorbance in the UV-Vis spectrum. From this, the original plan was to use a multiple linear regression model to analyse the concentrations of both the chlorite and hypochlorite ions, which appeared to have distinctly different characteristic peaks at 260 nm and 292 nm respectively(12). This model included the absorbances at various characteristic peaks within the region for each of the standards made. However, after further initial investigations this was deemed to no longer be a viable approach. This was due to inconsistent results given when testing the model, even after multiple layers of attempted improvements. For example, when testing mixtures of known concentrations, the model did not return the expected concentrations of the solutions. Therefore, a transition was made to use a simple linear regression model at hypochlorite's characteristic peak 292 nm instead, which focuses only on the concentration of the hypochlorite ions in the electrolyte. This led to a change in the objective of the investigation to concentrate on finding the optimal conditions to decrease the overall concentration of the hypochlorite ions in the electrolyte.

It was discovered that the pH had an effect on the absorbance of the hypochlorite ions, from the UV-Vis analysis of the standard solutions at varying pHs,

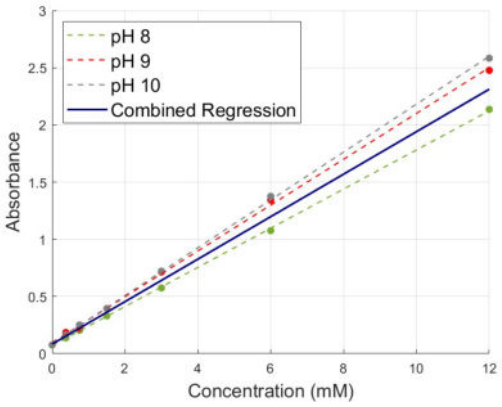


Figure 2: Linear regression relationship of concentration of hypochlorite ions against absorbance in the UV-Vis spectrum, at 292 nm varying the pH.

which was important to consider due the possible range of pHs in the electrolyte. The largest difference was seen in the pH 8 solution, shown in Figure 2, it is expected this decrease in the absorbance is due to the hypochlorite ions having an equilibrium relationship with hypochlorous acid (HOCl) in solution, shown both in Equation 6 and Figure 3.

$$\mathrm{HOCl} \rightleftharpoons \mathrm{H^+} + \mathrm{ClO^-} \tag{6}$$

At pH 8, a proportion of the ions exist as hypochlorous acid which has a different characteristic peak and a much lower absorbance(12), explaining why it was not seen on the UV-Vis spectra and why it would cause the peak to be diminished. This led to a decision being made to use a combined model of the pH 8 and pH 9 linear regressions because it was seen that the pH of the electrolytes tested fell within the range of 8 to 9.5. Whilst this model didn't necessarily give fully accurate values when it was tested, it was concluded the trend it would show was still correct. It was also decided that for consistency it was better to use one model for the analysis of the electrolyte rather than different models depending on the pH as there was not enough data at intermediate pHs.

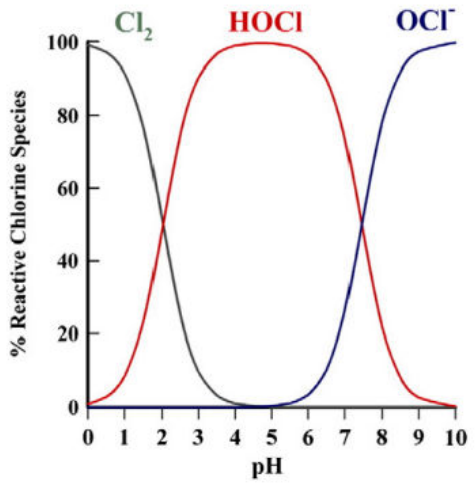


Figure 3: Diagram demonstrating thr relationship between pH and the concentrations of hypochlorite and hypochlorous acid in equilibrium.(13)

The R^2 value for this combined model (R^2=0.9855) indicated a strong linear relationship. The equation of the line was calculated as:

$$y = 0.1856x + 0.0844 \tag{7}$$

This was used by inputting the measured absorbance of the electrolyte at 292 nm for the y value and then concentration could then be found by solving for the x value, and multiplying by 10 to account for the dilution.

4.2. Electrolysis Analysis

The UV-Vis spectra did not indicate the presence of chlorite ions in the electrolyte, possibly because the chlorite ions were oxidised directly to chlorate ions. Hypochlorite initially reacts to form chlorite through a relatively slow first step. Once formed, chlorite undergoes a further reaction to produce chlorate, which proceeds at a faster rate (8). Chlorate is not active in the UV-Vis spectrum, meaning it cannot be identified using this method. Also, the chlorite ions have a lower absorbance than the hypochlorite ions, so it is possible if the concentration of the hypochlorite ions is much higher than the chlorite, it could mask the chlorite peak in the UV-Vis spectra(14). This further supported the notion to change the aim of the project to solely optimising the process to minimise the concentration of the hypochlorite ions present in the electrolyte.

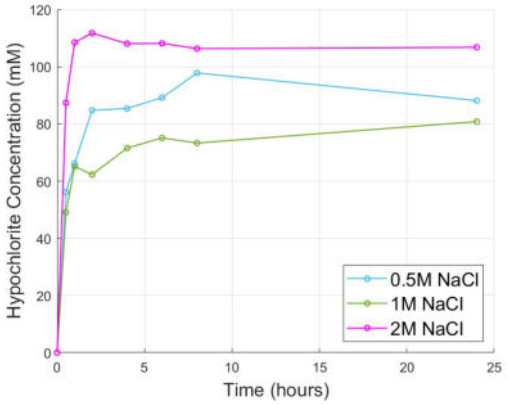


Figure 4: Plot of hypochlorite concentration (mM) over time at 25°C for various initial NaCl concentrations.

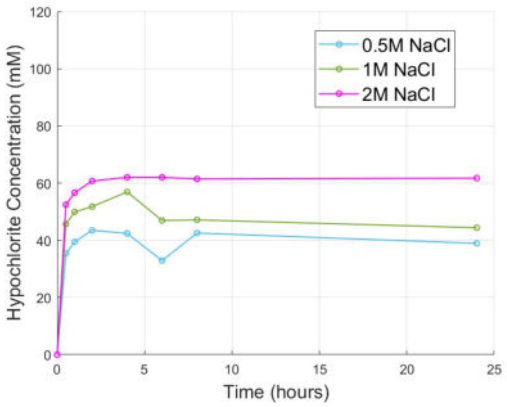


Figure 5: Plot of hypochlorite concentration (mM) over time at 50°C for various initial NaCl concentrations.

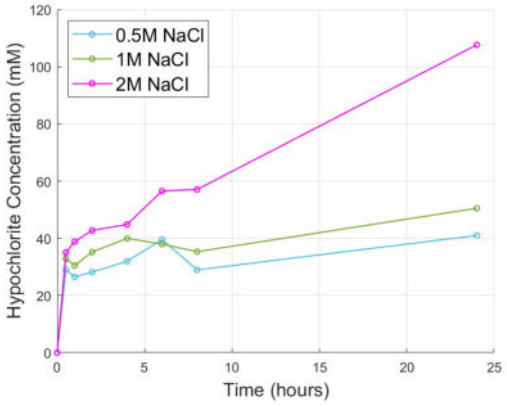


Figure 6: Plot of hypochlorite concentration (mM) over time at 75°C for various initial NaCl concentrations.

The general trend was similar across the various experiments of an initial increase in the concentration of hypochlorite ions followed by a plateau after 8 hours, shown for all groups in Figures 4- 6. The pH of the electrolytes remained stable over the 24 hour period with slight variation between experiments, but all remained within the range of pH 8 to 9.5.

It is clear to see from the data that increasing the time of electrolysis decreases the rate of formation of the hypochlorite ions, this was expected as throughout the running time the availability of chloride ions to be oxidised decreases as they are converted to hypochlorite reaching a point of saturation.

It is hard to draw to a precise conclusion for the relationship between the NaCl starting concentration and the concentration of the hypochlorite ions formed. The experiments at 50°C suggests a linear relationship considering the 24 hour readings, which would be a logical approach as the more available chloride ions there are at the beginning to oxidise, the more hypochlorite ions you would expect to be formed at the end of the experiment. The 25°C group contradicts this, but it should be noted that for this set, the temperature control was difficult due to the reaction itself heating up beyond 25°C to temperatures around 33°C. This deviation from the trend at small temperature changes could be due to hypochlorite's instability, at temperatures around 25°C to 35°C it is more likely to decompose(15).

It is shown that the higher the temperature the lower the concentration of the hypochlorite ions at the point where the plateau begins, implying the higher temperatures do not favour the production of hypochlorite ions, which agrees with what has been said in literature(9). The plateau region could suggest that the rate of formation of hypochlorite ions is equal to the rate of formation chlorite and chlorate ions , which may be the case where no drop from the plateau is seen.

The most promising result was seen in the 50°C group, as there was no further increase after the plateau in the 2 M NaCl concentration and for both the 1 M and 0.5 M NaCl concentrations there was a slight decrease. Which signifies a possibility of the presence of chlorite in the solution, as if the ions are being oxidised further it would show as a decrease in the hypochlorite ions. However, this decrease lies within the fluctuations seen at the previous time points, this potentially does not confirm that chlorite or chlorate ions are in the solution. Data points within the 8 hour to 24 hour period would be needed to verify this, ultimately paired with a testing technique suitable for chlorite and/or chlorate detection.

There is not much variation in the hypochlorite concentration when comparing the 50°C and 75°C variables for each NaCl concentration. This verifies that the lower temperature 25°C is not going to achieve the aim of this project, which suggests more research into the effect of the higher temperatures around 50°C and above 75°C needs to be done.

Whilst the data point of the 2 M NaCl at 75°C might seem like an outlier due to experimental issues, such as evaporation of the electrolyte; it is possible that this is an indication that the saturation point has in fact not been reached. If the region after the initial increase is considered in Figure 6, it appears to show a linear relationship between time and the increasing hypochlorite concentration. If this is the correct correlation it would lead to the conclusion that the experiment needs to be ran for a longer time period to be able to effectively see the saturation of the hypochlorite ions and the sequential drop as the ions are further oxidised. Since the oxidation of the ions occurs at the anode, it is possible within the system once the ions are initially oxidised, they leave the anode, and to be further oxidised they need to return

back. This process might be considerably time consuming due to the competition against the other ions which could be more favourably oxidised, considering the oxidation of the hypochlorite to chlorite is slower than other reactions in the system(8).

4.3. Box Behnken Analysis

The Box-Behnken Method was used to analyse the effects of NaCl concentration, temperature, and electrolysis time on hypochlorite concentration. This statistical approach evaluates linear, quadratic and combined effects of the factors on the response variable, which in this case is hypochlorite concentration. Linear effects are the direct, independent impact of each factor on the response. Quadratic effects describe the non-linear relationships, identifying curvature in the response surface Combined effects examine interactions between two or more factors, revealing how their simultaneous variation impacts the response. 3D plots of the effects of the factors on hypochlorite concentration can be seen in Figure 7- 9.

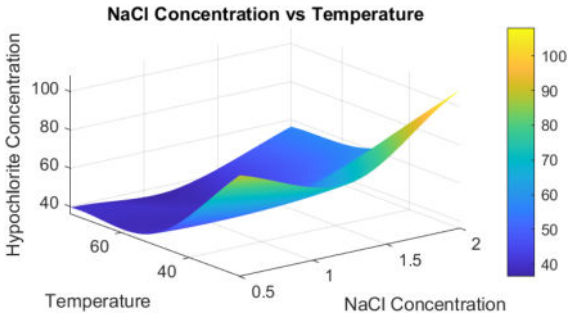


Figure 7: The combined effect of NaCl concentration and temperature on hypochlorite concentration

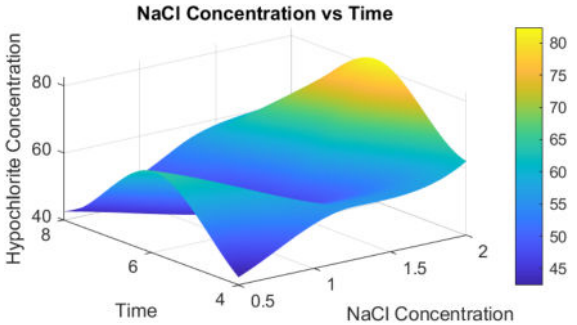


Figure 8: The combined effect of NaCl concentration and time on hypochlorite concentration

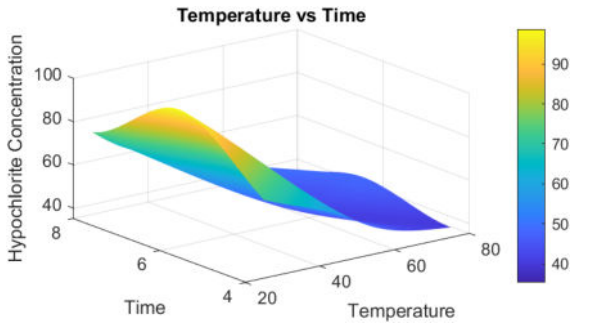


Figure 9: The combined effect of temperature and time on hypochlorite concentration

The factors were tested at three levels (low, medium, and high) with evenly spaced intervals, shown in Table 2. Although the statistical design specified the midpoint for NaCl concentration as 1.25 M, the actual experimental concentration used was 1 M, as the experiments were initiated before applying the Box-Behnken method for analysis.

Table 2: Factor levels for NaCl concentration, Temperature, and Electrolysis time.

Level	**NaCl Conc. (M)**	**Temp. (°C)**	**Time (hours)**
Low	0.5	25	4
Medium	1.25	50	6
High	2.0	75	8

The analysis revealed a statistically significant positive linear effect of NaCl concentration on hypochlorite concentration. The results in Table 3 show that increasing the NaCl concentration leads to higher hypochlorite production. This trend was consistent at 50°C and 75°C, where greater NaCl concentrations correlated with increased hypochlorite levels.

Temperature exhibited a highly significant negative linear effect, with a p-value of 0.000 and a coefficient of -21.37. This indicates that higher temperatures significantly decrease hypochlorite concentration, highlighting temperature as a key parameter for optimising the bleaching process.

Electrolysis time did not show a statistically significant linear effect on hypochlorite concentration as seen from the values in Table 3, suggesting it remains relatively constant within the tested range. However,

changes in hypochlorite levels observed at the 24 hour mark indicate that longer durations may be required to capture meaningful trends and fully understand time-dependent behaviour.

Table 3: Linear Effects of Factors on Hypochlorite Concentration.

Parameter	Coefficient	P-Value
Initial NaCl Concentration (M)	9.32	0.004
Temperature (°C)	-21.37	0
Electrolysis Time (hours)	-0.43	0.827

The factors also exhibit non-linear relationships with hypochlorite concentration, shown by their quadratic effects. NaCl concentration shows a statistically significant positive quadratic effect, shown in Table 4 reinforcing the increasing trend observed in its linear effect.

Temperature exhibits a complex relationship with hypochlorite concentration, as indicated by the statistically significant negative linear coefficient and positive quadratic coefficient. The negative linear term suggests that, overall, increasing temperature reduces hypochlorite concentration. However, the positive quadratic term introduces a curvature to this relationship, indicating that the rate of change in hypochlorite concentration is not constant and may vary depending on the temperature.

Electrolysis time displays a negative quadratic effect, though it is not statistically significant. Both the linear and quadratic effect are not statistically significant and this may be because of the selected time range for the analysis. To investigate the effect of electrolysis time better, selecting different time intervals that cover regions of variations may be of benefit.

Table 4: Quadratic Effects of Factors on Hypochlorite Concentration.

Parameter	Coefficient	P-Value
Initial NaCl Concentration^2 (M)	12.58	0.006
Temperature^2 (°C)	15.5	0.002
Electrolysis Time^2	-5.76	0.088

The interactions between NaCl concentration, temperature, and electrolysis time were found to have negligible effects on hypochlorite concentration, as indicated by the high p-values seen in Table 5. These results suggest that the combined variation of these factors does not significantly influence hypochlorite production within the tested range.

Table 5: Combined Effects of Factors on Hypochlorite Concentration.

Parameter	Coefficient	P-Value
NaCl Conc. x Temperature	-0.5	0.855
Temperature x Time	-1.62	0.563
NaCl Conc. x Time	-0.18	0.948

5. Conclusions

5.1. Conclusion

This study investigates the optimisation of key parameters influencing electrochemical bleaching, with a particular focus on minimising hypochlorite formation. Higher temperatures lead to a lower concentration of hypochlorite ions formed within shorter electrolysis times. Further investigation of this key parameter is needed to determine whether this trend continues at even higher temperatures. The initial sodium chloride starting concentration had a large effect on the concentration of the hypochlorite ions. The trend seen, an increase of formation of hypochlorite ions at higher initial sodium chloride concentrations, aligned with expectations, as the greater availability of chloride ions at the beginning means more oxidation products can be formed. Longer electrolysis times resulted in a decrease in the rate of formation of hypochlorite likely due to the subsequent oxidation of chlorite to chlorate, which reduces the hypochlorite concentration. This suggests that prolonged electrolysis promotes an increase in the formation of higher oxidation products.

5.2. Importance

These findings represent a promising first step toward developing a bleaching process that minimises

dioxin production. Eliminating dioxins from pulp mill effluents offers substantial benefits, including reducing exposure to these harmful compounds, which can cause cancer, suppression of the immune system, and reproductive issues. Furthermore, it mitigates long-term environmental contamination, as dioxins are highly persistent, with half-lives of 7 to 11 years. They also tend to bioaccumulate in the food chain, particularly in the fatty tissues of animals which has many additional consequences. Finally, a dioxin-free process could lower the burden on wastewater and solid waste treatment systems in pulp mills, presenting economic and environmental advantages(16).

5.3. Outlook

Considering the future implications of this project it is important to first look at the short-term improvements that can be made. Validation of the results should be completed by including more data points, for example in the 8-24 hour period, to identify whether the concentration does in fact plateau over this time, and at more frequent time intervals to see if the minor fluctuations are simply due to experimental error or if there is a significance to it. This could be related to the fluctuations seen in the pH values; this encourages the notion to explore the effect of pH further. To improve the regression model pH should be incorporated, as it has been shown that this affects the absorbance of the ions(17). For this to be done more data would need to be collected at a broader range of pHs to be able to find a correlation between these factors which could be used in a reformed regression model. One paper suggests there are two separate regions of linearity, one below pH 7 and then a different linear relationship from pH 8(17), this implies it could be possible to incorporate a pH relationship into the testing model. This would allow for a more accurate measurement of the concentration of the hypochlorite ions in the electrolyte.

Whilst it has been hypothesised that there are higher oxidation products present, inevitably further analysis needs to be conducted to verify this. One way this could be done is by using Ion Chromatography which has been shown to be an effective detection technique for both chlorite and chlorate as they have distinctively different retention times(18). Another part of this testing would include checking that dioxins were not present in the electrolyte. There is limited literature to say whether chlorite and chlorate form dioxins, but due to them being less reactive, because of their chemical structures(19), it was theorised they would be less likely to form dioxins. This analysis would be required to conclusively say that this process has a less negative impact.

Following on, it would be beneficial to carry out a full investigation into the bleaching power of these ions, and how they affect the stability of the biomass as it is known that the hypochlorite ions can damage the cellulose, which is not ideal for the pulp quality. The effects of the other parameters tested for electrolysis optimisation should also be evaluated, as it is possible that whilst they are beneficial in this sense, they may have adverse effect on the biomass. Both pH and temperature influence the bleaching capability and production of dioxins(2), so this a secondary area to optimise as well.

Finally, before this method of bleaching biomass could be applied to an industrial scale it is important to consider an economics analysis. For example, it has been suggested that increasing the temperature is favourable, however, this would cause an increase in both the capital and operating expenditure, which would have a negative impact on the economics of the process.

Acknowledgements

The authors of this report would like to express their gratitude to Ruihua Luo and Amir Dezashibi for their invaluable support and guidance throughout the course of this project.

References

[1] E.-S. Z. El-Ashtoukhy, "Electrochemical bleaching of kraft bagasse pulp using a new cell design," *The Canadian Journal of Chemical Engineering*, vol. 88, no. 1, pp. 126–132, 2010.

[2] O. o. T. A. U.S. Congress, "Technologies for reducing dioxin in the manufacture of bleached wood pulp," U.S. Congress, Tech. Rep., 1989.

[3] W. Klunklin, S. Hinmo, P. Thipchai, and P. Rachtanapun, "Effect of bleaching processes on physicochemical and functional properties of cellulose and carboxymethyl cellulose from young and mature coconut coir," *Polymers*, vol. 15, no. 16, 2023.

[4] P. Jour, K. Halldén, and E. Wackerberg, "Environmental systems analysis of alternative bleaching sequences with focus on carbon footprint," in *Proceedings of the ABTCP 2013: The 46th ABTCP International Pulp and Paper Congress*, ABTCP, Ed., ABTCP. Sao Paulo, Brazil: AkzoNobel, October 8–10 2013.

[5] S. of the Stockholm Convention, "Guidelines on best available techniques and provisional guidance on best environmental practices," United Nations Environment Programme, Tech. Rep., 2008.

[6] PulpPaperMill, "Hypochlorite bleaching," www.pulppapermill.com/hypochlorite-bleaching/.

[7] C. Schreiber, E. Zapp, C. R. de Aguiar, and P. Brondani, "Cotton fibers bleaching through in situ electrochemical generation of oxidant agents," *Research, Society and Development*, vol. 10, p. e318101018928, 08 2021.

[8] L. J. J. J. L. R. CZARNETZKI, "Formation of hypochlorite, chlorate and oxygen during nacl electrolysis from alkaline solutions at an ruo2 [tio2 anode," *JOURNAL OF APPLIED ELECTROCHEMISTRY*, vol. 22, no. 4, pp. 315–32, 1992.

[9] L. Czarnetzki, "Aspects of electrochemical production of hypochlorite and chlorate," Phd Thesis 1 (Research TU/e / Graduation TU/e), Chemical Engineering and Chemistry, 1989.

[10] M. Lister, "Decomposition of sodium hypochlorite: The catalyzed reaction," *Canadian Journal of Chemistry*, vol. 34, pp. 479–488, 02 2011.

[11] M. Saleem, M. H. Chakrabarti, D. B. Hasan, M. S. Islam, R. Yussof, S. A. Hajimolana, M. A. Hussain, G. M. Arifuzzaman Khan, and B. S. Ali, "On site electrochemical production of sodium hypochlorite disinfectant for a power plant utilizing seawater," *International Journal of Electrochemical Science*, vol. 7, no. 5, pp. 3929–3938, 2012.

[12] N. Kishimoto, "State of the art of uv/chlorine advanced oxidation processes: Their mechanism, byproducts formation, process variation, and applications," *Journal of Water and Environment Technology*, vol. 17, no. 5, pp. 302–335, 2019. [Online]. Available: https://www.jstage.jst.go.jp/article/jwet/17/5/17_19-021/_article

[13] U. D. of Agriculture, "Hypochlorous acid handling," U.S. Department of Agriculture, Tech. Rep., 2015.

[14] Y.-C. Chu, "Analysis of mixtures of chlorine, chlorine dioxide and chlorite," Ph.D. dissertation, Technische Universiteit Eindhoven, 1940.

[15] E. C. Agency, "Sodium hypochlorite," European Chemicals Agency, Tech. Rep., n/a.

[16] World Health Organization, "Dioxins and their effects on human health," World Health Organization, Fact Sheet, 2024, accessed: 2024-12-12. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/dioxins-and-their-effects-on-human-health

[17] T. Zayas, M. Vega, G. Soriano-Moro, A. Handal, M. Morales, and L. Salgado, "Electrogeneration of hypochlorite ions using a dimensionally stable anode-type (ti/ptpd(10%)ox) electrode," *Journal of Electrochemical Science and Technology*, vol. 15, no. 2, pp. 268–275, 2024.

[18] H. Fatunmbi and B. Fatunmbi, "Determination of chloride, chlorite and chlorate in drinking water samples by high performance ion chromatography," *SMT Separation Journals*, 2018.

[19] S. T. R. Thomas Myers, Keith B. Oldham, *Holt Chemistry*. Holt Rinehart & Winston, 2003.

Superhydrophobic Cotton and PU Sponge Modified by Silica Nanoparticles Coating with Octadecyltrimethoxysilane

Yijie Zhou and Zhuocheng Kong
Department of Chemical Engineering, Imperial College London, U.K.

Abstract

Among the five selected materials, cotton and PU sponges exhibit a superhydrophobic feature ($\theta_{CA,cotton} = 154°, \theta_{CA,PU} = 150.9°$) modification. The material modification process involves only two steps: pre-treating the materials with one molar sodium hydroxide solution at 60℃ for 20 minutes, followed by a simple solution-immersed modification method by using TEOS and ODTMS under an alkaline chemical environment to enable the bonding of silica nanoparticles with long alkyl chains on the pre-treated materials. A series of performance assessments including oil and water sorption ability tests, continuous and stationary oil-water separation, chemical and acid-base durability tests and reusability tests, showed excellent results. Referring to the result of single vegetable oil-water separation via NanoDrop, the separation efficiencies exceeded 99.75%. The reusability tests indicate that both modified materials maintain an oil-water separation efficiency of over 90% after 15 cycles. Continuous oil-water separations via flowmeter not only illustrate the operation preference for reducing water content in the final liquid collected but also highlight several possible modification improvements on the material, such as pore size design. Durability tests show that both modified cotton and PU sponges demonstrate special wettability under most conditions, and perform well except in extreme pH values, particularly under highly alkaline conditions. In conclusion, the modified cotton PU sponge can provide a cost-effective and efficient option for various oil-water separation applications.

1. Introduction

Hydrophobicity and superhydrophobicity are common in nature, for example, in the leaves of some wetland species[17]. Hydrophobicity usually refers to a surface with a water contact angle θ > 90º, and when the WCA θ > 150º[9], the material is superhydrophobic.

With the development of human society, oily wastewater from different industries such as petroleum refining, the petrochemical industries and the catering industries may not only affect human health but also endanger aquatic lives. To treat the oily wastewater and meet the emission standards, several traditional methods such as evaporation separation, gravitational separation, filtration and flotation are commonly applied[11]. However, these methods are either time-consuming, energy-intensive or costly, and emulsification of oil and water in certain industries makes oil-water separation even more challenging. To improve the situation, the special wettability of hydrophobic materials promotes a possible solution path.

Hydrophobic modification can take place in various kinds of materials, for example, polymers[7], wood, plastic, and some fabrics like cotton and cellulose[8]. A material with a more porous structure is preferred to enlarge the oil sorption capacity of each unit of hydrophobic materials. To produce hydrophobic materials, existing hydrophobic modification methods including layer-by-layer coating, laser modification, solution immersed method and sol-gel method[2] are utilised. However, these methods might be expensive, complex and time-consuming and they may not be suitable for every kind of material.

Among these raw materials and modification methods, five kinds of materials are chosen for further research in this paper, as these materials usually have a porous structure that results in a high oil sorption ability. Additionally, these materials are cheap and in large quantities. The method selected is a one-pot solution immersed method with the co-hydrolysis and co-condensation of octadecyltrimethoxysilane (ODTMS) and tetraethyl orthosilicate (TEOS)[1] in a basic chemical environment at room temperature and pressure.

After modification, cotton and PU sponge show a superhydrophobicity property in WCA measurement ($\theta_{CA,cotton} = 154°, \theta_{CA,PU} = 150.9°$) and excellent oil sorption in the experiment; hence, the nanostructure of two kinds of materials is investigated via SEM characterisation. Furthermore, several experiments are completed to assess the oil-water separation efficiency, continuous oil-water separation efficiency, reusability, chemical durability, and acid-base durability of the modified cotton and PU sponge.

The reaction mechanism in the modification process supports the results. The strong bonding formed between hydroxyl groups on the NaOH pre-treated raw materials with silanol groups and further long alkyl chain attachment generated from the co-hydrolysis and co-condensation of TEOS and ODTMS in a basic chemical environment. This bonding provides low surface energy to the modified materials which resulting in its superhydrophobicity properties. The robust chemical bonding between silica nanoparticles and the material enables it to perform well in durability and reusability testing.

2. Materials and Methods

2.1. Materials

Ethanol (≥ 99.8 %), acetone (≥ 99.8 %), hydrochloric acid (HCl, 37 %), sodium hydroxide (NaOH) pellets, sodium hydroxide-ammonia solution [NH_3(aq), 28 %] and ODTMS (90 %) were obtained from VWR (Lutterworth, UK). Cotton wool, tetraethyl orthosilicate (TEOS, ≥ 99 %), methylene blue solution, anhydrous magnesium chloride ($MgCl_2$), sodium chloride (NaCl, ≥ 99.0 %) and calcium chloride dihydrate ($CaCl_2 \bullet H_2O$) were purchased from Sigma Aldrich (Gillingham, UK). Cyclohexane (≥ 99.8%) is obtained from Fisher

Scientific (Loughborough, UK). Deionized (DI) water was supplied by a PURELAB Chorus 1 (ELGA Lab Water) purification system. Rapeseed oil, sunflower oil, olive oil and sesame oil were obtained from a local supermarket.

All chemicals and reagents were used without further purification.

2.2. Sample Pre-treatment

The 1.0 M NaOH solution was prepared by dissolving sodium hydroxide pellets in deionized (DI) water and preheating the solution to 60 °C. Cotton, PU sponge, cellulose sponge, wool and fabric were used for modification. Samples were fully immersed in the heated solution and etched for 20 minutes to enhance surface modification. The sample was retrieved and thoroughly rinsed five times with DI water after etching to ensure the removal of residual alkali and to achieve a neutral pH in the supernatant. The pretreated sample was then dried overnight.

2.3. Modification of Cotton and PU Sponge

To functionalize the sample with silica nanoparticles and form a hydrophobic shell, a solution was prepared by mixing ethanol (100 mL), deionized water (10 mL), tetraethyl orthosilicate (TEOS, 6 mL), and octadecyltrimethoxysilane (ODTMS, 3 mL). The mixture was stirred at 500 rpm for 5 minutes at room temperature. The pretreated sample was then dropped in the solution and stirred for an additional 3 minutes under the same conditions. Ammonia solution (10 mL) was subsequently added dropwise to initiate the condensation reaction and was allowed to proceed for 8 hours. After the reaction, ethanol or acetone is used to wash the silica-modified sample to remove residual reagents. The samples were left overnight, then compressed air was applied to remove any loose particles adhering to the surface.

2.4. Characterisation of Modified Materials

Cotton and PU sponge morphology was investigated using a scanning electron microscope (SEM) at 5 kV, namely the Leo Gemini 1525 (Carl Zeiss AG, Germany). The samples were coated with gold (15 nm) before imaging.

The water contact angle (WCA) is one of the main pieces of evidence that can indicate the hydrophobicity of materials. A 590 Goniometer measures the ramé-hart instrument Co. To start the measurement, the needle connected to the water source should be positioned in the middle of the image and the camera focuses on that. For the preparation of the testing sample, a piece of 1cm×1cm material is attached to the objective slide by double-side tape, press the material to make it flat enough to carry out the measurement, light intensity can also be adjusted to provide a clear position of three-phase contact point.

After setting up the baseline, a water droplet with a measurable size can be dropped on the modified material, it is supposed to inject around 6μl DI water from the needle, move the needle downward until it reaches the surface of the material and lift the needle to leave the water droplet on the materials' surfaces. However, in practical operation, this is extremely difficult to achieve because the needle is much more hydrophilic than the material, then a pipette is used instead. After dropping a water droplet, the WCA measurement will be operated by one horizontal line and two vertical lines, and two intersections of these lines should be placed at the three-phase contacting point of the water droplet and the material. Then the measurement will be operated by the software, and it is vital to confirm the blue line, which is the tangent initiating from the intersections mentioned above, staying at the same position during the whole process otherwise, the measurement will not be accurate. After the measurement, a data sheet including left, right and mean WCA numbers can be exported, it is necessary to keep the difference between left and right WCA no more than 5 degrees to obtain a more reliable mean WCA of the modified material.

2.5. Sorption Performance

2.5.1. Water Sorption

The modified samples were immersed in DI water and simulated hard water for an hour to assess their water sorption properties. The hard water was prepared using 0.1 M solution of CaCl2 dihydrate. After removal, the samples were allowed to drain until no visible water droplets remained, and then mass change was measured. The comparison of pre-treated and raw cotton and PU sponge was considered by using the same method.
The water sorption ability is calculated by

$$Sorption\ ability\ \left(\frac{mg\ oil}{mg\ material}\right) = \frac{m_2 - m_1}{m_1} \qquad (1)$$

where m_1 and m_2 is the mass of samples before and after water sorption, respectively.

2.5.2. Oil Sorption Ability

To evaluate the maximum oil sorption ability of the samples, the modified sample was immersed in pure oil for an hour. It was later removed and allowed to drain until no visible oil droplets fell. The maximum sorption ability was calculated based on the mass difference before and after oil sorption, the same as water sorption ability in Equation 1.

2.6. Oil-Water Separation

2.6.1. Oil-Water Separation Efficiency

Two samples were used for each test: A primary sample M1 is designated for repeated oil sorption and subsequent washing for regeneration. Residual sample M2 (which is around a quarter to half of the M1) is utilized to absorb any residual oil remaining after M1 treatment.

The primary sample M1 was immersed in the oil-water mixture for 5 minutes to absorb the oil. Afterwards, M1 was removed, and its mass was recorded as M1′. M2 was carefully held with tweezers and used to absorb the remaining oil on the water surface until no visible oil sheen was observed. The mass after was recorded as M2′.

The oil sorption efficiency η for each cycle was calculated using the following equation:

$$\eta(\%) = \frac{M2' - M2}{M_{oil}} * 100\% \quad (2)$$

where M2 is the initial mass and M2′ is the final mass of M2 (g), M_{oil} is the mass of 3 ml oil.

2.6.2. Oil-Water Separation via Nanodrop

NanoDrop is the device that can reveal the concentration of targeted content with unsaturated carbon bonds in a mixture by the spectrum absorbance, hence these experiments are for four kinds of pure vegetable oil only. The first step is to rinse the cuvette with absolute ethanol and dried with compressed air, soft tissue is used to wipe the transparent sides of the cuvette and keep it crystal clear. Secondly, blank the spectrum of cyclohexane, which is the chosen organic solvent. To carry out this, 3ml of cyclohexane is injected into the cuvette. Although the data collected are in the wavelength range of 200-750nm, only 260nm is the desired data point, and the sign shows blanking is successfully done if the absorbance on 260nm is 0.

After the blanking, calibration is required to obtain the linear relationship between vegetable oil concentration and absorbance. 2-4µl, 4-14µl 10-20µl and 3-8µl oil is added to the cuvette respectively. To have a better mixing, a 0-3ml pipette is used to soak and inject the mixture inside the cuvette to accelerate the mixing. After the linear relationship between oil concentration and absorbance is obtained, the oil-water separation efficiency using the modified PU sponge and cotton is tested. Around 200mg of modified cotton and 100mg of modified PU sponge were placed inside a bottle with 3ml of vegetable oil listed previously that floated on a reasonable amount of coloured water for experiments. After 5 minutes, the modified materials are carefully taken out and 2.5ml of cyclohexane is added to the liquid left in these bottles and close the lid to avoid evaporation of cyclohexane while waiting for the mixing of cyclohexane and vegetable oils. After more than 5 minutes, the oil phase becomes uniform and is ready to test. A 10-1000µl pipette is used to transfer 500µl of liquid to the cuvette until the cuvette has 2ml liquid in it. Then the Nanodrop after blanking the absorbance of cyclohexane can provide the absorbance data for these vegetable oils, and the average of 5 absorbance data of each group can be used to calculate the concentration of vegetable oils by considering two oil phases mixed completely.

All the data recorded will be available on a data sheet in the Supplementary Information.

2.6.3. Flowmeter Test of Modified PU Sponge

Compared with cotton, PU sponge has a more porous structure that enables fluid to pass through the structure more easily. To carry out the test, the inlet tube of the flowmeter is inserted into the modified PU sponge, and a mixture of 250ml sunflower oil and 750ml of coloured DI water is used to test the oil-water separation efficiency over a period. The oil-water separation efficiency is observed by compositions of water and oil content collected at the outlet of the Flowmeter.

Several variables including the rotation per minute (RPM) of the magnetic stirrer, flowrates of the flowmeter, and the mixing state of the oil-water mixture are investigated, and experiments are categorized into different groups by these standards. The first set of experiments is based on altering the flow rate of the flowmeter by adjusting the RPM of its motor inside while mixing the sunflower oil and water with a magnetic stirrer around 100RPM. After considering the information obtained from results and progress during the operation with 20RPM, 15RPM, 10RPM, 5RPM and 2.5RPM, separation path optimisation is attempted by switching the flow rate of the flowmeter by monitoring the water and oil composition in the tube during the operation.

The second set of experiments is based on the RPM of the magnetic stirrer, which are 500RPM, 300RPM, 200RPM and 100RPM at the same flow rate. The result is also evaluated by the pictures taken during the experiments.

The final set of experiments is the oil-water separation without mixing, and the same experiments are designed by varying the flow rate of the flowmeter and separation path optimisation is also explored. However, after analysing the pictures taken at different times during the operation, further optimisation that combines non-mixed and mixed separation is considered.

2.7. Reusability and Durability

2.7.1. Reusability Evaluation

The recyclability of the modified cotton and PU sponge samples was assessed through repeated oil-water separation and regeneration cycles. Similar to the oil-water separation test, the process involved repeatedly absorbing and then regenerating the original sample M1.

M1 was cleaned to restore its functionality after each test. The sample was cleaned three times in cyclohexane to dissolve the absorbed oil, and then rinsed with acetone to eliminate any residual cyclohexane and dried using compressed air until the weight stabilized, indicating complete drying.

The regenerated M1 was reused for subsequent cycles. Tests were terminated until either the number of cycles reached 15 or the oil sorption efficiency dropped below 90%.

2.7.2. Durability Testing

The modified cotton and PU sponge samples were evaluated by assessing their resistance to varying pH levels and chemical conditions. Samples were immersed in the test solutions for 1 hour, followed by rinsing three times with deionized (DI) water to remove residual solutions. The rinsed samples were then air-dried at room temperature before measuring the water contact angle (WCA) to observe changes in surface hydrophobicity.

Acidic and basic solutions of pH 2, 4, 6, 8, 10, 12 and 13.73 (which is the pH of one molar sodium hydroxide solution) were used to test the pH resistance to evaluate the samples' tolerance to acidic and basic conditions. Chemical resistance was assessed by immersing the samples in 1 M solutions of $CaCl_2$,

$MgCl_2$, NaCl, NaOH, and HCl to examine their stability in the presence of common chemicals.

3. Results and Analysis

3.1. Modification Process

Figure 1 below illustrates the sequential modification process for achieving a superhydrophobic surface on the material. The procedure involves alkaline pretreatment and surface functionalization with silica nanoparticles.

During the pretreatment step, the material is immersed in a 1.0 M NaOH solution at a moderate temperature to eliminate surface lignin and impurities[14], thereby enhancing surface roughness and creating active sites for subsequent modifications. However, excessive exposure to the alkaline solution should be avoided to prevent fibre aggregation or structural degradation[5].

Silica nanoparticles are deposited onto the pretreated cotton surface through the modification process. During this process, condensation reactions may occur between the silanol groups of the silica nanoparticles and the hydroxyl groups on the fibres, forming covalent bonds between the two components. In the presence of water, ODTMS degrade irreversibly into siloxane polymers. Ammonia is used to promote the hydrolysis of silane, generating more reactive intermediates in which methoxy groups are replaced by hydroxyl groups[6]. These newly formed hydroxyl groups condense with the silanol groups and any remaining hydroxyl groups on the fibres, introducing a hydrophobic layer and enhancing the water-repellence of the surface. Additionally, each silane molecule in ODTMS can link with neighbouring silane molecules through siloxane bonds, forming a Si–O–Si bridge network that further stabilizes the hydrophobic coating[18].

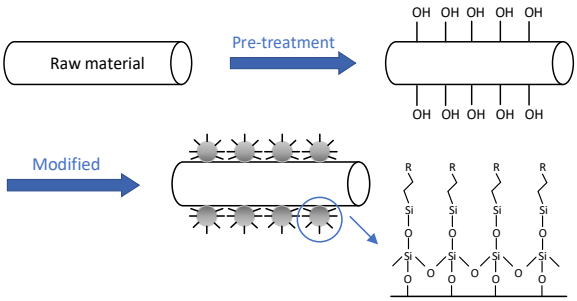


Figure 1. *Pretreated and Modification Process*

3.2. Superhydrophobicity of Modified Materials

From the result of past research, low surface energy can provide hydrophobicity[16] to the material surfaces and the term superhydrophobicity refers to a surface with a static water contact angle (WCA) over 150º[19]. This special type of wettability is very dependent on the surface topography, and surface topography is related to the surface roughness. This is because the surface of modified materials usually is not smooth. Under this condition, the Wenzel Model and the Baxter-Cassie Model can be applied to the situation and the mechanism is shown in Fig 4.

This theory can match the result from SEM scanning. Fig 2&3 indicates the surface of the modified cotton and PU sponge is coated with silica nanoparticles which do not spread on the surfaces uniformly hence creating surface roughness.

The Cassie-Baxter model shows the surface condition when the modified materials are not wetted or coated with liquid completely, hence the air pockets caused by surface roughness still can hold the water droplets on it and avoid being fully wetted. However, under the experiments of oil-water separation, low energy surface of modified materials will have higher selectivity towards low surface energy liquids, which as oil in oil-water separation. When sorption of oil takes place on the surface of the modified materials, the air pockets will be filled with the oil hence the feature of superhydrophobicity will be weakened, and water droplets will have the chance to contact with the material surface in depth refers to the Wenzel Model[20] in Fig 4. This might be able to explain the situation discovered in the water sorption, oil-water separation efficiency via mass difference and oil-water separation efficiency via Flowmeter. Specific scenarios and explanations will be provided in each related section respectively.

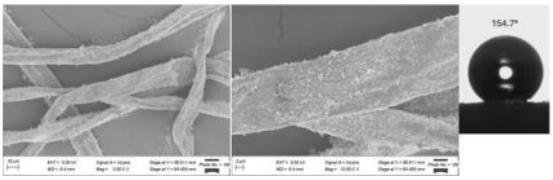


Figure 2. *SEM Images of Modified Cotton*

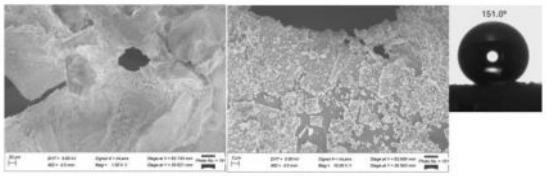


Figure 3. *SEM Images of Modified PU Sponge*

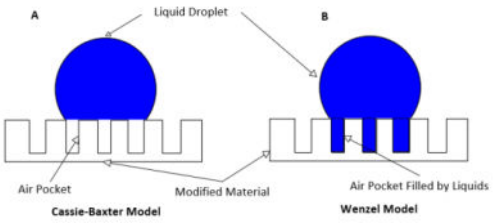


Figure 4. *Cassie-Baxter Model and Wenzel Model*

3.3. Material Selection

Before proceeding to further oil-water separation and durability testing, 5 kinds of materials are modified, including cellulose sponge, PU sponge, cotton, wool and one unknown fabric. The level of hydrophobicity of these materials is assessed by water sorption test, WCA measurement and oil sorption capacity. By considering the results of three experiments, PU sponge and cotton are chosen for the upcoming durability and oil-water separation exploration.

3.3.1. Water Sorption

Figure 5 presents the water sorption rates of raw and pre-treated cotton and PU sponge. It is evident that pre-

treatment significantly reduces the water sorption rate for both materials, with a more notable effect on cotton. Specifically, raw cotton absorbs 35 mg of water/mg cotton in an hour, which decreases substantially after pre-treatment. Similarly, PU sponge also shows a reduction in water sorption after treatment, though the change is less pronounced compared to cotton.

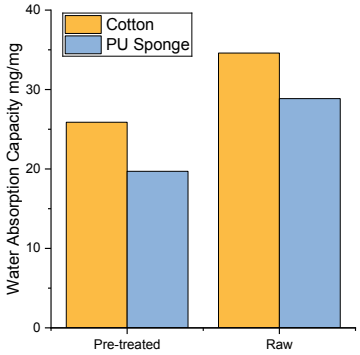


Figure 5. *Water Sorption Rate of unmodified samples*

Figure 6 compares the water sorption rates of cotton, PU sponge, cellulose sponge, wool, and fabric when exposed to DI water and simulated hard water prepared with 0.1 molar $CaCl_2$ solution with DI water. Among the materials tested, the cellulose sponge exhibits the highest water sorption rate, particularly in hard water. PU sponge and cotton both show higher sorption rates in hard water compared to DI water. In contrast, wool and fabric demonstrate almost constant sorption, with fabric showing the lowest rate across both water types. The interaction between the material structure and immersion in water weakens the superhydrophobicity of the materials, leading to some water sorption. Materials with more porous structures, such as cellulose sponge, PU sponge, and cotton, tend to absorb more water because water droplets become trapped within the pores. Notably, the sorption rate of fabric remains relatively stable, irrespective of water composition because of its thin layer and less porous structure.

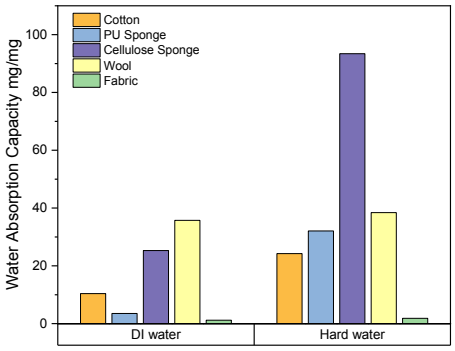


Figure 6. *Water Sorption Rate of modified samples*

3.3.2. WCA Measurement

From Fig 7, only modified cotton and PU sponges have a WCA of over 150º and reach the standard of superhydrophobic materials. Hence from the results of these tests, cotton and PU sponges are chosen for further exploration.

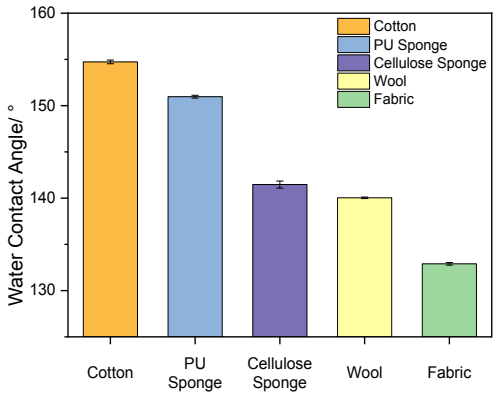


Figure 7. *Water Contact Angle of modified samples*

3.3.3. Oil Sorption Capacity

According to Fig 8, after 5 kinds of materials are soaked in rapeseed oil, sunflower oil which contains unsaturated fats and cyclohexane contains saturated fats for 1 hour. The absorbing ability is calculated, and the results show that modified hydrophobic materials can absorb more unsaturated fats than saturated fats[3],[13].

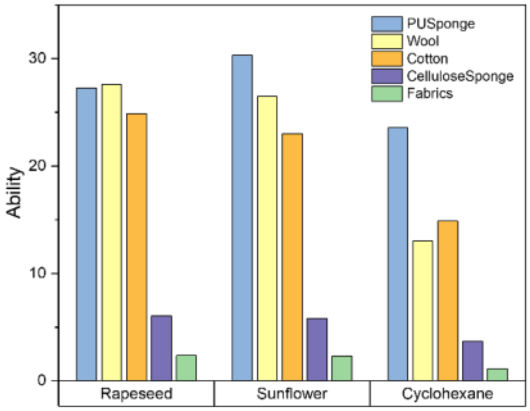


Figure 8. *Oil Sorption Ability*

This trend can be explained by the molecular structure of cyclohexane and other main unsaturated fats like linoleic fatty acid in sunflower oil. The Linoleic fatty acid is a C_{18} organic chemical that contains carbon double bonds and a hydroxyl group in its structure. These features enable linoleic fat to form a relatively strong binding with the material surface, and its long-chain structure allows the linoleic fat to adsorb on different activation sites on the surface of modified materials. Compared to linoleic fat, cyclohexane is a non-polar and small molecule, this may contribute to the weak binding strength and limited binding sites when absorbed by the modified materials.

By comparing the adsorption capacity of each oil among five kinds of modified materials, the modified PU sponge generally shows the largest oil capacity compared to the other four kinds of materials, which is around 25-30 times its original mass[12]. A high porosity structure increases the specific surface area of the modified PU sponge, and a larger specific surface area provides more possible sorption sites for the oil, hence resulting in a higher adsorption capacity.

3.4. Oil-Water Separation

3.4.1. Oil-Water Separation via Mass Difference

Figs 9 and 10 show the oil-water separation efficiency for single oil and mixtures of vegetable oil and organic oil, the modified cotton generally maintains a separation efficiency of 97-99%, which is better than 94-97% measured with the modified PU sponge.

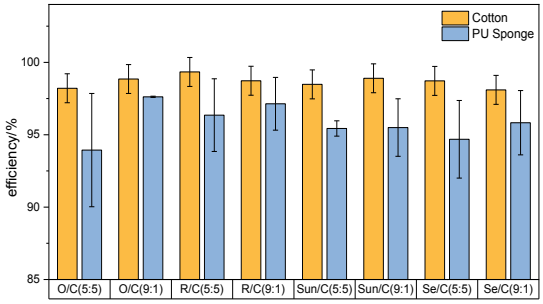


Figure 9. *Oil-Water Separation of Vegetable Oil Mixed with Cyclohexane. O, C, R stand for olive oil, cyclohexane and rapeseed oil separately. Sun and Se stand for sunflower oil and sesame oil respectively*

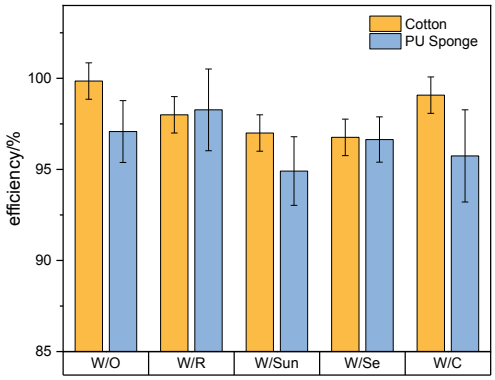


Figure 10. *Oil Sorption Efficiency of Vegetable Oil. W represents DI water*

Figure 11A shows how the oil-water separation is carried out by the materials. Unlike cotton, the PU sponges have a more rigid shape, which makes it difficult to fill up the liquid surface with the sponge and there is an uncovered region around it. The mass transfer rate of oil without contacting the material surface directly can be influenced, and the experimenting time is only 5 minutes, this difference might affect the results.

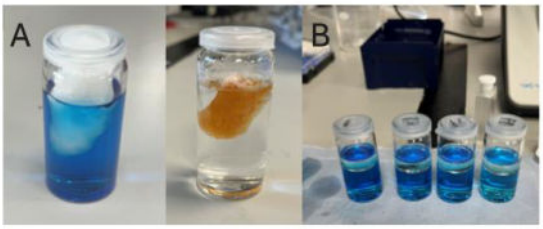


Figure 11. *Oil-Water Separation Methods. Left-hand side A shows the process of oil-water separation and right-hand side B shows the assessment using NanoDrop*

The state transition between Cassie-Baxter state to Wenzel state is mentioned in section 3.2 and can take place on the surface of the modified materials by absorbing the oil, and PU sponge has a larger proportion immersed in the liquid compared to cotton, so this might be able to explain the reason of the assessed PU sponge oil-water separation efficiency can be lower than cotton.

However, in Fig 10, a general trend of oil-water separation efficiency can be ranked as olive oil > rapeseed oil > sunflower oil > sesame oil in the single vegetable oil-water separation from cotton, and different fatty acid compositions in rapeseed, olive, sunflower and sesame oil may explain this[15].

The main two types of fatty acids in the four kinds of vegetable oils mentioned above are oleic and linoleic acid. Oleic acid is the main fatty acid found in rapeseed and olive oil, which is over 60%, and oleic acid and linoleic acid share the same composition that is around 40% in sesame oil, and linoleic acid has a composition of over 60% in sunflower oil. Compared to oleic acid, which is monounsaturated and linoleic acid is described as polyunsaturated, and from its structure, two carbon-carbon double bonds can be found. One more double bond indicates more intermolecular disorders may appear between linoleic acids than oleic acids. Referring to the relationship between surface energy and hydrophobicity in section 3.2, the increase in the level of disorder on the material surface can potentially increase the surface energy hence reducing the hydrophobicity of the material surface and resulting in the rank mentioned above.

The same trend can also be discovered in the vegetable oil-cyclohexane (9:1)-water separation efficiency from the PU sponge. However, in the mixed oil-water separation experiment, the composition of the oil mixture left in the container was not assessed due to technical limitations, so the 5:5 oil-water separation will not be discussed.

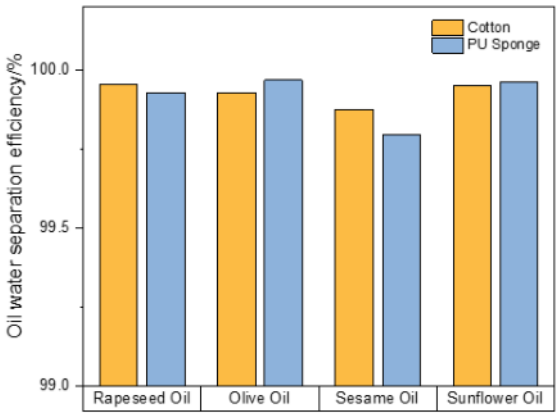


Figure 12. *Oil-Water Separation Efficiency via NanoDrop*

3.4.2. Oil-Water Separation via NanoDrop

The accuracy of the data collected via mass difference might be influenced by many reasons mentioned previously, hence NanoDrop is used to generate more accurate data through spectrum absorbance.

Figure 12 shows that an oil-water separation efficiency of over 99.75% is achieved in all sets of experiments with the modified cotton and PU sponge.

3.4.3. Oil-Water Separation via Flowmeter

Figures 13G, 13H, 13I and 13J show the composition of liquid collected in the outlet after operating for 20 minutes when the rotation speed of the magnetic stirrer

is decreased from 500RPM to 100RPM, the blue-coloured water content is reduced. Figures 13A-13F show the liquid mixture received at the end of the separation, and the composition of water content in the outlet is reduced with decreased flow rate. This scenario is explained by referring to the state transition mentioned in section 3.2 when the modified PU sponge is fully coated with oil, the special type of wettability is dismissed. Also, mixing and large pressure drop created by a high flow rate through the sponge's porous structure may result in encouraging more water content transport from the bulk water into the sponge instead of repelling water on the surface of the sponge.

Figures 13K to 13L are the result of separation optimization when the oil layer and water were kept stationary at first. Figure 13L is designed to fit the sequence of 10RPM-5RPM-2.5RPM without mixing through the whole process, which takes 35 minutes to complete and 13K has a sequence of 10RPM-5RPM-2.5RPM and starts to mix under 100RPM when operated with a flow rate of 2.5RPM, which takes 30 minutes to finish. The pictures also show that 13L's final liquid contained more water content than 13K. Figure 13M shows the sequence of 10RPM-5RPM-2.5RPM mixing throughout the process.

By comparing these three pictures, the oil collected with a non-mixing starting will be clearer because it contains less water content, and mixing may enhance the oil mass transportation when the oil concentration is low.

However, in this experiment, the technical limitations make the quantitative analysis of the water content in the outlet difficult, hence an exact time for flow rate switching will be hard to obtain, this might fully rely on observations about the water and oil composition in the tube. On the other hand, although purer oil can be separated when the oil layer is stationary, it is necessary to keep the main part of the sponge inside the oil layer, so the position of the sponge needs to be adjusted over time, which provides uncertainty to the experimental results.

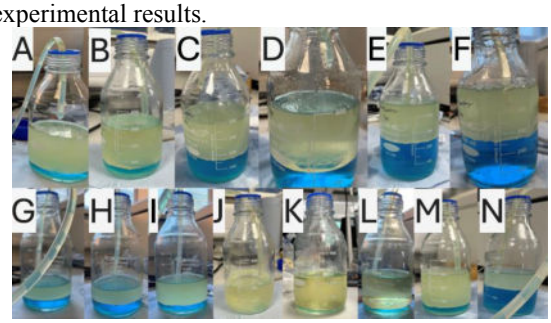


Figure 13. *Flowmeter tests with Various Conditions. A–F show the water and oil levels in the flowmeter test at flow rates of 2.5, 5, 7.5, 10, 15, and 20 rpm, respectively. G–J display the water levels at mixing rates of 500, 300, 200, and 100 rpm, respectively. K demonstrates the oil-water separation under the optimized path: 10 rpm, 5 rpm, and 2.5 rpm, with stirring maintained during the 2.5 rpm phase. L shows the water and oil levels under the same conditions as K but without stirring, while M represents the levels with continuous stirring across all states. N provides a comparison, showing the water level at 10 rpm with stirring for an unmodified PU sponge.*

3.5. Reusability and Durability

3.5.1. Reusability for Oil-Water Separation

A cycle performance test of cotton and PU sponge is conducted. Cotton showed a higher initial separation efficiency than the sponge, starting at 97.94%, while the sponge began at 95.7%. By the 15th cycle, cotton's efficiency drops to 93.2%, reflecting a decrease of approximately 4.7%. In contrast, the sponge demonstrated a more consistent performance throughout the cycles, with efficiency fluctuating within a narrower range, between 92% and 96%. Although the initial efficiency of the sponge is slightly lower than that of cotton, its stability over repeated cycles is better than cotton. After repeated regeneration with cyclohexane and acetone, the shape of the recycled cotton shrank significantly. However, the sponge maintains stable performance due to its resilience. Even after 15 cycles, its shape remained nearly unchanged.

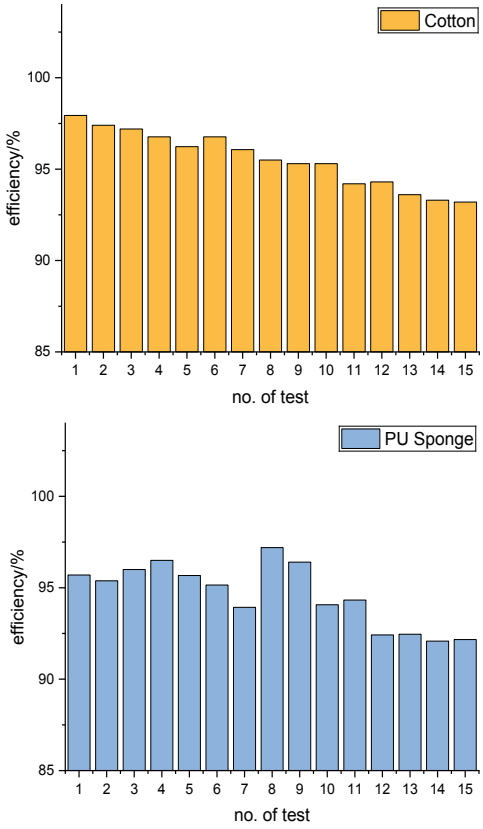


Figure 14&15. *Oil Sorption Efficiency of Cotton and PU Sponge through 15 cycles*

3.5.2. Chemical Durability via WCA Data

The chemical durability of modified cotton and PU sponge was evaluated by measuring their water contact angles (WCAs) after immersion in various chemical solutions. The results are summarized in Fig 16.

Both materials have superhydrophobicity (WCA > 150°) in their original state. However, after exposure to different chemical environments, notable differences in their durability were observed.

For cotton, exposure to NaOH resulted in the most significant decrease in WCA, dropping to 128.16° from 154.73°, indicating that the alkaline environment compromised its hydrophobic coating, which was much

greater than that of the other tested chemicals. At the same time, cyclohexane also destroyed the hydrophobic layer on the surface of modified cotton (138.58°). PU sponge demonstrated better chemical stability across most conditions.

Overall, the PU sponge exhibits superior chemical durability compared to cotton, maintaining a more consistent superhydrophobic performance across different chemical environments.

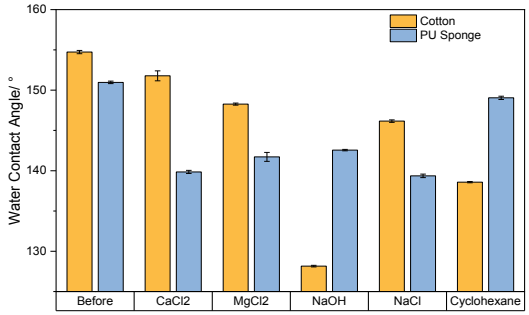


Figure 16. *Water Contact Angle after Chemical tests*

3.5.3. Acid-Base Durability via WCA Data

The stability of cotton and PU sponge under various pH conditions is assessed by measuring the WCA of samples immersed in the solution for an hour.

From Figure 17, both cotton and PU sponge showed superhydrophobicity with WCAs of 154.73° and 150.97°. While cotton is less able to tolerate extreme pH levels than PU sponge.

The WCA of cotton was 140.13° at pH 2 and even lower (128.16°) at pH 13.73, reduced by around 30° under basic conditions. PU sponge, however, exhibited greater stability in WCA across the pH range, with a maximum variation of 12°.

When the sample is fully immersed in the corrosive solution, a transition from the Cassie-Baxter state to the Wenzel state occurs. Therefore, the ions in the corrosive solution will directly contact the coating surface, destroying the surface microstructure and forming corrosion pits[10],[4]. The surface roughness of modified materials can be damaged irreversibly hence resulting in lower WCA measured. Additionally, the reduction in WCA of samples in strong alkaline solutions is expected, as the pretreatment process occurs under basic conditions, making it degrade more severely in similar environments.

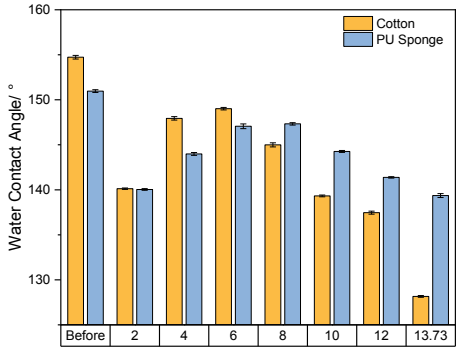


Figure 17. *Water Contact Angle after pH tests*

4. Conclusion and Outlook

The modified cotton and PU sponge produced by pre-treatment and co-hydrolysis and co-condensation of TEOS and ODTMS in an alkaline environment, both show a WCA over 150º ($\theta_{CA,cotton} = 154°, \theta_{CA,PU} = 150.9°$). In the oil-water separation, not only the high separation efficiency (>99.75%) is explored by using NanoDrop, but a wettability issue with surface state transition is also discovered, especially for the result of the PU sponge. To solve this, coating the material with a chemical with low surface energy or designing the pore size to limit undesired mass transportation might be applicable.

In the continuous water-oil separation with the modified PU sponge, the attempts of path optimisation do reduce the time consumed and collect purer oil in the outlet. However, due to technical limitations, the water content in the final liquid collected is not assessed quantitatively. If these can be achieved, the exact time to reduce the flow rate in operation can be determined by the composition of water and oil content collected over time, which can be used to design the automatic controller in practical applications.

The modified materials also have generally good reusability as oil-water separation efficiency maintained over 90% after 15 cycles. Chemical durability and Acid-Base durability are notable except for extreme alkaline conditions. Experiments can be repeated more times or last for a longer period to have more comprehensive data to obtain a more comprehensive assessment.

For the economic and sustainable factors of the modified materials, the chemicals used are mainly environmentally friendly and in large abundance in nature, which makes the synthesis method sustainable and cost-effective. However, the PU sponge itself is non-biodegradable, so material selection and waste disposal could be reconsidered.

In conclusion, sustainability, cost-effectiveness, excellent oil-water separation performance, acceptable durability and reusability of two modified materials give it strong potential for applications in industries.

Acknowledgements

The authors would like to express their gratitude to Dr. G. Tsaone for his diligent proofreading of the report and assistance in device training.

Reference

[1] Abu-Thabit, Nedal Y, et al. *Superhydrophobic Nanohybrid Sponges for Separation of Oil/ Water Mixtures*. Vol. 294, 1 May 2022, pp. 133644–133644, https://doi.org/10.1016/j.chemosphere.2022.133644.

[2] Ahmad, Darem, et al. "Hydrophilic and Hydrophobic Materials and Their Applications." *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, vol. 40, no. 22, 5 Sept. 2018, pp. 2686–2725, https://doi.org/10.1080/15567036.2018.1511642.

[3] Fang, Yuan, et al. "What Is the Driving Force behind the Adsorption of Hydrophobic Molecules on Hydrophilic Surfaces?" *The Journal of Physical Chemistry Letters*, vol. 10, no. 3, 2 Jan. 2019, pp. 468–473, https://doi.org/10.1021/acs.jpclett.8b03484.

[4] Ge, Yunyun, et al. "Durability and Corrosion Behaviors of Superhydrophobic Amorphous Coatings: A Contrastive Investigation." *RSC Advances*, vol. 12, no. 51, 2022, pp. 32813–32824, pubs.rsc.org/en/content/articlehtml/2022/ra/d2ra06073f, https://doi.org/10.1039/D2RA06073F.

[5] Gosiamemang, Tsaone, and Jerry Y.Y. Heng. "Simple One-Pot Fabrication of Durable Superhydrophobic Cotton Wool Using Octadecyltrimethoxysilane Modified Silica-Sol for Oil-Water Separation." *Separation and Purification Technology*, vol. 336, 30 Dec. 2023, p. 126185, www.sciencedirect.com/science/article/pii/S1383586623030939#b0170, https://doi.org/10.1016/j.seppur.2023.126185.

[6] ---. "Superhydrophobic Cotton for Addressing Fatbergs through Oily Wastewater Treatment." *International Journal of Biological Macromolecules*, vol. 277, 16 July 2024, p. 133863, www.sciencedirect.com/science/article/pii/S0141813024046683, https://doi.org/10.1016/j.ijbiomac.2024.133863.

[7] Hoang, Anh Tuan, et al. "Advanced Super-Hydrophobic Polymer-Based Porous Absorbents for the Treatment of Oil-Polluted Water." *Chemosphere*, vol. 277, Aug. 2021, p. 130274, https://doi.org/10.1016/j.chemosphere.2021.130274.

[8] Huang, Wenqi, et al. "Enhanced Washing Durability of Hydrophobic Coating on Cellulose Fabric Using Polycarboxylic Acids." *Applied Surface Science*, vol. 257, no. 9, Feb. 2011, pp. 4443–4448, https://doi.org/10.1016/j.apsusc.2010.12.087.

[9] Law, Kock-Yee. "Definitions for Hydrophilicity, Hydrophobicity, and Superhydrophobicity: Getting the Basics Right." *The Journal of Physical Chemistry Letters*, vol. 5, no. 4, 20 Feb. 2014, pp. 686–688, https://doi.org/10.1021/jz402762h.

[10] Mao, F. F., et al. "Preparation and Properties of a Strong and Durable Composite Superhydrophobic Coating." *Digest Journal of Nanomaterials and Biostructures*, vol. 17, no. 4, 30 Nov. 2022, pp. 1297–1311, chalcogen.ro/1297_MaoFF.pdf, https://doi.org/10.15251/djnb.2022.174.1297.

[11] Medeiros, Alexandre D'Lamare Maia de, et al. "Oily Wastewater Treatment: Methods, Challenges, and Trends." *Processes*, vol. 10, no. 4, 12 Apr. 2022, p. 743, https://doi.org/10.3390/pr10040743.

[12] Michot, L.J., and F. Villiéras. *Chapter 12.9 Surface Area and Porosity*. Vol. 1, Elsevier, 4 Sept. 2007, pp. 965–978, www.sciencedirect.com/science/article/pii/S1572435205010354?ref=pdf_download&fr=RR-2&rr=8f09ec9d887fbee9.

[13] National Sunflower Association. "Sunflower Oil Fatty Acid Profiles." *Sunflowernsa.com*, 2019, www.sunflowernsa.com/health/sunflower-oil-fatty-acid-profiles/.

[14] Nguyen-Tri, Phuong, et al. "Robust Superhydrophobic Cotton Fibers Prepared by Simple Dip-Coating Approach Using Chemical and Plasma-Etching Pretreatments." *ACS Omega*, vol. 4, no. 4, 30 Apr. 2019, pp. 7829–7837, www.ncbi.nlm.nih.gov/pmc/articles/PMC6648567/, https://doi.org/10.1021/acsomega.9b00688.

[15] Orsavova, Jana, et al. "Fatty Acids Composition of Vegetable Oils and Its Contribution to Dietary Energy Intake and Dependence of Cardiovascular Mortality on Dietary Intake of Fatty Acids." *International Journal of Molecular Sciences*, vol. 16, no. 6, 2015, pp. 12871–90, www.ncbi.nlm.nih.gov/pubmed/26057750, https://doi.org/10.3390/ijms160612871.

[16] Shim, Myoung Hee, et al. "The Effects of Surface Energy and Roughness on the Hydrophobicity of Woven Fabrics." *Textile Research Journal*, vol. 84, no. 12, 6 Feb. 2014, pp. 1268–1278, https://doi.org/10.1177/0040517513495945.

[17] Sikorska, Daria, et al. "Variation in Leaf Surface Hydrophobicity of Wetland Plants: The Role of Plant Traits in Water Retention." *Wetlands*, vol. 37, no. 5, 12 June 2017, pp. 997–1002, https://doi.org/10.1007/s13157-017-0924-2.

[18] Wang, Jiaqi, et al. "Highly Efficient Reusable Superhydrophobic Sponge Prepared by a Facile, Simple and Cost Effective Biomimetic Bonding Method for Oil Absorption." *Scientific Reports*, vol. 11, no. 1, 7 June 2021, www.nature.com/articles/s41598-021-91396-9.pdf?origin=ppub, https://doi.org/10.1038/s41598-021-91396-9.

[19] Wang, S., and L. Jiang. "Definition of Superhydrophobic States." *Advanced Materials*, vol. 19, no. 21, 5 Nov. 2007, pp. 3423–3424, https://doi.org/10.1002/adma.200700934.

[20] Zeng, Qinghong, et al. "Review on the Recent Development of Durable Superhydrophobic Materials for Practical Applications." *Nanoscale*, vol. 13, no. 27, 2021, pp. 11734–11764, https://doi.org/10.1039/d1nr01936h.

Screening the chemical space: An ML Approach to Predicting Synthesis Costs of Chemical Compounds

David Ong and Wi Rong Ng
Department of Chemical Engineering, Imperial College London, U.K.

Abstract Accurately predicting the synthesis cost of chemical compounds is crucial for improving efficiency and reducing production expenses in chemical manufacturing, particularly in drug discovery. This paper presents a novel approach to predicting synthesis costs using a XGBoost machine learning model, which leverages key features extracted from AiZynthFinder, and a hybrid approach unlike previous literature. The final optimised model yielded suboptimal results with a RMSE of 900.02 ($/kg) and R^2 of 0.37, indicating challenges in generalisation due to the lack of data points. Further analysis revealed key features affecting cost predictions such as functional groups, precursor count and certain reaction types. Despite these challenges, the findings provided valuable insights into synthesis cost prediction and establish a foundation for future improvements, including incorporating additional features and expanding the dataset to enhance robustness.

Keywords: Chemical Synthesis, Machine Learning, Retrosynthesis, Industrial Applications, Feature Engineering

Introduction

Effective design of new chemical compounds remains a significant challenge for industries ranging from pharmaceuticals to material sciences. Pharmaceutical companies face the dual challenge of developing effective drugs that address the wide range of diseases while ensuring commercial viability to remain profitable and serve shareholders' needs. Drug discovery and development often involve an iteration-heavy process that require substantial investment in time, money and resources. This leaves behind an expensive legacy entailing vast costs in time and money for the purpose of therapeutic discovery, with the mean cost of discovering and testing new synthetically viable drugs estimated at $800M USD [17].

As technology and computational power advances, paired with the growing interest in machine learning (ML) implementation, the use of retrosynthesis tools emerges as a new paradigm of chemical compound discovery [28]. These tools can optimise multiple properties simultaneously, dramatically reducing the search space of millions of synthesis routes down into tractable form and making drug development more efficient. The ability to conduct virtual screening in silico enables faster discovery, helping bring new drug compounds into the commercial space at significant time and cost savings contradictory to traditional costly and laborious manual synthesis [21].

However, despite these advancements, retrosynthesis tools are not perfect. While many retrosynthesis tools are available, some of which are commercialised as industrially usable, such as SYNTHIA, only three are fully open-source and well-established - ASKCOS, Lillymol, and AiZynthFinder (AZF). These tools can efficiently generate theoretically valid synthetic pathways faster than even the most skilled chemists. However, they often fail to account for critical factors such as industrial scalability, environmental consideration, or equipment constraints, which are essential in industrial applications resulting in a significant number of impractical routes being identified as 'optimal' synthesis pathways [8]. To move forward and further advance the paradigm of chemical search space, there is a clear need for a solution that can identify more targeted, industrially relevant routes – those that meet not just chemical feasibility but also practical requirements of large-scale manufacturing processes.

Aims And Objectives

The primary aim of this paper is to apply a machine learning technique, specifically XGBoost, to predict synthesis cost of chemical compounds. This will be achieved by leveraging key features derived from retrosynthesis tools like AZF, with a focus on understanding the factors that influence chemical synthesis cost.

Key objectives include identifying and analysing important features, such as molecular structural components, reaction types, reaction steps and precursor counts. Gaining an understanding of how these features affect costs will provide valuable insights for cost optimisation in chemical production. The project will also ensure that the model has practical applicability in a real-world setting by accounting for industrial feasibility and practicality constraints, which is achieved by using a curated list of industrially relevant reactions to guide the model's identification of feasible and practical synthetic routes. Ultimately, this paper aims to contribute a novel machine learning framework that can aid in optimising chemical synthesis, offering a more efficient and cost-effective approach to chemical production.

Methodology

Of the three open-source retrosynthesis tools that we considered, AZF was selected for the project. This is attributed to AZF being mainly a Python command-line interface, as opposed to ASKCOS' web-based interface and Lillymol's C++-based workflow [26]. Using Python, therefore allows us to take advantage of Anaconda's pre-installed machine learning and data science packages, enabling a more streamlined investigation as well as opening up frontiers for further engagement with the data science ecosystem in the context of future research.

Retrosynthesis algorithms can be classified into template-based, template-free, or sometimes both. Template-based retrosynthesis algorithms use known

reaction rules and templates that describe chemical transformations in a reaction to deconstruct a molecule, whereas template-free algorithms learn reaction patterns and can propose new reactions.

AZF utilises a template-based retrosynthesis analysis. When applied to a Monte-Carlo tree search with an expansion policy network, AZF iteratively deconstructs a target molecule into a range of possible retrosynthesis pathways from a USPTO database of millions of known reactions and molecules [14]. It selects the best retrosynthesis pathway and generates child nodes until no known templates can be applied, or purchasable precursors found from the ZINC database, a repository of commercially available molecules, are found.

We present a unique approach that aims to optimise retrosynthetic predictions for industrial chemistry, by utilising a curated list of 80 reaction templates that represent the most industrially common, unpatented and relevant reaction types [7][25]. Additionally, a hybrid reaction pathway assessment incorporates both Tanimoto Similarity (TS) and RxnFP-based Cosine Similarity (RFP) as post-processing techniques to ensure product-level and reaction transformation similarity [28]. This framework successfully met both theoretical and practical requirements, enabling more industrially relevant route identification.

To ensure consistency with data collection and adherence to the research objective, a systemic methodology was developed consisting of the following steps.

Data Gathering

The foundation of our retrosynthesis analysis lies in data gathering, which is divided into two key sections: molecules and reactions.

In order to assess the cost of producing chemicals, we compiled a database of approximately 850 relevant, unpatented pharmaceutical molecules along with their associated data from publicly available, reputable sources, including BMJ Global Health, eEML, PharmaCompass, and the FDA Orange Book [3][19][23]. These data consisted of essential details such as molecule name, active pharmaceutical ingredient (API) cost, and production quantity, were used for evaluating the production costs of the chemicals that we considered in our analysis [16].

Our criteria for determining the reaction types that constituted industrially relevant reactions should generally satisfy criteria such as (but not limited to):

1.The reaction has a track record of successful industrial implementation.

2.Commonly used in the industry, reflecting its reliability and practicality.

3.Carried out in existing industrial equipment, without the need for specialised apparatus.

4.Well-documented in literature, indicating consistent results and broad applicability.

Examples of the recorded data are presented in Table 1 where reactions were categorised into 'Reaction Type'. This process was applied across a curated set to obtain a diverse and representative reaction set, encompassing only commonly employed industrial reactions such as esterification, Friedel-Crafts acylation, etc [7][25].

Table 1 - Data Recording of Literature Review

Reaction Type	Reactants	Reagents	Products
Amide Bond Formation	CH_3COOH, NH2	-	CH_3CONH_2
Aromatic Substitution (Nucleophilic)	C_6H_5Cl, NaOH	-	C_6H_5OH
…	…	…	…

Data Processing

To begin our process of data processing, we started by generating machine-readable retrosynthetic data for all the identified molecules. Thus, SMILES notations and Morgan Fingerprints were generated for all recorded molecules. SMILES is the conventional notation used to represent molecular structures using ASCII strings, whilst the Morgan Fingerprint, as shown in Figure 1, encodes the presence or absence of specific substructural features in a binary vector, providing a fixed-size representation of a molecule to be processed.

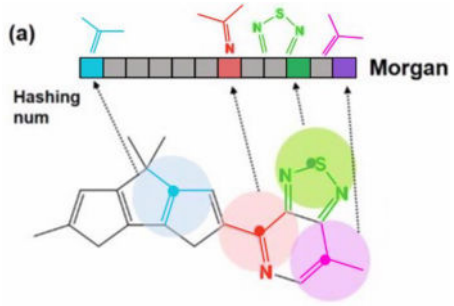


Figure 1 – Morgan Fingerprint Binary Vector Representation [11]

Subsequently, the data from Table 1 were first converted into SMILES notation and merged into reaction representations. These reaction types and their 'SMILES Reaction' in Table 2, were further enhanced with atom-mapped labels using AiZynthTrain's rxnutils package pipelines to produce generalised SMILES reactions. These were in a form of a generalised format that can be directly parsed and converted into reaction objects, facilitating future hybrid approach calculations and enabling precise tracking of atom transformations across reactions.

Table 2 - Smiles Reaction Generation

Reaction Type	SMILES Reaction	GeneralisedSmiles Reaction
Amide Bond Formation	CC(=O)O.[NH2] ≫ CC(=O)N	[C:1][C:2](=[O:3])[NH2:4] ≫ [C:1][C:2](=[O:3])O.[NH2:4]
Aromatic Substitution (Nucleophilic)	c1ccccc1Cl.[OH-] ≫ c1ccccc1[OH]	[cH:1]1[cH:2][cH:3][cH:4][cH:5][c:6]1[X:7] ≫ [cH:1]1[cH:2][cH:3][cH:4][cH:5][c:6]1[H].[X:7]
…	…	…

Finally, the complete molecule dataset was processed using AZF, which contained USPTO patented reactions as the expansion policy, to generate retrosynthesis routes and their associated output features. The output features generated for all molecules included metrics such as the total number of reactions, precursors count, and number of precursors in stock which were subsequently used as model input features in Table 3. To enhance both the accuracy and efficiency of data generation, modifications were made to the AZF configuration. Each molecule was analysed until all solved routes were extracted, ensuring ample time for tree search iterations which minimised the risk of overlooking any synthetic pathways. Additionally, the compiled synthesis routes were filtered to exclude solutions with unrecognised precursors that were not present in the ZINC database.

Production Cost Estimation

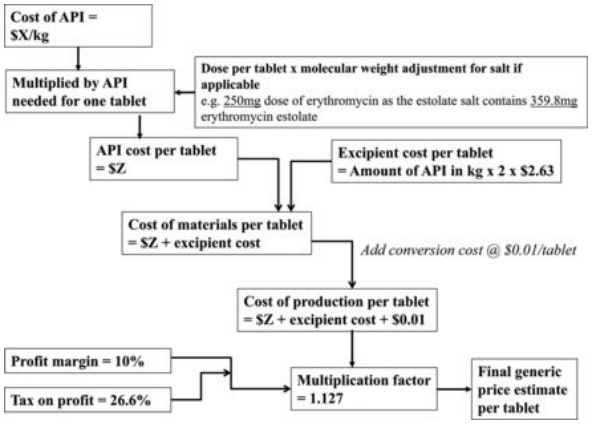


Figure 2 - Algorithm used to estimate generic prices

The methodology outlined in Figure 2, was used to calculate generic price estimates of a molecule, derived from several factors such as Cost of API, dosage per tablet, profit margin that were obtained from PharmaCompass. Subsequently, we worked backwards from the price estimates to determine an estimated production cost.

Hybrid Approach

The hybrid approach was adopted to evaluate all AZF-generated routes and identify the most industrially relevant routes.

For TS evaluations, reaction templates were extracted from the atom-mapped reactions listed in Table 2 and stored as SMARTS strings. These templates generalise patterns that occur at the reaction centre for each molecule, storing information on chemical transformations – such as bonds broken, formed or altered in the immediate surroundings. Extracted reaction templates were iteratively applied to each target molecule (reactant) and the generated product(s) were structurally compared to the target AZF product [4] via TS. Where applicable, an aggregate similarity score was calculated across all reaction steps in the route. The algorithm was configured to return the highest TS scoring route along with the corresponding reactant-template pair and reaction type.

RFP evaluations on the other hand does not rely on reaction templates and extracts transformation patterns directly from reactions itself [28]. Thus, RxnFP reaction fingerprints, trained on the Pistachio dataset, were generated for both the external reactions list and the generated AZF SMILES reactions. Cosine Similarity was subsequently used to evaluate similarity between the reactions. Likewise, the algorithm returned the highest RFP scoring route along with the corresponding reaction types used.

In the hybrid approach, TS is treated as the primary criterion for selecting optimal synthetic route. However, if the TS score falls below a predetermined threshold score of 0.65 and the RFP score is significantly higher, the RFP route is chosen instead. This ensures that the framework can benefit from TS's precision when focusing on structural similarity, whilst also relying on RFP's reaction-level context when the first falls short. By using this hybrid approach, we mitigate the potential shortcomings of each method when used independently and improve selection of relevant synthetic routes. The results from this processing stage, was compiled with other relevant features and used to train the model. Table 3 illustrates an example sample dataset for metoclopramide. A full methodology workflow is also summarised in Figure 3.

Table 3 - Example Sample Dataset (metoclopramide)

Compound	Morgan Fingerprint	Reaction Steps	Precursors Count	Reaction Type	Total Number of Functional Groups	Production Volume (kg/year)	Production Cost ($/kg)
metoclopramide	0,0,0,....,0,0,0	1	2	Acylation	5	89200	182.26

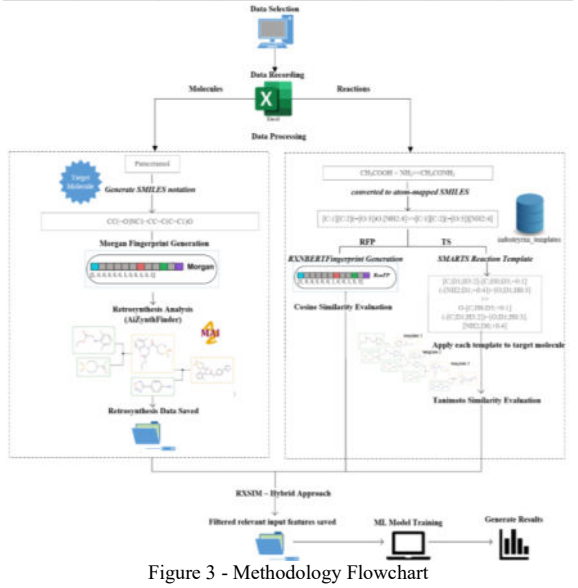


Figure 3 - Methodology Flowchart

Machine Learning

Various machine learning models were employed to predict the target molecule's synthesis cost. These models

range from interpretable statistical methods such as linear and polynomial regression, to advanced deep learning methods like artificial neural networks which can capture high-dimensional relationships.

The input features being used to train the model included reaction steps, precursor count, reaction types, and total number of functional groups. The selection of the optimal machine learning model for analysing input features was determined by the size and complexity of the dataset. Both statistical and deep learning models were explored and their performance metrics evaluated, to identify the model best suited for estimating synthesis cost of chemical compounds.

Data Pre-Processing

To ensure the machine learning models performed reliably and as intended, it was imperative that the dataset and input features were relevant and easily interpreted by the computer. This was achieved through data pre-processing using Python's built-in library, scikit-learn, through a process that can be summarised into the following 3: data cleaning, data normalisation and data transformation [34].

We began the cleaning stage by ensuring that missing, invalid data, and duplicate data points were removed. A 95 % confidence interval filter around mean production cost and production volume was taken to remove outliers, which could disproportionately bias the model, causing overfitting or underfitting issues. Duplicate molecules, collected from overlapping databases, were also removed to avoid bias. Data Normalisation, especially for linear regression models, is recommended to improve performance and prevent features of a different magnitudes from skewing the model [2]. A Z-score normalisation was applied using StandardScaler(), a function in the scikit-learn library, which resulted in a uniformly scaled dataset. Finally, categorical data, such as reaction type, was transformed to a numerical format via one-hot encoding. This method creates a binary column for each unique category in the data, an example of which is shown in Table 4.

Table 4 - One-Hot Encoded Categories of Reaction Types

Reaction Type	→	Acylation	Reductive Amination
Acylation, Reductive Amination		1	1
Acylation		1	0

Model Design and Architecture

Statistical-based Modelling

Several statistical models were initially explored due to their ease of implementation. The Simple Linear Regression (SLR) model assumes and predicts a linear relationship between an independent and dependent variable, providing a baseline comparison for other, more complex models [18]. Polynomial Regression (PR) is an extension of SLR that predicts the dependent variable as polynomials of the independent variables, offering better predictions for non-linear relationships [22].

However, both SLR and PR models struggle with multicollinearity, where highly correlated, independent variables lead to unreliable coefficient estimates. They also lack mechanisms to handle irrelevant features. Principal Component Analysis (PCA) resolves these issues through dimensionality reduction, transforming the original features into new, uncorrelated principal components. Although PCA mitigates multicollinearity and overfitting, it also reduces interpretability and faces significant information loss if too many components are reduced.

Models that employ regularisation techniques, such as Ridge Regression (RR), Lasso Regression (LsR), and Elastic Net, address these limitations more effectively. While Elastic Net generally provided superior results for large datasets, it underperformed against LsR for low sampled datasets (<10,000), making LsR an ideal choice for this project [1].

LsR is another extension of SLR that employs L1 regularisation. L1 regularisation introduces a penalty term, λ, on the absolute sum of the regression coefficients, $\sum_{i=1}^{p}|\beta_j|$, as an additive term in the loss function, $L(x)$, shown in (1). This penalty prevents overfitting by shrinking or zeroing certain coefficients, retaining only important features.

$$L(x) = \text{RMSE} + \lambda \sum_{i=1}^{p}|\beta_j| \qquad (1)$$

Tree-based Modelling

Tree-based models - Random Forest (RF) and Gradient Boosting (GB) - were also explored due to their robustness against multicollinearity and better handling of non-parametric, complex datasets [10].

RF is a tree-based, ensemble learning algorithm, meaning predictions are made through the combination of multiple models [6]. As seen in Figure 4, RF consists of independent, individual, decision trees each trained on a random subset of data (bagging) in parallel, and their predictions are combined and averaged to provide the final predicted output value. This combination reduces the risk of overfitting and can capture complex relationships between features.

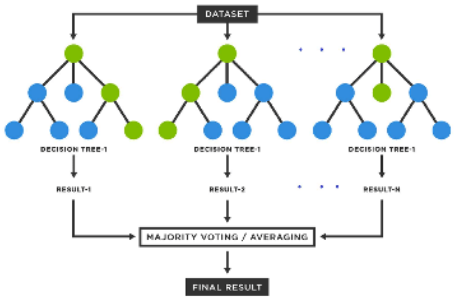


Figure 4 - 3 decision trees trained using bagging, with the final result obtained via the average of their individual outputs [31].

RF is often praised for its 'out of box' performance. To further enhance the model's predictive capabilities, hyperparameter tuning was explored. Key hyperparameters included:

- max_depth: controls the depth of each tree; greater depth captures more information

- min_samples_split: minimum number of samples required to split an internal node
- max_leaf_nodes: limits tree size in terms of the number of final leaves

Unlike RF, GB constructs trees (weak learners) sequentially, with each new tree correcting the errors of its predecessor (Figure 5). Predictions are refined iteratively which help reduce bias, while the loss function is minimised using gradient descent. Among many GB algorithms available, XGBoost was chosen due to its consistent performance, generally matching or surpassing other gradient boosting methods across different scenarios [12].

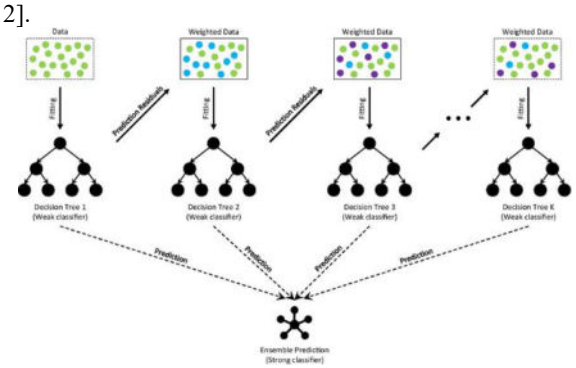


Figure 5 - Iterative Refinement for Enhanced Prediction Accuracy

These models are better suited for capturing non-linear interactions between features relative to SLR and PR, providing a more robust approach to predictive modelling.

Neural Network Modelling

While Neural Networks can outperform XGBoost for data sets with complex relationships between features. XGBoost was chosen as the main model due to the built-in feature importance metric with tree-based models, making it more straightforward to extract insight on variable most critical for cost prediction. Additionally, with small datasets, neural networks often struggle. XGBoost is more robust for low data sets as the underlying architecture automatically employs regularisation techniques, thus outperforming neural networks when the data set is small but involves complex relationships [27].

Results and Discussion

An effective machine learning model should exhibit low prediction errors between the actual and predicted production costs, quantified using root mean squared error (RMSE) [15]. This metric represents the average deviation between predicted and actual costs and are used to evaluate accuracy of the trained models. Additionally, the R^2 statistic is incorporated to assess how well the model explains variance in the data, especially when using linear regression [5]. An ideal model would consist of a high R^2 combined with low RMSE. Further discussions will address the results of the hybrid approach, the machine learning model outcomes, and a detailed analysis of the model's performance.

Evaluation of Synthetic Route Predictions

A preliminary analysis was performed to assess the contributions of TS and to route selections. As visualised in Figure 6, there was a stark contrast between both methods when selecting the optimal synthetic route. TS favours specific top routes like Route 1, which appears most frequently in contrast to RFP who shows a broader distribution of route selections. This divergence highlights the limitations of relying on either method independently, supporting the decision to adopt a hybrid approach.

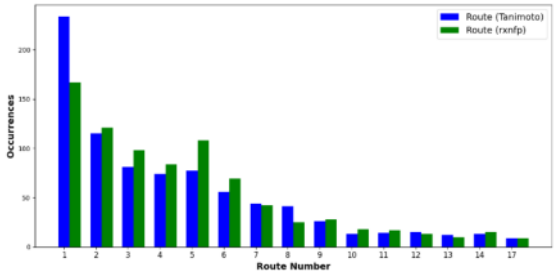


Figure 6 - AZF Top 15 Routes Identical Matches with Both Methods

The hybrid approach demonstrated potential improvements in route relevance and feasibility. An analysis of the top 5 routes identified for each molecule, found that just only 65% had matching routes, suggesting that while AZF's default method identifies plausible routes, it may overlook other viable alternatives that are more aligned with real-world constraints.

By focusing on just the most common and industrially relevant reaction types, we can assert with greater confidence that the hybrid approach generated paths that were better suited to meet industry standards for process scalability, feasibility, and cost effectiveness. However, Figure 7 depicts approximately 35% of molecules had an increase in reaction steps compared to the AZF baseline. This can be attributed to industrial reactions often involving additional steps to balance reaction efficiency and the practical feasibility and meet practical, regulatory and safety requirements [20][29].

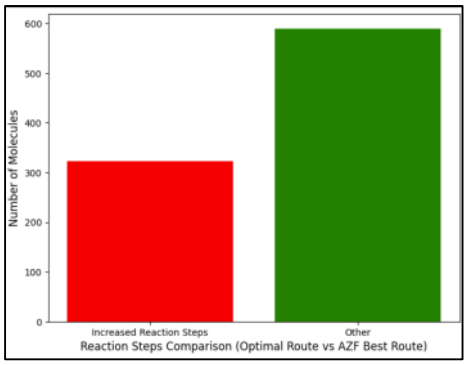


Figure 7 - Reaction Steps Comparison (Optimal Route vs AZF Route 1)

Model Training

After applying data pre-processing, the initial molecules from the data set were culled to 192 unique molecules, and we identified 44 relevant input features.

Statistical Model Training

SLR, employing PCA, was used as a baseline, against which other machine learning models would be compared to. Optimal PCA components were investigated using Cumulative Explained Variance (CEV), the running total of variance by the first n principal components, and RMSE. CEV asymptotes as number of principal components increases. Balancing this with predictive accuracy, as shown in Figure 8, is crucial. After analysis, the optimal PCA dimensionality was deduced to be 30 components. This choice provided a balance between a low Test RMSE of 4187.23 ($/kg), a sufficiently high CEV of 98%, and Test R^2 of -1.02. While the negative R^2 suggest limitations in the linear model's predictive power, the PCA chosen offered a suitable baseline.

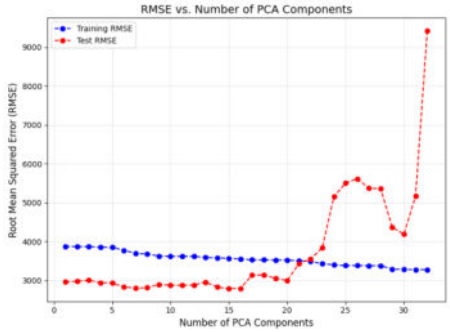


Figure 8 - Test and Validation RMSE Plot against Number of PCA Components

When investigating PR, it was observed that polynomials of degree 2 or greater would cause extreme overfitting of the training data, even after applying PCA. It was observed that the Train RMSE approached 0 with a corresponding R^2 of 1, indicating the model was memorising training data, rather than learning generalisable relationships. This approach was very susceptible to noise and minor fluctuations in data and thus not deemed a good fit for our model, as evident in Figure 9.

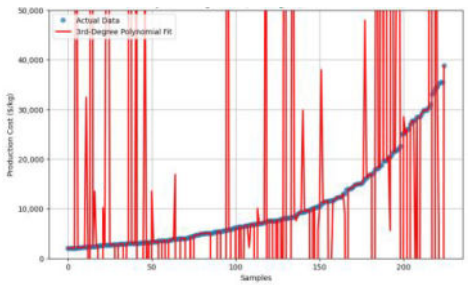


Figure 9 – Comparison of Actual vs. Predicted Production Cost (Polynomial Regression) for 3rd Degree Polynomial

The final statistical model evaluated was LsR. Due its ability to shrink coefficients down to zero, LsR did not require PCA reduction. Its performance remained competitive with SLR, producing a Test RMSE and R^2 of 3720.02 ($/kg) and 0.09 respectively. Although these figures are not a dramatic improvement from the baseline, LsR's strength lies in feature selection, a process that will be discussed later.

Tree-based Model Training

RF hyperparameters were tuned using scikit-learn's RandomSearch() function, which samples a specified number of random combinations of hyperparameters. These combinates include min_samples_split = [30,70), max_leaf_nodes = [3,9), and max_depth = [2,6).

RandomSearch() was set to 500 iterations to ensure that any optima within the range were captured. Additionally, cross-validation, which splits the data into test and training subsets, was employed to ensure that the trees were not 'learning' the patterns of the data set (overfitting). In Figure 10, the red dot denotes the 'optimal' combination of hyperparameters as min_samples_split = 57, max_leaf_nodes = 4, and max_depth = 3. The minimum sample split of 57 indicates the model favours larger, more stable splits, avoiding the noisy smaller splits that may result in decisions that only applicable to outliers. A max_leaf_nodes of 4 restricts the tree to produce very general predictions. Since each leaf represents a specific decision outcome, having 4 suggests the model was forced toward more general patterns. A max_depth of 3 prevents the tree from overfitting by stopping it from delving deeper into niche patterns that do not generalise well. Therefore, the model is restricted to broad, fundamental decisions that govern the entire dataset, rather than finely tuned segments. Ultimately, the RF model obtained a Test RMSE of 1101.34 ($/kg) and R^2 of 0.19.

Even with conservative hyperparameters, RF (Figure 11) was still prone to overfitting, evident from the low Train RMSE of 716.83 ($/kg) and R^2 of 0.34.

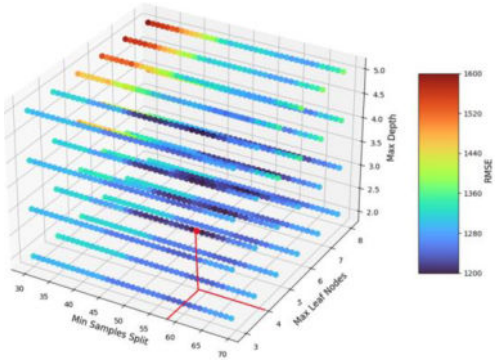


Figure 10 - 4D Heat Map of min_samples_split, max_leaf_nodes, and max_depth, with a RMSE Scale for RF

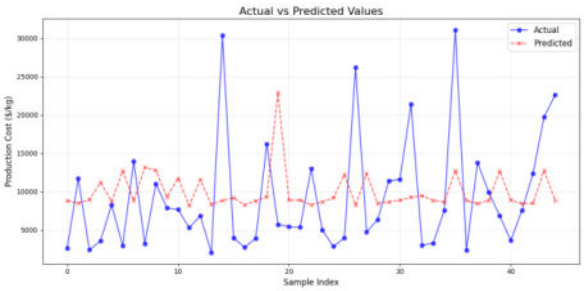


Figure 11- Actual vs. Predicted Production Cost using Random Forest.

XGBoost was the lastly evaluated with the tuned hyperparameters listed in Table 5. Similar to RF, RandomSearch() and cross validation were used to find the optimal hyperparameters (Figure 12). However, due to the increased number of hyperparameters, RandomSearch() was set to 700 iterations to better model the interactions between datapoints.

Table 5 - Hyperparameters, their Range and Function

Hyperparameter	Range	Function
n_estimators	[100, 400)	Control number of gradient-boosted trees
learning_rate	[0.001, 0.2)	Controls predictive power, updates more/less conservative
max_depth	2, 10)	Maximum depth of each tree
subsample	[0.6, 0.9)	Fraction of training data sampled to build each tree
gamma	[2, 6)	Minimum loss required to extend a leaf node

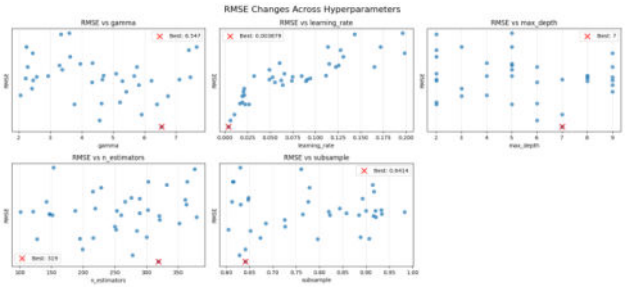


Figure 12 – Scatter plot of RMSE values against hyperparameters. With the red crosses denoting the optimal RMSE values.

n_estimators = 319 suggests that the model requires many steps to refine predictions, suggesting nuanced relationships in the data that are not immediately obvious. While many trees come with the risk of overfitting, they are controlled by other parameters. small learning_rate = 0.003879 implies each tree only makes tiny adjustments, preventing the model from reacting to noise or overfitting by preventing large shifts in predictions. A max_depth = 7 indicates that XGBoost recognises the data intricacies. However, like RF, deeper trees can lead to overfitting if not controlled. A subsample = 0.6414 indicates 64.14% of the data is used per tree, introducing variability which prevents a single tree from fitting too tightly. A gamma = 6.547 shows the model demands a substantial improvement in objective before splitting, suggesting that while the model favours deeper trees, it only allows meaningful splits to occur under strict criteria.

Compared to RF, XGBoost (Figure 13) is less sensitive to fluctuations and sudden peaks in data, which can be explained by the choice of hyperparameters. This is also reflected in its test RMSE of 900.02 ($/kg) and R^2 of 0.37.

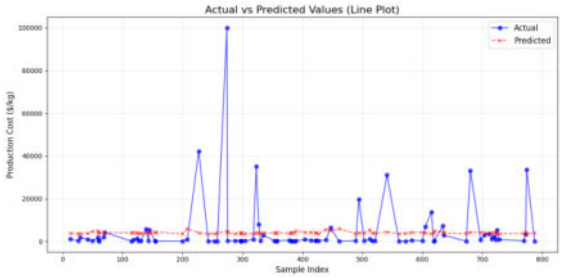


Figure 13 – Actual vs. Predicted Production Cost using XGBoost.

Feature Analysis

Table 6 - Top 5 Feature Importance for Various Models

LsR	RF	XGB
Oxidation	Reduction of Phenol	Reduction of Phenol
Reduction of Phenol	Reduction of amide to CH2-amine	Reduction of amide to CH2-amine
Michael Addition	Michael Addition	Hydrolysis of Cyclic Imides
Number of Precursors	Number of Precursors	Reduction (Nitro to Amine)
Deprotection (Boc)	Total Number of Functional Groups	Hydrolysis (Ester)

Feature importance, as seen in Table 6 was extracted from LsR, RF, and XGB, due to their mechanisms for evaluating individual features. 'Reduction of Phenol' was initially thought to be highly significant due to its common occurrence in our dataset, which each model detected. However, this was not the case, as most top reactions did not appear in the common reactions seen in Figure 14.

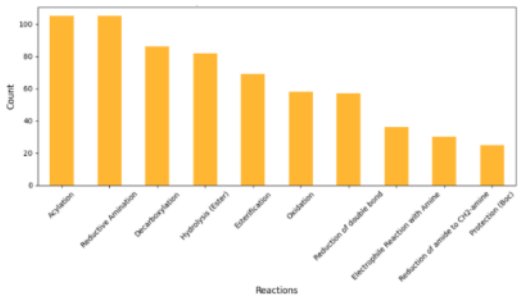


Figure 14 – Top 10 Most Common Reaction Types in the Dataset

Further analysis revealed that most of these reactions had low production volumes (around 0.1 kg/year). This can be attributed to various pharmaceutical drugs being produced at widely varying scales, or the lack of data from PharmaCompass. Due to the skewed data, the median

production volume was taken as a better representation of the 'average' data points. Hence by taking a 99.7% confidence interval of the median, only molecules with production volume >400kg/year were kept. Consequently, XGBoost was utilised to re-calculate feature importance and its results are shown in Figure 15 .

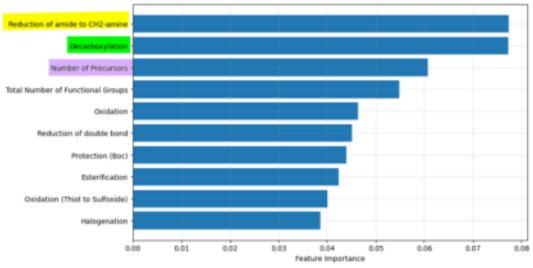


Figure 15 - Re-calculated Top 10 features of XGBoost

Table 7 - Compounds Undergoing Reduction of Amide to CH2-Amine

Compound	Number of Reaction Steps	Number of Precursors	Total Number of Functional Groups
almotriptan malate	5	3	5
prucalopride succinate	1	2	7
nintedanib esylate	2	3	7
glycopyrrolate	2	2	4
avatrombopag maleate	5	5	8
tedizolid phosphate	7	7	5
rifabutin	4	3	8
saxagliptin hydrochloric	7	5	5
ruxolitinib phosphate	1	2	5
elagolix sodium	3	4	8
vidaza	5	4	6
zolmitriptan	4	3	3
ramelteon	2	2	2
tofacitinib citrate	4	3	7
olaparib	6	4	5

The new features shown in Figure 15 were investigated by combing through the data set. Table 7 lists compounds containing the 'Reduction of amide to CH2-amine' one-hot encoding. It was concluded that compounds that have undergone 'Reduction of amide to CH2-amine' were the greatest predictor of cost as often, these molecules often undergo additional reactions which is seen by the large number of reaction steps and number of pre-cursors (another top feature). A correlation matrix, in Figure 16 was plotted to verify that none of the above features were correlated with each other.

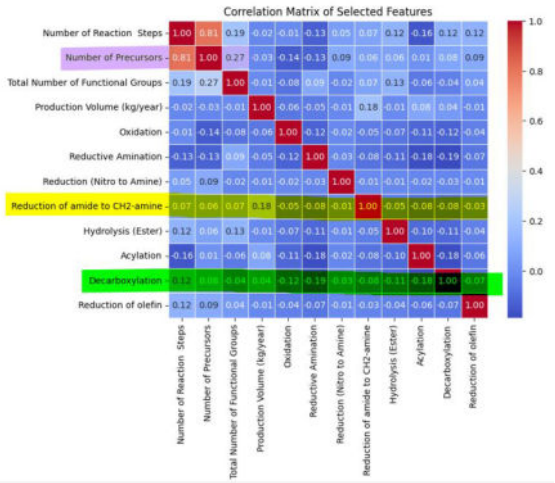


Figure 16. Correlation matrix of selected features. Highlighted rows correspond to the top 3 features in Figure 15

Test on Unseen Data

The predictive power of the optimal XGBoost model was evaluated on two unseen molecules- dexamethasone and budesonide- [33]. Following the procedures outlined in the Methodology section, the data was subsequently pre-processed and run through the XGBoost model, achieving an R^2 of 0.37 and a RMSE of 79.04 ($/kg). The observed RMSE produced was significantly lower than that obtained on the test RMSE, suggesting potential overfitting or more likely that these data points were insufficient to represent the chemical space sampled during model training. It is also important to note that these results were derived from only 2 compounds and may not indicate the model's true generalisability. To confirm the model's generality and predictive performance, further validation with a diverse set of unseen compounds is essential.

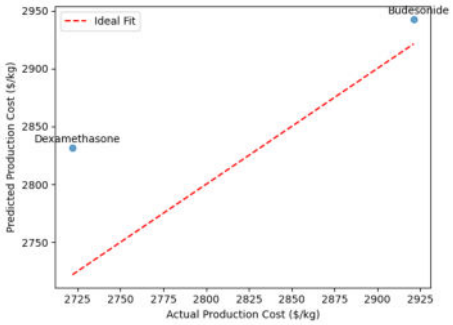


Figure 17 - Predicted vs Actual Plot

Conclusions

A novel machine learning framework to predict synthesis costs was developed with AiZynthFinder, a retrosynthesis algorithm employed to extract key statistical features such as reaction steps and precursor count. These features were refined using a hybrid approach to ensure industrial relevance and practical applicability. With additional input features sourced from publicly available sources, a range of statistical and tree-based models were trained to predict synthesis cost, and further fine-tuned through hyperparameter optimisation to enhance performance. Despite the extensive efforts, XGBoost's model was found to the best performing, with an underwhelming accuracy that achieved an RMSE of 900.02 ($/kg) and a R^2 of 0.37. When tested on unseen molecules, dexamethasone and budesonide, a RMSE of 79.04 ($/kg) and a R^2 of 0.37 was achieved, reflecting the challenges the model faced to when encountering unseen data. This primarily stems from the niche focus of the project upon existing and documented industrially relevant reactions, which resulted in a small, skewed dataset which hindered the model's ability to capture patterns across a broader range of synthesis costs. Furthermore, the lack of a centralized, high-quality dataset for data gathering heavily impacted the reliability and consistency of the training set. To understand the model's performance better, a range of analysis techniques were employed, which included

feature importance and dependency analysis. Feature analysis revealed key features that heavily influenced cost predictions, such as number of functional groups, number of precursors, and specific reaction types like amide reduction, decarboxylation and oxidation. However, a correlation matrix analysis that followed highlighted a high correlation between the number of reactions and precursor count. Although this was acknowledged, it was not fully addressed due to the limitations imposed by the small dataset size. In conclusion, despite the model's performance being restricted by data quality and size, the framework developed provides a novel approach for predicting chemical synthesis costs, offering valuable insights into the major cost factors in chemical synthesis. The framework lays the foundation for future applications in cost optimisation and supports the growing integration of machine learning in chemical engineering and manufacturing.

Research Recommendations

The conclusions of this research highlight several promising directions for future exploration. Of utmost importance, to improve model performance, it is crucial to train and test the model on larger, more diverse datasets, with a focus on unpatented molecules to ensure industrial relevance. Further efforts to expand the AiZynthTrain dataset size discussed earlier was deferred for future work due to project time constraints. Therefore, looking into expanding the dataset by including more reactions of the same type can enhance the model's generalisation ability, addressing current limitations faced with the small dataset size, and potentially replacing the hybrid approach. Subsequently, incorporating additional factors such as process conditions, catalytic factors, reaction kinetics, yield, raw material costs, market demand, and geographical location could further enhance the model's applicability in various scenarios and improve feature engineering. Geographical data, for example, would offer insights into how supply chain and transportation costs vary by region, while market demand would reflect how pricing structures shift with additional demand. However, the necessity for a centralised, accessible database is needed to consolidate such data, which would require integration of existing, limited databases and collaboration with industry, academia and government bodies to refine feature selection, as well as validate model predictions in real-world scenarios. Improvements to the filtering process are also recommended, as close to half of the molecules were deemed ineligible for model training. Ways to address this include expanding the ZINC Stock database and configuring AZF to use deeper search trees to improve route identification, though this would demand an increase in computational resources, advanced hardware, and algorithmic optimisations. Given the non-linearity of the problem as demonstrated by the trained machine learning model, further exploration of neural network models, with a significantly larger data set, is suggested as they may better capture complex relationships between the features and its impact on synthesis cost. Broader applications could focus on expanding the model's scope to predict process efficiency, environmental impact, and sustainability metrics alongside costs, providing a more holistic understanding of chemical synthesis.

References

[1] Altelbany, S. (2021). Evaluation of Ridge, Elastic Net and Lasso Regression Methods in Precedence of Multicollinearity Problem: A Simulation Study. Journal of Applied Economics and Business Studies, 5(1), pp.131–142. doi:https://doi.org/10.34260/jaebs.517.

[2] Anggoro, D.A. (2019). Improving Accuracy by applying Z-Score Normalization in Linear Regression and Polynomial Regression Model for Real Estate Data. International Journal of Emerging Trends in Engineering Research, 7(11), pp.549–555. doi:https://doi.org/10.30534/ijeter/2019/247112019.

[3] Approved Drug Products with Therapeutic Equivalence Evaluations (2023). Available at: https://www.fda.gov/media/71474/download?attachment.

[4] Bajusz, D., Rácz, A. and Héberger, K. (2015). Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? Journal of Cheminformatics, 7(1). doi:https://doi.org/10.1186/s13321-015-0069-3.

[5] Bobbitt, Z. (2021). RMSE vs. R-Squared: Which Metric Should You Use? [online] Statology. Available at: https://www.statology.org/rmse-vs-r-squared/.

[6] Breiman, L. (2001). Random Forests. Machine Learning, 45(1), pp.5–32. doi:https://doi.org/10.1023/a:1010933404324.

[7] Brown, D.G. and Boström, J. (2015). Analysis of Past and Present Synthetic Methodologies on Medicinal Chemistry: Where Have All the New Reactions Gone? Journal of Medicinal Chemistry, 59(10), pp.4443–4458. doi:https://doi.org/10.1021/acs.jmedchem.5b01409.

[8] Chen, B., Li, C., Dai, H. and Song, L. (2020). Retro*: Learning Retrosynthetic Planning with Neural Guided A* Search. [online] arXiv.org. doi:https://doi.org/10.48550/arXiv.2006.15820.

[9] Chen, J. (2022). What is a Neural Network? [online] Investopedia. Available at: https://www.investopedia.com/terms/n/neuralnetwork.asp.

[10] Chowdhury, S., Lin, Y., Liaw, B. and Kerby, L. (2021). Evaluation of Tree Based Regression over Multiple Linear Regression for Non-normally Distributed Data in Battery Performance. [online] arXiv.org. doi:https://doi.org/10.48550/arXiv.2111.02513.

[11] Data-based Polymer-Unit Fingerprint (PUFp): A Newly Accessible Expression of Polymer Organic Semiconductors for Machine Learning - Scientific

Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Schematic-diagrams-of-several-molecular-fingerprints-a-Morgan-Morgan-fingerprints_fig1_362145207 [accessed 9 Dec 2024]
[12] Davide Boldini, Grisoni, F., Kühn, D., Friedrich, L. and Sieber, S.A. (2023). Practical guidelines for the use of gradient boosting for molecular property prediction. Journal of Cheminformatics, [online] 15(1). doi:https://doi.org/10.1186/s13321-023-00743-7.
[13] Genheden, S., Norrby, P.-O. and Engkvist, O. (2023). AiZynthTrain: Robust, Reproducible, and Extensible Pipelines for Training Synthesis Prediction Models. Journal of Chemical Information and Modeling, [online] 63(7), pp.1841–1846. doi:https://doi.org/10.1021/acs.jcim.2c01486.
[14] Genheden, S., Thakkar, A., Chadimová, V., Reymond, J.-L., Engkvist, O. and Bjerrum, E. (2020). AiZynthFinder: a fast, robust and flexible open-source software for retrosynthetic planning. Journal of Cheminformatics, 12(1). doi:https://doi.org/10.1186/s13321-020-00472-1.
[15] Hanane D (2022). Evaluation Metrics For Regression Models - Hanane D. [online] Hanane D. Available at: https://machinelearning-basics.com/evaluation-metrics-for-regression-models/ [Accessed 9 Dec. 2024].
[16] Hill, A.M., Barber, M.J. and Gotham, D. (2018). Estimated costs of production and potential prices for the WHO Essential Medicines List. BMJ Global Health, [online] 3(1), p.e000571. doi:https://doi.org/10.1136/bmjgh-2017-000571.
[17] Ipi.org. (2024). From Inception to Ingestion:The Cost of Creating New Drugs. [online] Available at: https://www.ipi.org/ipi_issues/detail/from-inception-to-ingestionthe-cost-of-creating-new-drugs [Accessed 10 Dec. 2024].
[18] James, G., Witten, D., Hastie, T., Tibshirani, R. and Taylor, J. (2023). Linear Regression. An Introduction to Statistical Learning, pp.69–134. doi:https://doi.org/10.1007/978-3-031-38747-0_3.
[19] list.essentialmeds.org. (n.d.). eEML - Electronic Essential Medicines List. [online] Available at: https://list.essentialmeds.org/.
[20] Newman, S.G. and Jensen, K.F. (2013). The role of flow in green chemistry and engineering. Green Chemistry, 15(6), p.1456. doi:https://doi.org/10.1039/c3gc40374b.
[21] oied.osu.edu. (n.d.). AI-Enabled Retrosynthesis for Drug Development | Office of Innovation and Economic Development, The Ohio State University. [online] Available at: https://oied.osu.edu/technologies/ai-enabled-retrosynthesis-drug-development.
[22] Ostertagová, E. (2012). Modelling using Polynomial Regression. Procedia Engineering, 48, pp.500–506. doi:https://doi.org/10.1016/j.proeng.2012.09.545.
[23] PharmaCompass.com. (n.d.). PharmaCompass – Grow Your Pharma Business Digitally. [online] Available at: https://www.pharmacompass.com/.
[24] PubChem (2024). PubChem. [online] Nih.gov. Available at: https://pubchem.ncbi.nlm.nih.gov/.
[25] Roughley, S.D. and Jordan, A.M. (2011). The Medicinal Chemist's Toolbox: An Analysis of Reactions Used in the Pursuit of Drug Candidates. Journal of Medicinal Chemistry, 54(10), pp.3451–3479. doi:https://doi.org/10.1021/jm200187y.
[26] Saigiridharan, L., Hassen, A.K., Lai, H., Torren-Peraire, P., Engkvist, O. and Genheden, S. (2024). AiZynthFinder 4.0: developments based on learnings from 3 years of industrial application. Journal of Cheminformatics, 16(1). doi:https://doi.org/10.1186/s13321-024-00860-x.
[27] Sanja Stevanović, Dashti, H., Marko Milošević, Salem Al-Yakoob and Dragan Stevanović (2024). Comparison of ANN and XGBoost surrogate models trained on small numbers of building energy simulations. PLoS ONE, 19(10), pp.e0312573–e0312573. doi:https://doi.org/10.1371/journal.pone.031257
[28] Schwaller, P., Probst, D., Vaucher, A.C., Nair, V.H., Kreutter, D., Laino, T. and Reymond, J.-L. (2021). Mapping the space of chemical reactions using attention-based neural networks. Nature Machine Intelligence, [online] 3(2), pp.144–152. doi:https://doi.org/10.1038/s42256-020-00284-w.
[29] Schwan, J. and Christmann, M. (2018). Enabling strategies for step efficient syntheses. Chemical Society Reviews, 47(21), pp.7985–7995. doi:https://doi.org/10.1039/c8cs00399h.
[30] Smith, M.B. and March, J., 2007. March's Advanced Organic Chemistry: Reactions, Mechanisms, and Structure. 6th ed. Hoboken, NJ: Wiley.
[31] Spotfire. (n.d.). Spotfire | Demystifying the Random Forest Algorithm for Accurate Predictions. [online] Available at: https://www.spotfire.com/glossary/what-is-a-random-forest.
[32] Spotfire. (n.d.). Spotfire | Demystifying the Random Forest Algorithm for Accurate Predictions. [online] Available at: https://www.spotfire.com/glossary/what-is-a-random-forest.
[33] Wang, J., Levi, J., Ellis, L. and Hill, A. (2021). Minimum Manufacturing Costs, National Prices, and Estimated Global Availability of New Repurposed Therapies for Coronavirus Disease 2019. Open Forum Infectious Diseases, [online] 9(1), p.ofab581. doi:https://doi.org/10.1093/ofid/ofab581.
[34] www.purestorage.com. (n.d.). What Is Data Preprocessing for Machine Learning? | Pure Storage. [online] Available at: https://www.purestorage.com/uk/knowledge/what-is-data-preprocessing.html.

Incorporating Manufacturing Uncertainty and Sustainability into Supply Chain Modelling for Autologous CAR-T Cell Therapy

Chiara Mauri and Claudia Torres Quiros
Department of Chemical Engineering, Imperial College London, U.K.

Chimeric Antigen Receptor (CAR) T-cell therapy involves genetically modifying a patient's T-cells to target and destroy cancer cells. Since its 2017 clinical approval, it has rapidly evolved, with six products authorized for targeting B-cell malignancies. High costs and logistical challenges, driven by its autologous nature, remain significant barriers. Computational tools have been developed to optimize supply chain decision-making, with a focus on patient scheduling and the critical timing of this life-saving therapy. Until now, optimization has been purely deterministic. In this paper we propose a framework to quantify manufacturing uncertainty and integrate it into supply-chain optimisation. We employ Global Sensitivity Analysis (GSA) and Uncertainty Analysis (UA) to identify critical manufacturing bottlenecks and obtain distributions for Key Peformance Indicators (KPIs). We then perform deterministic optimization informed by patient-dependent variability of KPIs. Finally, multi-objective optimization (MOO) via epsilon-constraint is used to address trade-offs among cost, therapy return time, and transport emissions. Environmental impact considerations were integrated throughout our work to align with broader sustainability commitments of the biopharma industry.

Keywords: CAR-T Cell Therapy, Supply Chain Modelling, Multi-Objective Optimisation (MOO), Global Sensitivity Analysis (GSA), Environmental Impact

1 Introduction

Chimeric Antigen Receptors (CAR) are recombinant antigen receptors genetically engineered to redirect a patient's T-cells to target and eliminate cancer cells[1,2]. CAR-T cell therapy follows an autologous supply chain comprising of three key stages: (1) leukapheresis to collect the patient's cells, (2) manufacturing of the therapeutic product, and (3) administration of the therapy back to the same patient[3].Throughout medical history, cancer treatment has relied on established modalities such as surgery, chemotherapy, and radiotherapy. Although they have had undisputed effectiveness in reducing cancer mortality, they are often accompanied by the significant drawback of causing collateral damage to healthy tissues. Since its first clinical approval in 2017, CAR-T cell therapy has emerged as a promising way to further enhance overall survival while reducing systemic toxicity[4], in many cases significantly increasing survival time from the low overall median of 6 months. Currently, there are a total of six CAR-T cell therapy products approved by the U.S. Food and Drug Administration Agency (FDA) and European Medicines Agency (EMA) for the treatment of seven different B-cell malignancies[4], and the market is expected to grow to 16 billion by 2032.[5]

The growing demand for CAR-T cell therapy has intensified interest in the field while introducing increased complexity and pressure on existing facilities. To ensure successful therapies and manage growth effectively, the entire supply chain must be managed. Given the autologous nature of the process, efficient patient scheduling is critical[3], and digital tools are essential for effective supply chain modelling and optimisation. Significant progress has been made in this area[6–8], but continuous refinement of these tools remains critical. In particular, incorporating uncertainties into the modelling process is required to assess the supply chain's robustness.

As a life-saving therapy, addressing financial feasibility and timely delivery is essential to improve treatment accessibility, affordability, and outcomes. Therapies are sold at high list prices ranging from $373,000 to $475,000 per treatment[9]. Moreover, prolonged therapy return times pose a serious risk, increasing the likelihood of mortality and disease progression before CAR-T cell infusion[10]. The manufacturing process represents the most critical, costly, and time-intensive stage in the CAR-T cell lifecycle[3]. While the literature on CAR-T cell manufacturing has quantified cost ranges[11–13], it has not conducted systematic uncertainty analysis as done in related contexts[14–16]. Building on the Global Sensitivity Analysis (GSA) framework of Sarkis et al.[15], we extend such methods to CAR-T cell manufacturing to derive quantified distributions for manufacturing KPIs (Costs and Time) and identify key bottlenecks.

An extensive number of methodologies have been developed for optimisation under uncertainty[17]. Sarkis et al.[18] have applied stochastic optimisation in biopharma, in the context of viral vectors, but no uncertainty considerations have been made for CAR-T cell therapy. In this work, we augment the deterministic optimisation of CAR-T supply chains presented in Triantafyllou et al.[6] and integrate information on patient-dependent variability at the manufacturing level. This allows a quantification of the impact of manufacturing uncertainty on optimal supply chain network structures and schedules and a consideration of whether it is worth incorporating through stochastic programming.

The biopharma industry, contributing 4.4% of global emissions, is increasing its focus on sustainability to meet Net Zero targets[19]. In the CAR-T cell manufacturing step, while specific studies on

the environmental impact are limited, related research suggests single-use systems reduce energy use and water consumption by up to 80% compared to traditional bioreactors[20].Additionally, single-use facilities show a 30% lower global warming potential due to reduced electricity usage, offsetting plastic use[21]. Multi-objective optimization models of pharmaceutical supply chains have been proposed in literature, considering the three sustainability pillars: economic, environmental and social[22]. In our research we augment the supply-chain model of Triantafyllou et al.[6] by tracking environmental impact in the manufacturing and transport steps, and perform Multi-Objective Optimisation (MOO) to investigate trade-off.

2 Methodology

Our methodology, outlined in Figure 2.2.2.1 , can be subdivided into three main sections:

i) Defining the input uncertainty space for the manufacturing process and establishing the corresponding Key Performance Indicators (KPIs) to evaluate cost, time, and environmental impact.
ii) Performing Global Sensitivity Analysis (GSA) to determine critical parameters, and Uncertainty Analysis (UA) to generate distributions for the chosen manufacturing KPIs.
iii) Incorporating the outputs from step (ii) into a supply-chain simulation model to perform optimisation under uncertainty and evaluate trade-off among therapy cost, return time, and environmental impact via epsilon-constrained Multi Objective Optimisation (MOO).

2.1 Defining Manufacturing Input Uncertainties and Key Performance Indicators

2.1.1 Key input manufacturing uncertainty

In the first step, we define the input uncertainty space for the manufacturing process, comprising variables that fluctuate within ±50% of their nominal values. These nominal values are sourced from SuperPro's 'Autologous Cell Therapy' techno-economic process flowsheet, leveraging its comprehensive industry-specific databases. Process scale is set at 150 mL of frozen leukopaks, with key parameters (e.g., temperature, pressure, and volume) maintained at industry-optimized standards for this scale.

Input uncertainties are identified in the cost and throughput space. On the throughput side, we consider process times. Process times in SuperPro are specified for all sub-processes within one operation unit; only times in the order of 10^3 seconds were selected. These include: Activation, Transduction and Batch Phase time within Batch Kinetic Fermentation; Cell Washing-Concentration Time; and Micro-filtration Time. Perfusion Phase Time (a sub-process of Cell Expansion) was not included as it is set equal to the Micro-filtration Time, thus not being an independent parameter.

Only resource costs were considered for cost-related inputs, as labor, quality control, and equipment costs are inherently less variable. The input space was narrowed to five key resource costs—Medium Solids, Viral Stock, HSA, Ab Complex Mix, and Leukopak—based on the variability of raw material and stock mixture costs.

Viral vectors and Human Serum Albumin (HSA) are seen as the most uncertain. Viral vectors constitute a major portion of CAR T-cell therapy costs and bring significant uncertainties due to manufacturing susceptibilities[16]. The industry's immaturity and challenges in rapid scale-up can lead to shortages during sudden demand increases, as occurred during the COVID-19 pandemic. Human Serum Albumin (HSA) can be either human-derived plasma HSA (pHSA) or recombinant HSA (rHSA). Both production methods carry challenges, which make HSA subject to shortages. pHSA is limited especially by supply-chain dependence on blood donations, whilst rHSA is limited due to the complexity of protein production, folding, and function.[23]

Leukopak cost, incorporating leukapheresis costs into manufacturing, was treated as an uncertainty to propagate through the supply chain model. Medium solids and Antibody Complex Mix, while not inherently uncertain, were included due to their high costs to evaluate the impact of cost variability.

2.1.2 Key Performance Indicators (KPIs)

The model outputs being monitored are classified into three primary categories based on key performance indicators (KPIs): cost-related metrics, cycle processing time, and environmental parameters.

Cost-related outputs are further subdivided into annual capital costs, annual fixed operating costs and variable operating costs.

$$\text{Capital Cost} = \text{Direct Fixed Capital Costs} + \text{Start-Up Costs} + \text{Working Capital} \quad (2.1.2\text{-}1)$$

Capital costs are calculated as shown in Equation 2.1.2-1 . In this context, Direct Fixed Capital (DFC) refers to the initial investment in fixed assets, which include equipment, process piping, instrumentation, and infrastructure such as buildings and facilities. Working capital comprises expenses associated with labor, raw materials, utilities, and waste management. In SuperPro, start-up costs are estimated as a percentage of the DFC. To account for the temporal aspect of capital expenditures, the total capital cost is annualized by dividing the sum by the expected operational lifespan of the manufacturing facility.

Operating costs include expenditures for resource consumption during production, such as raw materials, consumables, utilities, labor, quality control (QC), quality assurance (QA), and facility-specific operational expenses.

Cycle time is defined as the total duration required to process a batch, from the upstream unit operations (USP) to the final fill-and-finish stage.

Environmental impact is evaluated by quantifying carbon dioxide (CO_2) emissions, water usage, and waste generation.The E-factor is used to compute the amount of waste produced relative to the amount of desired product generated, as outlined in Equation (2.1.2-2).

$$\text{E-factor} = \frac{\text{Total Waste Produced}}{\text{Product Produced}} \quad (2.1.2\text{-}2)$$

2.2 Global Sensitivity Analysis and Uncertainty Analysis

Global Sensitivity Analysis (GSA) and Uncertainty Analysis (UA) are performed to: (1) quantify the effect of uncertain inputs on KPI variability, and (2) quantify the distribution of the chosen KPIs. Both steps are performed with a SobolGSA-MatLab-Component Object Model (COM)-SuperPro interface.

2.2.1 Global Sensitivity Analysis (GSA)

In Global Sensitivity Analysis (GSA), inputs are varied simultaneously to observe the effect on output variance. Random sampling-high dimensional model representation (RS-HDMR) is used to generate 4096 quasi-random input samples that sufficiently represent the input space. Input contribution to output variance is quantified with Sobol indices, where first order indices (S_i) show the contribution of a single input i, and second order indices (S_{ij}) show the joint contribution of inputs i and j. The overall contribution of a single input (S_T) is calculated as $S_T = S_i + S_{ij}$, neglecting higher order contributions. Inputs with total Sobol sensitivity indices $S_T \geq 0.05$ were classified as critical inputs.

2.2.2 Uncertainty Analysis (UA)

For Uncertainty Analysis (UA) the techno-economic model was simulated for 4069 quasi-random input samples, considering uniform input distributions. KPI distributions are quantified and visualised through violin plots. Median and interquartile ranges for each KPI were extracted to compare variability across process parameters.

2.3 Supply Chain Scenario Modeling

The supply chain is modelled using the i-SHIPMENT platform for Individualised Supply chain Optimization in Personalised Medicine Treatments, as developed by Tryantafyllou et al.[6]. This is a Mixed Integer Programming (MIP) model which determines potential supply-chain configurations that minimise total cost, whilst meeting demand and a target therapy return time constraint. The model is structured with demand based across four major UK cities - represented by the leukapheresis and hospital sites (c_i and h_i, $i = 1...4$); whilst manufacturing sites ($m_i, i = 1...6$) are located across the US and Europe. There are two options ($j = 1,2$) for transport between each node: a lower-cost 2-day standard delivery, or a higher-cost 1-day express delivery service. Storage is not included, so each therapy p travels sequentially from the leukapheresis site, through the manufacturing site, to the hospital. In this context, therapy return time is defined as the time steps elapsed from when therapy enters the leukapheresis node to when it is administered to the patient at the hospital node.

In this work, we focus on updating the manufacturing-step of the supply chain model, adjusting the values of four key manufacturing KPIs - capital cost, fixed operational cost, variable operational cost, and manufacturing time - based on the outputs from the uncertainty analysis outlined in Section 2.2.2. This results in an updated objective function, shown below:

$$\min \text{TOTAL COST} = \sum_m \left(\frac{\text{FCAP}_m}{\text{CAPMAX}_m} \cdot \text{CIM}_m + \text{E}_m \cdot \text{CFVM}_m \right) + \sum_p \text{CVMQC}_p + \sum_p \text{TTC}_p \quad (2.3\text{-}1)$$

We modify variable operational costs ($CVMQC_p$) and manufacturing time to be patient-indexed, to assess the impact of patient-specific uncertainty on optimisation results. Simulations are run for a trimester, considering a demand of 50 patients per trimester. Given the nature of Sobol sampling, extracting the first 50 output samples generated through the uncertainty analysis outlined in Section 2.2.2 represents the landscape of the input space sufficiently.

Then, we assess the impact of uncertainty in facility-dependent costs (capital investment CIM_m and fixed operational cost $CFVM_m$) comparing three scenarios: best-case (with parameters held at lower quartile values), median-case, and worst-case (with parameters held at upper quartile values).

To investigate the trade-off between therapy return time (TRT) and cost, a multi-objective optimisation was performed by integrating the epsilon constraint method. The problem is reformulated to having a single objective: minimising cost, subject to constrained maximum TRT values.

Four scenarios were simulated:

1. TRT $\leq$ 25 days.
2. TRT $\leq$ 23 days.
3. TRT $\leq$ 23 days, with the average TRT $\leq$ 21 days.
4. TRT $\leq$ 23 days, with the average TRT $\leq$ 20 days.

Note that all previous simulations use the return time constraints outlined in point 3 above.

2.3.1 Environmental considerations

To consider the environmental impact of the manufacturing process, we add patient-indexed parameters for CO_2, water use and waste production, to track each metric at the supply chain level. Beyond manufacturing, the other key aspect of the supply chain that has large environmental footprint is transport. Hence, a methodology was developed to estimate CO_2 emissions for transport. Two transport steps have to be considered: transport from leukapheresis site to manufacturing site and transport from manufacturing site to hospital; however, as hospital and leukapheresis sites are assumed to be co-located, emissions mirror each other.

There are six manufacturing sites, outlined in Table 2.3.1.1; for those present in the UK (m_1, m_6) we assume transport via freight lorry, whilst for those outside of the UK (m_2, m_3, m_4, m_5) we assume air transport. They are assumed to have emissions of respectively 8.52083 $(10^{-4} kg_{CO2})/km$ and 1.428055 $(10^{-4} kg_{CO2})/km$. Both modes of transport are assumed to operated with a reefer (refrigerated container), which

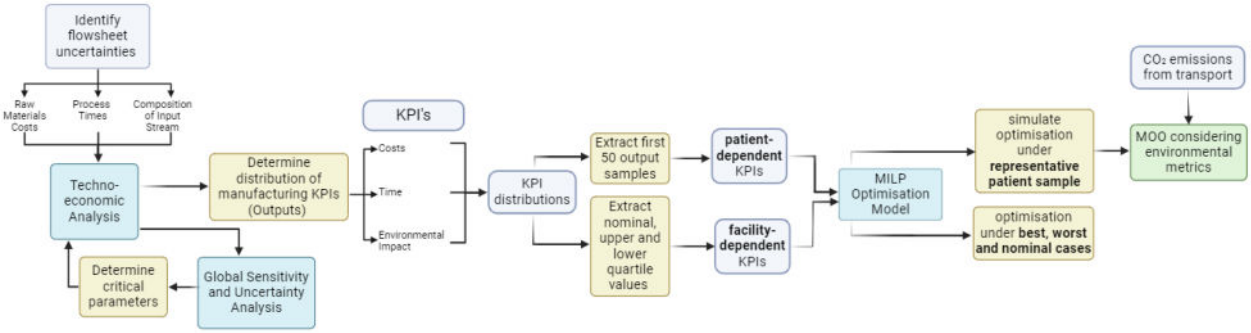


Figure 2.2.2.1 : Overview of Research Workflow

has operation mode of freezing (temperature between -35℃ and -18℃).[24]

Table 2.3.1.1 : Location of Manufacturing Facilities

Index	Location	Manufacturing Lines
m1	UK/Stevenage	4
m2	UK/Glasgow	31
m3	EU/Berlin	10
m4	EU/Belgium	4
m5	USA/Pensylvania	31
m6	USA/Virginia	10

We ran a multi-objective optimisation (MOO) utilising the epsilon-constraint method. The MOO problem was reformulated as a single-objective optimisation problem where the objective was to minimise cost subject to constraining the second objective, bounded by an epsilon value, maximum CO_2 emissions. Total CO_2 emissions from manufacturing and transportation sections where constrained iteratively from 250 kg to a minimum of 20 kg, to assess its impact on network structure and average therapy costs. To then include water usage and E-factor into the model's decision making process, we develop an overall environmental metric, and constrain it to a maximum value per patient.

$$\text{overall}_{\text{env_metric}} = W_1 \cdot \text{CO}_2^{\text{scaled}} + W_2 \cdot E_{\text{factor}}^{\text{scaled}} + W_3 \cdot \text{WATER_USE}^{\text{scaled}} \quad (2.3.1\text{-}1)$$

3 Results & Discussion

3.1 Global Sensitivity Analysis and Uncertainty Analysis

3.1.1 Economic Analysis of the Manufacturing Process

A comprehensive global sensitivity and uncertainty analysis was conducted to quantify the influence of raw material costs and operational process times on critical economic outputs, including annualized capital costs, fixed annual operating costs, and variable operating costs per batch.

The sensitivity of capital costs to uncertain inputs is illustrated in Figure 3.1.1.1. Capital costs exhibit a pronounced sensitivity to micro-filtration time and viral vector stock expenses, with first-order sensitivity indices of 0.31 and 0.56 respectively, showing the interdependence between time-related inputs and cost-outputs. Despite this, capital costs display limited variability, as a significant proportion of these costs are associated with fixed expenditures. The annual capital cost has a median value of $\$2,972,636$, with inter-quartile ranges of \$ 2,968,393 - \$ 2,976,943, reflecting a maximum percentage change of just 0.3%. This stability can be attributed to equipment purchase costs, which remain constant and represent approximately 70% of the total capital investment required for the manufacturing process.

Annualized fixed operating costs, calculated in units of \$/yr, further delineate the economic implications of bottleneck operations. As depicted in Figure 3.1.1.1, first-order sensitivity indices highlight the relatively weaker contribution of cost-related inputs compared to parameters influencing process bottlenecks. Key drivers, such as microfiltration time, significantly impact bottleneck lengths and, by extension, the number of batches produced annually. Specifically, the first order sensitivity indices for microfiltration, activation in the gas permeable bag, and batch phase duration were determined to be $S_i = 0.64, 0.16,$ and 0.16, respectively. These findings underscore the critical role of optimizing process bottlenecks to improve scalability and reduce fixed operating costs. The recorded fixed operating cost has a median value of $\$6,990,991$, with inter-quartile ranges spanning $\$6,600,486 - \$7,379,639$ resulting in a percentage change of 11.8%. The violin plot can be seen in Figure 3.1.1.1.

The variability in variable operating costs per batch is predominantly driven by viral stock costs, with a first-order sensitivity index of $S_i = 0.73$. The viral vector stock has the highest contribution among the raw material costs classified as uncertain. The direct correlation between the magnitude of the cost contribution of each raw material per batch and the sensitivity indices is highlighted in Table 3.1.1.1. In contrast, medium solids exhibit a weaker first-order sensitivity ($S_i = 0.03$), despite their larger throughput volume, due to their nominal cost being approximately 25 times lower than that of viral stock. The variable operating costs per batch have a median value of $\$8,859$ with interquartile ranges spanning $\$8,181 - \$9,531$, resulting in a percentage change of 17% as displayed in Figure 3.1.1.1.

These results underscore the interdependence between process

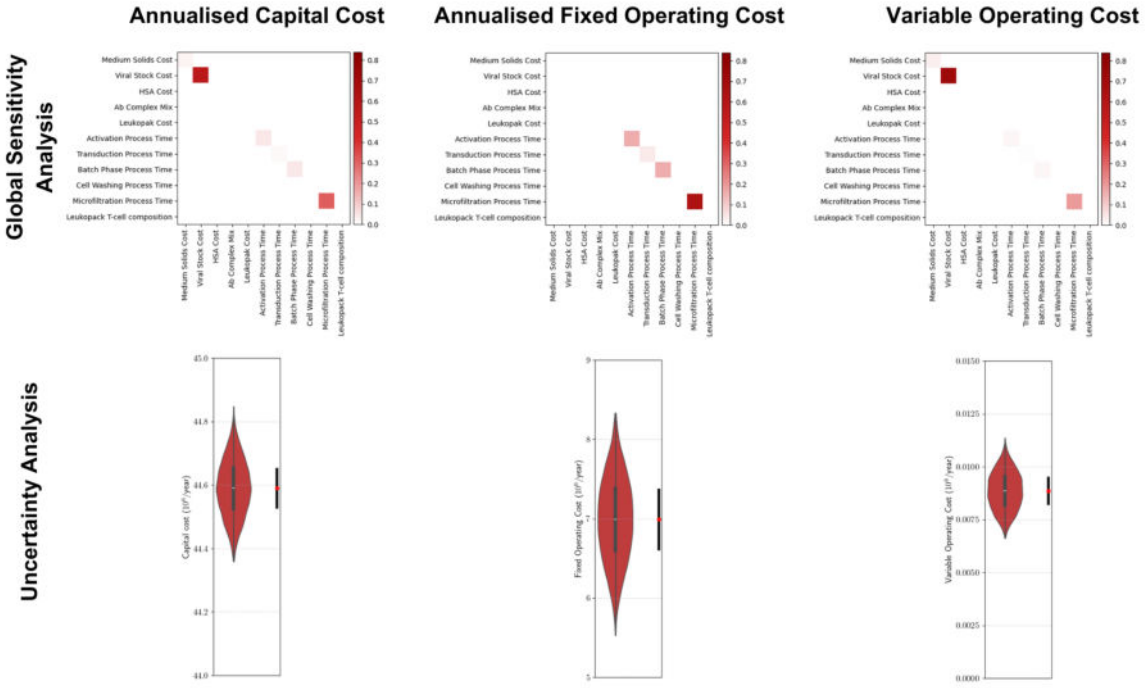


Figure 3.1.1.1 : Global Sensitivity and Uncertainty Analysis on Economic KPIs

Table 3.1.1.1 : Cost Contribution per Batch of Raw Materials

Item	Cost Contribution ($/batch)	Sensitivity Indices
Viral Stock	2666	0.73
Medium Solids Costs	612.41	0.032
Leukopak Costs	1.56	6.91×10^{-12}
HSA Costs	57.54	8.57×10^{-12}
Ab Complex Mix	1.25	2.82×10^{-11}

inputs and cost-related outputs, highlighting how raw material costs and process time uncertainties propagate through the system. Furthermore, the proportional relationship between the nominal value of inputs and their corresponding sensitivities emphasizes the criticality of high-cost materials, such as viral stock, in determining the overall economic variability of CAR-T manufacturing.

3.1.2 Time Analysis of the Manufacturing Process

The analysis of the results presented in Figure 3.1.2.1 provides critical insights into the dynamics of CAR-T manufacturing. It demonstrates that micro-filtration time exerts the greatest sensitivity on the overall process, as its duration significantly surpasses that of other operations. This finding highlights the pivotal role of upstream processing in determining cycle time efficiency. Table 3.1.2.1 quantitatively confirms this relationship, showcasing the disproportionately high sensitivity index for micro-filtration relative to other operations.

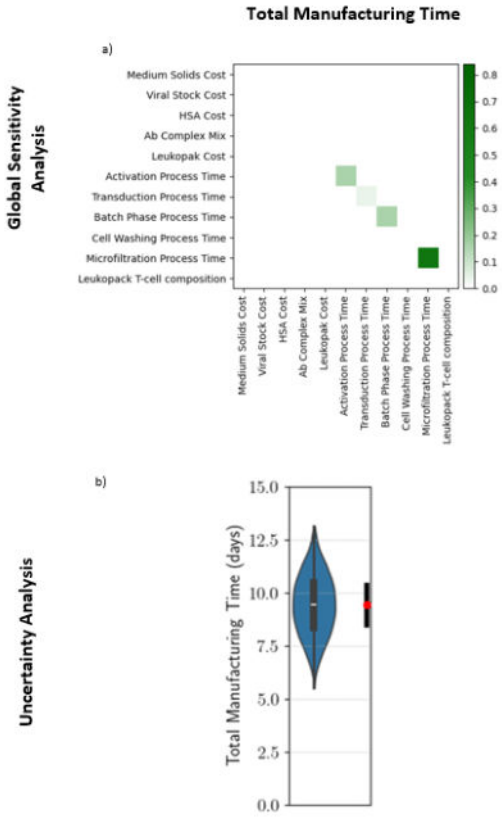


Figure 3.1.2.1 : Global Sensitivity and Uncertainty Analysis on Total Manufacturing Time

Furthermore, the duration of the perfusion phase, a subset of the cell expansion process, is set equal to the micro-filtration time as modelled by SuperPro Designer, hence leading to substantially greater sensitivity index for micro-filtration time compared to the other procedures. The observed sensitivity indices for activation in a gas-permeable bag and in the batch phase during cell expansion further highlight the correlation between nominal values and sensitivity indices. Since both procedures share the same nominal value, they exhibit identical sensitivity indices. The manufacturing process time has a median value of $815,034s$ (9.43 days) with interquartile ranges spanning $723,546s - 906,510s$ (8.37 - 10.49 days), resulting in a percentage change of 25% as displayed in Figure 3.1.2.1

Table 3.1.2.1 : Sensitivity Indices of Process Time Lengths

Operation	Duration (s)	Sensitivity Indices
Micro-filtration Time	345600	0.64
Activation in a Gas Permeable Bag	172800	0.16
Batch Phase in Cell Expansion	172800	0.16
Transduction in a Gas Permeable Bag	86400	0.04
Cell Washing Concentration Time	1200	7.72×10^{-6}

3.1.3 Environmental Analysis of the Manufacturing Process

Sensitivity analysis for total water usage and CO_2 emissions are depicted in Figure 3.1.3.1 . The high sensitivity of these metrics to micro-filtration time arises from the interdependence between the micro-filtration and cell expansion processes. Notably, cell expansion accounts for 95% of total emissions and approximately 63% of total water consumption in the manufacturing process.

The total water usage per batch has a median value of $1.87kg$ with interquartile ranges spanning $1.55kg - 2.19kg$,resulting in a percentage change of 42% as displayed in Figure 3.1.3.1 . Moreover, the CO_2 emissions per batch have a median value of $0.01kg$ with interquartile ranges spanning $0.007kg - 0.014kg$, resulting in a percentage change of 100%, as displayed in Figure 3.1.3.1 .

As outlined in Section 2.1.2, the E-factor, a critical metric for evaluating the sustainability of manufacturing processes, is directly proportional to the waste generated by the system. In this analysis, the output stream S-26 from the micro-filtration procedure is identified as the predominant contributor, generating approximately 1600 mL per batch, which accounts for 62% of the total waste produced. Finally, the E-factor across all samples has a median value of 29.16 with interquartile ranges spanning $25.18 - 33.13$, resulting in a percentage change of 32% as displayed in Figure 3.1.3.1

The global sensitivity analysis determined that the manufacturing steps with the biggest impact on the environment were micro-filtration and cell expansion. Their significant contribution highlights the need for targeted optimisation of these steps to improve the sustainability of the manufacturing process.

3.2 Supply Chain Model

A key objective of this work was to assess how variability in manufacturing KPIs propagate to the supply chain and affect optimal decisions. Using the i-SHIPMENT platform[6], we introduced patient-dependent variability in manufacturing time, variable operational costs and environmental metrics, and scenario-based variability in facility costs.

3.2.1 Therapy Cost and Return Time analysis

Introducing KPI variability results in cost distributions across patients for all three scenarios shown in Figure 3.2.1.1 . While the average variability is $6,002, the high magnitude of the costs limits the maximum percentage variation from the nominal value to just 0.5%. The supply chain model facilitates the calculation of an average cost per patient; extending our methodology to larger demand scales can provide industry with a foundational metric to estimate the minimum marketable price of the therapy, where each patient is charged a uniform amount to distribute the cost uncertainty.

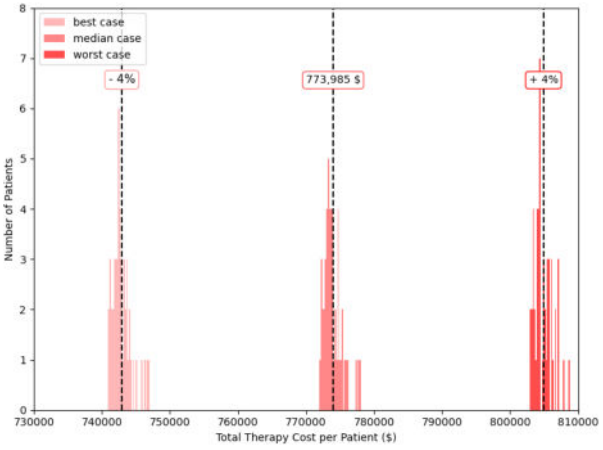


Figure 3.2.1.1 : Total Cost variation at best (lower quartile), median and worst case (upper quartile) values for facility dependent costs.

Facility-dependent costs were varied to evaluate their impact on total cost per therapy under worst-case (upper quartile) and best-case (lower quartile) scenarios, as shown in Figure 3.2.1.1 . Capital costs and fixed operational costs deviate by ±0.14% and ±5.55% from their nominal values, respectively. This uncertainty propagates through the supply chain, leading to a ±4% variation in the total cost per therapy.

Figure 3.2.1.2 shows the transport cost for four scenarios that constrain the therapy return time (TRT) at different values. At a TRT of 25, average cost per therapy is at its minimum, as the TRT constraint remains inactive, because all patients are able to receive their therapy within the set time limit whilst selecting the cheapest, yet longer, transport options at both nodes. Constraining TRT to 23 days increases the cost only marginally, as only one therapy (with the highest manufacturing time, 13 days) is affected by this. To constrain other therapies' return times whilst ensuring model feasibility, an average return time variable is added and constrained to 21 and 20 days. As

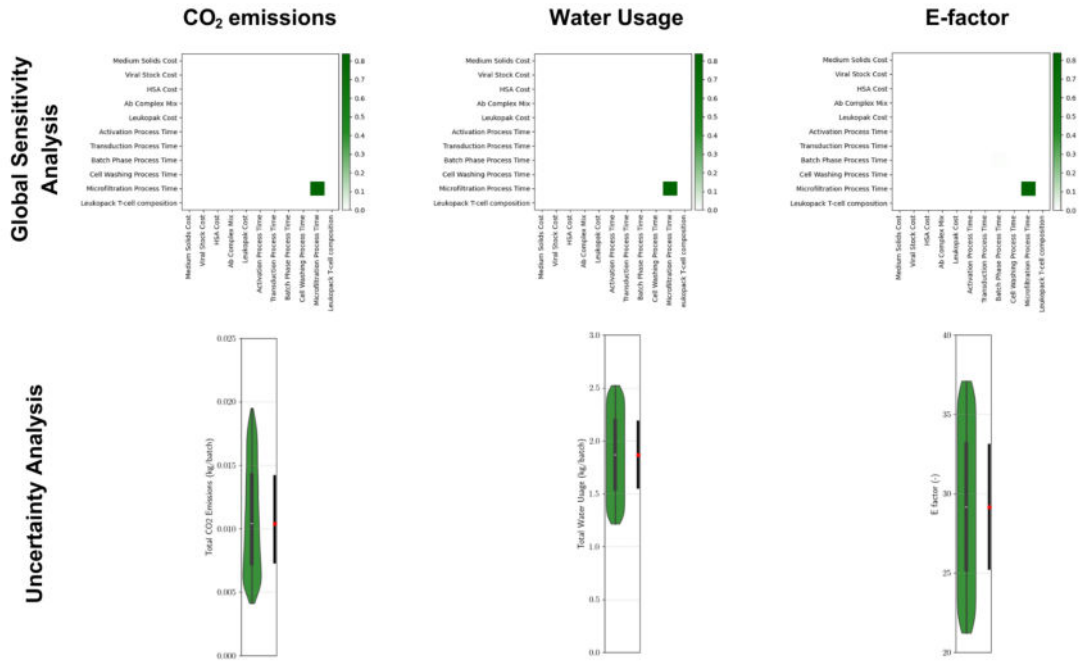


Figure 3.1.3.1 : Global Sensitivity and Uncertainty Analysis for Environmental Metrics

can be seen, the transport cost increases significantly in the 20-day average return time case. Network structure is driven by minimising total costs, but given the higher magnitude of capital costs as opposed to transport costs, the resulting network minimises facilities built, essentially not considering that higher transport costs are incurred to ship the therapy to further away locations. As shipping time does not depend on distance, the network structure remains the same for all model realisation, with all therapies being shipped to a single manufacturing facility in Berlin (m_3). This is true at low demand scales, where facilities are always underutilized. However, with higher demand, changes could be observed, as facilities are more likely to be fully utilised. Stricter time constraints may require faster transport to ensure patients arrive earlier at facilities, which restricts the optimization model's flexibility in patient-scheduling. This can potentially lead to facility overutilization at specific times. As a result, patients may be redirected to alternative facilities, driving changes in the network structure.

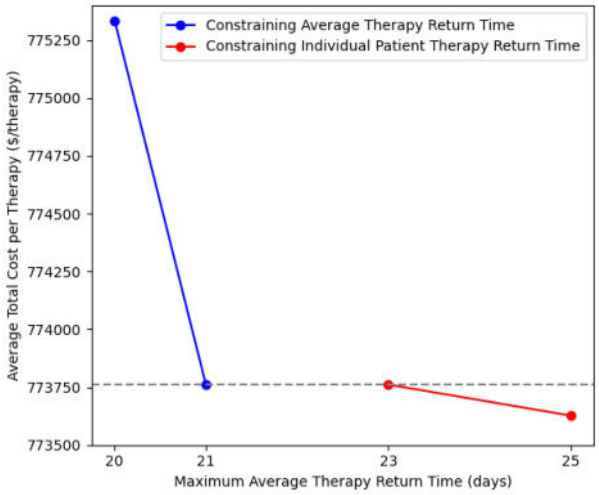


Figure 3.2.1.2 : Effect of Return Time Constraints on Average Transport Cost

3.2.2 Supply Chain Environmental Considerations

In this section we investigate the trade-off between environmental impact and therapy costs. Figure 3.2.2.1 shows that constraining total CO_2 emissions increases average cost per therapy. This is entirely dependent on the feasible space of network structures available to the optimisation model. When CO_2 emissions are constrained to 250 kg, the constraint remains inactive, and the network is built choosing facility m_3 (EU/Berlin). This facility has 10 manufacturing lines and is able to cover all demand; hence, the cost of only one facility is distributed across all patients. If we constrain emissions further,

in the range of 70-150 kg, some therapies are sent to facility m_1 (UK/Stevenage), as transport emissions are lower due to lower distances and less polluting mode of transport (truck). Further constraining emissions below 70 kg increases the cost significantly to use only facilities located within the UK (m_1 and m_2); m_2 incurs higher capital costs due to its 31 parallel lines, far exceeding demand requirements. This leads us to a discussion based on how ethically sound it is to increase the cost of a life saving drug, where accessibility is decreased, based on environmental grounds purely.

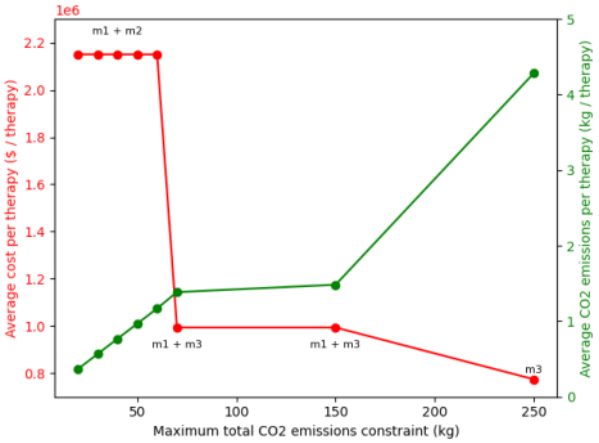


Figure 3.2.2.1 : Variation of Total Cost with Constraining Total CO_2 Emissions

An overall environmental metric has been developed to incorporate the effects of water usage and waste production into the model's decision-making process. Each metric is scaled and weighted by W_i, representing its relative contribution to global greenhouse gas (GHG) emissions. Waste contributes 12% of global GHG emissions, driven by landfill methane, industrial waste management, and improper disposal practices[25]. CO_2 accounts for 82% of global emissions, according to the IEA and IPCC[26], while water usage contributes 6%, accounting for energy-related water usage[27]. The resulting environmental metric integrates these contributions, weighted accordingly.

$$\text{overall}_{\text{env_metric}} = 0.82 \cdot \text{CO}_2^{\text{scaled}} + 0.12 \cdot E_{\text{factor}}^{\text{scaled}} + 0.06 \cdot \text{WATER}^{\text{scaled}} \quad (3.2.2\text{-}1)$$

Figure 3.2.2.2 shows the distribution of the overall environmental metric for our patient sample, when total CO_2 emissions are constrained to 70 kg. When unconstrained, the overall environmental metric ranges from 0.375 to 0.984 per therapy. Under stricter constraints, the distribution shifts left as transport options with lower emissions are prioritized. Future research should incorporate water usage and waste production throughout the supply chain, going beyond the manufacturing stage, ensuring that CO_2 emissions from transport are not the sole decision variables.

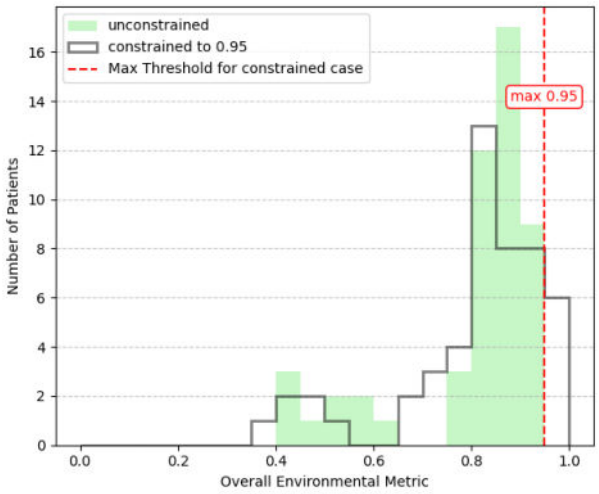


Figure 3.2.2.2 : Distribution of Patients Across Overall Environmental Metric

4 Conclusion

CAR-T cell manufacturing poses significant challenges that limit scalability, affordability, and accessibility. The highly personalized nature of CAR-T therapy, relying on autologous cell sources, introduces complexity in the supply chain. The industry faces hurdles in achieving consistent product quality due to variability in patient-derived cell material and difficulties in rapidly scaling manufacturing infrastructure to meet increasing demand. These challenges are further compounded by the environmental impact of resource-intensive processes, which remain a growing concern as the biopharma industry strives to align with sustainability targets.

Through the application of Global Sensitivity Analysis (GSA) and Uncertainty Analysis (UA), critical bottlenecks in the manufacturing process were identified, highlighting the substantial influence of micro-filtration time and viral vector costs on both time and economic outputs. Furthermore, we augmented the manufacturing uncertainty analysis done in previous literature to account for environmental footprint, quantifying E-factor, water-usage and CO_2 emissions. Microfiltration and cell-expansion were found to be the manufacturing steps with the greatest environmental footprint.

Table 4.1 summarizes the nominal values and associated uncertainty ranges for key performance indicators (KPIs). The

Table 4.1 : Summary of Process KPIs Nominal Value and Uncertainty Ranges

Process KPIs	Nominal Value	Percentage Range
Annual Capital Cost	$2,972,636/year	0.3%
Annual Fixed Operating Cost	$6,990,991/year	11.8%
Variable Operating Cost	$8,859/patient	17%
Total Manufacturing Time	9.43 days	25%
E-factor	29.16	32%
Water Usage	1.87 kg/batch	42%
CO_2 Emissions	0.01 kg/batch	100%

percentage variabilities indicates that financial metrics are more stable, while environmental and efficiency-related KPIs are

subject to greater uncertainty, potentially highlighting areas for process optimization and risk mitigation.

The supply-chain model by Triantafyllou et al.[6] was enhanced to include patient-dependent variability in operational costs and manufacturing times, enabling population-representative cost estimates. Additionally, simulated quartile scenarios revealed that facility-dependent cost variability causes a ±4% variation in average therapy costs.

Furthermore, multi-objective optimization (MOO) using the epsilon-constraint method was performed to assess trade-off between total return time and CO_2 emissions against average therapy costs. Constraining therapy return time at lower demand scales only increases transport costs, because: (1) shipping time is independent of distance, and (2) facilities remain under-utilised and hence unaffected by reduction in patient-scheduling flexibility. The trade-off between cost and therapy return time highlights the importance of quantifying the value of a one-day reduction in return time for patients. This value varies significantly based on individual cases, as return times are influenced by manufacturing durations, which depend on patient-specific factors and the urgency dictated by the disease stage. The trade-off between costs and emissions prompts an ethical debate on raising the cost of life-saving drugs, reducing accessibility solely for environmental reasons. Supply chain emissions ranged from 18 kg to 214 kg CO_2 per therapy depending on the network structure, but decreasing emissions significantly increased costs ($773,985-$2,150,665). Additionally, to integrate all environmental factors into decision-making, an overall environmental metric was formulated by combining scaled values of CO_2 emissions, water usage, and waste production weighted by their relative impact.

5 Outlook

With growing demand and interest in CAR-T cell therapy, our work has addressed key gaps in the literature, though significant further efforts are required to unlock its full societal potential.

Future research should capture variability more accurately and widely. At the manufacturing level, our sensitivity analysis varied uncertain inputs by $\pm 50\%$ of their nominal values and did not account for patient-dependent variations (i.e. inlet blood stream composition) due to the stoichiometric nature of the flowsheet. Addressing this requires conducting industry-specific studies to obtain accurate parameter variability and transitioning to a kinetic model. At supply chain level, a tiered system could link cancer stage with treatment urgency, assigning time constraints and resource priorities based on disease severity. Moreover, research should consider how reductions in manufacturing time could impact therapy quality. Shorter manufacturing times, while beneficial for logistical efficiency, may compromise the stability or efficacy of sensitive treatments. Understanding this trade-off is crucial for optimizing operational timelines without sacrificing therapeutic success rates.

To form more representative conclusions, scaling model simulations to reflect real-world demand is crucial. Additionally, uncertainty considerations should be enhanced by performing stochastic optimisation. Taking manufacturing time variability as an example, this would encompass each patient's manufacturing time being sampled from its uncertainty distribution, to run several discrete scenarios with the aim of identifying the network structure that gives the best overall performance. Due to the high computational tractability of this formulation, the model would have to be simplified, sacrificing patient-variability and reducing the time horizon.

Finally, to build on the analysis of the direct environmental impacts of therapy manufacturing and transport considered here, future work should prioritize developing a comprehensive Life Cycle Assessment (LCA). Considering the potential supply chain structures discussed in the literature[3,8], LCA should be conducted across all configurations to ensure sustainability remains a key factor in future logistics decisions. Studies could further explore multi-objective optimization techniques, balancing competing goals such as cost efficiency, environmental responsibility, and clinical effectiveness.

References

[1] M. Sadelain, R. Brentjens, I. Rivière and J. Park, 'Cd19 car therapy for acute lymphoblastic leukemia,' *American Society of Clinical Oncology Educational Book*, no. 35, e360–e363, 2015, PMID: 25993197. doi: 10.14694/EdBook_AM.2015.35.e360, eprint: https://ascopubs.org/doi/pdf/10.14694/EdBook_AM.2015.35.e360, [Online]. Available: https://ascopubs.org/doi/abs/10.14694/EdBook_AM.2015.35.e360.

[2] B. L. Levine, J. Miskin, K. Wonnacott and C. Keir, 'Global manufacturing of car t cell therapy,' *Molecular Therapy - Methods Clinical Development*, vol. 4, pp. 92–101, Mar. 2017. doi: 10.1016/j.omtm.2016.12.006, [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5363291/.

[3] M. M. Papathanasiou, C. Stamatis, M. Lakelin, S. Farid, N. Titchener-Hooker and N. Shah, 'Autologous car t-cell therapies supply chain: Challenges and opportunities?' *Cancer Gene Therapy*, vol. 27, pp. 799–809, Jan. 2020. doi: 10.1038/s41417-019-0157-z.

[4] R. C. De Marco, H. J. Monzo and P. M. Ojala, 'Car t cell therapy: A versatile living drug,' *International Journal of Molecular Sciences*, vol. 24, p. 6300, Jan. 2023. doi: 10.3390/ijms24076300, [Online]. Available: https://www.mdpi.com/1422-0067/24/7/6300.

[5] fortune, *Car-t cell therapy market size, share | analysis report [2032]*, 2024. [Online]. Available: https://www.fortunebusinessinsights.com/car-t-cell-therapy-market-108455.

[6] N. Triantafyllou, A. Bernardi, M. Lakelin, N. Shah and M. M. Papathanasiou, 'A digital platform for the design of patient-centric supply chains,' *Scientific Reports*, vol. 12, Oct. 2022. doi: 10.1038/s41598-022-21290-5. (visited on 21/11/2024).

[7] P. Karakostas, N. Panoskaltsis, A. Mantalaris and M. C. Georgiadis, 'Optimization of car t-cell therapies

supply chains,' *Computers Chemical Engineering*, vol. 139, pp. 106 913–106 913, Aug. 2020. doi: 10.1016/j.compchemeng.2020.106913. (visited on 29/04/2023).

[8] K. Wang, C.-Y. Tseng, Z. Li *et al.*, 'A simulation-based comparison of centralized and point-of-care supply chain strategies for autologous cell therapy,' *Cytotherapy*, vol. 25, pp. 1370–1379, Sep. 2023. doi: 10.1016/j.jcyt.2023.08.007. (visited on 10/12/2024).

[9] Y. Li, Z. Hu, Y. Li and X. Wu, 'Charting new paradigms for car-t cell therapy beyond current achilles heels,' *Frontiers in Immunology*, vol. 15, p. 1 409 021, 2024. doi: 10.3389/fimmu.2024.1409021. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/38751430/ (visited on 08/06/2024).

[10] N. Ahmed, W. Wesson, M. U. Mushtaq *et al.*, '"waitlist mortality" is high for myeloma patients with limited access to bcma therapy,' *Frontiers in Oncology*, vol. 13, Aug. 2023. doi: 10.3389/fonc.2023.1206715. [Online]. Available: https://www.frontiersin.org/journals/oncology/articles/10.3389/fonc.2023.1206715/full (visited on 02/09/2024).

[11] M. Khang, S. Suryaprakash, M. Kotrappa, W. Mulyasasmita, S. Topp and J. Wu, 'Manufacturing innovation to drive down cell therapy costs,' *Trends in Biotechnology*, vol. 41, no. 10, pp. 1216–1219, 2023, issn: 0167-7799. doi: https://doi.org/10.1016/j.tibtech.2023.04.006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167779923001270.

[12] H. Kumar, A. Arunachalam, M. Yasar *et al.*, 'Decentralized manufacturing of anti cd19 car-t cells using clinimacs prodigy®: Real-world experience and cost analysis in india,' *Bone Marrow Transplantation*, vol. 58, pp. 1–8, Nov. 2022. doi: 10.1038/s41409-022-01866-5.

[13] A. G. Lopes, R. Noel and A. Sinclair, 'Cost analysis of vein-to-vein car t-cell therapy: Automated manufacturing and supply chain,' *Cell and Gene Therapy Insights*, vol. 6, pp. 487–510, Jun. 2020. doi: 10.18609/cgti.2020.058.

[14] N. Triantafyllou, N. Shah, M. M. Papathanasiou and C. Kontoravdi, 'Combined bayesian optimization and global sensitivity analysis for the optimization of simulation-based pharmaceutical processes,' *Computer Aided Chemical Engineering*, pp. 381–386, 2023. doi: 10.1016/b978-0-443-15274-0.50061-5. (visited on 10/12/2024).

[15] M. Sarkis, N. Shah and M. M. Papathanasiou, 'Characterization of key manufacturing uncertainties in next generation therapeutics and vaccines across scales,' *Journal of Advanced Manufacturing and Processing*, vol. 5, Apr. 2023. doi: 10.1002/amp2.10158. (visited on 10/12/2024).

[16] R.-M. Comisel, B. Kara, F. H. Fiesser and S. S. Farid, 'Lentiviral vector bioprocess economics for cell and gene therapy commercialization,' *Biochemical Engineering Journal*, vol. 167, p. 107 868, Mar. 2021. doi: 10.1016/j.bej.2020.107868.

[17] N. V. Sahinidis, 'Optimization under uncertainty: State-of-the-art and opportunities,' *Computers Chemical Engineering*, vol. 28, pp. 971–983, Jun. 2004. doi: 10.1016/j.compchemeng.2003.09.017.

[18] M. Sarkis, N. Shah and M. M. Papathanasiou, 'Resilient pharmaceutical supply chain: Assessment of optimisation strategies under uncertainty in network design problems,' *Computers & Chemical Engineering*, 2024.

[19] M. Sarkis, A. T. Fyfe, C. Kontoravdi and M. M. Papathanasiou, 'Towards a net zero, socially sustainable and eco-efficient biopharma industry: How far are we?' *Current opinion in chemical engineering*, vol. 44, pp. 101 027–101 027, Jun. 2024. doi: 10.1016/j.coche.2024.101027. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2211339824000285.

[20] M. Sarkis, A. T. Fyfe, C. Kontoravdi and M. M. Papathanasiou, 'Towards a net zero, socially sustainable and eco-efficient biopharma industry: How far are we?' *Current Opinion in Chemical Engineering*, vol. 44, pp. 101 027–101 027, Jun. 2024. doi: 10.1016/j.coche.2024.101027.

[21] A. Sinclair, L. Leveen, M. Monge, J. Lim and S. Cox, *The BioPharm International Guide: The Environmental Impact of Disposable Technologies - Can disposables reduce your facility's environmental footprint?* IQPC, 2008. [Online]. Available: https://www.iqpc.com/media/7763/11363.pdf.

[22] I. Duarte, B. Mota, T. Pinto-Varela and A. P. Barbosa-Póvoa, 'Pharmaceutical industry supply chains: How to sustainably improve access to vaccines?' *Chemical Engineering Research and Design*, Apr. 2022. doi: 10.1016/j.cherd.2022.04.001.

[23] M. Latta, M. Knapp, P. Sarmientos *et al.*, 'Synthesis and purification of mature human serum albumin from e. coli,' *Bio/Technology*, vol. 5, no. 12, pp. 1309–1314, Dec. 1987. doi: 10.1038/nbt1287-1309. [Online]. Available: https://www.nature.com/articles/nbt1287-1309.

[24] M. Spielmann, T. Kägi, P. Stadler and O. Tietje, *Life cycle inventories of transport services*, Ecoinvent Association, 2023. [Online]. Available: https://www.ecoinvent.ch.

[25] UNFCCC, *Food loss and waste account for 8-10% of annual global greenhouse gas emissions; cost usd 1 trillion annually*, 2020. [Online]. Available: https://unfccc.int/news/food-loss-and-waste-account-for-8-10-of-annual-global-greenhouse-gas-emissions-cost-usd-1-trillion.

[26] IPCC, *Chapter 16: Innovation, technology development and transfer*, n.d. [Online]. Available: https://www.ipcc.ch/report/ar6/wg3/chapter/summary-for-policymakers/.

[27] S. G. S. A. Rothausen and D. Conway, 'Greenhouse-gas emissions from energy use in the water sector,' *Nature Climate Change*, vol. 1, no. 4, pp. 210–219, Jun. 2011. doi: https://doi.org/10.1038/nclimate1147.

An imaged-based workflow to characterise multi-scale porosity in rocks

Dam Irene and Ghafir Yasmine

Department of Chemical Engineering, Imperial College London, U.K.

Abstract

In many applications including geological carbon storage, catalysis, separation processes, and contaminant transport, it is critical to characterise the porous space in porous media. An image-based workflow is presented to quantify microporosity distributions on a grain-by-grain basis using Ketton limestone (which has dual porosity) as the prototype. 3 sets of images acquired by X-ray tomography at 3 volume fractions (20, 40 and 100%) of the tracer sodium iodide (NaI) and background fluid sodium chloride (NaCl) were provided. The average microporosity found was 35%. The workflow also allows extraction of other key petrophysical properties such as micro- porosity distributions, sphericity, and pore throat size distribution. They were found to be in agreement with the literature and data obtained through mercury injection capillary pressure (MICP), reinforcing the reliability of the workflow.

1. Introduction

X-ray tomography relies on the assembly of two-dimensional pictures taken at different orientations, to form three-dimensional images. An X-ray beam is sent through the sample, which rotates in hundreds of different angles, and the attenuation of the X-ray source beam is measured by a Computational Tomography (CT) scanner. The two-dimensional images are then collected by an X-ray detector, and numerical algorithms are used in reconstruct the 3D images from this data (Ducheyne, 2017). Using a synchrotron X-ray radiation enables the emission of high energy electrons, which travel at speeds approaching light speed, forcing them to move in circular orbits. This acceleration results in the emission of a monochromatic radiation (Lett and Sinclair, 1993), which is much more intense than more commonly used laboratory X-ray generators (Brahme and Panetta, 2014). This leads to high X-ray sources, which result in short acquisition times (Feldman, 2018) and obtention of very high signal to noise ratio, which corresponds to the relative amount of signal compared to undesirable noise present in images (Semmlow, 2018). These recent improvements in X-ray tomography techniques have opened the door to the extraction of materials quantitative parameters. It also is a powerful and non-destructive tool providing accurate representation of the material's geometrical structure.

To limit the increase in temperature to only 1.5 ℃ above pre-industrial levels, many countries turned towards Carbon Capture and Storage (CCS) processes, which play a key role in the decrease of atmospheric carbon. Due to the challenges represented by carbon storage tanks which require extremely high pressures, geological carbon storage is being explored. Therefore, gaining a deep understanding in geological structures and properties is key to render gas storage processes safe and efficient (Zhong et al., 2024).

The main challenge is the dual porosity of rocks. In facts rocks usually present a microporosity and a macroporosity. Macroporosity is the pore space between the grains and in this study, it is larger than a few microns (Naresh *et al.*, 2020) There is also a microporosity which is the pore space within the grains themselves that are less than 1 micrometre. (Naresh *et al.*, 2020)

X-ray based techniques have been developed but they are inherently limited to the characterisation of macropores. This makes the characterisation of sub-micron pores especially challenging. In this work, a workflow was developed to segment the intergranular pore space and quantify the porosities for the sub resolution pore phase. The workflow was applied to a larger proportion of the rock (a quarter) than usually attempted in literature around a fifth (Lin *et al.*, 2016) and (Singh *et al.*, 2018) and further investigation into physical properties of the rock such as how sphericity may affect porosity was conducted. A greater depth quantitative analysis was done where key findings were compared to various literature sources and established techniques to test reliability and accuracy.

Key applications and motivation behind resolving microporosity includes the ability to fully describe flow and transport processes in porous media. For

example, a study found that accounting for sub-resolution porosity, changed the permeability of a system by one order of magnitude (Foroughi *et al.*, 2024), therefore also affecting the flow properties. Numerical transport models like pore network models often incorporate only the macropores, neglecting the impacts of micropores (Foroughi *et al.*, 2024). The work from this workflow can therefore be used to improve transport models and simulations to predict properties like permeability such that they align with experimental measurements.

2. Background

In literature, there are multiple reported methods to obtain porosity. This includes direct measurements such as helium porosimetry and mercury intrusion capillary pressure (MICP) measurements. While they can give accurate porosity values, they are unable to effectively resolve the entire pore space, particularly the micropores (Zhong et al., 2024). MICP specifically is also a destructive technique which means that the sample cannot be used for further analysis or experimentation. Imaging methods can achieve higher resolution. For example, scanning electron microscopy (SEM) can achieve resolutions of order sub nanometres (De Haan *et al.*, 2019) and nuclear magnetic resonance (NMR) which can operate up to 1.2GHz frequency (Wenchel, Gampp and Riek, 2024). However, this often means that the size of the sample imaged is quite small presenting limitations in analysing pore connectivity.

3. Methodology

3.1 MICP Experiments

Mercury Injection Capillary Pressure (MICP) consists of injecting mercury (Hg), a non-wetting gas into the porous material. The volume of the intruded mercury in the pore space in mL/g, is measured as a function of the pressure which is increased up to 32 000 psia. The initial data points where lower amounts of mercury penetrate the pores corresponds to mercury filling up void space in the penetrometer before reaching the sample and was hence disregarded for further calculations to increase accuracy. Since the mercury intrudes pores throats with a specific radius at a specific pressure, it allows to obtain a pore throat size distribution for the sample (Peng et al., 2017). The radius weighted-volume fraction was hence plotted as a function of the pore throat radius. The data was obtained for various Ketton samples, as well as Bentheimer sandstone samples, which was used as a control. This technique confirmed the presence of pores below 1 micrometre for the Ketton sample. The next step was then to establish a workflow to evaluate the intragrain microporosity of the studied Ketton sample.

3.2 Image Processing Workflow

3.2.1 Available Materials

The Ketton images provided were obtained through X-ray tomography performed at Diamond Source lab, using synchrotron radiation. Three image datasets were provided, all of dimensions 2560 x 2130 x 2560 $voxels^3$. Each were injected with a tracer fluid, containing 15wt% of NaI in 14.5wt% of carrier fluid NaCl. The iodide present acts as an X-ray contrasting agent, since its high molecular weight allows to absorb X-rays more effectively, leading to high pore-grain contrast. This mixture was injected at different volume fractions, obtaining three different datasets containing respectively 20,40 and 100 % volume fraction of tracer, which acted as an X-ray contrasting agent. Each image set contained grains, pores and air bubbles naturally present in the sample, as seen in Figure 1. as well as rubber englobing the rock. The resolution of the obtained images was 3 μm. These images were processed on the commercial software Thermo Scientifc Avizo version 2023.2.

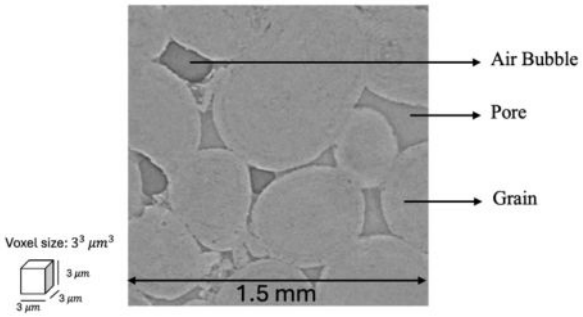


Figure 1. Slice of the extracted sub volume obtained through Avizo, with dimensions of 500 x 500 $voxels^2$

3.2.2 Grain mask

In order to identify the optimal mask, the image file corresponding to 20% volume fraction of NaI, which offers better contrast between pores and grains. This is since high volume fractions of tracer lead to denser fluid, which then tends to have similar greyscales to the grains.

A preliminary sub-volume corresponding to a cube of dimension 500^3 $voxels^3$ (3.375 mm^3) was

extracted to identify the optimal mask that would then be applied to the larger volume, as seen in Figure 2. (a).

The first step corresponded to the identification of the optimal filter parameters that would eliminate the noise present in the images, while minimising the blurriness of the boundaries between the grains and the pores, shown in Figure 2. (b). The choice was made to apply a Non-Local Mean (NLM) filter to our images. This filter assigns each voxel to a greyscale value obtained through the weighted average of the neighbouring voxels. The weighting is done through similarity: neighbouring voxels with higher similarity with the central voxel have a higher weight. The NLM filter is documented as particularly well suited to remove Gaussian noise (noise randomly distributed following a normal distribution of greyscale values), which is the predominant kind of noise present in images, while preserving grain edges (Wang et al., 2018a). Once the filter type was identified, the parameters were then selected. The main parameter was the Intensity Standard Deviation, which as it gets higher, smoothers out the image. The optimal value which allowed to remove the noise while preserving edges was found to be 0.8. The Spatial Standard Deviation and Search Window in pixels, respectively corresponding to the space in which the filter would be sensible to variations and the neighbouring region used to adjust each voxel, were both used at the intermediate value of 10. This allows to effectively capture the noise, while avoiding over-smothering the image.

The next step was to perform an image segmentation, shown in Figure 2. (c), to separate the grains from the pores. To do so, a grey value threshold was applied to identify grain and pore regions. Only small seed points in each material that could be unambiguously assigned to grains or pores were selected to mark each material, using a grey-scale value histogram. In this case, extremely low grey scale values were assigned to pores, while the highest greyscale values were assigned to grains. After marking both objects, a Watershed segmentation was performed to group the different voxels based on their intensity similarities, by filling the materials starting from the established markers (Rogowska, 2000). The watershed algorithm identifies local minimums in intensity functions. Then, these minimum points grow up through intensity, forming catchment basins until they reach neighbouring basins. Once the interface stops moving, all the voxels on the interface are assigned to a material (Andrew, 2014). The watershed algorithm is widely used in image processing, since it is a fast process which can separate materials even for poor contrasts (Aayushi, 2014). Once the pores and grains were identified, the pore material was removed to only keep the grains.

The Object Separation function, as seen in Figure 2. (d), was then applied to separate individual grains. This step requires the selection of an optimal marker extent to adequately separate each grain. Smaller marker extent would result in more separation while higher ones would merge grains together. After experimenting various marker extents, it was decided that the best grain separation which better captured the grain shape was 5.

The Remove Small Spots function was then applied to remove crushed grains, since Ketton contains a significant amount of grain fragments (Andrew, 2014). The 3D Volume Rendering was performed to the entire sub volume illustrated by Figure 2. (e). Then, the Border Kill function was also applied to only keep whole grains, as seen in Figure 2. (f), removing cropped ones, to ensure accurate further microporosity calculations. Finally, the label analysis was applied to label each grain.

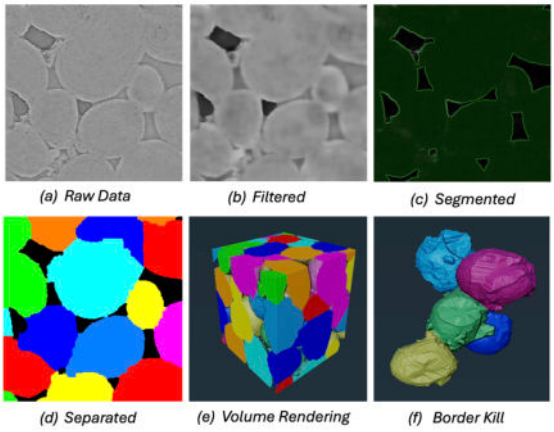


Figure 2. Workflow showing the different image processing steps performed for 500^3 *voxels*^{3}

3.2.3 Pore mask

The workflow was then applied in similar way for the pores, using the same filter parameters. The main difference resided in the segmentation, where a third material was introduced, corresponding to air bubbles present in the raw images. Once the

segmentation was performed in a similar way using grey-scale thresholding and watershed algorithm, the grains and air bubbles were removed to only account for the macropore space. The macropores were then separated and after the removing of small spots, volume rendering and border kill were applied, individual pores were applied. The volume fraction assigned to the pore material was used to determine the macroporosity of the rock sample. Once the macroporosity was identified using the segmentation, the next step was to quantify the intraparticle microporosity.

3.2.4 Microporosity quantification

When the tracer is injected into the rock samples, a concentration gradient is expected to be seen in the macropores, as well as within the grains themselves since the tracer penetrates the micropores. This leads to a change in X-ray attenuation within the grains, as well as different X-ray attenuations for the different tracer volume fraction image sets. These were then averaged for each grain for the three different tracer volume fractions. Since the Ketton images containing a tracer volume fraction of 20% offered optimal contrast between grains and pores, the latter was chosen as the mask to locate the labelled grains. The three raw files were then exported as well as the masked Tracer 20 file on Matlab. The choice was made to export them as raw files, in order capture as accurately as possible the details in the original image. The code calculated the average X-Ray attenuation throughout the three different concentrations for each individual label and traced the X-Ray attenuation fit line as a function of the tracer volume fraction for each grain label. The same method was applied for the pores, using the filtered and labelled pore segmentation. The microporosity of each grain was then calculated using the following formula:

$$\emptyset_i = \frac{\Delta I_{i,g}}{\Delta I}$$

Where $\emptyset_i$ corresponds the microporosity of the grain labelled i, $\Delta I_{i,g}$ corresponds to the gradient of the X-Ray attenuation against tracer concentration line for the grain labelled i, and ΔI corresponds to the average of all the slopes of the X-Ray attenuations versus tracer concentration obtained for the pores.

Having obtained the microporosity for each individual grain, the sphericity of each grain was also calculated using the following formula:

$$\psi_i = \frac{\pi^{\frac{1}{3}}(6V_i)^{\frac{2}{3}}}{A_i} \quad \text{[Equation 1]}$$

Where V_i and A_i correspond respectively to the volume and area of the labelled grain i, obtained through the label analysis done on Avizo.

3.2.5 Uncertainty calculations

In order to quantify the microporosities uncertainties, the rubber present in the raw datasets was used as a control, shown in Figure 3.

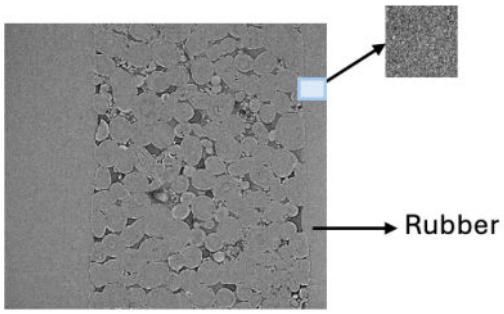


Figure 3. Slice of the raw images with dimensions of 2560 x 2560 $voxels^2$

Since it doesn't absorb any tracer, it was expected to have a constant X-ray attenuation value across all the datasets, independently of the tracer volume fraction. It was then assumed that any natural deviation present in the rubber would be propagated to the grains and pores. For the grains, several sub volumes similar to grain volume were extracted at different locations and depths, and the X-Ray Attenuation was plotted against the tracer concentrations for each of these locations. The standard deviation and mean of X-Ray Attenuation values for each tracer volume fraction were used to perform a Monte Carlo simulation, generating 10000 points, which allowed us to obtain a normal distribution that was used to get the interval of confidence of the X-ray attenuation lines. The same methodology was applied for the pores, with extracted sub-volumes corresponding to average pore volume. Therefore, for each labelled grain, the uncertainty in microporosity was calculated by applying the Standard Rule of error propagation below:

$$\delta f(x,y)^2 = \left(\frac{\partial f(x,y)}{\partial x} \times \delta x\right)^2 + \left(\frac{\partial f(x,y)}{\partial y} \times \delta y\right)^2$$

[Equation 2]

Since

$$\emptyset_i = \frac{\Delta I_{i,g}}{\Delta I} \quad \text{[Equation 3]}$$

Therefore, individual grain microporosity uncertainties were obtained:

$$\delta\emptyset_i = \left(\frac{1}{\Delta I}\right) \times \sqrt{\delta\Delta {I_{i,g}}^2 + \emptyset_i^2 \times \delta I^2} \quad \text{[Equation 4]}$$

4. Results and Discussion

The MICP experiment results were first analysed in Section 4.1. Trends in X-ray attenuation vs NaI volume fraction graphs were discussed in Section 4.4. A distribution of the sub-micron porosities was plotted and compared to literature in Section 4.5. Trends in sphericity are explored in Section 4.8 before interpreting connectivity of the macropores including the distribution of pore throats in Section 4.9.

4.1. MICP Analysis

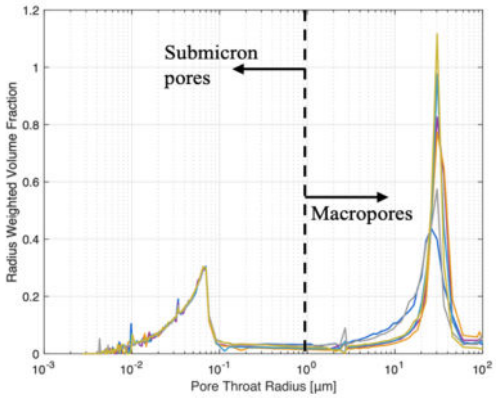


Figure 4. Ketton pore throat size distribution obtained through Mercury Injection Capillary Pressure (MICP)

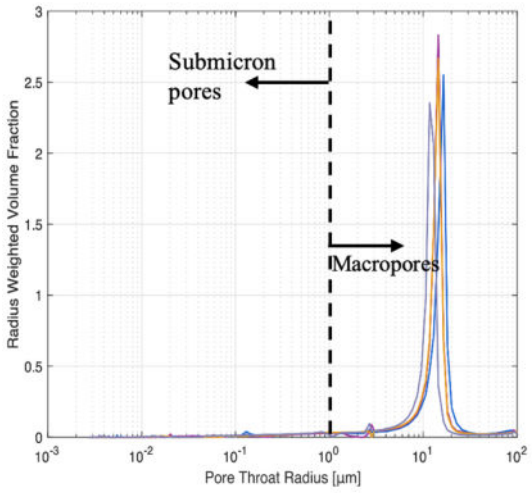


Figure 5. Bentheimer pore throat size distribution obtained through Mercury Injection Capillary Pressure (MICP)

The plotting of data obtained through MICP confirmed the presence of pores below 1 micrometre for the Ketton sample, as shown in Figure. 4. The determination integral of the microporosity peak gave a first estimate of the microporosity which represented 31% of the total rock porosity. A clear scale separation was observed between the macropore and micropores throat sizes. It is worth noting a thin a narrow macro-porosity peak, indicating a uniform macropore throat size distribution across each sample. It was noted that for the five different experiments, the peaks corresponding to microporosity agreed perfectly, suggesting high reliability of the experiment. As expected, the Bentheimer samples showed no microporosity peak, as seen in Figure 5.

4.2. Volume Rendering

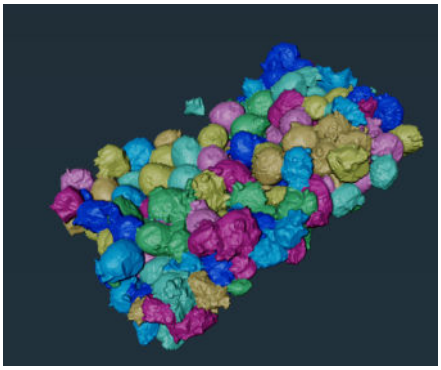


Figure 6. 3D volume rendering of the grains with dimensions 1401 x 2018 x 500 $voxels^3$ (width x height x depth)

The previous workflow was applied to a larger sub volume of 2.86 mm^3, and Figure 6. shows the 3D volume rendering that was obtained.

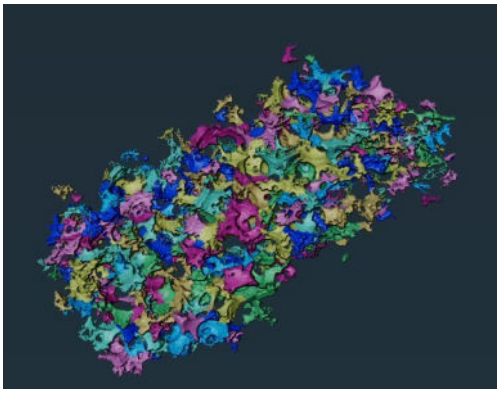


Figure 7. 3D volume rendering of the grains with dimensions 1401 x 2018 x 500 $voxels^3$ (width x height x depth)

It was also applied for the pores using the same sub volume and a macropore representation was obtained, as shown in Figure 7. The volume fraction of the pore material generated from Avizo allowed to determine the macro-porosity, which was found represent 14% of the entire rock volume.

4.3 Filter Importance

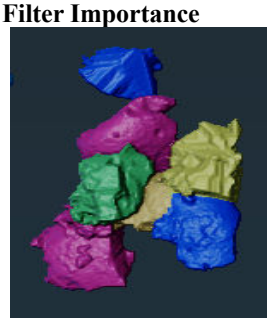


Figure 8. 3D volume rendering of the unfiltered sub volume of dimensions 500x500x500 $voxels^3$ (height x width x depth)

The importance of applying a filter for visualisation purposes was explored using the initial sub-volume. The 3D representation seen in Figure 8. was obtained. The results clearly showed the importance of a filter since the non-filtered image resulted in low sphericity values of 0.66, which is lower than what is expected for Ketton (Kurotori et al., 2020a). This is due to the presence of noise which can be misidentified as either a grain or a pore leading to unreliable segmentation.

4.4 X-ray Attenuation Trends

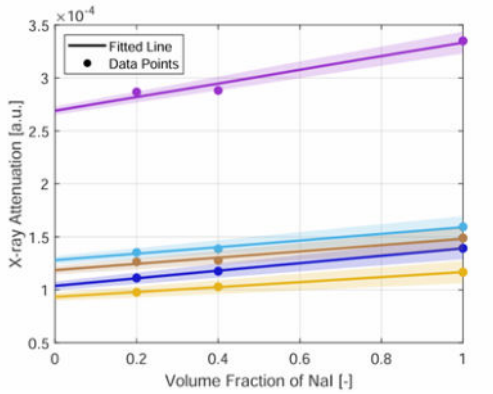


Figure 9. X-ray attenuation against volume fraction of NaI for 5 out of 136 grains plotted from Matlab

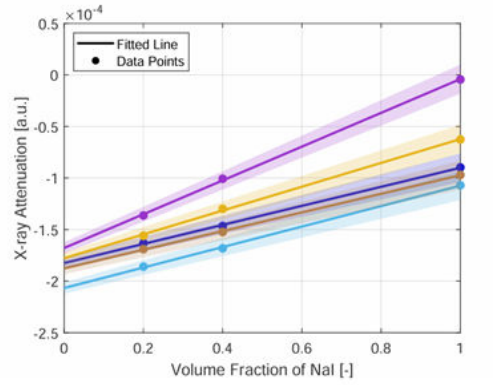


Figure 10. X-ray attenuation against volume fraction of NaI for 5 out of 570 pores plotted from Matlab

Figures 9. and 10. respectively show the X-ray attenuation graphs against tracer concentration for 5 selected grains and pore locations. When a tracer with a higher concentration penetrates the pores, they can absorb X-rays more strongly. Since attenuation is also directly proportional to contrast agent concentration (Zhuang *et al.*, 2022), the trend of attenuation values increasing linearly with the NaI volume fractions is as expected. A comparison of Figures 9. and 10. shows that the grains have larger attenuation values than the pores. Attenuation in CT scanning is directly related to the physical density of the material (Bernstein *et al.*, 2016). The grains consist of denser materials than pores, therefore they attenuate X-rays more effectively than fluids filling pores. It is worth noting that the gradients for the pores are a factor 3 times steeper than the grains. This is because the tracer fluid can access the macropores more easily than the micropores in the grains.

It is observed from Figure 9. there is a data point that has a gradient that is a factor of 2 times larger than

its counterparts. After investigating the raw images of the grains, this data point corresponded to a crushed grain. Structural details like crushed and cracked grains allow more tracer fluid to penetrate the grain, giving the impression of a higher microporosity. This example exemplifies that direct imaging combined with image analysis allows one to understand spatially the intensity contributions on a grain-by-grain basis.

4.5 Microporosity distribution

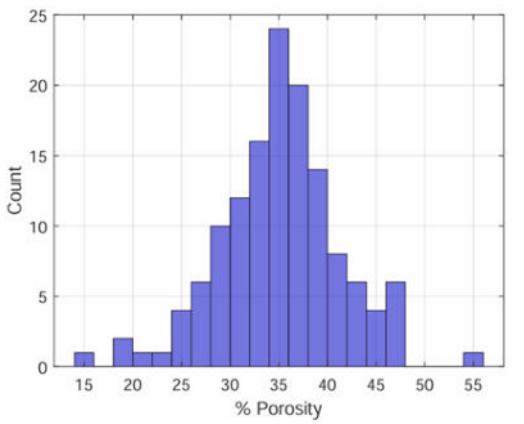


Figure 11. Histogram distribution of microporosities on a grain-by-grain basis for the 136 grains

A distribution of the micro-porosities was plotted for 136 grains where the average is 35%. The obtained microporosities were found to be significantly higher than what would be expected according to literature which states a microporosity of around 10% for Ketton (Andrew, 2014). A new mask was created by overlapping the pore masks for each tracer concentration to determine if the inaccuracy was from the pores. With this mask, it can be more certain that any bubbles which may appear at different locations when the images were taken at different tracer concentrations are excluded. However, the result was a minimal difference in the final microporosity values obtained.

The implication of this finding indicates the source of inaccuracy may come from the segmentation caused by imaging artifacts arising at the interfaces between materials of different densities (between pores/ tracer and grains). Specifically, adjacent voxels to tracer that were captured in the grain material. To minimize the impact of these artifacts on microporosity calculations, a practical approach is to exclude these interfaces by applying dilation and erosion functions. These functions remove unwanted tracer film while preserving the bulk grain structure for analysis.

4.6 3D Microporosity Map

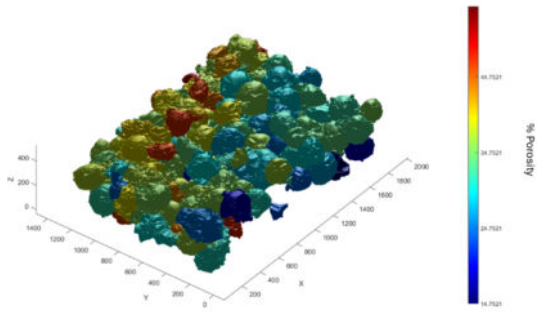


Figure 12. 3D volume rendering of the grains of dimension 1401 x 2018 x 500 (width x height x depth). The colourmap indicates each grain microporosity value

Figure 12. shows a 3D rendering of the grains which was plotted in Matlab coloured by their porosity values. Variability is observed which suggests spatial heterogeneity in the microporosities. This alludes to the non-homogeneous internal structure of the grains. It is important to point out that this spatial diversity is likely due to the nature of the rock which will be different to chemically engineered/ industry-produced porous materials where properties can be controlled.

4.7 Uncertainty calculations

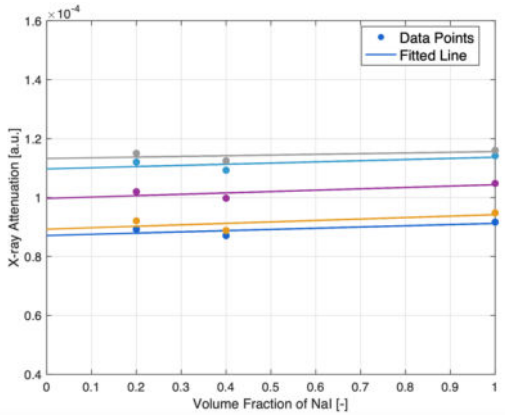


Figure 13. X-ray attenuation against volume fraction of NaI for 5 grain-size rubber sub volumes plotted from Matlab

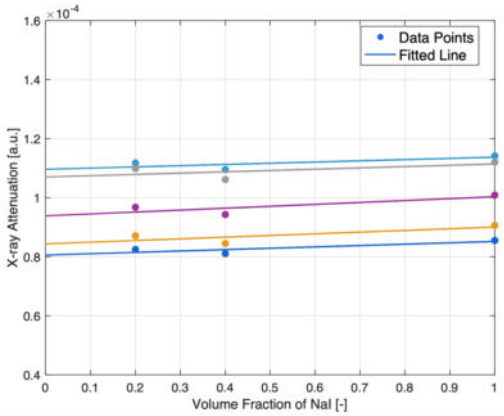


Figure 14. X-ray attenuation against volume fraction of NaI for 5 pore-sized rubber sub volumes plotted from Matlab

Figure 13. shows the X-ray attenuation values for grain sized sub volumes (147^3 $voxels^3$) extracted from the rubber for the different samples and Figure 14. shows the same for pore-sized sub volumes (32^3 $voxels^3$) also extracted from the same rubber. It is worth noting that the change in X -Ray attenuation for the different samples is very minimal for both, showing practically no change in intensity. This was in accordance with what was expected since the experiments were conducted at the same conditions, by only varying the tracer volume fraction. Since the rubber absorbs no tracer, it was therefore expected to not show any deviations throughout the samples. Moreover, it can be noted that the pore-size sub volumes showed slightly higher deviations than the grain-size sub volumes. This can be due to local heterogeneity present in the rubber, due to variations in its structure such as changes in density or presence of voids. Smaller sub-volumes would capture these smaller deviations more than the larger sub volumes which would average out the local variations, leading to smaller standard deviations. The uncertainties obtained were within an acceptable range and the average error was found to be 5.5%. This value is relatively low and therefore further supports the reliability of the established workflow.

4.8 Sphericity

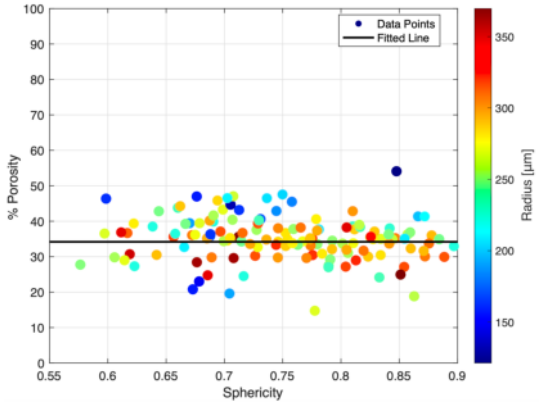


Figure 15. Porosity plotted against the sphericity for the 136 grains. The colourmap indicates each grain radius in μm

Figure 15. shows the microporosity of each grain that was plotted against the respective sphericity. It was found that the average sphericity across all the grains was 0.75, which is relatively high, suggesting homogeneity in grain shape and confirming the oolithic structure of Ketton (Andrew, 2014). The average diameter of the grains was found to be 523 μm, which is consistent with literature (Kurotori et al., 2020a). It is also worth highlighting that the microporosity doesn't vary with sphericity, suggesting that micropore formation is not influenced by grain shape, but might be a result of the material intrinsic properties.

4.9 Connectivity

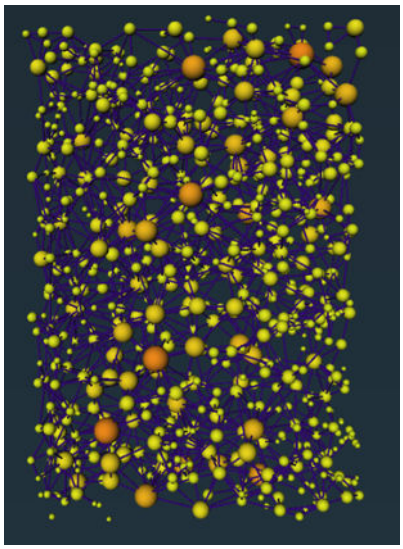


Figure 16. Visualisation of the macropore connectivity using the Generate Pore Network Model function in Avizo with dimensions of 1401 x 2018 x 500 $voxels^3$ (width x height x depth)

Connectivity is another important parameter to fully describe flow and transport. In avizo using the generate pore network model module, a visualisation of the macropore space was produced as shown in Figure 16. where the pores are represented using spheres, coloured and scaled according to its equivalent radius while the throats are shown as cylinders; scaled by its channel length. The model reveals that the sample is well connected, with macropores linked through multiple pathways.

Table 1. Summary of element count generated from Avizo

Element	Count
Pores (PNM)	768
Pore throats (PNM)	2041
Pores (Label Analysis)	570

However, it should be noted that a drawback of the generated pore network module is the division of one large pore into smaller pores. This explains the discrepancy in the pore count (PNM and Label analysis) in Table 1.

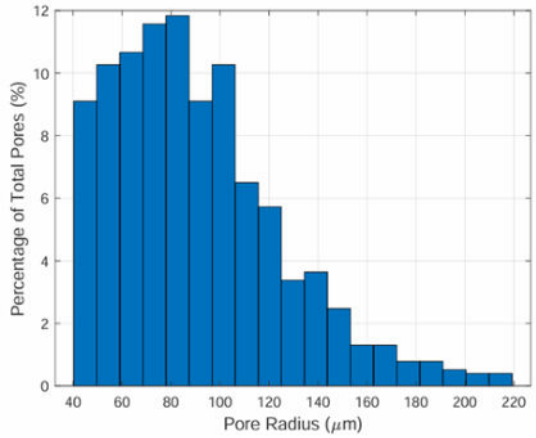


Figure 17. Histogram showing pore radii generated from the Pore Network Model function on Avizo.

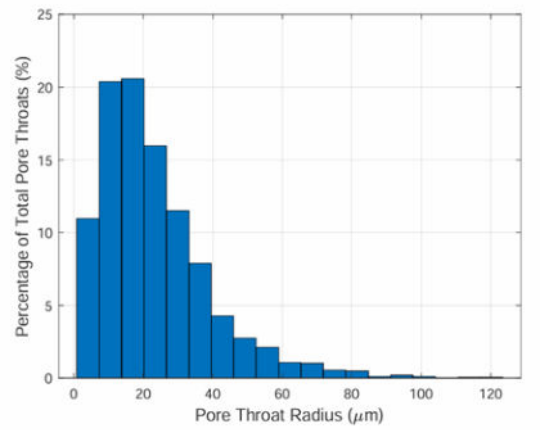


Figure 18. Histogram showing pore throat radii generated from the Pore Network Model function on Avizo.

Figures 17. And 18. presents a distribution of the pore and pore throat radii obtained from the pore network model. The equivalent pore radius is mainly distributed between 70-90μm which accounts for 23.41% of the total number of pores. The equivalent pore throat radius is concentrated between 7-20μm which contributes to 40.94% of the total number of pore throats. It should be noted that both cases present a positive skewed distribution..

In this work, only the macropore connectivity is considered, but further steps are required to resolve the sub- resolution pore region. This is required to fully simulate fluid flow as a study has found that the connectivity of macropores may increase due to connections through sub- resolution pores (Lin *et al.*, 2016).

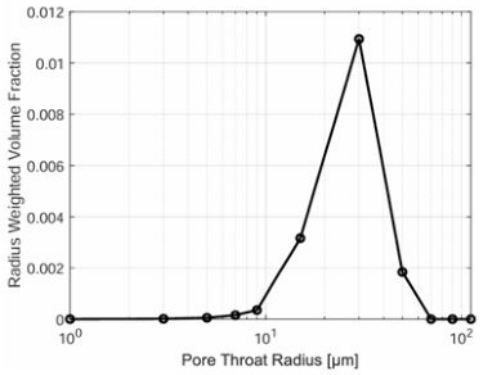


Figure 19. Pore throat size distribution

Figure 19. shows a plot of the radius weighted volume fraction against the pore throat radii obtained from the PNM. It is interesting to highlight that the peak at 30 μm co- aligns with that of the MICP data plot (Figure 4.). This agreement confirms the effectiveness of the segmentation and increases

confidence that the PNM can be used to characterise the pore space. This can lead to further applications in predictive analyses, such as permeability or fluid flow simulations.

5. Conclusion

In order to further improve the workflow, improvement in the segmented label mask by performing the erosion and dilation steps. The use of a dry scan (with no tracer injected) would improve the linear fit obtained for X-ray attenuation plots, leading to more accurate gradients that would result in more reliable microporosity values. Moreover, literature suggests that the use of an improved non-local mean (INLM) filter would allow a more precise segmentation. As opposed to the NLM filter, the INLM filter would also the spatial distance of the neighbouring voxels in the weighting average, with voxels closer to the central voxel would have higher weight. Literature suggest that this would outperform the NLM filter in the removal of noise (Wang et al., 2018a). Even though this work would be extremely relevant for CCS purposes, further applications could be explored such as optimisation microporous gas adsorption and membrane gas separation processes.

Acknowledgement

The authors would like to thank their co-supervisors, Takeshi Kurotori and Juan Carlos Roa Pérez for their continuous support throughout the project.

References

[1] Ducheyne, P. (2017). Comprehensive Biomaterials II. 7

[2] Lett, J.T. and Sinclair, W.K. (1993). DNA and Chromatin Damage Caused by Radiation. *Advances in Radiation Biology*, 17, pp.1–507.

[3] Brahme A. and Panetta, D. (2014). Physical Basis of X-ray imaging. *Comprehensive Biomedical Physics*, 2, pp.65–88.

[4] Feldman , D. (2018). *Vitamin D: Volume One: Biochemistry, Physiology and Diagnostics*. Fourth ed.

[5] Semmlow, J. (2018). *Circuits, Signals, and Systems for Bioengineering*. Third ed.

[6] Zhong, H., Wang, Z., Zhang, Y., Suo, S., Hong, Y., Wang, L. and Gan, Y. (2024). Gas storage in geological formations: A comparative review on carbon dioxide and hydrogen storage. *Materials Today Sustainability*, pp.100720–100720. doi:https://doi.org/10.1016/j.mtsust.2024.100720.

[7] Naresh, K., Khan, K.A., Umer, R. and Cantwell, W.J. (2020). The use of X-ray computed tomography for design and process modeling of aerospace composites: A review. *Materials & Design*, [online] 190, p.108553.

[8] Lin, Q. *et al.* (2016) 'Quantification of sub-resolution porosity in carbonate rocks by applying high-salinity contrast brine using X-ray microtomography differential imaging,' *Advances in Water Resources*, 96, pp. 306–322.

[9] Singh, K. *et al.* (2018) 'Partial dissolution of carbonate rock grains during reactive CO2-saturated brine injection under reservoir conditions,' *Advances in Water Resources*, 122, pp. 27–36.

[10] Foroughi, S. *et al.* (2024) 'Incorporation of Sub-Resolution porosity into Two-Phase flow models with a multiscale pore network for complex microporous rocks,' *Water Resources Research*, 60(4).

[11] De Haan, K. *et al.* (2019) 'Resolution enhancement in scanning electron microscopy using deep learning,' *Scientific Reports*, 9(1)

[12] Wenchel, L., Gampp, O. and Riek, R. (2024) 'Super-resolution NMR spectroscopy,' *Journal of Magnetic Resonance*, 366, p. 107746.

[13] Wang, G., Liu, Y., Xiong, W. and Li, Y. (2018b). An improved non-local means filter for color image denoising. *Optik*, [online] 173, pp.157–173. doi:https://doi.org/10.1016/j.ijleo.2018.08.013.

[14] Rogowska, J. (2000). Overview and Fundamentals of Medical Image Segmentation. *Handbook of Medical Imaging*, pp.69–85. doi:https://doi.org/10.1016/b978-012077790-7/50009-6.

[15] Andrew, M.G. (2014). Reservoir-Condition Pore-Scale Imaging of Multiphase Flow.

[16] Aayushi, A.K. (2014). Image Segmentation using Watershed Transform. *International Journal of Soft Computing and Engineering (IJSCE)*, 4(1).

[17] Kurotori, T., Zahasky, C., Benson, S.M. and Pini, R. (2020b). Description of Chemical Transport in Laboratory Rock Cores Using the Continuous Random Walk Formalism. *Water Resources Research*, 56(9). doi:https://doi.org/10.1029/2020wr027511.[

[18] Zhuang, L. *et al.* (2022) 'A novel method for estimating subresolution porosity from CT images and its application to homogeneity evaluation of porous media,' *Scientific Reports*, 12(1).

[19] Bernstein, A.L. *et al.* (2016) 'Improved sensitivity of computed tomography towards iodine and gold nanoparticle contrast agents via iterative reconstruction methods,' *Scientific Reports*, 6(1).

IMPERIAL

Investigating Liquid-Side Mass Transfer in Wavy-Laminar Falling Films

Alexander Nies and Álvaro Echeverría de Encio
Department of Chemical Engineering, Imperial College London, U.K.

Abstract

In this study, we present a full three-dimensional numerical analysis of a thin liquid falling film on a vertical surface by solving the full three-dimensional Navier–Stokes equations together with the advection-diffusion equation for CO_2 concentration. Flow conditions are based on the experimental study of Park and Nosoko[1] and numerical simulations from Dietze *et al.*[2] Direct Numerical Simulations (DNS) were performed on a high-performance computer (HPC), using code BLUE.[3] Our results show that both the inertia-dominated region and the visco-capillary dominated region that were first identified by Dietze *et al.* contribute to enhancing overall liquid-side mass transfer. Moreover, wave dynamics resulting in enhanced mixing and renewal of fluid elements at the gas-liquid interface from internal recirculation play a bigger role in mass transfer improvement than the increase in interfacial area. Diffusion hence becomes more limiting to overall mass transfer for more laminar flow (lower Reynolds number) in falling films, where the recirculation zone does not strongly interact with the underlying base film.

Keywords: Falling film, Multiphase flows, Computational fluid dynamics, Direct numerical simulations, Mass transfer

1. Introduction

Falling films play a vital role across a wide range of natural and industrial processes due to their ability to facilitate efficient heat and mass transfer through their thin liquid layers and large surface areas.[4] These advantages are provided by the formation of unique waves on the surface of the liquid films, a phenomenon which is practically always detected even for low liquid flow rates.[4][5] In natural settings, falling films occur widely, from gravity currents and mud flows to snow avalanches and volcanic lava flows.[6] They also manifest in biological systems, including the pulmonary airway linings where mucus flow is crucial in influencing breathing efficiency and airway opening and closure.[7] Similarly, thin corneal tear films covering the surface of the human eye are essential for maintaining ocular health and providing a clear optical surface.[8] This widespread prevalence underscores the significance of studying falling films to better understand geophysical dynamics, bio-adhesion, and physiological functions to predict and optimise natural and biological processes.

In industry, falling films are commonly used in chemical reactors, condensers, heat exchangers, microreactors, and even advanced microelectromechanical devices.[9][10] Moreover, modern and emerging fields such as microfluidics and nanofluidics rely on the principles of ultra-thin film flow to enhance performance and precision.[8] As thin-film structure offers superior heat and mass transfer rates, they improve operational efficiency in applications such as distillation, evaporation, and absorption processes.[11] In such energy-intensive systems, corrugated and modified surfaces are often employed over smooth surfaces as they have been shown to enhance gas-liquid mass transfer efficiency.[12] The literature attributes these improvements to three factors: increased gas-liquid

interfacial area, liquid residence time[13] and the formation of internal recirculation zones.[14]

The enhancement of mass transfer in falling films using modified surfaces has been a focal point of industrial optimisation efforts.[12] Experimental investigations have demonstrated the existence of optimum structures (at baffle spacing to height ratios of 10) attributed to the formation of flow eddies with a high value turbulent kinetic energy (TKE) distribution near the gas-liquid interface.[15] Further complimentary quantitative analysis was conducted by Düll *et al.* at higher Reynolds numbers, inclination angles and structure geometry to understand the underlying hydrodynamics.[16]

Apart from experiments, falling films can also be accurately simulated with analytical and numerical methods, unlike other types of open-flow hydrodynamic systems. Falling film problems are generally more tractable due to the long wave nature of the instability of falling films, where the length of waves is significantly longer compared with the film thickness, and the free surface deformations are weak.[17] Most simulations of wavy falling films are done in two-dimensional slices, but experimental work has demonstrated that the evolving waves are clearly 3D in structure.[1] Dietze *et al.* developed full Direct Numerical Simulations (DNS) of three-dimensional falling films for the first time, showing an accurate representation of fine interfacial details.[2]

In this paper, we elaborate on these works by integrating the advection-diffusion equation for the concentration of CO_2 into previous simulations completed by Kahouadji *et al.* to identify further the dominant factors contributing to mass-transfer enhancement in wavy-laminar falling film flows. We attempt to replicate the experimental work conducted by Park and Nosoko (case 1)[1] through DNS, as well as integrating mass transfer to the DNS completed by Dietze *et al.* (cases 2 and 3).[2] In section 2, we begin with developing the governing equations as the basis for DNS. In section 3, we explore the factors contributing to the mass transfer enhancement in wavy-laminar falling films. We conclude in section 4, aiming also to gain insight into how different wave dynamics can enhance mass transfer in applications such as falling film membrane contactors for CO_2 capture applications.[18]

2. Problem Formulation

This study examines the evolution of a thin liquid film falling down an isothermal vertical plane in a three-dimensional domain. We utilise a Cartesian coordinate system (x, y, z) where x is the streamwise coordinate, y the spanwise coordinate and z the cross-stream coordinate normal to the plane: $\mathbf{x} = (x, y, z) \in [0, \lambda_x] \times [0, \lambda_y] \times [0, \lambda_z]$. Figure 1 shows the domain setup, and table 1 contains the physical parameters used in this study.

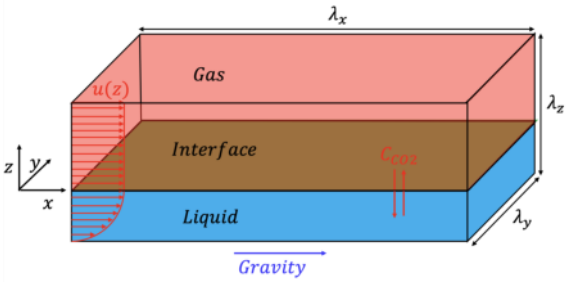


Figure 1: Computational domain of a 3D falling film configuration with gas absorption.

Table 1: Simulation Parameters.

Variable	Notation	Unit
Plane incline angle	α	rad
Gravitational constant	g	m/s^2
Liquid density	ρ_l	kg/m^3
Liquid viscosity	μ_l	Pa s
Liquid diffusion	$\mathscr{D}_l$	m^2/s
Gas density	ρ_g	kg/m^3
Gas viscosity	μ_g	Pa s
Gas diffusion	$\mathscr{D}_g$	m^2/s
Surface tension	σ	N/m

2.1. Governing Equations

The governing equations, namely continuity, momentum (Navier-Stokes) and mass transfer (advection-diffusion) can be written as:

$$\nabla \cdot \mathbf{u} = 0 \tag{1}$$

$$\rho \frac{D\mathbf{u}}{Dt} = -\nabla p + \nabla \cdot \mu\left(\nabla \mathbf{u} + \nabla \mathbf{u}^{\mathrm{T}}\right) + \rho \mathbf{g} + \mathbf{F} \tag{2}$$

$$\frac{DC}{Dt} = \mathscr{D}_l \nabla^2 C \tag{3}$$

where $\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla$ is the material total derivative operator, $\mathbf{u} = (u, v, w)$ is the fluid velocity vector, C and p are solute concentration and total pressure of

the fluid respectively, $\mathbf{g} = (g\sin(\alpha), -g\cos(\alpha), 0)$ is the gravitational acceleration, which is reduced to $(g, 0, 0)$, as we have a vertical plane ($\alpha = \frac{\pi}{2}$ rad). Finally, $\mathbf{F}$ is the local surface tension force at the interface and is defined as follows:

$$\boldsymbol{F} = \int_{A(t)} \sigma\kappa\widehat{\boldsymbol{n}}\delta_D\left(\boldsymbol{x} - \boldsymbol{x}_f\right) dA \tag{4}$$

where $A(t)$ is the interfacial surface area (as a function of time), σ is the surface tension between the liquid and the gas, κ the local curvature of the interface, $\widehat{\boldsymbol{n}}$ is the unit normal vector to the interface, and $\delta_D\left(\boldsymbol{x} - \boldsymbol{x}_f\right)$ is the Dirac delta function which vanishes everywhere except on the interface.

2.1.1. Main Assumptions

The main assumptions are stated below.

H1) the fluid is non-compressible and has constant properties.
H2) the plate is rigid (non-porous), hence there is no-slip and no-penetration condition on the surface of the plate.
H3) the liquid is nonvolatile, hence evaporation effects can be neglected.
H4) the gas acts as an infinite reservoir of CO_2, and is also maintained at constant pressure ($p_\infty =$ 1 atm) and temperature ($T_\infty = 294.15$K).

2.2. Boundary Conditions (BCs)

2.2.1. At the Surface of the Plate

The equations 1 and 2 are subject to no-slip and no-penetration BCs (H2):

$$\mathbf{u} = \mathbf{0} \tag{5}$$

Furthermore, there is also a no-flux condition for concentration:

$$\frac{\partial C}{\partial z} = 0 \tag{6}$$

2.2.2. At the Domain Boundaries

Periodic BCs are applied in the streamwise and spanwise directions. At the top of the domain ($z = \lambda_z$), a Dirichlet BC is applied to the concentration, for a no-flux condition:

$$\frac{\partial C}{\partial z} = 0 \tag{7}$$

A stress-free condition is also applied:

$$\frac{\partial u}{\partial z} = v = w = 0 \tag{8}$$

2.3. Initial Conditions

The initial spanwise and cross-stream velocities are set to zero:

$$v(t=0) = w(t=0) = 0 \tag{9}$$

The initial concentration profile is defined in the entire domain for the gas phase using the Heaviside indicator function as $\mathscr{H}(\mathbf{x}, t)$:

$$C(\mathbf{x}, t=0) = C_{g,0} + (C_{l,0} - C_{g,0})\mathscr{H}(\mathbf{x}, t=0) \tag{10}$$

where $\mathscr{H}(\mathbf{x}, t)$ vanishes everywhere except in the liquid phase. The gas acts as an infinite source of solute with diffusion much faster than in the liquid, $\mathscr{D}_g \gg \mathscr{D}_l$ (H4). Thus, we ensure constant gas saturation with mass transfer limited by diffusion in the film:

$$C(\mathbf{x}, t) = C_{g,0} + (C_l - C_{g,0})\mathscr{H}(\mathbf{x}, t) \tag{11}$$

The initial velocity of the falling film is described using the Nusselt theory, obtaining a parabolic velocity profile:

$$u_0(z) = \frac{\rho_l g \sin(\alpha)\delta_0^2}{\mu_l}\left[\frac{z}{\delta_0} - \frac{1}{2}\left(\frac{z}{\delta_0}\right)^2\right] \tag{12}$$

where δ_0 is the initial film thickness. Using the mean value theorem, the average velocity within the film can be obtained:

$$\overline{u}_0 = \frac{\rho_l g \sin(\alpha)}{3\mu_l}{\delta_0}^2 \tag{13}$$

From the definition of the film Reynolds number, the film thickness as a function of Reynolds number can be obtained as follows:

$$\delta_0 = \left(\frac{3\mu_l^2}{\rho_l^2 g \sin(\alpha)}\right)^{\frac{1}{3}} Re^{\frac{1}{3}} \tag{14}$$

An initial disturbance is set to the film interface to speed up the DNS by allowing a quasi-steady state to be reached sooner:

$$\delta|_{t=0} = \overline{\overline{\delta}}\left[1 + \varepsilon_x \cos\left(k_x x\right) + \varepsilon_y \cos\left(k_y y\right)\right] \tag{15}$$

where $\overline{\overline{\delta}}$ is the average film thickness, $k_x = \frac{2\pi}{\lambda_x}$, $k_y = \frac{2\pi}{\lambda_y}$, $\varepsilon_x = 0.2$ and $\varepsilon_y = 0.05$ are taken as the same values as Dietze *et al.*[2]

2.4. Nondimensionalisation

To nondimensionalise the governing equations, the following scalings obtained from the Nusselt flat film solution are used:

$$\tilde{\mathbf{x}} = \frac{\mathbf{x}}{\delta_0} \quad (16) \qquad \tilde{t} = \frac{t}{\delta_0/\overline{u}_0} \quad (18)$$

$$\tilde{\mathbf{u}} = \frac{\mathbf{u}}{\overline{u}_0} \quad (17) \qquad \tilde{p} = \frac{p}{\rho_l \overline{u}_0{}^2} \quad (19)$$

The dimensionless forms of the governing equations (equations 1 to 3) can be expressed as follows:

$$\nabla \cdot \tilde{\mathbf{u}} = 0 \quad (20)$$

$$\frac{D\tilde{\mathbf{u}}}{D\tilde{t}} = -\nabla\tilde{p} + \frac{1}{\mathrm{Re}}\nabla^2\tilde{\boldsymbol{u}} + \frac{1}{\mathrm{Fr}^2}\hat{\boldsymbol{k}} + \frac{1}{\mathrm{We}}\int_{\tilde{A}(\tilde{t})} \tilde{\sigma}\tilde{\boldsymbol{\kappa}}\hat{\boldsymbol{n}}\delta_D\left(\tilde{\boldsymbol{x}} - \tilde{\boldsymbol{x}}_f\right) d\tilde{A} \quad (21)$$

$$\frac{D\tilde{C}}{D\tilde{t}} = \frac{1}{\mathrm{Pe}}\nabla^2\tilde{C} \quad (22)$$

where Re is the Reynolds number, which is the ratio of inertial to viscous forces, Fr is the Fraude number, which is the ratio of inertial to gravitational forces, We is the Weber number, which is the ratio of capillary forces to inertial forces, and Pe is the (mass) Peclet number, which is the ratio of advective transport to diffusive transport.

$$\mathrm{Re} = \frac{\rho_l u_0 \delta_0}{\mu_l} \quad (23) \qquad \mathrm{We} = \frac{\sigma}{\rho_l \delta_0 u_0^2} \quad (25)$$

$$\mathrm{Fr} = \frac{u_0}{(g\delta_0)^{1/2}} \quad (24) \qquad \mathrm{Pe} = \frac{u_0\delta_0}{\mathscr{D}_l} = \mathrm{Re}\cdot\mathrm{Sc} \quad (26)$$

Sc represents the Schmidt number, which is the ratio of viscous to molecular diffusion rates:

$$\mathrm{Sc} = \frac{\mu_l}{\rho_l \mathscr{D}_l} = \frac{\mathrm{Pe}}{\mathrm{Re}} \quad (27)$$

Other relevant dimensionless numbers are the Kapitza (Ka) and Sherwood (Sh) numbers. The first describes the ratio of surface tension to gravitational forces, and the latter is defined as the ratio of the total mass transfer rate (in the liquid) to the rate of diffusive mass transport.

$$\mathrm{Ka} = \frac{\sigma}{\rho_l (g\sin\alpha)^{1/3} \nu_l^{4/3}} \quad (28) \qquad \mathrm{Sh} = \frac{k_l}{\mathscr{D}_l/\delta_0} \quad (29)$$

k_l is the overall mass transfer coefficient in the liquid phase. The global dimensionless numbers for all the cases, except for Sh which will be calculated in section 3, are shown in table 2.

Table 2: Dimensionless numbers for each case.

Case	Re	Ka	We	Sc	Pe
1a	59.3	3923.0	6.3	510	30090
1b				51	3009
2a	15.0	509.5	8.1	1610	24190
2b				161	2419
3a	6.2	17.8	1.2	11440	70930
3b				1144	7093

The fluid physical properties of the three cases studied are summarised in table 3. For each case, two simulations are performed: one with diffusion coefficient $\mathscr{D}_l = 1.75 \times 10^{-9}\,\mathrm{m}^2/\mathrm{s}$ (case a) corresponding to the diffusivity of CO_2 in water at 294.15K[19], and the other with $\mathscr{D}_l = 1.75\times10^{-8}\,\mathrm{m}^2/\mathrm{s}$ (case b). It should be noted that the aim is to explore the effect of hydrodynamics on the absorption of CO_2, and not to compare the effect of different solvents on the absorption of CO_2. Hence, all the cases used the same two values of diffusivity described above, and the same saturation concentration of CO_2 at the interface ($C_{sat} = 37.8\ \mathrm{mol/m}^3$).

Table 3: Fluid properties for each case. Water[1], DMSO (dimethylsulphoxide) and water mixture[2], DMS-T12 (dimethylsiloxane)[2].

Case	μ[Pa s]	ρ[kg/m^3]	σ[N/m]
1	8.9×10^{-4}	997.07	0.07201
2	3.1×10^{-3}	1098.3	0.0484
3	1.9×10^{-2}	949.0	0.020
Gas	1.8×10^{-5}	1.185	-

The DNS formulated were solved using code BLUE[3] on a high-performance computer to obtain the results.

3. Results and Discussion

3.1. Mass Transfer Analysis

The overall liquid-side mass transfer coefficient (k_l), is defined as:

$$\frac{d\overline{C}}{dt} = -k_l \frac{A}{V_l}\left(C_{sat} - \overline{C}\right) \quad (30)$$

where $\overline{C}$ is the average concentration of species in the liquid, $C_{sat} = 37.8$ mol/m^3 is the saturation concentration of CO_2 in the fluid, A is the gas-liquid interfacial area, and V_l is the volume of liquid. Equation 30 can be integrated using the boundary conditions $\overline{C}(t=0)=0$ and $\overline{C}(t=t_s)=\overline{C}$:

$$-\ln\left(1-\frac{\overline{C}}{C_{sat}}\right)\cdot\frac{V_l}{A}=k_l t \tag{31}$$

Equation 31 can be plotted using the data obtained from DNS to estimate the overall liquid-side mass transfer coefficient, k_l. Figure 2 shows an example linear regression plot to estimate k_l for case 2b.

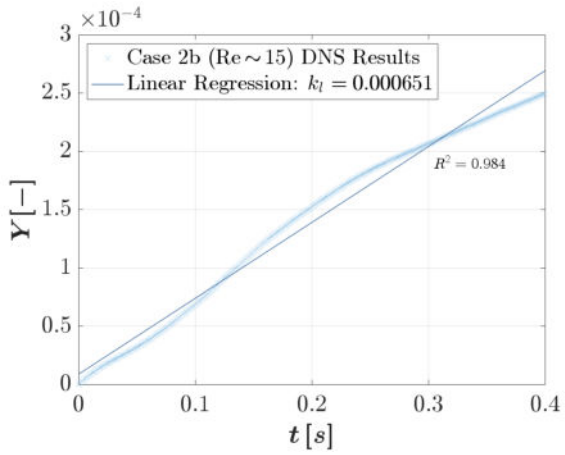


Figure 2: Linear regression plot to estimate k_l for case 2b with $R^2 = 0.984$.

k_l can then be substituted into equation 29 to calculate the Sherwood number. To assess the enhancement of overall mass transfer due to wave formation, penetration theory was used to simulate a base case in which no waves develop at the interface of a falling film. This hypothetical mass transfer coefficient serves as a reference for comparison with the actual mass transfer coefficient under wavy conditions. The key assumptions of the penetration theories are that i) unsteady mass transfer occur to a liquid element on the interface, ii) the gas-liquid interface is at equilibrium, iii) each liquid element on the interface stays in contact with the gas for the same period, and iv) the streamwise flow velocity is constant. The overall liquid-side mass transfer coefficient for a base case in which no waves develop ($k_{l,P}$) is defined as:

$$k_{l,P}=2\sqrt{\frac{\mathscr{D}_l}{\pi t}} \tag{32}$$

where $t = 0.4s$ is the overall contact period between the gas and the liquid at the interface, corresponding to the simulation time used in the DNS. We can now define the mass enhancement factor ε_k as:

$$\varepsilon_k=\frac{k_l}{k_{l,P}} \tag{33}$$

All the liquid-side mass transfer parameters are summarised in table 4.

Table 4: k_l, ε_k and Sherwood numbers for each case.

Case	Re	k_l [m/s]	ε_k	Sh
1a	59.3	8.67×10^{-4}	8.55	120.74
1b		9.26×10^{-4}	3.95	12.89
2a	15.0	5.69×10^{-4}	7.67	107.94
2b		6.51×10^{-4}	2.85	12.35
3a	6.2	6.48×10^{-4}	8.72	341.70
3b		7.48×10^{-4}	3.02	39.43

3.1.1. Liquid Diffusion

A few interesting trends can be seen from the results in table 4. Comparing the enhancement factor between cases a and b, it can be seen that as the diffusivity $\mathscr{D}_l$ increases by tenfold, the mass enhancement is reduced. This result implies that wave-induced enhancement of mass transfer might not be as relevant for heat transfer, where the diffusivity is much larger (by a factor of ~ 100).

Another trend is that k_l does not increase with increasing nor decreasing Reynolds number, as shown by case 2 exhibiting the lowest k_l. This result may imply that for the different cases, different mass transfer mode may be rate limiting. We will first investigate whether diffusion in the liquid is more limiting for higher Reynolds number or lower Reynolds number.

Unsurprisingly, as diffusivity increases tenfold, k_l increases. However, the ratio between k_l for the same case but with different diffusivity is higher for cases with lower Reynolds number, as shown clearly in the table below.

Table 5: Ratio between k_l for each case as diffusivity is increased by a factor of 10.

Case	Re	Ratio between k_l
1	59.3	1.07
2	15.0	1.14
3	6.2	1.15

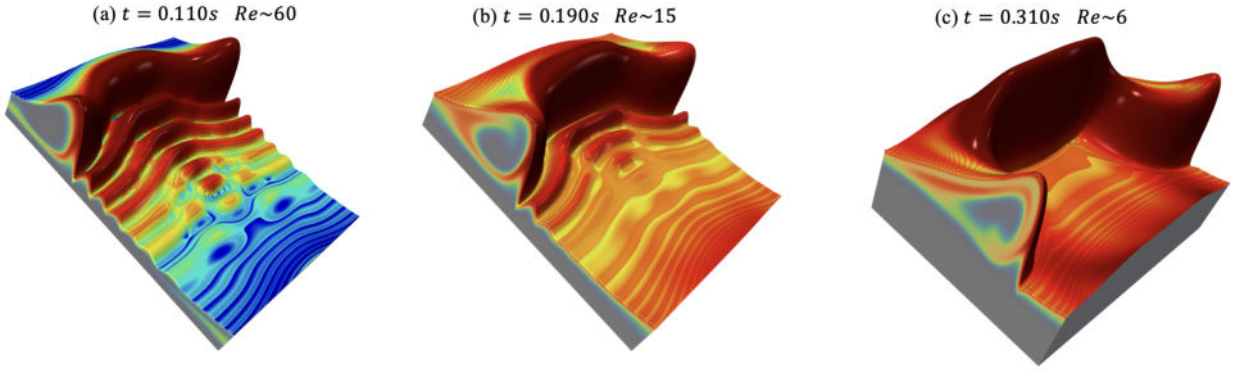


Figure 3: DNS results for a) case 1a, b) case 2a, and c) case 3a, for normalised global concentration of CO_2 in liquid film at ~ 0.2.

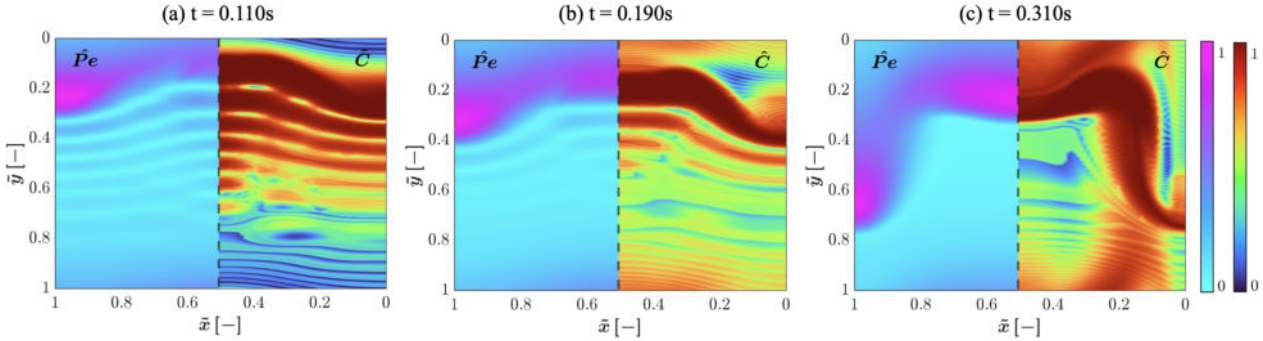


Figure 4: Contours of normalised local mass Peclet number, $\hat{Pe}$ (left half-plane) and normalised local CO_2 concentration, $\hat{C}$ (right half-plane) for a) case 1a, b) case 2a, and c) case 3a, for normalised global concentration of CO_2 in liquid film at ~ 0.2.

3.1.2. Interfacial Area Enhancement

The mass transfer could be augmented due to an enhancement in the interfacial surface area as waves form. Figure 5 shows the interfacial area enhancement, $\tilde{A}$, plotted against time.

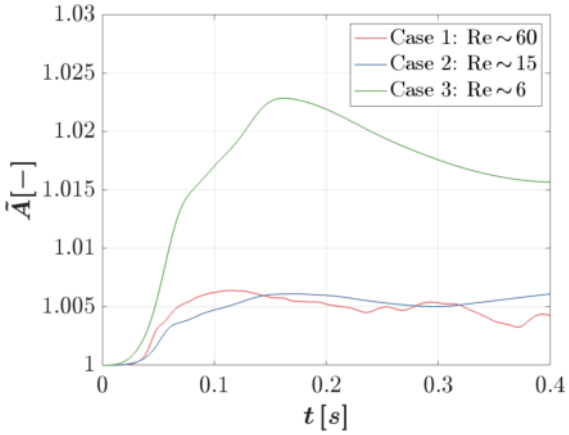


Figure 5: Plot of area enhancement against time.

From table 4, we see that the magnitude in ε_k is in the range of $\sim 10^0$. Surprisingly, figure 5 shows that the area enhancement is only in the magnitude of $\sim 10^{-2}$. This indicates that other factors, namely the wave dynamics at the free surface and recirculation within the waves, are the more dominant contributors in enhancing mass transfer.

3.1.3. Interface Dynamics

To analyse whether advection or diffusion is more limiting on the interface, we plot the contours of the normalised local mass Peclet number, $\hat{Pe}$ on the left half-plane, and normalised local CO_2 concentration, $\hat{C}$ on the right half-plane, for all cases where the ratio of CO_2 absorbed is approximately 0.2 (shown in figure 4). $\hat{\text{Pe}}$ is defined as:

$$\widehat{\text{Pe}} = \frac{\|\boldsymbol{u}\|\delta}{\mathscr{D}_l} \tag{34}$$

where $\|\boldsymbol{u}\|$ is the local velocity magnitude and δ is the local film height. Figure 4a and 4b shows that two regions are responsible for the absorption of CO_2: i) leading primary wave humps with high $\hat{\text{Pe}}$, and ii) capillary ripples. Similarities can be drawn between the current work focusing on mass transfer and the work completed by Dietze *et al.* focusing on wave dynamics and instabilities.[2] In Dietze's work, the local Weber number was used to determine the different regions of the flow field. The primary

wave region is classified as inertia-dominated, and the capillary waves region is classified as viscous (or capillary) dominated. In this work, the mass transfer in these two regions are further explored using the local Peclet number.

For the primary wave, it is unsurprising that there exists a high mass transfer rate, as higher local Peclet numbers suggest that there is greater advection in this region compared with other regions. However, it is surprising that the capillary waves region, with much lower local Peclet numbers, is just as saturated in CO_2 compared to the primary wave hump. These capillary ripples may be the reason in why case 1 exhibits a larger k_l than case 3, despite case 3 exhibiting the greatest Sherwood number (shown in table 4). The high Sherwood number for case 3 is primarily due to the fact that the average film thickness is greater due to a larger leading wave compared with other cases. But the formation of smaller capillary waves that are not as high in Pe number is also significant in enhancing mass transfer. Moreover, we will show that recirculation occurs within both the main leading wave and the capillary waves for case 1, which could also explain how this case exhibits the highest k_l.

3.1.4. Wave Dynamics

Figures 6ai, 6bi, 6ci present a three-dimensional view of a single wave for all cases. A slice is taken in the x-z plane through the center of the film, bisecting the trailing wave (yellow line). Figures 6aii, 6bii, 6cii, highlight this yellow slice, outlining the wave's profile and are further colored to indicate the concentration of solute absorbed in the liquid. Additional streamlines are displayed within the reference frame of the moving wave, allowing for the identification of a re-circulation zone towards the crest of the wave. This internal re-circulation zone aligns with the concentration gradient of CO_2, further demonstrating the close interaction between the fluid dynamics and the solute absorption in the falling film. It should be noted that the "b cases" with enhanced diffusion depict the same hydrodynamics, and hence their slices are not shown for redundancy.

To further investigate the phenomena, three additional vertical slices are taken at specific locations: i) the peak (pink), ii) the capillary waves (green) and iii) the trough of the main wave (black). Figure 6aiii, 6biii, 6ciii plot the normalized concentration, $\hat{C}$, against the normalized film thickness, $\hat{\delta}$, illustrating the concentration profile at these particular x-locations and revealing distinct transport mechanisms across the liquid film.

Additionally, we obtain local Peclet numbers at the free surface for each of the three slices being examined. By comparing this to the average Peclet of the film from table 2, we define the local Peclet enhancement, $\hat{\varepsilon}_{Pe}$ as:

$$\hat{\varepsilon}_{Pe} = \frac{\hat{Pe}_{int}}{Pe} \tag{35}$$

Table 6: Local enhancement of Peclet numbers at the gas-liquid interface for the peak, capillary and trough.

Case	$\hat{\varepsilon}_{\mathbf{Pe,peak}}$	$\hat{\varepsilon}_{\mathbf{Pe,capillary}}$	$\hat{\varepsilon}_{\mathbf{Pe,trough}}$
1	43.2	22.8	6.31
2	66.1	11.3	9.51
3	60.6	-	5.50

At the trough of the wave, a sharp decline in concentration is observed as we move away from the interface, resulting in a thin and well-defined boundary layer. The streamlines in this region are smooth and parallel, indicating that mass transfer is predominantly governed by molecular diffusion. Specifically, for case 1 (figure 6aii), a local maximum in concentration appears within the film, suggesting that in the least laminar scenario, some advection occurs even at the trough.

In contrast, at the peak of the wave, particularly within the internal recirculation zone, the boundary concentration layer is noticeably smoother. The local enhancement of the Peclet number at the peak ($\hat{\varepsilon}_{Pe,peak} \sim 43-66$) indicates a significantly greater influence of advection on mass transport, even in the most laminar case. Interestingly, for case 3 (figure 6ciii), the concentration profile reveals two distinct local maxima followed by a sharp decline, suggesting two key phenomena. First, while internal recirculation enhances mixing, diffusion remains a limiting factor, preventing the development of a flat, fully saturated profile near the interface. Second, the recirculation zone does not strongly interact with the underlying base film. Over time, the liquid recirculated toward the gas-liquid interface is already saturated, hence diminishing the positive effect of the recirculation zone on mass transfer.

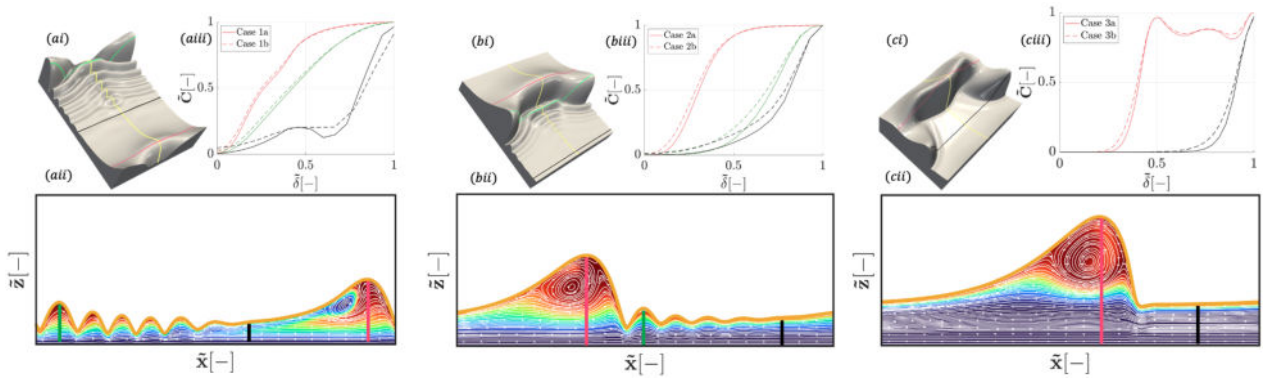


Figure 6: Concentration profile within the liquid film for (a) case 1, (b) case 2, and (c) case 3, at times $t = 0.1s$, $t = 0.2s$, and $t = 0.35s$ respectively. Subfigures (i) show the full three-dimensional wave, with the investigated planes marked for reference: streamwise bisection (yellow), spanwise main peak wave (pink), spanwise capillary wave (green), and spanwise trough (black). Subfigures (ii) depict the yellow slice (streamwise), including streamlines in the reference frame of the moving wave. Subfigures (iii) illustrate the concentration profile at the peak (pink), capillary waves (green), and trough (black) within the streamwise slice (yellow).

Dietze reached similar conclusions, showing that the intensification of scalar transfer by vortices diminishes as the fluid within the vortex becomes saturated.[20] He further demonstrated that introducing corrugation bumps on the plate can counteract this natural decay. These bumps redistribute liquid within the wave hump, allowing the moving-frame vortex to replenish with fresh liquid, thereby reactivating the intensification mechanism.

Additionally, capillary waves were analyzed for cases 1 and 2. For case 1 (figure 6aiii), the concentration profile closely resembles that observed at the wave peak, corroborated by a higher local enhancement of the Peclet number. This indicates that these capillary waves are more advection-dominated and contribute significantly to overall mass transfer. In contrast, the capillary waves in case 2 (figure 6biii) exhibit a boundary concentration profile more akin to that of the trough. This suggests that these waves are predominantly diffusion-controlled, as evidenced by the similarity in the local enhancement of the Peclet number to that of the trough (table 6).

Figure 7 illustrates the spanwise distribution of concentration and the corresponding velocity fields, serving as a visual summary of these phenomena. To quantify the degree of spanwise advection, we define the ratio of the maximum spanwise to streamwise velocity R_{vu} as:

$$R_{vu} = \frac{\max(v)}{\max(u)} \tag{36}$$

Table 7: Comparison of the ratio of spanwise to streamwise velocity at the wave peak, capillary waves, and wave trough for all cases. For the capillary waves and trough in case 3, the domain used to determine the maximum local velocities was reduced from that shown in Figure 7 to exclude the influence of the primary wave near the domain edges.

Case	$R_{vu,peak}$	$R_{vu,capillary}$	$R_{vu,trough}$
1	0.23	0.11	0.02
2	0.25	0.09	0.04
3	0.35	-	0.04

At the wave trough, mass transfer is diffusion-dominated. Velocity fields in this region are predominantly vertical, with minimal velocity ratios ($R_{vu} < 0.04$) highlighting the diffusion-dominated nature of this region.

Large spanwise recirculation patterns at the wave peak are observed across all cases, with significantly higher velocity ratios ($R_{vu} \sim 0.23 - 0.35$), underscoring the dominance of advection-driven transport. Capillary waves further highlight this variability: those in case 1 show recirculation patterns resembling the main wave, while those in case 2 mirror the diffusion-dominated profiles of the trough. These findings underscore the role of localised hydrodynamics in shaping solute transport mechanisms.

In summary, case 1 emerges as the most effective for enhancing mass transfer due to its higher Reynolds number and resultant turbulence. Internal recirculation at the wave peaks facilitates the upward transport of fresh liquid, maintaining a high concentration gradient at the gas-liquid

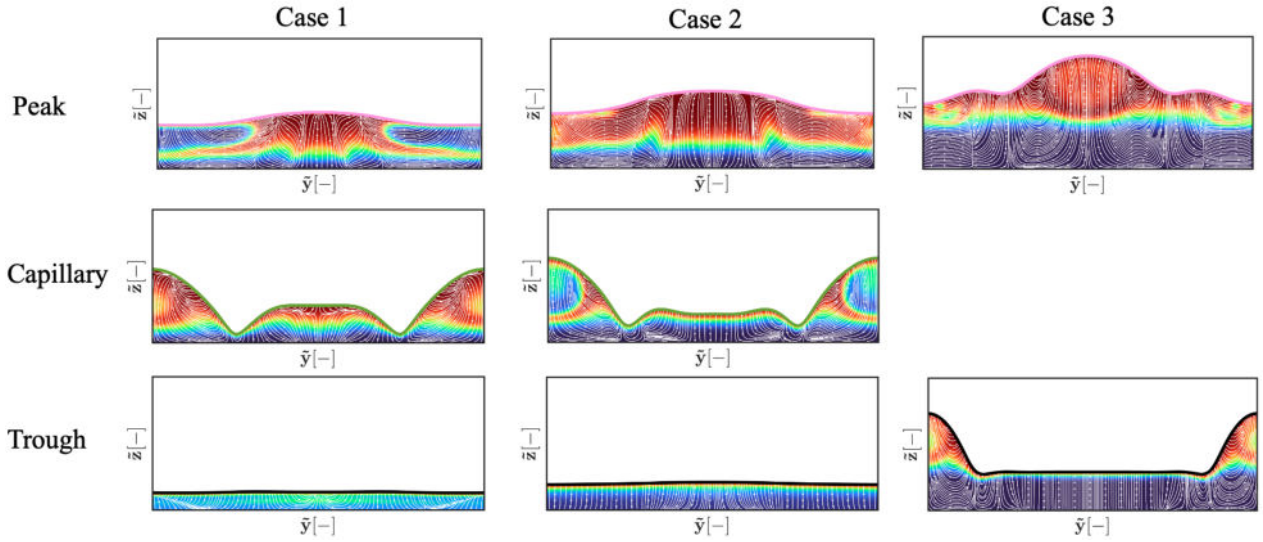


Figure 7: Spanwise slices of the concentration profile for the liquid film with streamlines at at times $t = 0.1s$, $t = 0.2s$, and $t = 0.35s$ respectively. Slices correspond to the main peak (pink), capillary waves (green), and trough (black) for (a) case 1, (b) case 2, and (c) case 3. Refer to the figure 6 for the full three-dimensional waves and plane locations.

interface. Additionally, the advection-dominated capillary waves in case 1 further enhance localized absorption. However, further increasing the Reynolds number comes with a trade-off: excessively high velocities shorten the residence time of a liquid particle at the interface, potentially reducing overall solute absorption efficiency. This analysis suggests that while the examined turbulence levels were beneficial, further studies at higher Reynolds numbers are needed to identify the optimal balance between enhanced mixing and residence time. Understanding this trade-off is essential for maximizing mass transfer in practical applications.

4. Conclusion

We have presented full DNS of 3D falling liquid films coupled with the advection-diffusion equation for the first time. We show that both the primary wave and capillary ripples are key contributors in enhancing overall mass transfer in the liquid-side, and wave dynamics have a more significant impact on mass transfer enhancement compared with interfacial area augmentation. Internal liquid diffusion and recirculation allow for deeper absorption into the liquid film, enhancing overall film saturation. Three cases have been presented with low, moderate, and large Reynolds numbers, which allowed us to establish that diffusion becomes more limiting to overall mass transfer for more laminar flows.

In the initial stages of this study, an effort was made to replicate the experimental work conducted by Düll et al.[16] with higher Re numbers and corrugated surfaces on an inclined plane. A quasi-2D model was constructed for DNS, with the boundary and initial conditions closely matching those observed in this study. This setup aimed to capture the fundamental dynamics of more turbulent flows and its effect on mass transfer enhancement, while having experimental results to validate the simulations against. However, the computational cost of advancing the timesteps to reach statistically significant results was prohibitively high. The resource demand required to simulate turbulent flows, particularly in the context of corrugated surfaces with enhanced numerical complexity, is very intensive.

Firstly, the ability of oscillatory forcing to induce beneficial waveforms represents an exciting avenue for further exploration. By varying the frequency and amplitude of the forcing, it may be possible to selectively generate specific wave types that optimise mass transfer in industrial applications. Secondly, extending the analysis to more turbulent flows is essential for understanding the tradeoff between enhanced mixing and reduced residence time. Lastly, applying the principles observed in this study to larger domains like how large-scale ocean and sea waves influence the absorption of atmospheric CO_2 could enhance our understanding of these processes,

contributing to efforts to model and mitigate climate change.

5. Acknowledgements

The authors would like to express their gratitude to the Matar Fluids Group for providing the opportunity to conduct this research. Special thanks go to Dr. Lyes Kahouadji and visiting PhD Andrea Düll from Karlsruhe Institute of Technology (KIT) for their valuable technical assistance and insights.

6. Supplement Information

Supplementary animations of the DNS results are freely available at https://doi.org/10.6084/m9.figshare.28016498.

References

[1] C. D. Park and T. Nosoko, 'Three-dimensional wave dynamics on a falling film and associated mass transfer,' vol. 49, no. 11, pp. 2715–2727, Nov. 2003. DOI: https://doi.org/10.1002/aic.690491105.

[2] G. F. Dietze, W. Rohlfs, K. Nährich, R. Kneer and B. Scheid, 'Three-dimensional flow structures in laminar falling liquid films,' *Journal of Fluid Mechanics*, vol. 743, pp. 75–123, Mar. 2014. DOI: https://doi.org/10.1017/jfm.2013.679.

[3] S. Shin, J. Chergui and D. Jurić, 'A solver for massively parallel direct numerical simulation of three-dimensional multiphase flows,' *Journal of Mechanical Science and Technology*, vol. 31, pp. 1739–1751, Apr. 2017. DOI: 10.1007/s12206-017-0322-y.

[4] L. Kahouadji, A. Batchvarov, I. T. Adebayo *et al.*, 'A numerical investigation of three-dimensional falling liquid films,' *Environmental Fluid Mechanics*, vol. 22, pp. 367–382, Mar. 2022. DOI: 10.1007/s10652-022-09849-2.

[5] S. V. Alekseenko, V. E. Nakoryakov and B. G. Pokusaev, 'Wave formation on a vertical falling liquid film,' *AIChE*, vol. 31, pp. 1446–1460, Sep. 1985. DOI: 10.1002/aic.690310907.

[6] N. J. Balmforth and A. Provenzale, *Geomorphological fluid mechanics*. Springer, 2011, ISBN: 9783642076930.

[7] J. B. Grotberg, 'Pulmonary flow and transport phenomena,' *Annual Review of Fluid Mechanics*, vol. 26, no. 1, pp. 529–571, Jan. 1994. DOI: 10.1146/annurev.fl.26.010194.002525.

[8] R. V. Craster and O. K. Matar, 'Dynamics and stability of thin liquid films,' *Reviews of Modern Physics*, vol. 81, pp. 1131–1198, Aug. 2009. DOI: 10.1103/revmodphys.81.1131.

[9] S. Chen, T. Zhang, L. Lv, Y. Chen and S. Tang, 'Simulation of the hydrodynamics and mass transfer in a falling film wavy microchannel,' *Chinese Journal of Chemical Engineering*, vol. 34, no. 6, pp. 97–105, Oct. 2020. DOI: 10.1016/j.cjche.2020.09.014.

[10] M. Zanfir, A. Gavriilidis, C. Wille and V. Hessel, 'Carbon dioxide absorption in a falling film microstructured reactor: Experiments and modeling,' *Industrial & Engineering Chemistry Research*, vol. 44, no. 6, pp. 1742–1751, Feb. 2005. DOI: 10.1021/ie049726k.

[11] L. Ma, Y. Ding, X. Zhu, H. Wang and Q. Liao, 'Numerical investigation on falling film flow and heat transfer characteristics over corrugated plates,' *International Journal of Thermal Sciences*, vol. 198, pp. 108 882–108 882, Jan. 2024. DOI: 10.1016/j.ijthermalsci.2024.108882.

[12] Y.-q. Guo, W.-r. Hong and J.-U. Repke, 'Hydrodynamics of new structured packings: An experimental and micro-scale cfd study,' *Microgravity Science and Technology*, vol. 30, no. 6, pp. 911–924, Sep. 2018. DOI: 10.1007/s12217-018-9648-z.

[13] L. Zhao and R. Cerro, 'Experimental characterization of viscous film flows over complex surfaces,' *International Journal of Multiphase Flow*, vol. 18, pp. 495–516, Jul. 1992. DOI: 10.1016/0301-9322(92)90048-l.

[14] Y. Lu, F. Stehmann, S. Yuan and S. Scholl, 'Falling film on a vertical flat plate – influence of liquid distribution and fluid properties on wetting behavior,' *Applied Thermal Engineering*, vol. 123, pp. 1386–1395, Aug. 2017. DOI: 10.1016/j.applthermaleng.2017.05.110.

[15] H. Lu, L. Lu and X. Gao, 'Mass transfer enhancement of falling film liquid desiccant dehumidification by micro-baffle plates,' *International Journal of Heat and Mass Transfer*, vol. 169, p. 120 945, Apr. 2021. DOI: 10.1016/j.ijheatmasstransfer.2021.120945.

[16] A. Düll, A. C.-L. Lagadec, J. Buchmüller, T. Häber, C. Ates and M. Börnhorst, 'Intensifying interfacial oscillations in falling film flows over rectangular corrugations,' *Physics of Fluids*, vol. 36, no. 9, Sep. 2024. DOI: 10.1063/5.0222760.

[17] S. Kalliadasis, C. Ruyer-Quil, B. Scheid and M. G. Velarde, *Falling Liquid Films*. Springer Science & Business Media, Sep. 2011.

[18] N. Ghasem, 'Efficient co2 absorption through wet and falling film membrane contactors: Insights from modeling and simulation,' *Scientific Reports*, vol. 13, Jul. 2023. DOI: 10.1038/s41598-023-38249-9.

[19] G. F. Versteeg and W. P. M. Van Swaaij, 'Solubility and diffusivity of acid gases (carbon dioxide, nitrous oxide) in aqueous alkanolamine solutions,' *Journal of Chemical & Engineering Data*, vol. 33, pp. 29–34, Jan. 1988. DOI: 10.1021/je00051a011. (visited on 07/08/2020).

[20] G. F. Dietze, 'Effect of wall corrugations on scalar transfer to a wavy falling liquid film,' *Journal of Fluid Mechanics*, vol. 859, pp. 1098–1128, Nov. 2018. DOI: https://doi.org/10.1017/jfm.2018.851.

Evaluating future pathways for decarbonization in the UK's Road Vehicles sector through agent-based modelling

Ka Chun Lai and Louisa Tai
Department of Chemical Engineering, Imperial College London, UK

12th December 2024

Abstract To combat climate change and achieve net zero by 2050, the UK has put efforts in decarbonizing the transport sector which contributes to 20-26% of direct CO_2 emissions. In particular, various policies have been put in place in the road vehicle sector to promote technological development of zero emission vehicles. Evaluating whether these interventions have their intended impact requires the use of Integrated Assessment Models (IAMs). Existing models rely on cost optimization methods to calculate investment decisions but lack representation of different population groups and their behaviors. Therefore, this study uses ModUlar Energy Systems (MUSE), an agent-based model to evaluate future pathways for decarbonization by considering different consumer behaviors in relation to their uptake of cars. Scenarios examined include Business-As-Usual (BAU), Fuel Efficiency Development, Carbon Budget implementation and Technological Development. The combination of Technological Development and Carbon Budget resulted in full decarbonization by 2040, with Battery Electric Vehicles (BEVs) comprising 100% of the vehicle supply. Further analysis indicates that fast development of BEVs is crucial towards early decarbonization. Finally, a sensitivity analysis was performed to provide producers a minimum capital cost target for BEVs, estimating that a cost of 1968.00 MGBP $bvkm^{-1}$ is sufficient to achieve complete decarbonization by 2040. The findings reaffirm that the use of effective policies to discourage emission-intensive vehicles and the improved affordability and performance of zero-emission vehicles are essential to achieving widespread adoption and a faster rate of decarbonization.

1. Introduction

Global warming has increased the Earth's surface temperature by approximately 1.2°C since the late 1800s, primarily driven by human activities such as the burning of fossil fuels. This has led to significant environmental challenges, including rising sea levels, extreme weather events, melting polar ice, and declining biodiversity. To address these impacts, the UK has committed to achieving net-zero carbon emissions by 2050, positioning itself as a leader in climate action. Compared to others, it is the first major economy to halve greenhouse gas emissions between 1990 and 2022 while growing its economy by 79% [1]. These advancements have been made by decarbonizing and promoting technological development of emission intensive sectors, which includes transport.

In 2021, the transport sector accounted for 26% of the UK's total greenhouse gas emissions, with cars contributing over half (52%) of these emissions [2], indicating that emission abatement is crucial within these technologies. Thus, recent technological advancements in the production of Zero Emission Vehicles (ZEVs), which includes battery electric vehicles (BEVs) and Hydrogen Fuel Cell Vehicles (HFCs) have offered a pathway towards decarbonization. In 2023, the number of EVs in the UK has been accelerating, with new registrations in 2023 doubling that in 2020 [3]. Given the rapid progress, it is crucial to assess whether these developments are sufficient to achieve decarbonization and make steps towards achieving net-zero emissions.

Typically, Integrated Assessment Models (IAMs) are common models that are used to provide the UK government with predictions and pathways for decarbonization. However, these models do not fully capture consumption patterns from different population groups. An exception for this is MUSE, an IAM that uses agent-based modeling (ABM) to capture the diverse behaviors and decision-making processes of consumer groups, to provide socio-economic insights on energy consumption, technology demand, supply, and emissions.

Therefore, this study focuses on using cars as a representation of the road vehicles sector and as a key consumer good to leverage the strengths of Agent-Based Modeling (ABM). The research aims to provide a realistic and comprehensive analysis of whether governmental policies and technological advancements on cars can drive the decarbonization of road transport in the UK, through consideration of consumer behavior. Additionally, the research seeks to offer economic insights for producers and governmental agencies on potential pathways to achieve a low-carbon, sustainable future in the automotive industry.

2. Background

2.1 UK transport

The UK government has implemented several key initiatives to accelerate the transition to low-carbon transport. These include a ban on the sale of new petrol and diesel vehicles by 2030 and hybrids by 2035 [4], the launch of the Transport Decarbonization Plan in 2021 and financial incentives such as EV purchase subsidies and grants for home and public charging infrastructure. With the number of resources being invested, there is a need to understand whether these policies are effective in reaching their environmental and economic targets.

Previous research has emphasized the critical role of zero emission vehicles in transport decarbonization and the importance of policies, such as carbon budgets, using UK energy system model (ESME) [5]. While these studies have provided valuable insights for decarbonization, they are

broad and often address the entire economy. This approach lacks the specificity needed to implement practical solutions within individual sectors. To address this gap, this study focuses specifically on the road vehicle sector, aims to evaluate to potential pathways to decarbonization and offer sector-specific recommendations.

2.2 Overview of IAMs

IAMs are comprehensive tools designed to simulate complex interactions and feedback mechanisms between socioeconomic systems and natural systems over a long period of time. These models are explicitly developed to support climate policymaking by analyzing the implications of various strategies [6]. Over time, many IAMs have been developed with unique techniques tailored to specific contexts. Among the most prominent are energy systems models (ESM), which analyzes energy production, transformation, and consumption. Examples include partial equilibrium models such as MARKAL and TIMES [7], both of which utilize optimization techniques to compute the least-cost energy system pathways.

However, Lamperti *et al* mentions that IAMs rely on general equilibrium models, where representative sectors (such as producers and consumers) interact in perfectly efficient markets [8]. This simplification overlooks the diversity and complexity of real-world systems, particularly within consumer-based sectors. Consumers' heterogeneity, including their investment capacities and risk appetites, critically influence their ability and their preference for a specific technology.

2.3 MUSE

ModUlar energy systems Simulation Environment (MUSE) is an agent-based model that supports flexible characterization of agent decision-making. MUSE integrates this agent-based approach in a partial-equilibrium framework and enables a technology-rich description of the energy systems with an unprecedented degree of flexibility for including technological, temporal, and geographical granularity. To further reflect real-world scenarios, MUSE also implements partial foresight. A property which limits agents' knowledge on future techno-economic parameters to reflect real-world uncertainty during their decision-making process. These features offer a more realistic understanding of social behaviours within energy systems, aligning closely with the objectives of this research.

MUSE's agent-based approach adds another layer of sophistication by incorporating the behavior of decision-makers with varying investment strategies. Agents can optimize for different objectives, including minimizing costs, maximizing comfort, reducing emissions etc, while adapting to factors such as demand fluctuations, price changes, and technology availability. This behavioural modelling is critical for exploring how the rate of decarbonization in the road vehicle sector evolves under different scenarios.

3. Methodology

3.1 MUSE Algorithm and MCA

The simulation starts when an initial exogenous demand and initial price was fed into each sector. Based on these initial values, agents chose a combination of end-use technologies depending on the exogenous values, agent objectives, constraints, and decision strategies. The chosen technologies are aggregated to form a commodity demand that is fed into the Market Clearing Algorithm (MCA). The MCA subsequently passes these demands to the supply or conversion sectors who supplies the required energy commodities and creates a price for them. The MCA then determines whether the demand and supply from different sectors matches with each other, if not then the price will then be fed back into the demand sector, where an iterative process occurs. In each iteration, MCA compares prices until they converge. At this point sectors agreed on the quantity of each commodity exchanged and an equilibrium was reached.

The model used in this study contains two sectors: the road vehicles and distribution sector. The road vehicles sector which contained all car technologies, grouped under the commodity of 'cars' and included the fuel consumption of each technology. The distribution sector represented all fuel types, with each one representing a separate commodity. Additionally, files containing exogenous demands for cars and price projections for fuels were included.

3.2 Technology Section

To reflect technologies that were developed and available in the car market, the model included 5 main vehicle types: petrol cars, diesel cars, plug-in hybrid vehicles (PHEVs), hybrid vehicles (HEVs), battery electric vehicles (BEVs), and hydrogen fuel cell vehicles (FCEVs). Since PHEVs and HEVs can run on both petrol and diesel, they were treated as separate technologies. Additionally, FCEVs were divided into two categories: standard hydrogen fuel cell vehicles (HFCs) and plug-in hydrogen fuel cell vehicles (PHFCs).

3.3 Parameters and Temporal Granularity

Each technology included within MUSE requires a variety of technical and economic data for agents to consider during their investment decisions, these parameters included but were not limited to, capital costs, fixed costs, fuel costs, maximum capacity growth and lifetime of the technology.

Most of this was derived from the UK TIMES energy model, ensuring alignment with established UK energy system projections. Other sources of data included government projections, censuses, existing databases, and literature sources. All techno-economic data was summarized within the supplementary material.

The benchmark years for this model range from 2020 to 2050, in 5-year intervals, with input data required from 2020 onwards in 5 or 10-year intervals. If data gaps existed between these years, the model used cubic interpolation to fill them. Partial foresight for agents was set to 5 years.

3.3.1 Capital Cost

Capital cost (C_t) represented the investment cost for the technology. It was calculated using two parameters, cap_par and cap_exp, as shown in the equation below:

$$C_t = cap_{par} * (Capacity)^{cap_exp} \quad (1)$$

cap_par was estimated from the sum of capital and operational costs given in the UK TIMES Model, in units of MGBP bvkm^{-1} (Million Great British Pounds per billion vehicle kilometers). For simplification, cap_exp assumed to be 1, so that capacity growth was independent to capital costs.

3.3.2 Emission cost and Carbon Budget

When a carbon budget was implemented in the model, cumulative emissions for the road vehicle sector were calculated and compared against the budget to determine if they exceeded or remained within the limit. If the budget was exceeded, MUSE calculated a carbon price which was fed back towards the same sector. This carbon price added towards the total cost during an agent's investment decision; therefore, it was useful in driving down the consumption emission intensive vehicles.

3.3.3 Levelized Cost of Transport (LCOT)

The levelized cost of transport was the cost per unit of transport output over the entire lifespan of a vehicle, making it a long-term cost. This value is in units of MGBP bvkm^{-1}.

$$LCOT = \frac{\sum_{t=1}^{n}(C_t + F_t + E_t)}{\sum_{t=1}^{n} T_t} \quad (2)$$

Here, t represented the year count, F_t represented the fixed cost, E_t represented the emission costs, both calculated in MGBP. T_t represented the total vehicle output within a year and was displayed as the distance travelled for the entire lifetime of a vehicle technology, expressed in units of bvkm.

3.3.4 Efficiency (Fuel Efficiency)

Efficiency (η) represented the amount of fuel (PJ) required to produce 1 PJ worth of energy within each car technology. It served as a measure of how effectively a technology converted fuel energy into electrical energy. Therefore, it can be adjusted by modifying input and output values.

$$\eta = \frac{1}{Fuel\ Energy\ Input\ (PJ)} \quad (3)$$

Efficiency was also converted into a percentage and added within the techno-economic data of a specific technology, making it an objective for agents to consider.

3.3.5 Existing Capacity

Existing capacity referred to the initial stock of vehicles available at the beginning of each investment year, where it can be used to satisfy agent demands. For simplification, the study assumed that after 2020 there would be no existing capacity, ensuring that all cars produced in an investment year would be allocated to satisfy demand.

3.4. Scenario Development

Four scenarios were explored to understand the impact of policies and the current paths for road vehicle decarbonization in the UK.

3.4.1 Scenario 1: Business-as-Usual (BAU)

Business-as-Usual (BAU) represented a baseline projection of future trends without the implementation of policies aimed at reducing emissions. Conservative capital cost projections for the period 2010 to 2050 were based on the UK TIMES model, with fuel efficiency assumed to remain unchanged at 2020 levels. These inputs helped to reflect a scenario with minimal technological advancements from 2020 onward. This scenario provided a clear reference point for evaluating the potential impact and effectiveness of alternative cases.

3.4.2 Scenario 2: Fuel Efficiency Development

In a partial step to understand the impact of technological development, this scenario has allowed fuel efficiencies to increase linearly from 2020 to reach a final target set in 2050. As efficiency was both a techno-economic parameter that impacted fuel inputs, it was assumed that the percentage change in efficiency corresponds directly to the percentage change in fuel consumption. Fuel efficiencies for BEVs, HFCs, HEVs, diesel and petrol vehicles in 2020 and 2050 were sourced from a report produced by Transport and Environment [9] and can be summarized in table 1.

Table 1 Fuel Efficiency values in 2020 and projected values in 2050 [9]

Technologies	Efficiency (%) (2020)	Efficiency (%) (2050)
Petrol Vehicles	16.00	18.00
Diesel Vehicles	20.00	22.00
HEV (Petrol)	36.00	38.00
HEV (Diesel)	40.00	42.00
PHEV (Petrol)	69.00	73.00
PHEV (Diesel)	76.00	79.00
BEV	77.00	81.00
HFC	33.00	42.00
PHFC	68.75	75.25

3.4.3 Scenario 3: Carbon budget

An additional step towards reflecting reality was integrating the UK's current carbon budget framework with fuel efficiency development. This scenario aimed to assess the effectiveness of policies in driving decarbonization.

To calculate the carbon budget for cars, the overall carbon budget was first multiplied by the percentage of emissions attributed to transportation for each benchmark year, then multiplied by the percentage of transport emissions attributed to cars to obtain the final value. As governmental data for transport and car emissions were only available up to 2030, the carbon budget was then extrapolated linearly towards net-zero emissions by 2050.

3.4.4 Scenario 4: Technological Development

This scenario examined the full impact of technological advancements in ZEVs by considering further reductions in capital costs. The capital costs of other vehicles would follow

Table 2 Explanation of different types of technological growth and their representative mathematical models

Types of Growth	Explanation	Function Type	Equation
Fast	Rapid technological development in the beginning, leading to fast reduction to its final target cost, plateaus after target is met	Exponential Decrease	$C(t) = C_0 e^{-kt}$
Slow	Progressive technological development, slow reduction in costs in early years, followed by a sharp decline at later years	Logistic (S-curve)	$C(t) = \frac{C_0}{1 + e^{-k(t-T)}}$
Linear	Linear technological development, constant rate of cost reduction to its final target cost	Linear	$C(t) = C_0 - kt$

their original cost trajectories established in previous scenarios, as their technologies were assumed to be matured. Furthermore, as an indirect effect of growth, fuel efficiency development was also included in the scenario.

Cost reductions for ZEVs were modelled based on the conclusions drawn on their current technological development rate. For BEVs, the established mass production of batteries, increased automation, and ease of manufacturing through the "unboxed process" has driven rapid price declines [10]. Therefore, within this model, they were assumed to experience a faster initial cost reduction, with the rate of decrease plateauing sooner. Conversely, expensive fuel technologies, lack of infrastructure and high costs of transportation for hydrogen has caused fuel cell vehicles to remain at their high prices [11]. Thus, within this model, HFCs and PHFCs were modeled with a slower initial cost decrease. However, HFCs had significant potential for innovation, which could lead to breakthroughs that reduce costs significantly after 2040.

Technological Development Pathways

There was a possibility that technological growth did not follow the trends seen in the section above, therefore, a further extension of this scenario explored different cost reduction rates for BEVs, HFCs and PHFCs, and whether different combinations could yield an optimal future pathway for decarbonization in the road transport sector. Three types of technological growth were summarized in table 2.

Mathematical Modelling

As shown above, simple mathematical functions were used to model these types of cost reductions, where capital cost values between 2020 and 2050 could then be obtained. Within these equations, $C(t)$ was the capital cost, C_0 was the initial capital cost in 2020, k is the annual reduction rate and t is time and T is total time or sigmoid point.

To model the equation, the parameter k and C_0 was required to be found for the exponential and linear model. In this case, two points were used to fit the model. The initial point was determined using the 2020 data from the UK TIMES model, and the end point was assumed to be half of the 2020 value by 2050 [12]. For the logistic curve, the sigmoid was assumed to occur in 2040 to represent late development.

According to the Advanced Propulsion Centre report, the capital costs of PHFCs and HFCs were unlikely to drop below those of BEVs [11]. Therefore, to simplify the modeling process and to ensure this happens, the PHFC and HFC cost curves were assumed to have the same gradient.

Hence, HFC and PHFC cost curves were generated based on the annual reduction rate of the BEV curve, whilst ensuring alignment at their respective initial and final cost points.

3.5 Agent Characterization

To simulate consumer behavior in response to policy changes and technological development, five representative agents have been introduced to model UK consumer groups based on preferences, income levels and socio-economic status.

To categorize the population into specific consumer groups and model their respective uptake of cars each benchmark year, each agent was assigned a quantity weighting which is calculated based on their car ownership rates according to their income bracket, reflecting the average number of vehicles they own or purchase [13].

Each agent was then modeled with two or three objectives based on their specific preferences. This was done by employing a weighted sum decision-making approach, where the weights for each objective are determined based on the percentage of survey respondents according to their income level. The survey assessed the UK public's motivations for adopting new technologies [14]. Some agents were assigned spending limits to better reflect their willingness for purchasing a car, which was calculated based on their annual income and the average frequency at which they replaced their cars [15].

Table 3 Agent characterization with the relevant quantity, objectives, weightings and spend limit

Agent (Quantity)	Objectives	Weights	Spend Limit
Environmentalists (4%)	Emission cost, Efficiency	0.77, 0.23	Unlimited
Early Adopters (12%)	Emission cost, Efficiency	0.55, 0.45	Unlimited
Followers (37%)	Emission cost, Efficiency, LCOT	0.33,0.33, 0.33	Unlimited
Practical Commuters (36%)	Fuel consumption cost, LCOT	0.60, 0.40	2500
Low-Income (10%)	Capital cost, Fuel Consumption cost	0.75, 0.25	2000

Environmentalists
Environmentalists have a high-income level and are financially secure. This group consists of households that are enthusiastic about new technologies and are typically open to embracing forward-thinking solutions. They view ZEV ownership as a social responsibility and a status symbol, valuing their efficiency and emissions reduction over cost considerations.

Early Adopters
This group is similar to Environmentalists with a high-income level and preference to reduce emissions. However, they are driven by a passion for advanced technology. Therefore, they pay premium prices for vehicles that offer high fuel efficiency and enhanced performance.

Followers
This group is driven by a combination of environmental awareness and financial considerations, whilst still being open to the adoption of new technologies. They are willing to reduce emissions in a cost-effective manner and tend to follow the lead of Environmentalists and Early Adopters.

Practical Commuters
This group consists of budget-conscious households that purchase cars primarily for everyday use and are apathetic to their environmental impact. They are willing to adopt new technologies if they are affordable and practical for daily use.

Low-Income
This group has limited purchasing power and cannot afford new vehicles. They focus on prioritizing short-term costs as this takes up most of their available budget. Consequently, they are more likely to opt for cheaper and used vehicles.

4. Results and Discussion

4.1 Scenario 1: Business-As-Usual (BAU)

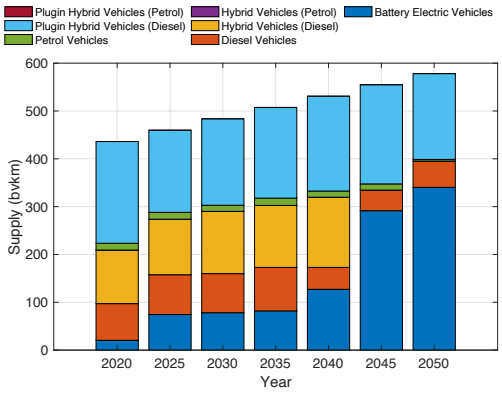


Figure 1. Overall of different road vehicles from 2020 to 2050 (BAU)

In this scenario, where the present situation does not change and no policies are applied, plugin hybrid diesel vehicle (PHEV (diesel)) and hybrid diesel vehicles (HEV (diesel)) make up most of the supply from 2020 to 2040, starting from 74.47% to 65.00% respectively. This can be attributed to their high fuel efficiencies which are at a value of 40.00 and 77.00% respectively, as well as being cheaper or equivalent to the price of BEVs. The uptake of these technologies is primarily driven by followers and practical commuters. Their consumption behavior revolves upon purchasing vehicles which are cost-effective in the long term, suggesting that they prioritize vehicles that travel further and have higher fuel efficiencies. As they also represent 73.00% of the total demand of the population, it aligns with reason why these technologies dominate the supply.

The uptake of BEVs also starts slowly due to its high starting capital costs and the lack of existing capacity in the market. In 2020, BEVs only take up 4.63% of the total , only experiencing significant growth from 2040. Prior to this point, all consumption of BEVs can be attributed to environmentalists and early adopters, due to their primary objective of maximizing efficiency, as well as having no constraints in minimizing costs. Both agents start with purchasing BEVs and PHEV (diesel) in 2020 but shift their consumption fully towards BEVs after. After 2040, practical commuters start to invest in BEVs, as they will become more affordable and become a superior option compared to hybrid vehicles due to its high fuel efficiency and lower LCOT. The uptake of BEVs is also hindered by the consumption behavior of low-income agents. Due to their lack of disposable income, they are only restricted to investing in diesel and petrol vehicles.

In terms of HFCs and PHFCs, it is seen that there is no development or investment within these technologies. Compared to BEVs, they are both more expensive in terms of capital cost as well as being less fuel efficient, at a value of 68.75%. Hence, BEVs serve as the superior technology for zero-emission vehicles.

4.2 Scenario 2: Fuel Efficiency Development

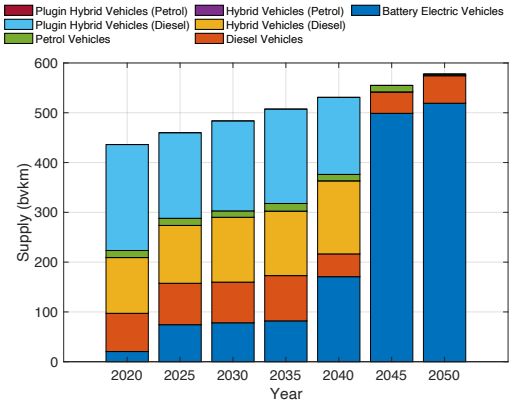


Figure 2. Overall supply of different road vehicles from 2020 to 2050 (Fuel Efficiency Development)

In comparison to BAU, this scenario assumes that fuel efficiencies for each technology will increase to different values in the future, which presents a more realistic situation. The only change as shown in figure 2, is the significant increase in uptake of BEVs from 2045, attributed by the replacement of PHEV (diesel) by followers. In contrast to the BAU scenario, investment in BEVs have started earlier in

2040. This uptake can be explained by section 3.4.2, where BEV fuel efficiencies are predicted to increase the most out of all other technologies due to further development, rising from 77% to 81% by 2050, whilst PHEV (diesel), only grows from 76% to a value of 79%. Thus, this makes BEVs become more competitive against PHEV (diesel), hence it becomes the primary source of investment for the follower agent in later years.

HFCs and PHFCs also experience a growth in fuel efficiency from 33 to 42% and 68.75 to 75.25% respectively. However, like the scenario above, their capital costs are still relatively high compared to other alternatives with similar efficiencies, such as PHEV (Diesel) (40-42%), HEV (diesel) and BEVs.

4.3 Scenario 3: UK Carbon Budget

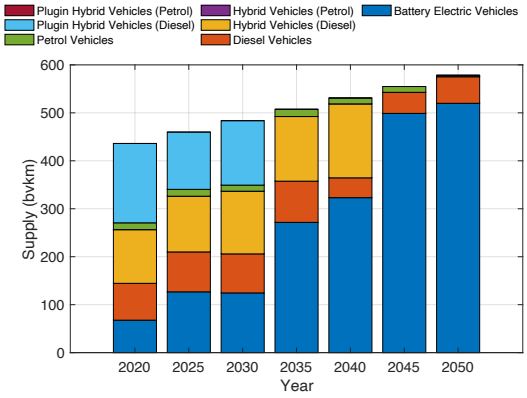


Figure 3. Overall supply of different road vehicles from 2020 to 2050 (Carbon Budget)

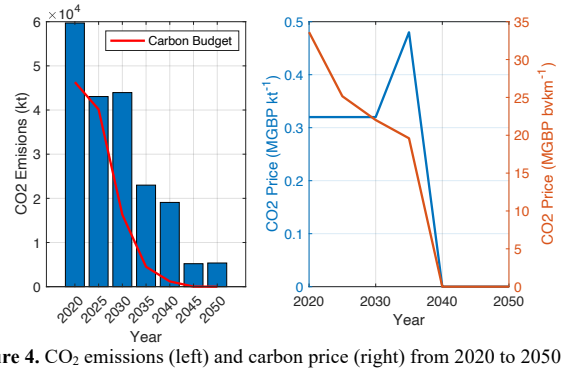


Figure 4. CO_2 emissions (left) and carbon price (right) from 2020 to 2050

From previous scenarios, achieving complete decarbonization of the road vehicle sector still remains challenging. Thus, this scenario explores the impact of policy intervention, specifically the UK's carbon budget. The implementation of such creates a carbon price, creating an additional cost to emission producing vehicles, hence encouraging the investment of emission free alternatives. As such, this is reflected in the increased uptake of BEVs from 2020 to 2040. Although it is expected that this change can impact environmentalists, early adopters, and followers, due to their objective in minimizing emission costs, it is only followers that has driven the increase in supply. Environmentalists and early adopters have already put all their investments into BEVs regardless of the implementation of the carbon budget, as seen from previous scenarios. With BEVs having zero emissions, it means that their cost is already minimized and there is no incentive to switch technologies. Thus, introducing a carbon budget does not affect their uptake over time.

For practical commuters and low-income users, no change in consumption patterns can be observed. Although practical commuters minimize their LCOT which includes emission costs, it can be seen from figure 4 that the carbon price only reaches a maximum price of 0.48MGBP kt^{-1} (equivalent to 19.59 MGBP $bvkm^{-1}$) and is small compared to the capital costs of each technology per year. Moreover, the additional emission cost for HEV (diesel) is not sufficient to allow BEVs to appear more affordable. For low-income users, as they prioritize minimizing capital and fuel costs, they are oblivious to changes in carbon pricing, hence their investment decisions are not affected.

From the above, it is evident why followers are the only contributor to the increased uptake of BEVs. As these agents have previously invested in large amounts of PHEV (diesel) which produce emissions, they have the incentive and ability to shift their investments towards BEVs starting from 2020. Compared to the fuel efficiency development scenario, complete investment in BEVs starts earlier at 2035. Overall, this change in behavior is due to the follower's incentive in maximizing efficiency and minimizing their emission cost, which will outweigh LCOT.

4.4 Scenario 4: Technological Development

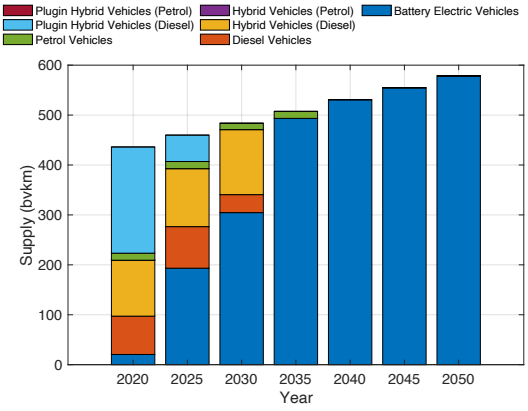


Figure 5. Overall supply of different road vehicles from 2020 to 2050 (Technological Development)

Evidently, the issue with implementing a carbon budget within this model is that the cost of emission vehicles is relatively unaffected and only one agent contributes to the increase in BEV uptake. Hence, this scenario explores the idea of technological development.

Decreasing the cost for BEVs has increased the affordability of the technology to different agents, reflecting the increase in uptake and the complete adoption of the technology by 2040. Agents that are affected the most are followers, practical commuters, and low-income agents, as they all aim to either minimize their LCOT or capital costs.

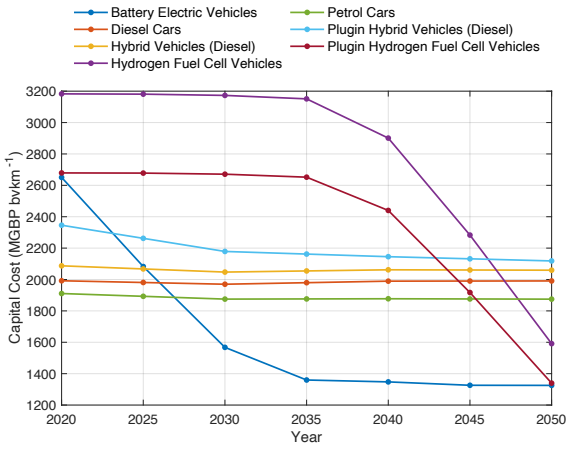


Figure 6. Capital costs for different technologies from 2020 to 2050, BEVs (blue) experiencing fast cost reductions, HFCs (burgundy) and PHFCs (purple) experiencing slow cost reductions (**Specific capital costs can be found in supplementary information**)

Environmentalists and early adopters are unaffected due to their cost insensitivity. For followers, they start investing in BEVs in 2025, which is 5 years later compared to the carbon budget scenario. This is because no emission costs are involved, hence the agent only prioritizes minimizing its LCOT. Thus, as the cost of BEVs which is at 2670.78 MGBP bvkm^{-1} is more than PHEV (diesel) which is at 2345.88 MGBP bvkm^{-1}, the latter is still the preferred option in 2020. However, complete investment occurs earlier in 2030, as the capital cost of BEVs which is at 1567.60 MGBP bvkm^{-1} becomes significantly cheaper than PHEV (diesel) which is at 2178.73 MGBP bvkm^{-1}.

For practical commuters, compared to scenario 2 and 3, investment in BEVs start earlier at 2030, and there is complete adoption by 2035. Similar reasoning to above, in 2020, the capital costs for BEVs were more expensive than HEV (diesel) which had a cost of 2087.47. In 2025, the prices become similar and only after this BEVs become more affordable. Low-income agents who were initially resistant to adopting new technologies as shown in scenarios above, start to invest in BEVs by 2035 with complete adoption occurring in 2040, as BEVs become more affordable than petrol and diesel cars. Their later investment compared to other agents can be explained in figure 6, where the cost of BEVs (blue) is higher than petrol and diesel vehicles (green and orange) until after 2025. Compared to other agents, they are still the slowest adopters of BEVs.

As expected, HFCs and PHFCs are not adopted by any agents within this scenario, as their costs fall slowly compared to BEVs, it is only after 2040 that their technologies become affordable to most agents.

4.5 Comparisons of Each Scenario (Emissions)

By gathering results from above, figure 7 shows the effectiveness of each scenario in achieving decarbonization for the road vehicle sector. As shown, technological development reduces CO_2 emissions far more effectively than other scenarios, reaching 100% decarbonization by 2040 and generally having lower CO_2 emissions compared to other scenarios. Changes in capital costs influences most agents' investment decisions, as non-emission vehicles become more affordable in the short and long term. A separate simulation on technological development has been run without fuel efficiency development, however the emission profile remains the same, indicating that it plays a small impact in decarbonization when combined with changes in capital costs.

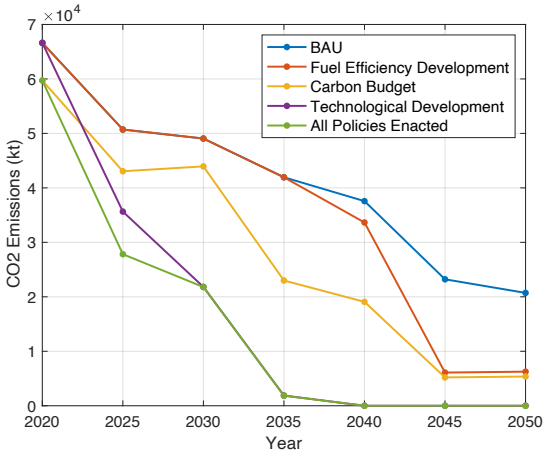


Figure 7. CO_2 emission comparisons of different policies from 2020 to 2050

Interestingly, CO_2 emissions for technological development are higher by 7407.39 kt compared to the carbon budget scenario in 2020. Without a carbon budget, agents have no incentive in reducing emissions as it doesn't incur an add-on cost to their vehicles. Hence, from 2020 to 2030, where BEVs are still more expensive, CO_2 emitting vehicles are still preferred. Thus, a combination of both policies is required, yielding the optimal result as highlighted in figure 7.

In contrast, the sole implementation of a carbon budget only achieves a 91% reduction in CO_2 emission in 2050. However, it still achieves lower emissions every year when compared to BAU and fuel efficiency development scenarios, still highlighting its importance in incentivizing the uptake of BEVs. Again, a separate simulation on carbon budget has been run without involving fuel efficiency development, the same supply profile was produced. However, to incentivize uptake, carbon prices increase exponentially to a maximum value of 33 MGBP kt^{-1}.

4.6 Technological Development Pathways

As seen, technological development plays a significant role in affecting investment decisions for different agents. Thus, in this section, different pathways are explored to identify the optimal path for decarbonization. Six different cases were modelled, with each case being a different possibility of technological growth between HFCs, PHFCs and BEV is summarized in table 4.

Table 4. Six different cases showing the different possibilities of technological growth between BEVs, HFCs and PHFCs

Case	BEVs	HFCs	PHFCs
Base	Fast	Slow	Slow
Linear	Linear	Linear	Linear
Alternative 1	Linear	Fast	Fast
Alternative 2	Fast	Linear	Linear
Alternative 3	Linear	Slow	Slow
Alternative 4	Fast	Fast	Fast

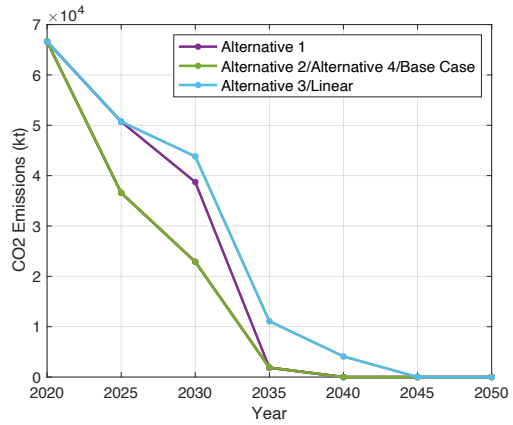


Figure 8. CO_2 emission comparisons for different technological development pathways

After running these simulations within MUSE, the optimal paths are identified to be Alternative 2, base, and Alternative 4, where complete decarbonization occurs in 2040. The commonality between these cases are that BEVs fall quickly to their final costs, increasing their affordability and hence their uptake by all agents, leading to their complete adoption in 2040, as seen in section 4.4. Their equivalence for technology uptake and CO_2 emissions can also be explained by the same reason. As BEVs fall quickly to their final costs which are lower than that of HFCs and PHFCs, in addition to the fact that each function type is modelled under the same rate of decrease for each technology, PHFCs and HFCs are always more expensive than BEVs in every single year. Thus, the different technological development possibilities for HFCs and PHFCs do not affect the result of the decarbonization profile.

Likewise, for Alternative 3 and Linear cases, regardless of whether PHFCs and HFCs follow slow and linear technological growth rates, their capital costs are higher than BEVs across all years. However, the difference is that BEVs now follow a linear path for technological growth, which is a slower rate of cost reduction. Consequently, complete decarbonization occurs later in 2045, as agents opt to replace their emission intensive vehicles in later years.

Despite having higher CO_2 emissions from 2025 to 2035 when compared to the optimal path, Alternative 1 matches the optimal profile after 2035, achieving decarbonization by 2040. These results can be explained by the cost and efficiency of PHFCs. Although the fast development of this technology allows for increased affordability relative to others, it still has an initial efficiency of 68.75% which is smaller compared to BEVs and PHEV (diesel) that both have an initial efficiency of 77%. Agents such as followers take this into consideration, as PHEV (diesel) still make up all their uptake in 2025, despite PHFCs being cheaper by 161.21 MGBP bvkm^{-1}. Followers only start replacing their consumption with PHFCs in 2030 and fully adopt the technology until 2040, as the cost gap between widens.

A similar effect can be observed after 2040, where followers replace PHFCs with BEVs. At this point, the cost of both technologies is relatively similar with the prior being 1340.00 MGBP bvkm^{-1} and the latter being 1546.18 MGBP bvkm^{-1}. Thus, the differentiating factor becomes fuel efficiency, making BEVs the superior option. Consequently, PHFCs act as a suitable substitute for the short term due to its temporary affordability, however as the costs for other technologies become cheaper it is replaced by superior alternatives such as BEVs.

Considering all factors above, alternative 2, Alternative 4 and base case provide the optimal path for decarbonization. To further discern them apart, the LCOT is considered. Intuitively, the Alternative 4 becomes the best option as it lowers the capital costs for both non-emission vehicles at the fastest rate. However, the technological development for hydrogen vehicles is not expected to mature quickly. Therefore, the base case is recommended as the only optimal case, showing that the same rate of decarbonization can be achieved even with a slow rate of development for hydrogen powered vehicles.

4.7 Sensitivity analysis on capital cost

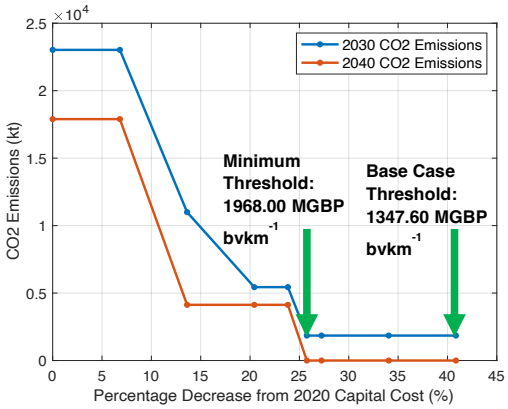


Figure 9. Sensitivity analysis on capital costs for years 2030 and 2040, with minimum threshold and base case threshold capital costs

Although the base case shows the most realistic and optimal path for decarbonization of the road vehicle sector, there is a possibility that cost reduction targets for BEVs might not be achieved. Therefore, a sensitivity analysis is conducted on BEV capital costs, to identify the minimum threshold that producers should target for achieving complete decarbonization by 2040.

The analysis assumes that capital costs will only decrease over time. The upper limit for each investment year's sensitivity analysis is set to the capital cost of the

previous investment year (e.g., for 2030, the upper limit is the 2020 cost, which is 2650.60 MGBP bvkm^{-1}), whilst the lower limit is fixed to the base case cost target for the current year. Additionally, a carbon budget is applied to ensure that agents will also minimize their emission costs to reflect reality. In this case, the sensitivity analysis will be performed by adjusting the capital cost for 2030 until either complete decarbonization is achieved by 2040 or the lower limit is reached. After the minimum threshold is determined, additional sensitivity analysis is performed on the capital cost for 2040 and 2050 to evaluate potential changes in CO_2 emissions.

Figure 9 illustrates the result of the analysis, showing that complete decarbonization in 2040 can be achieved with a 25% capital cost reduction, which corresponds to a value of 1968.00 MGBP bvkm^{-1}. This is larger than the base case limit of 1567.60 MGBP bvkm^{-1} which reflects a cost reduction of 40.86%. Using the new threshold, further sensitivity analyses was conducted in 2040, where between the limits of 1968.00 and 1347.60 MGBP bvkm^{-1}, the amount of CO_2 emissions remain at zero. Similar results were then obtained in 2050. Thus, the minimum cost reduction threshold for producers to target remains at 1968.00 MGBP bvkm^{-1} from 2030 onwards. Notably, these thresholds are only slightly lower than the capital cost for diesel cars. The improved affordability allows low-income agents to purchase BEVs, enabling complete decarbonization by 2040. Evidently, these agents are the most resistant to adopting decarbonization technologies due to their limited purchasing power and the prioritization of short-term investments influenced by vehicle replacement expectations. Thus, from the producer's perspective, aligning BEV capital costs with diesel cars serves as an effective benchmark to incentivize widespread decarbonization.

5. Conclusion

Within the UK, the transport sector has been a key focus for pollution abatement, where the development of emission free vehicles has been incentivized through various subsidies and investment, and policies have been put in place to reduce consumption on combustion vehicles. Whether these actions are impactful and effective in achieving their intended purpose for the future is difficult to quantify and predict. Therefore, this study has used MUSE, to evaluate and assess pathways for decarbonization within the road vehicles sector in the UK by considering consumer behavior in relation to their uptake of cars.

From the scenarios explored, the combination of the UK's carbon budget and technological development of emission free vehicles created the fastest path to decarbonization. The increased affordability of BEVs, PHFCs and HFCs targets population groups with limited purchasing power to replace their investments in petrol and diesel cars with these emission free alternatives, allowing them to completely decarbonize. The simultaneous implementation of a carbon budget also helps to reduce emissions prior to the years when BEVs become largely affordable, as it discourages consumption of pollution emitting vehicles by indirectly increasing its cost. This especially targets practical commuters and followers due to their cost minimizing objectives, incentivizing them to increase their uptake of BEVs earlier.

Further exploration into different combinations of technological development rates for non-emission vehicles was conducted to account for potential future variations. Alternative 2, Alternative 4 and the Base Case were identified as the optimal pathways for development, emphasizing that rapid cost reductions of BEVs are crucial in achieving complete decarbonization by 2040. Ultimately, the Base case is evaluated as the most effective pathway, as it reduces the cost-reduction burden on PHFCs and HFCs, while maintaining the same rate of decarbonization as the other optimal pathways.

A sensitivity analysis was then performed, to identify a suitable cost target for car producers to meet, if the base case pathway is infeasible. The results indicate that adhering strictly to the Base Case pathway is not required, as the minimum threshold to achieve decarbonization in 2040 is 1968.00 MGBP bvkm^{-1} from 2030 onwards. This target, being slightly lower than the cost of diesel cars, is essential for enabling low-income agents to adopt BEVs.

In summary, this study demonstrates that the complete decarbonization of the UK road vehicles sector by 2050 is feasible. The evaluation of future pathways from above highlights the requirement of key strategies for achieving early and effective decarbonization. This includes ensuring the affordability of zero-emission vehicles for low-income populations, implementing policies to discourage consumption of emission-intensive vehicles, and providing subsidies to reduce production costs. These measures will help zero-emission vehicles gain a competitive edge over emission-intensive alternatives, encouraging widespread adoption.

6. Limitations and Future Development

Whilst MUSE incorporates an agent-based approach to simulate human behavior, its results are heavily dependent on the input data that it is given. For instance, as the main objectives for several agents are designed to minimize cost and maximize efficiency, the model preferentially selects diesel powered vehicles over petrol ones. In reality, petrol cars are often preferred for their comfort, lower noise levels, and reduced nitrogen oxide (NOx) emissions. Therefore, over-simplification of certain objectives and poor design of weightings can cause the model to over-rely on a certain piece of technology to satisfy objectives.

Another limitation to MUSE lies in the decision-making process. For example, the model assumes that consumers do not consider emission costs unless a carbon budget is implemented. In the absence of the budget, the emission cost remains at 0, causing agents to prioritize their other objectives. This approach fails to capture consumer awareness for environmental sustainability, which can independently influence the consideration of carbon prices in vehicle purchasing decisions.

Data limitations also present a significant challenge. Input data from the UK TIMES model, inherently carries uncertainties, as well as future projections based on literatures sources. Specifically, within our study, cost reductions in BEVs, HFCs and PHFCs, rely on projections that may not be materialized and mathematical models that may over-simplify results.

To address these limitations, future research should prioritize developing more dynamic behavioral models within the MUSE framework to model reality. One improvement would be to incorporate dynamic weighting systems for agent objectives to capture shifts in consumer behavior over time, due to policy changes, technological development, or emerging consumer trends. This approach would reduce reliance on discrete data sources such as projections or surveys that may introduce biases.

Another area for development involves integrating the power sector into the model. This creates an opportunity to understand how the growth and integration of renewable energy can impact road vehicles and fuel sectors. This also provides a more holistic view on decarbonization by including indirect emissions and add-on costs.

Additionally, technologies such as biofuel and nuclear-powered vehicles can be incorporated into the model. As more data becomes available from the literature regarding the cost projection pathways of these technologies, this enhancement will help to estimate their potential impact and provide deeper insights into alternative decarbonization pathways that may occur.

Acknowledgements

We would like to express our sincere gratitude to Dixit Sharwari and Tom Bland for their invaluable guidance and support throughout this project.

References

[1] Department for Energy Security and Net Zero (2024) *UK first major economy to halve emissions*, *GOV.UK*. Available at: https://www.gov.uk/government/news/uk-first-major-economy-to-halve-emissions#:~:text=The%20UK%20is%20the%20first,USA%20between%201990%20and%202021. (Accessed: 11 December 2024).

[2] UK Parliament (2024) *Electric vehicles and infrastructure*, *House of Commons Library*. Available at: https://commonslibrary.parliament.uk/research-briefings/cbp-7480/ (Accessed: 11 December 2024).

[3] Carlier, M. (2024) *Annual new electric car registrations in the United Kingdom between 2015 and 2023*, *Statista*. Available at: https://www.statista.com/statistics/1411261/uk-annual-electric-vehicle-sales/ (Accessed: 11 December 2024).

[4] Department for Transport (2021) *Decarbonising Transport – a better, Greener Britain*, *Department for Transport*. Available at: https://assets.publishing.service.gov.uk/media/610d63ffe90e0706d92fa282/decarbonising-transport-a-better-greener-britain.pdf (Accessed: 11 December 2024).

[5] Pye, S., Sabio, N. and Strachan, N. (2015) *An integrated systematic analysis of uncertainties in UK Energy Transition Pathways*, *Elsevier*. Available at: https://www.sciencedirect.com/science/article/pii/S030142151400706X#s0100 (Accessed: 11 December 2024).

[6] van Beek, L. *et al.* (2020) *Anticipating futures through models: The Rise of Integrated Assessment Modelling in the climate science-policy interface since 1970*, *Elsevier*. Available at: https://www.sciencedirect.com/science/article/pii/S0959378020307743 (Accessed: 11 December 2024).

[7] UCL (2020) UK Times, UCL ENERGY INSTITUTE MODELS. Available at: https://www.ucl.ac.uk/energy-models/models/uk-times#:~:text=UK%20TIMES%20is%20the%20successor,to%20research%20and%20policy%20questions. (Accessed: 11 December 2024).

[8] Lamperti, F. *et al.* (2018) *Faraway, so close: Coupled climate and economic dynamics in an agent-based integrated assessment model*, *Elsevier*. Available at: https://www.sciencedirect.com/science/article/abs/pii/S0921800917314623 (Accessed: 11 December 2024).

[9] Transport & Environment (2021) *Efficient pathways to electrifying UK transport*, *Transport & Environment*. Available at: file:///C:/Users/User/Downloads/T_2024-04-29-191042_zeng%20(5).pdf (Accessed: 11 December 2024).

[10] Orangi, S. *et al.* (2023) *Historical and prospective lithium-ion battery cost trajectories from a bottom-up production modeling perspective*, *Journal of Energy Storage*. Available at: https://www.sciencedirect.com/science/article/pii/S2352152X23031985#s0065 (Accessed: 12 December 2024).

[11] Advanced Propulsion Centre (2023) *Battery and fuel cell future cost comparison*, *Advanced Propulsion Centre*. Available at: https://www.apcuk.co.uk/wp-content/uploads/2023/02/Battery-and-Fuel-Cell-Cost-Comparison-report.pdf (Accessed: 11 December 2024).

[12] *Nissan launches the arc* (2024) *Nissan*. Available at: https://careersatnissan.co.uk/nissan-launches-the-arc/#:~:text=By%20developing%20EVs%20in%20families,and%20ICE%20models%20by%20fiscal (Accessed: 11 December 2024).

[13] *Household car availability by household income quintile: England, 2002 onwards* (2024) *National Travel Survey, Department for Transport statistics*. Available at: https://assets.publishing.service.gov.uk/media/66ce119f8e33f28aae7e1f7c/nts0703.ods (Accessed: 11 December 2024).

[14] Williams, B. and Watkins, A. (2023) *Climate change insights, families and households, UK: 14 June to 9 July 2023*, *Office for National Statistics*. Available at: https://www.ons.gov.uk/economy/environmentalaccounts/articles/climatechangeinsightsuk/august2023 (Accessed: 11 December 2024).

[15] *Household income* (2022) *GOV.UK*. Available at: https://www.ethnicity-facts-figures.service.gov.uk/work-pay-and-benefits/pay-and-income/household-income/latest/#download-the-data (Accessed: 11 December 2024).

A Deep Deterministic Policy Gradient for Distributed Residential Energy Control

Leonidas Christie and William Ho

Department of Chemical Engineering, Imperial College London, UK

Abstract

This study explores Reinforcement Learning to optimise the real time control of distributed residential energy within a simulated energy environment. A Deep Deterministic Policy Gradient algorithm, with its capabilities in high dimensions and continuous state and action spaces was evaluated for its effectiveness in tackling such an environment. Initial research into Q-Learning with Q-tables, Linear Function Approximation, Deep Q-Networks and Deep Deterministic Policy Gradients in a grid-world case study provided the step up to the final implementation. The results demonstrate that the Deep Deterministic Policy Gradient excels in adapting to this environment with complex dynamics, providing 5.79% cost reduction against heuristics and ensuring a 50% reduction in constraint violations. The environment was designed to be compatible with OpenAI's Gymnasium allowing for the future implementation and analysis of multiple algorithms and standardised testing. This approach holds significant promise for the implementation of Reinforcement Learning methods to optimise and reduce costs in distributed residential energy systems.

Keywords: *Reinforcement Learning, Machine Learning, Neural Networks, Residential Energy Optimisation*

1 Introduction

The transition towards a future of sustainable energy is one of the defining challenges of our time. The energy demands and limitations of renewable energy have introduced unprecedented challenges in real-time grid optimisation and management. While these challenges may seem insurmountable, the large advancements in Artificial Intelligence (AI) and Machine Learning (ML) provide unprecedented opportunities to overcome them. By enabling data-driven decision-making and real-time optimisation, these technologies are well suited to manage the variability and uncertainty inherent in distributed residential energy systems.

Traditional optimisation methods, though effective in static environments, struggle to adapt to the dynamic and continuous nature of energy systems [1], leading to highly expensive computation and bottlenecks after unforeseen disturbances. This paper looks into Reinforcement Learning (RL), a subset of ML. It can tackle the environment despite the stochasticity and partial observability caused by weather, or restrictions in real-time communication of private data, making decentralised decisions. RL can address key challenges in balancing supply and demand, managing flexible loads, and minimising costs and environmental impact [2]. Specifically, in distributed residential energy systems with renewable energy and variable demands, RL can optimise the use of assets like electric vehicles, batteries, and electric heating systems to reduce grid dependency and carbon emissions [3].

This paper addresses a similar problem statement tackled by *Charbonnier et al.* [3], moving away from their Q-table Multi Agent RL (MARL) approach to investigate a Deep Deterministic Policy Gradient (DDPG) algorithm [4]. It takes a step-by-step approach tackling the basics of Q-learning via Q-tables, Linear Function Approximation (LFA), Deep Q-Networks (DQNs) and finally a DDPG. This initial research is performed within a grid-world environment, a classic test case for RL algorithms. The analysis then extends to a simulated distributed residential energy environment using the DDPG, where the dynamic interplay of energy supply, demand and control presents a realistic challenge for the RL-based solutions.

The primary objective of this study is to assess the performance of DDPGs in training an agent (an autonomous decision-maker that interacts with an environment) to solve complex energy optimisation tasks and to evaluate their potential for broader applications in residential energy systems. By benchmarking DDPG against the linear Q-learning approach and other heuristics, the study aims to uncover its strengths, limitations, and implications for future research and real-world deployment, offering insights into how these methods could support the transition to sustainable and intelligent energy grids.

2 Background

2.1 Reinforcement Learning

Reinforcement Learning (RL) trains agents to make sequential decisions by interacting with an environment to maximise rewards over time. Inspired by behavioural psychology, RL mimics trial-and-error learning, where actions are reinforced by rewards or penalties [5]. RL has become a powerful tool for solving complex decision making problems, mostly through its iterative approach which allows it to solve dynamic problems by learning adaptive strategies in uncertain and evolving environments. It has been successfully applied to varying areas, such as video games, anomaly

detection, robotics and financial management [6]. **Figure 1** encapsulates the basic idea of a RL algorithm. The agent receives a state, s_t, at time t, and outputs an action, a_t. The agent then receives the next state, s_{t+1}, through a transition function, and a scalar reward, r_{t+1}, updating its decision process based on this reward. The agent will continuously take steps until reaching a terminal state where the agent restarts the process.

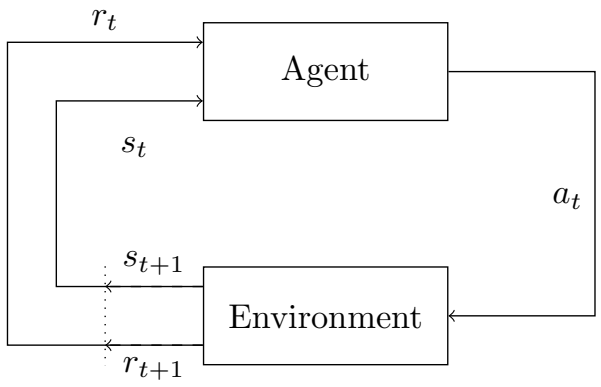


Figure 1: Agent-Environment Interaction Diagram. The agent takes actions in the environment and receives states and rewards as feedback.

The process is modelled as a Markov Decision Process (MDP), which consists of states, actions, transition probabilities, and rewards. The Markov Property states that the future is independent of the past given the present, i.e. the current state [5]. An agent's policy, π, is used to provide a mapping from the observation to the action and can be deterministic or stochastic. The agent also has to balance exploration (trying new actions to find better rewards) and exploitation (taking known actions that provide the most reward). The return, G_t, shown in **Equation 1**, is the total discounted reward from time-step t onwards and γ is the discount factor ($0 \leq \gamma \leq 1$) which scales the importance of immediate or future rewards [5].

$$G_t = r_{t+1} + \gamma r_{t+2} + \cdots = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \tag{1}$$

$$v(s) = \mathbb{E}\left[G_t \mid S_t = s\right] \tag{2}$$

Value functions, $v(s)$, are then defined as the expected total discounted reward and predict the total future reward given a starting state as seen in **Equation 2**. The agent's objective is to find the optimal policy, maximising the cumulative discounted reward for each episode.

2.2 Q-learning

Q-learning is a cornerstone of RL, known for its off-policy nature, where learning an optimal policy is decoupled from the behaviour policy used to generate data. Q-learning estimates the value of state-action pairs as a Q function, $Q(s, a)$, which represents the expected cumulative reward when an agent takes action, a, from state, s, and then follows the optimal policy. This approach is grounded in the Bellman optimality equation, a fundamental principle that recursively defines the optimal Q-values as seen in **Equation 3** [7].

$$Q^*(s, a) = \mathbb{E}_{s' \sim P}\left[r + \gamma \max_{a'} Q(s', a') \mid s, a\right] \tag{3}$$

$Q^*(s, a)$ is the optimal action value function representing the maximum cumulative expected reward starting from (s, a) and following the optimal policy after, with r as the immediate reward after (s, a). The next state-action pair is denoted by (s', a'). $\mathbb{E}$ is the expectation and P is the transition function.

Q-learning iteratively applies **Equation 4** using temporal difference (TD) learning, which combines the principles of dynamic programming and Monte Carlo methods [5], allowing the Q-function to be updated incrementally based on observed transitions. The update rule, known as the temporal Q-learning update, is expressed in **Equation 4**, where α is the learning rate controlling the magnitude of the updates.

$$Q(s, a) \leftarrow Q(s, a) + \alpha[r + \gamma \max_{a'} Q(s', a') - Q(s, a)] \tag{4}$$

In discrete state-action spaces, Q-learning is traditionally implemented using a Q-table, a tabular representation, where each cell corresponds to a specific state-action pair. It is usually stored as a matrix or look-up table where each row represents a state, and each column represents an action. The values in the table correspond to the expected cumulative reward for taking an action in the state and following the learned policy thereafter. During training, the Q-values in the table are iteratively updated to show improved estimates of future rewards.

However, in high-dimensional or continuous environments, maintaining a table becomes unfeasible due to the curse of dimensionality [8]. This led to the development of function approximation methods, including Linear Function Approximation (LFA)[8]. LFA offers a computationally efficient means to scale Q-learning to environments with continuous or high-dimensional state-action spaces. Instead of relying on a Q-table, this method approximates $Q(s, a)$ using a linear combination of the feature vector, $\mathbf{f}$, and the weights vector, $\boldsymbol{\theta}$, shown in **Equation 5** [9].

$$Q(s, a; \theta) = \sum_{i=0}^{n} f_i(s, a)\theta_i \tag{5}$$

The weights are adjusted iteratively in **Equation 6**, using the temporal difference error, δ, shown in **Equation 7**.

$$\theta \leftarrow \theta + \alpha \cdot \delta \cdot f(s, a) \tag{6}$$

$$\delta = r + \gamma \max_{a'} Q(s', a'; \theta) - Q(s, a; \theta) \tag{7}$$

These equations form the fundamentals of how Q-learning will be applied in this paper.

2.3 Deep Q-Learning

Deep Q-networks (DQNs) [10] are a crossover between Q-learning in RL and Neural Networks (NNs) and are known to excel in high dimension observation spaces as well as handle more computationally complex problems. For example, DQNs achieve expert human level performance in many Atari games [11]. However, one major drawback is that DQNs can not handle continuous or high dimensional action spaces, due to the structural limitation in only having one NN [4]. Therefore, this paper looks at Deep Deterministic Probability Gradients (DDPGs) [4], an extension of DQNs. A

DDPG uses two NNs and therefore can handle continuous action spaces, thereby applicable to the distributed residential energy environment.

The two networks are the the actor network and the critic network. The actor network learns the policy whilst the critic network evaluates the actor's action by estimating the return for the action, behaving like a Q function. These networks combined, achieve the state-action pairs seen in normal Q-learning.

There are a few techniques used in the algorithm to improve stability and performance. Firstly the implementation of target networks is a vital part of a DDPG. Target networks are taken from DQNs and offer many bonuses, including stabilising training by reducing oscillations in target Q-values. It facilitates convergence by decoupling the main network and target networks, and it provides consistent targets for the main function [12].

Next, the use of replay buffers [13] are used to decrease correlation between data points. This works by decoupling updates from sequential data, allowing random sampling of past transitions. This approach improves data reuse, ensures diverse training samples, and facilitates learning from rare but critical events [14][12].

To introduce exploration, noise is sampled from a mean zero Gaussian distribution with a standard deviation input. Noise is vital to make the DDPG policy explore better and improve its learning during training as the policy is deterministic, meaning each state is mapped directly to a given action with no randomness [15]. Furthermore, the noise sampled is decayed over training in order to focus on the optimal path closer to the end of an episode.

Lastly, the inputs into the networks, are normalised between -1 and 1 using the variable's minimum and maximum as bounds. Input normalisation is essential in deep RL, ensuring that each feature contributes equally to the learning process. Without it, features of larger magnitudes can dominate, slowing convergence and destabilising training. Normalisation smooths gradients, accelerates convergence by creating a more balanced optimisation landscape, and prevents issues such as vanishing or exploding gradients [16].

3 Environment Description

The foundational environment, as outlined by *Charbonnier et al.*[3], serves as the basis for the environment in this paper. The fundamental assumption on which this environment is built, is that each prosumer (a producer and consumer of electricity) has both photovoltaic cells (PV), and an electric vehicle (EV). The objective is for each prosumer to learn to control the flexibility offered by these, in conjunction with electric space heating and flexible loads.

3.1 Variables

A set of n prosumers, $i \in P = \{1, ..., n\}$, are considered over a set of time steps $t \in T = \{t_0, ..., t_{end}\}$. At each time step, each prosumer can take an action, which consists of three local decisions: the energy flows in and out of the battery, $b^t_{in,i}$ and $b^t_{out,i}$, and the required heating consumption, h^t_i.

Each prosumer's electricity demand, d^t_i, at each time step is modelled by sampling from data of the UK's electricity consumption, separated into peak and off-peak hours. A similar strategy is used to model each prosumer's PV production, $p^t_{PV,i}$, at each time step. External temperature, T^t_e, is also given by real data. EV availability, μ^t_i, is a boolean variable (1 if available, 0 if unavailable). EV demand, $d^t_{EV,i}$, and solar heat flow (SHF) rate, ϕ^t, are initialised randomly within a specified range.

The impacts of the decisions made by the prosumers are both local and system-wide. Local impacts include battery energy level, E^t_i, prosumer import, p^t_i, and, both building mass temperature (BMT), $T^t_{m,i}$, and indoor air temperature (IAT), $T^t_{air,i}$. The impacts on the system are seen as changes to the total system wide costs. Total costs are made up of grid (c^t_g), distribution (c^t_d) and storage (c^t_s) costs. The objective is to minimise these costs whilst also minimising the deviation from comfortable temperatures and the grid imports for each prosumer. This multi-part objective is reflected in the reward function of the RL process outlined later in this paper.

$$c^t_g = C^t_g(g^t + \epsilon_g) \tag{8}$$

$$\epsilon_g = \frac{R}{V^2}(g^t)^2 \tag{9}$$

The grid cost coefficient at time t, is the sum of the grid electricity price and the product of the carbon intensity of generation and the Social Cost of Carbon [3]. The total grid imports to the prosumers, g^t, and the grid losses, ϵ_g, are scaled by the time-dependent grid cost coefficient. The grid losses are calculated in **Equation 9**, where the nominal root mean square grid voltage is V, and the average resistance between the main grid and the distribution network is R [17].

$$c^t_d = C_d \sum_{i=1}^{P} max(0, -p^t_i) \tag{10}$$

The distribution cost coefficient C_d is used to scale the total exports of energy from prosumers to the grid.

$$c^t_s = C_s \sum_{i=1}^{P} (b^t_{in,i}, b^t_{out,i}) \tag{11}$$

Storage battery depreciation costs are calculated based on the method outlined in the foundational environment, where they are proportional to the throughput of the battery, using depreciation coefficient C_s [17]. A detailed summary of the initialisation of all parameters is found in **Table 1**.

3.2 Constraints

The following constraints are again based on the foundational environment. E_0, $\underline{E}$, $\overline{E}$ are the initial, minimum and maximum battery energy levels respectively. $\overline{b_{in}}$ is the maximum battery charge in each time step. η_{ch} and η_{dis} are the battery charging and discharging efficiencies.

It is assumed that the demand for each prosumer, d_i^t, is equal to that prosumer's consumption. This highlights one of the main differences in the energy balances from the foundational environment. The energy balance about each prosumer is defined in **Equation 12**.

$$p_i^t = d_i^t + h_i^t + \frac{b_{\text{in},i}^t}{\eta_{\text{ch}}} - \eta_{\text{dis}} b_{out,i}^t - p_{\text{PV},i}^t \tag{12}$$

The total grid import to the prosumers, g^t, is the sum of each prosumer's import.

$$\sum_{i \in \mathcal{P}} p_i^t = g^t \tag{13}$$

Following on, the battery energy balance at each time step t is defined in **Equation 14**.

$$E_i^{t+1} = E_i^t + b_{in,i}^t - b_{out,i}^t - d_{\text{EV},i}^t \tag{14}$$

Equation 15 shows the terminal battery energy balance and **Equations 16 - 18** are the battery and charging/discharging constraints.

$$E_0 = E_i^{t_0} = E_i^{t_{end}} + b_{in,i}^{t,end} - b_{out,i}^{t,end} - d_{EV,i}^{t_{end}} \tag{15}$$

$$\mu_i^t \underline{E}_i \leq E_i^t \leq \overline{E}_i \tag{16}$$

$$b_{in,i}^t \leq \mu_i^t \overline{b_{in}} \tag{17}$$

$$b_{in,i}^t \leq \mu_i^t \overline{E_i} \tag{18}$$

The heating model used is a Crank-Nicholson scheme [18], where κ is a 2 x 5 matrix of temperature coefficients. $\underline{T}_i^t$ and $\overline{T}_i^t$ are the indoor air temperature lower and upper bounds, respectively.

$$\begin{bmatrix} T_{m,i}^{t+1} \\ T_{air,i}^{t+1} \end{bmatrix} = \kappa \left[1, T_{m,i}^t, T_e^t, \phi^t, h_i^t\right]^T \tag{19}$$

$$\underline{T}_i^t \leq T_{air,i}^t \leq \overline{T}_i^t \tag{20}$$

The variables subject to non-negativity constraints are shown in **Equation 21**.

$$d_i^t,\ h_i^t,\ E_i^t,\ b_{in,i}^t,\ b_{out,i}^t \ \geq\ 0 \tag{21}$$

A representation of the equations and energy balances that define the environment is show in **Figure 2**.

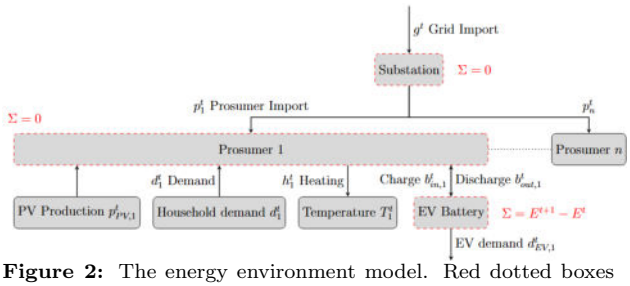


Figure 2: The energy environment model. Red dotted boxes represent energy balances.

4 Case Study: Grid-World

4.1 Discrete Grid-World and Q-Tables

The grid-world exercise was implemented in order to gain a fundamental understanding of Q-learning before delving into the energy environment. Initially, this environment was set up as a discrete 4 by 4 'grid-world' space of 16 states with 4 possible actions out of each state, corresponding to the four cardinal directions. This led to a fixed size Q-table with 64 state-action pairs, where those associated with moves out of the environment were invalid. The object of this case study was to train an RL agent to traverse this environment from the start state [4, 4] to the terminal state [1, 1] in as efficient a route as possible. An obstacle was placed in the state [3, 2]. The agent received a highly positive reward for entering the terminal state (+100), and a negative reward for entering the obstacle (-30). Every other action was rewarded by a small negative reward (-1), which trained the agent to take the most direct path. For the action selection, the agent used an ϵ-greedy strategy to balance exploitation and exploration. In an ϵ-greedy strategy, an agent selects random actions with a defined probability, ϵ, and follows the optimal policy derived from the Q-table in all other instances. In early iterations, ϵ is higher, and in this case, is reduced exponentially with iterations. In this way, the agent is highly explorative in early iterations and behaves more and more exploitatively with more iterations.

Algorithm 1: Q-Learning Algorithm - Fixed Size Q-tables

```
Initialise Q-table with random integers
for episode = 1 to N_episodes do
    Decay ε
    while state ≠ terminal state do
        Select action with ε-greedy strategy
        Update state
    end
    Update Q-table (Equation 4)
    Update total reward
end
```

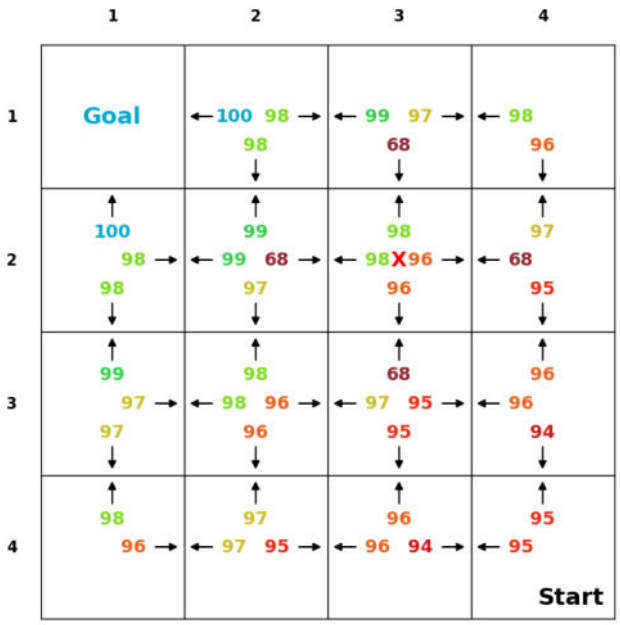


Figure 3: A Visualisation of the Q-Table indicating possible moves with arrows and the corresponding total expected future reward for taking that move.

The general structure of the Q-learning algorithm

is outlined in **Algorithm 1**. The values in the Q-table were updated with **Equation 3**. **Figure 3** shows a visualisation of the Q-table after a sufficient number of episodes after which the values had converged. These values represent the total expected future reward if the agent takes that action from its current state, then follows the learned policy.

4.2 Continuous Grid-World

The discrete grid-world case study was built upon by developing the environment into a continuous decision space. The decision space was a 10 by 10 space, where the agent could take actions of between -0.1 and 0.1 in both the x and y directions. Similarly, there was a target state and an obstacle, which was circular in this iteration of the case study. The case study was further developed by experimenting with different locations for the start, goal, and obstacle.

4.2.1 Linear Function Approximation

Whilst the epsilon-greedy strategy remained unchanged, the Q-table was replaced with the linear function approximation (LFA) method for the Q-function. In the final algorithm, five features were selected: the distance to the goal (d_1), the distance from the obstacle (d_2), the difference between the two, the ratio between the two and the boundary feature (BF), giving the Q-function seen in **Equation 22**.

$$Q(s,a;\theta) = i \cdot (2 \cdot d_1 \cdot \theta_1 + d_2 \cdot \theta_2 + (d_1 - d_2) \cdot \theta_3 + (d_1/d_2) \cdot \theta_4) + (1-i) \cdot BF \cdot \theta_5 \tag{22}$$

where:

$$i = \begin{cases} 1 & \text{if within defined distance to a boundary} \\ 0 & \text{otherwise} \end{cases}$$

The BF is a function of the distance to the nearest boundary and the defined boundary threshold.

The criteria under which the agent received rewards was also developed further. As before, a strongly positive or negative reward was awarded for reaching the goal or entering the obstacle respectively. However, instead of simply not allowing the agent to move out of the defined decision space, a strongly negative reward was introduced to inhibit this behaviour. For all other actions, a reward scaled by the distance of the agent to the goal multiplied by -1 was returned, providing the agent's distance from the obstacle was above a defined threshold. If the distance to the obstacle was below this threshold, the reward was scaled instead, by the distance to the obstacle. In this way, the agent is encouraged to travel efficiently towards the goal, unless it is near the obstacle, in which case it temporarily prioritises avoiding the goal, until the 'safe' distance is once again respected. The framework of the algorithm used is very similar to **Algorithm 1** but is shown in **Algorithm 2** for clarity.

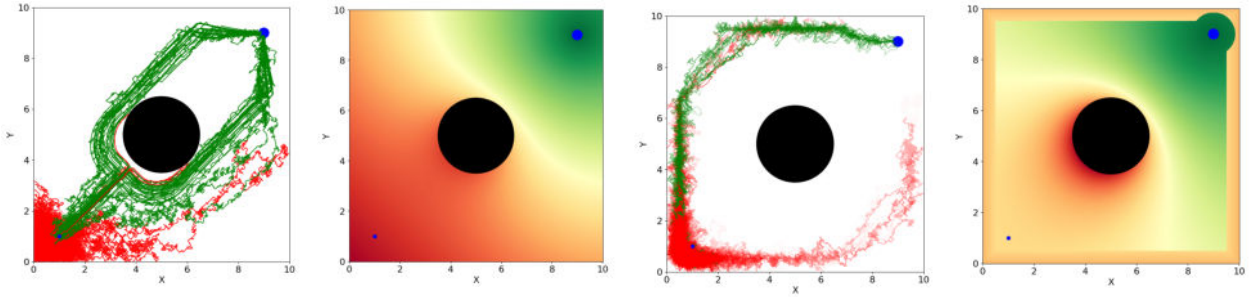
Algorithm 2: Q-Learning Algorithm - Linear Function Approximation

```
Initialise environment
Initialise agent
for Episodes do
    Reset environment
    for Steps do
        Get state
        Choose action
        Find next state
        Get reward for the move
        Update weights (Equation 6)
        if terminal state then
            break
        end
        State ← next state
    end
end
```

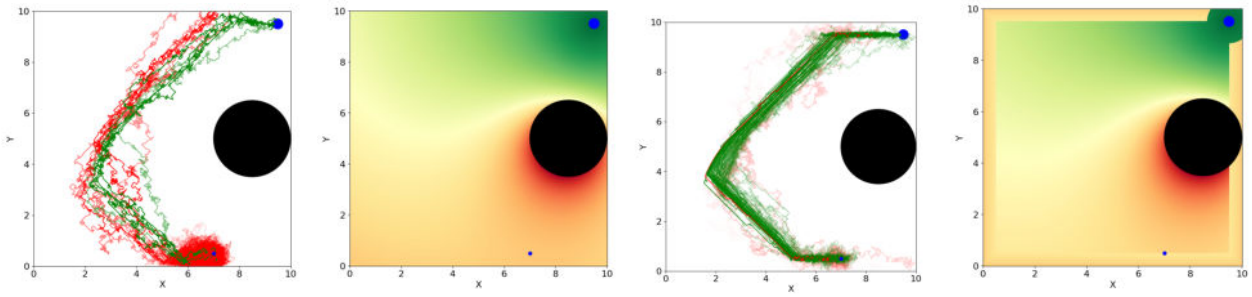
4.2.2 Continuous Grid-World: LFA Results

It was found that the efficacy of these features was highly sensitive to the initialisation of the locations of the start, goal and obstacle. **Figure 4** shows the training results and LFA visualisations for two scenarios with different initialisations of the start, goal and obstacle locations. These results show that in the instance where the optimal route did not approach the boundaries of the decision space, the 'boundary' feature had a significant negative impact on the learning and performance of the agent. However, this same feature had extremely positive impacts in instances where the agent's path around the obstacle took it near to one of the boundaries. This therefore implies that features in the LFA would have to be rigorously engineered to ensure consistent success.

This process is called feature engineering, which involves transforming raw data into meaningful input features that enhance the performance of ML models. It involves selecting, modifying, or creating features that make algorithms more effective and efficient. Based on this inconsistency in performance, it was concluded to not be beneficial to extend linear function approximation to the higher-dimensional energy environment. Applying LFAs would necessitate extensive feature engineering, which is both challenging and beyond the scope of this work and would prevent applicability of the algorithm to future environments. Additionally, research has shown that the use of linear approximation function strategies can diverge with increased learning [19]. As a result, with LFAs no longer applicable, a more complex Q-learning method was explored. This initially led to a DQN being built, however due to the limitations mentioned before, was similarly not applicable. Therefore, a DDPG was built in the continuous grid world in order to test hyperparameters and gain understanding of the algorithm.



(a) No Boundary Feature - Training **(b)** No Boundary Feature - Q-Function **(c)** Boundary Feature - Training **(d)** Boundary Feature - Q-Function



(e) No Boundary Feature - Training **(f)** No Boundary Feature - Q-Function **(g)** Boundary Feature - Training **(h)** Boundary Feature - Q-Function

Figure 4: Comparison of training and Q-functions for scenarios 1 (a)-(d) and 2 (e)-(h) with and without boundary features. The small blue circle is the starting location of the agent and the larger blue circle is the goal. The large black circle is the obstacle. In the training plots, green paths denote episodes that ended with the goal being reached, and red paths denote a failed episode. The opacity of the paths increase with later episodes. In the Q-function visualisations, the red to green gradient represents increasing Q-values of the LFA.

5 Environment Architecture

5.1 Input Data

The simulation of the energy environment spans 3871 1-hour time steps, from early May to mid October 2024. For electricity demand and PV production, a mean and standard deviation were calculated for subsets of the real data set, and then each prosumer's value for this variable was generated from a distribution defined by these parameters. The electricity demand data was accessed from historical national grid data [20], and was split into two subsets, peak and off-peak demand, so that the hourly variation was reflected in the values sampled. The solar production data utilised was a dataset collected by M. Marsh, and was split into two subsets (day and night) in order to reflect the hourly variation. Electricity demand and carbon intensity data was taken from real historical data for the time frame over which the simulation runs [21] [22].

Table 1 shows the initialisation of variables and parameters within the environment. A number of these values are taken from *Charbonnier et al.*[3]. Others, such as solar heat flow rate and EV demand were initialised based on reasonable values of these variables. The EV availability is defined by the probability $\mu_i^t \sim \{0 : 0.3,\ 1 : 0.7\}$. The Social Cost of Carbon is given as 70 £/tCO_2. The heating coefficients are defined by the matrix in **Equation 23**.

$$\kappa = \begin{bmatrix} 6.84\text{e-}2,\ 9.08\text{e-}1,\ 9.15\text{e-}2,\ 2.62\text{e-}4,\ 2.52\text{e-}1 \\ 2.40\text{e-}1,\ 8.80\text{e-}1,\ 1.20\text{e-}1,\ 3.460\text{e-}4,\ 1.46 \end{bmatrix} \quad (23)$$

In addition to the environment parameter setup, the energy environment was constructed to be compatible with OpenAI's Gymnasium [23] framework. This represents a significant milestone in this work. Gym compatibility is crucial in RL research, as it provides a standardised interface that facilitates reproducibility, scalability, and interoperability with state-of-the-art algorithms and tools. By adhering to the Gymnasium framework, the environment was designed with a robust structure, including clearly defined observation and action spaces, a well-specified reward function, and mechanisms for episodic termination and reproducibility. The environment was developed with the capability to handle any window length, allowing efficient optimisation of this parameter. The window length is the number of previous states that are considered when the algorithm selects the next action.

Table 1: Table of Parameters and Values [3][17]

Parameter	Value	Parameter	Value
$\overline{E}$	75 kWh	C_d	0.01 £/kWh
E_0	$0.5\overline{E}$	C_s	20 $/MWh
$\underline{E}$	$0.1\overline{E}$	d_{EV}	$d_{EV} \in [0, 3]$
$\overline{b_{in}}$	22 kW	μ_i^t	$\{0, 1\}$
R	0.084 Ω	Φ^t	$\Phi^t \in [0, 5]$
V	415 V	η_{ch}	0.932
η_{dis}	0.932		

This compatibility enables seamless integration of the environment, with a wide range of reinforcement learning libraries, such as Stable-Baselines3 (SB3) and Ray, ensuring that future researchers can readily adopt, modify, and extend the environment without addi-

tional implementation overheads. As the Gym framework is widely regarded as the de facto standard for RL experimentation [24], this work aligns with best practices in the field, allowing it to be easily benchmarked against other environments and algorithms.

5.2 Deep Deterministic Policy Gradient

Algorithm 3: Deep Deterministic Policy Gradient (DDPG) training algorithm

Initialise actor A_θ and critic Q_ϕ
Initialise target as clones of the main networks $\theta_{targ} \leftarrow \theta, \phi_{targ} \leftarrow \phi$
Initialise empty replay buffer $\mathcal{D}$
for *episode* $= 1$ **to** $N_{episodes}$ **do**
  Reset the environment
  Receive initial state s_1
  for *steps* **do**
    Select action $a \leftarrow A_\theta(s) + \epsilon$
    Execute action a
    observe r, s', and d
    Store (s, a, r, s', d) in $\mathcal{D}$
    if s' terminal, break loop
    if *update time* **then**
      Randomly sample batch $B \leftarrow (s, a, r, s', d)$ from $\mathcal{D}$
      Compute targets: $y(r, s', d) \leftarrow r + d \cdot \gamma \cdot Q_{\phi_{targ}}(s', A_{\theta_{targ}}(s'))$
      Update critic: $\nabla_\phi \frac{1}{|B|} \sum_{(s,a,r,s',d) \in B} (Q_\phi(s, a) - y(r, s', d))$
      Update actor: $\nabla_\theta \frac{1}{|B|} \sum_{s \in B} Q_\phi(s, A_\theta(s))$
      Update target networks: $\theta_{targ} \leftarrow \tau\theta + (1 - \tau)\theta_{targ}$, $\phi_{targ} \leftarrow \tau\phi + (1 - \tau)\phi_{targ}$
    end
    set $s \leftarrow s'$
  end
end

The outline of a DDPG is shown in **Algorithm 3**. Training begins with initialising the parameters θ of the actor A_θ and parameters ϕ of the critic Q_ϕ. The replay buffer $\mathcal{D}$ and decayed variables are also reset. The target networks are initialised as clones of the main actor and critic networks. For each episode in the loop, the state is reset. Commencing each step, the current state is observed, and inputted into the actor network, returning the optimal action $a = A_\theta(s) + \epsilon$, bounded by $[a_{Low}, a_{High}]$. The noise ϵ is added to the action to introduce exploration, where $\epsilon \sim \mathcal{N}$. Next, the next state s', reward r, and termination flag d is observed from the environment. The state, action, reward, next state and done flag is then added to the replay buffer as a tuple. If the state is terminal, the episode ends and restarts anew; otherwise, the network update loop is commenced, if it's the correct time. Following on, a randomly sampled batch from the replay buffer is taken $B = (s, a, r, s', d)$. The next actions and their Q-values are calculated using the target network, and the RHS of the bellman equation, is set to y. Next, the current Q-value is calculated, $Q_\phi(s, a)$ to find the Mean Squared Bellman Error (MSBE) and one step of gradient descent is performed to minimise the error and update the critic network. The actor network is updated by maximising the mean Q-values of the actions it outputs for the batch, using a gradient ascent. To end the network update, a soft update of the target network parameters is performed. Finally, back in the step loop, the state is updated to be the next state found in that iteration and the step loop repeats. This process will carry on until the specified number of training episodes is reached.

5.3 Algorithm Setup

With the environment structured to align with OpenAI's Gymnasium, this enabled the seamless integration of the Stable-Baselines3 [25] implementation of the DDPG algorithm. This compatibility represents a critical advantage, as it uses a well-established library that has been extensively tested and optimised for performance. By utilising SB3, the implementation benefits from strong foundations, streamlined experimentation, and access to advanced features such as parallelisation. This approach not only ensures the reliability and scalability of the algorithm within the energy environment, but also significantly reduces development overhead, including quick tuning of hyperparameters. This allows the focus to remain on evaluating and refining the RL methodology, such as the reward function, in the energy environment.

The specifics of the DDPG imported from SB3 began with using the MlpPolicy and standard parameters outlined in the original DDPG paper by *Lillicrap et. al.* [4] and [26]. The buffer size was initialised at 1,000,000 and used a batch size of 256. The network was designed using two hidden layers for both the critic and actor networks, each layer having 400 and 300 nodes respectively, resulting in around 130,000 parameters including the final layer weights. *ReLU* activation functions between layers were used, excluding the action output where the *tanh* activation function was implemented to bound the actions. The network's updates used Adam as the optimiser to learn the parameters in the network, with a learning rate of 10^{-3} for the actor network and 10^{-4} for the critic network. A value for τ of 0.005 was used for the soft updates on the target networks. A discount factor of $\gamma = 0.99$ in the Q-equation, emphasises the importance of future rewards. The noise to encourage exploration used a normal distribution with mean 0 and standard deviation $\sigma = 0.2$.

The reward function is made up of three terms. The first is the negative of the cost at time t; the second is a quadratic penalty for the deviation of the IAT from the desired value of 20°C; the final term is a penalty for prosumer import, encouraging the agent to use the energy from PV production or energy stored in the battery, where possible. Each of these are scaled by coefficients to balance their prioritisation. The tuning

of these coefficients formed a substantial part of the reward function shaping.

6 Results and Discussions

Figure 5 shows the learning curve for 5 different window lengths. It was found that with a window length of 1, the algorithm demonstrated poor learning. This is because the environment is subject to considerable noise, due to the large fluctuations in variable values. Therefore, the current state alone did not provide sufficient information to select an effective action. A repeated action is applied to the next n states, with n equal to the window length, a larger n resulted in poor performance. This is due to the dynamics of the environment being faster than the window length. As a result, a middle-ground of window length of 5 hours was found to maximise the learning of the agent and balance effects from reduced noise and ineffective decision making. However, the addition of a window length is a violation of the Markov Property, due to this inclusion of historical data [5]. Nevertheless, this is a common practical technique in RL, to improve performance and is normal in this type of model due to it being a Partially Observable Markov Decision Process (POMDP).

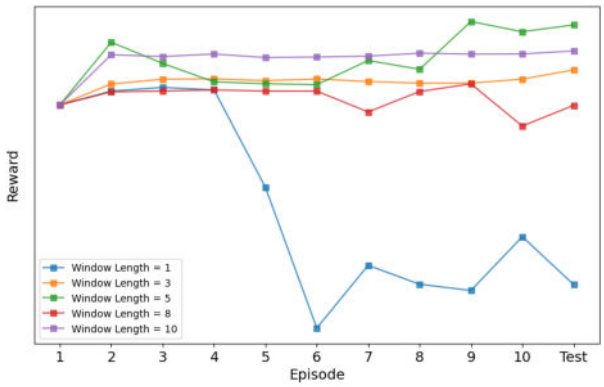


Figure 5: Learning curves for varying window lengths. Rewards are normalised to the same starting values.

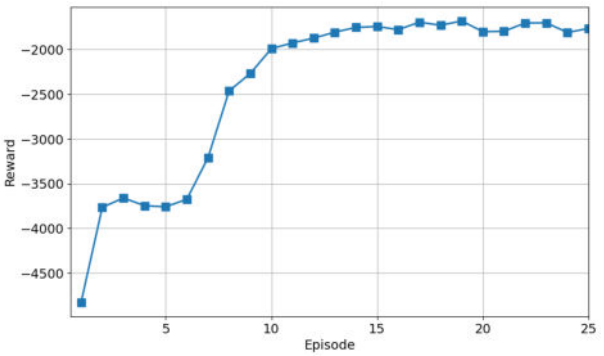


Figure 6: Learning curve for DDPG in energy environment, with window length = 5.

The learning curve of the agent over 25 episodes (20,000 steps) is plotted in **Figure 6**. The agent reaches fairly stable behaviour within 12-15 episodes (∼10,000 steps). In order to compare the performance of the DDPG, it is compared with some simple heuristics. These heuristics select actions based on simple logic similar to bang-bang control. For example, if the temperature is below a particular threshold, heat by a certain amount, and if the PV production exceeds demand, charge the battery by a certain amount, or else discharge. **Figures 7 and 8** plot the mean costs and bounds for 10 prosumers. Not only is the DDPG's cost per prosumer lower at £732 over the testing period, compared with £777 for the heuristics, there is a reduction of 5.79%. The consistency between each prosumer is better demonstrated by visually smaller bounds and a standard deviation of 2.59 for the DDPG and 28.48 for the heuristics.

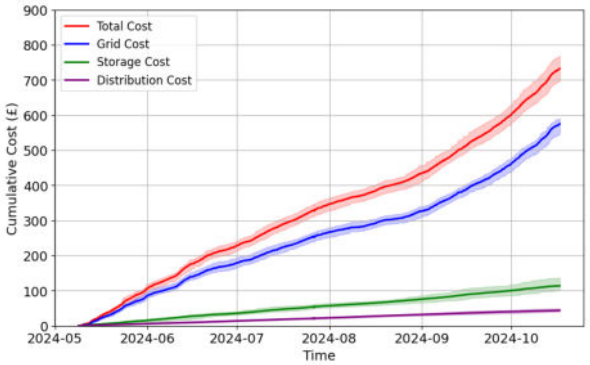


Figure 7: DDPG mean cumulative costs for 10 prosumers.

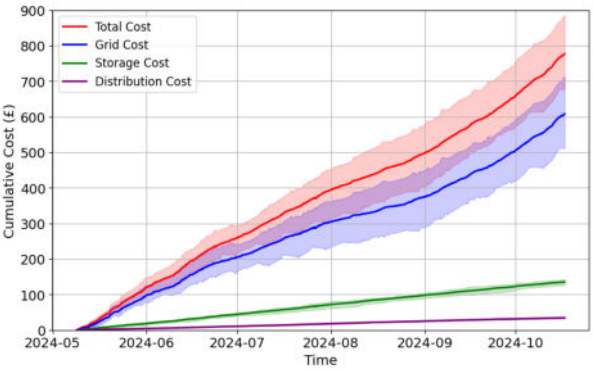


Figure 8: Heuristics mean cumulative costs for 10 prosumers.

Another comparison made between the DDPG and the heuristics was the IAT. These results are plotted in **Figures 9 and 10**. The raw temperature data is overlaid with a 12-hour rolling average to diminish the high level of noise seen on the graph. It can be seen that the DDPG's control does not violate the temperature's lower bound constraint, but it does violate the upper bound constraint during the summer months. This is explained by the lack of cooling capabilities, such as air conditioning, in the environment model, therefore making this violation hard to avoid. The heuristics, however, does violate the lower bound constraint in the colder months at the end of the testing period. This superior performance in optimisation by the DDPG agent is reflected in the empirical constraint violation probability values; 2.69% for the DDPG and 5.4% for the heuristics.

Finally, **Figure 11** demonstrates how the heating actions selected by the agent are related to the external temperature in the model. Again, a rolling average is plotted for clarity. It is clear that the agent has successfully learnt to select higher heating energy consumption actions whenever the external temperature

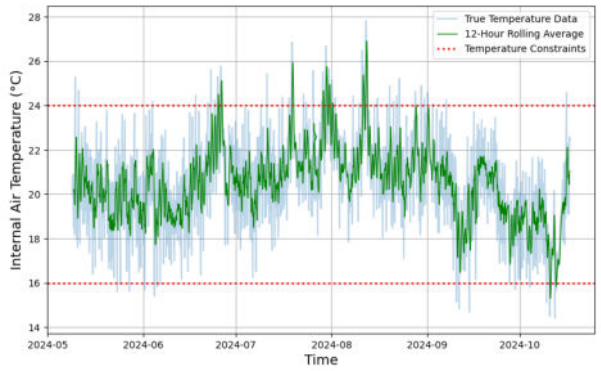


Figure 9: Internal Air Temperature as a result of heuristic based actions.

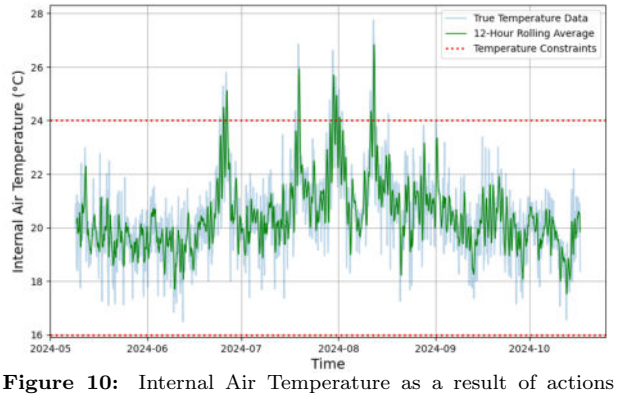


Figure 10: Internal Air Temperature as a result of actions selected by the trained DDPG agent.

is lower.

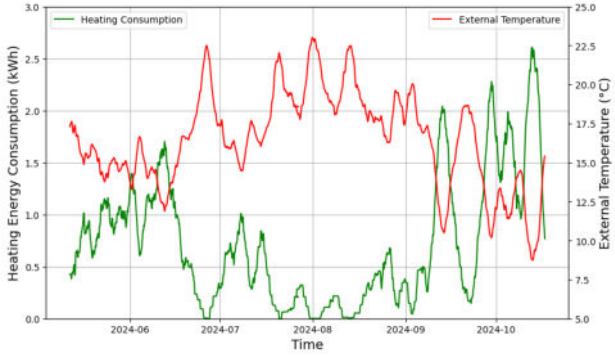


Figure 11: Heating energy consumption: actions selected by the trained DDPG agent.

7 Conclusions and Outlook

7.1 Conclusions

This paper aimed to build a DDPG to tackle a complex energy environment and build upon the MARL Q-learning approach used in *Charbonnier et al*[3]. Initial research within a grid world environment provided the foundations for the LFA, DQN and DDPG algorithms. Through this, LFA was found to be highly sensitive to feature engineering. Due to this being a difficult task to manage in more complex environments, the feasibility of implementing NNs (specifically DQNs) was explored. DQNs can handle more dynamic problems and have increased computational capability due to the hundreds of thousands of parameters within its network. However, the significant drawback is they can not handle continuous action spaces due to the structure of the neural network. DDPGs, with their actor-critic network design, provided a solution. A DDPG was initially built and implemented in the grid world to explore its hyperparameters and capabilities before building the final energy environment.

The design of the energy environment was a very positive part of this project. It is an open source code accessible to anyone in the world and designed to the Gymnasium standard. The environment used real data to train the algorithms allowing for much more rigorous and accurate assessment. This alongside specially constructed reward functions and heuristics provided the baseline for the evaluation of any algorithms.

The final iteration of the algorithm used the DDPG adapted from the SB3 and achieved promising results. It demonstrated effective learning, achieving consistent cumulative rewards after 12-15 episodes. Sensitivity analysis to the window length concluded that a length of 5 optimised the learning of the agent. Against heuristics, the DDPG produced results that provided cost savings of 5.79% for the individual, and as a result, the whole system. These total costs with the DDPG were distributed much more closely around the mean, meaning a more consistent performance, and were achieved whilst also controlling the indoor air temperature more effectively. However, the temperature upper bound constraint was violated in the summer months of the testing period, indicating a need to introduce constrained RL [27]. The final values of the costs are highly sensitive to input parameters, and therefore should be considered relative to the costs of the heuristics rather than their absolute value. The most important conclusion is that cost reductions are possible in the implementation of a DDPG as the controlling agent in this environment. This paper indicates the successful implementation of DDPGs in distributed residential energy control, and proves and shows the viability of RL methods in energy system optimisation

7.2 Outlook

RL methods are always rapidly developing. Other algorithms, such as SB3's TD3 [28] uses a DDPG setup but with two or more critic networks to more accurately optimise the Q-value. Other networks such as Recurrent NNs (RNNs) [29] could be used due to the environment's sequential data, or Multi agent DDPGs (MADDPGs) [30] with each prosumer having an individual agent instead of one agent controlling all prosumers. Furthermore, SB3 offers multiple policies per algorithm, which could be explored. In addition, more rigorous sensitivity analysis of the hyperparameters can be conducted to further optimise the DDPG in this environment.

Model improvements can also be made, beginning with the heating model. More accurate representation of house heat efficiencies and other metrics could be included in the environment. Some parameters, nominally ϕ^t and d_{EV}, were initialised randomly between a specified range. The accuracy of the model would be improved if real data sets were sampled to generate values for these two variables. Constrained RL can be considered to eliminate constraint violations. More ac-

curate results can be achieved through larger and more up-to-date data sets which the algorithms can train on. Lastly, whilst feature engineering was beyond the scope of this paper, it is still a viable option. Specifically designed features, in enough quantity, within such an environment could potentially lead to better results than complex NNs such as DDPGs, however, they may suffer from long running times.

8 Acknowledgments

The authors express their utmost gratitude to M. Bloor for his extremely valuable technical assistance and insightful discussions. Also, M. Marsh for his collaboration in building the energy environment. Lastly, python scripts for the 5 algorithms and 3 environments can be made available upon request.

References

[1] A. Sun and D. Phan, "Modern Optimization Models and Techniques for Electric Power Systems Operation," 2012.

[2] Y. Li, C. Yu, M. Shahidehpour, T. Yang, Z. Zeng, and T. Chai, "Deep Reinforcement Learning for Smart Grid Operations: Algorithms, Applications, and Prospects," *Proceedings of the IEEE*, vol. 111, pp. 1055–1096, Sept. 2023.

[3] F. Charbonnier, T. Morstyn, and M. D. McCulloch, "Scalable multi-agent reinforcement learning for distributed control of residential energy flexibility," *Applied Energy*, vol. 314, May 2022.

[4] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep rl," July 2019.

[5] D. Silver, "Lectures on RL," 2015.

[6] K. Arshad, R. Ali, A. Muneer, I. Aziz, S. Naseer, N. Khan, and S. Mohd Taib, "Deep Reinforcement Learning for Anomaly Detection: A Systematic Review," *IEEE Access*, vol. PP, Nov. 2022.

[7] B. Jang, M. Kim, G. Harerimana, and J. Kim, "Q-Learning Algorithms: A Comprehensive Classification and Applications," *IEEE Access*, vol. PP, pp. 1–1, Sept. 2019.

[8] T. Miller, "Mastering Reinforcement Learning, Q-function approximation." Accessed 10.11.2024, (Available at gibberblot.github.io/rl-notes/single-agent/function-approximation.html).

[9] Z. Wang, Q. Fu, J. Chen, Q. Liu, Y. Lu, H. Wu, and F. Hu, "LinFa-Q: Accurate Q-learning with linear function approximation," *Neurocomputing*, vol. 611, p. 128654, Jan. 2025.

[10] V. Mnih, K. Kavukcuoglu, and D. Silver, "Human-level control through deep rl," *Nature*, vol. 518, pp. 529–533, Feb. 2015.

[11] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing Atari with Deep RL," Dec. 2013.

[12] M. Roderick, J. MacGlashan, and S. Tellex, "Implementing the Deep Q-Network," Nov. 2017.

[13] L.-J. Lin, "Self-improving reactive agents based on reinforcement learning, planning and teaching," *Machine Learning*, vol. 8, pp. 293–321, May 1992.

[14] T. Schaul, J. Quan, I. Antonoglou, and D. Silver, "Prioritized Experience Replay," Feb. 2016.

[15] A. I. TutorMaster, "What is Gaussian Noise in Deep Learning? How and Why it is used? | In Plain English." Accessed 28.10.2024, (Available at plainenglish.io/blog/what-is-gaussian-noise-in-deep-learning-how-and-why-it-is-used).

[16] "Normalizing Inputs of Neural Networks | Baeldung on Computer Science." Accessed 19.11.2024, (Available at baeldung.com/cs/normalizing-inputs-artificial-neural-network).

[17] T. Morstyn and M. D. McCulloch, "Multiclass Energy Management for Peer-to-Peer Energy Trading Driven by Prosumer Preferences," *IEEE Transactions on Power Systems*, vol. 34, pp. 4005–4014, Sept. 2019.

[18] "Energy performance of buildings — energy needs for heating and cooling, internal temperatures and sensible and latent heat loads." ISO 52016-1, 2017.

[19] D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller, "Deterministic policy gradient algorithms," in *Proceedings of the 31st International Conference on Machine Learning* (E. P. Xing and T. Jebara, eds.), vol. 32 of *Proceedings of Machine Learning Research*, (Bejing, China), pp. 387–395, PMLR, 22–24 Jun 2014.

[20] A. Vidal, "Electricity Consumption UK 2009 - 2024," 2024. Accessed 14.11.2024, (Available at kaggle.com/datasets/albertovidalrod/electricity-consumption-uk-20092022/data).

[21] openweathermap.org, "One call api 3.0," 2024. (Accessed 14.11.2024, (Available at openweathermap.org/api/one-call-3).

[22] NESO, "Carbon intensity - national, statistics," 2024. Accessed 14.11.2024, (Available at carbon-intensity.github.io/api-definitions/statistics).

[23] "Gymnasium Documentation." Accessed 27.11.2024, (Available at gymnasium.farama.org).

[24] "The Optical RL-Gym: An open-source toolkit for applying reinforcement learning in optical networks," in *2020 22nd International Conference on Transparent Optical Networks*, (Bari, Italy).

[25] "DDPG Stable Baselines3 2.5.0a0 documentation." Accessed 30.11.2024, (Available at stable-baselines3.readthedocs.io.

[26] A. Raffin, J. Kober, and F. Stulp, "Smooth Exploration for Robotic RL," June 2021.

[27] S. Miryoosefi and C. Jin, "A Simple Reward-free Approach to Constrained RL," 2021.

[28] "TD3 Stable Baselines3 2.5.0a0 documentation." Accessed 30.11.2024, (Available at stable-baselines3.readthedocs.io).

[29] L. Sehovac and K. Grolinger, "Deep Learning for Load Forecasting: Sequence to Sequence Recurrent Neural Networks With Attention," *IEEE Access*, vol. 8, pp. 36411–36426, Feb. 2020.

[30] X. Zhang, Y. Liu, J. Duan, G. Qiu, T. Liu, and J. Liu, "DDPG-Based Multi-Agent Framework for SVC Tuning in Urban Power Grid With Renewable Energy Resources," *IEEE Transactions on Power Systems*, vol. 36, pp. 5465–5475, Nov. 2021.

Numerical Approximation of Multivariate Functions using Kolmogorov-Arnold Representation Theorem

Chun Wai Fung and Hiu Chung Tang
Department of Chemical Engineering, Imperial College London, U.K.

Abstract Nonlinear and high-order interactions pose significant challenges to traditional parametric models, often resulting in overfitting and misrepresentation. Our objective is to develop a computationally efficient, accurate, interpretable, and scalable framework to address high-dimensional, nonlinear problems for chemical engineering applications. Our model framework is based on the Kolmogorov-Arnold Representation Theorem (KART) which suggests that all multivariate functions can be decomposed and represented as a superposition of continuous univariate functions. The representation reduces dimensionality while preserving the reconstruction of complex mappings, aiding optimisation tasks in multivariate problems. Our implementation begins with the Special Ordered Sets of type 2 (SOS2) formulation to approximate univariate functions, followed by the integration of strengthening constraints and Big-M formulations which demonstrates superior runtime performance and scalability. Both formulations are categorised as Mixed Integer Quadratically Constrained Programming (MIQCP) problems, and we used Gurobi version 12.0.0 as our solver. We further adapted the framework into machine learning (ML) by developing a sparse Multi-layer Perceptron (MLP). The performances of our models are validated by comparative runtime and mean squared error against the equation of state and the standard Artificial Neural Networks (ANN) in the context of thermodynamics and dynamical systems. The results highlight the reduction in computational cost, improved approximation accuracy, and enhanced scalability. This study concludes that the Big-M formulation of MIQCP is more interpretable, accurate and efficient in smaller datasets and theoretically can guarantee finding the global optimum whereas, our sparse neural network model (MLP) is more efficient for larger dataset problems due to its high scalability.

Keywords*: Kolmogorov-Arnold Representation Theorem, Machine Learning, Sparse Neural Network, MIQCP, Multilayer Perceptron*

1. Introduction

High-dimensional functions often govern the dynamic behaviour of systems in chemical engineering, physics, and even socio-economic disciplines. It is not always viable to find an analytical solution to discover the intricate relationships between all variables. In the field of chemical engineering, modelling reaction kinetics or thermodynamic properties involves capturing the interactions between multiple variables such as temperature, pressure, concentration, time, and other coefficient parameters. These dynamic behaviours require functional representations that exhibit nonlinear and high-order interactions.

The traditional parametric model gives assumptions to the system based on known existing pre-defined functions. Nonetheless, these functions might only apply to a certain range or threshold and might lose their validity in other conditions and cannot consistently generalise outside their specified domains which limits their utility in dynamic or untested scenarios. Non-parametric methods, such as kernel regression[1] or Gaussian processes[2], do not require a pre-defined functional form, but they often provide little traceable insight into their underlying mechanisms and treat the system as a black box, making optimisation and troubleshooting challenging. For kernel methods, the kernel matrix grows as n^2 with the dataset size n. For Gaussian processes, inverting a covariance matrix involves ($O(n^3)$) time complexity. These methods also encounter severe computational bottlenecks when applied to high-dimensional datasets; hence, the computational costs in both runtime and memory requirements grow exponentially with increasing variables, making them essentially impractical for large-scale problems.

1.1 Background

The Kolmogorov-Arnold Representation Theorem (KART) [3] suggests that all multivariate nonlinear functions can be decomposed and represented as a superposition of finite continuous univariate functions with the binary operation of addition.

$$f(X) = f(x_1, \dots, x_n) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^{n} \phi_{q,p}(x_p) \right) \quad (1)$$

Where $\phi_{q,p}$ and Φ_q are the continuous inner and outer univariate functions. The expression was first established by Andrey Kolmogorov in 1957 and later refined by Vladimir Arnold in 1963, who solved one of the more constrained forms of Hilbert's thirteenth problem. Arnold's work generalised the result by providing explicit bounds on the number of univariate functions required and simplifying its applicability to practical problems. The theorem states that for a dimensionality of the input space n (number of variables), the basic function index $q = 2n + 1$ and the variable index $p = n$. The input variables are first transformed by the ϕ functions. The outputs from the ϕ function are summed with respect to the variable index p. The result of the summation becomes the input of the outer function Φ. Finally, all the outputs from the outer functions are summed, producing the final output.[3]

The Kolmogorov–Arnold representation theorem has drawbacks despite its adaptability. There is no clear instruction on how to create the univariate functions. Φ_q and $\phi_{q,p}$ in the theorem. Furthermore, it is still difficult to guarantee the smoothness or differentiability of these functions. To overcome these problems, recent research has constructed the univariate functions using computer methods such as splines and piecewise linear

approximations. Recent developments have applied KART's decomposition principles in designing Kolmogorov-Arnold Networks (KAN). Each KAN layer is constructed as a matrix of trainable 1D functions $\phi_{p,q}$, where p corresponds to the input dimensions and q corresponds to output dimensions for that layer.[4] These functions are often modelled using B-splines to flexibly interpolate between input datapoints. A single KAN layer with n_{in} input dimensions and n_{out} output dimensions can be expressed as:

$$\left[\Phi = \{\phi_{q,p}\}, p = 1,2,\dots,n_{in}, q = 1,2,\dots,n_{out}\right] \quad (2)$$

KAN processes input data through successive transformations and the output is defined as:

$$KAN(x) = \left(\Phi_{q-1} \circ \Phi_{q-2} \circ \cdots \circ \Phi_0\right)(x) \quad (3)$$

Here, Φ_k represents the univariate mappings at the $\boldsymbol{k}$-th layer. This layer-wise construction simplifies the multivariate function into a series of summation and compositional operations. The architecture of a KAN is represented as an array of integers, where each integer indicates the number of nodes in the corresponding layer. For an input vector x_0 with n_0 dimensions, the output is computed through transformations across multiple layers and then parameterised by learnable coefficients during the training phase. The final output Z is given by:

$$Z = \Phi_q\left(\Phi_{q-1}\left(\dots\left(\Phi_1(y_0)\right)\dots\right)\right)X \quad (4)$$

Where the univariate functions are refined by activation functions (f_k) to generate the layer output (y_k) and contribute to the overall model response (Z).
During training, KAN uses a forward pass to compute outputs layer by layer. The error between predicted (Z) and true responses is backpropagated to update the parameters of the univariate functions. This iterative process ensures convergence toward optimal performance while maintaining the interpretability of the model's internal workings. [4]

The practical application of KAN is showcased in recent studies by Ansar and Ashraf (2024) where they introduced a modification to the KAN architecture by incorporating the Pearson Correlation Coefficient (PCC) into the loss function. This adjustment introduced the KAN_PCC variant and was benchmarked against the traditional KAN and multi-layer perceptron (MLP). This enhanced architecture is tested in the optimisation of energy systems, including energy efficiency cooling (ENC) and energy efficiency heating (ENH) in buildings and power generation operation of a 660MW thermal power plant. KAN_PCC only showed marginal improvements over KAN in ENC, ENH, turbine heat rate, and thermal efficiency, with PCC integration benefiting MLPs more consistently. KAN demonstrated strong interpretability through integration with SHAP (SHapley Additive exPlanations) analysis. Its adaptability was showcased in its use in a nonlinear programming framework for multi-objective optimisation. However, they suggested that its performance advantage over MLPs remains context dependent. They did not extend their findings to other domains or thoroughly explore scalability to high-dimensional datasets.[4]

1.2 Project objective and contributions

This project aims to build upon these insights by shifting the focus to developing a computationally efficient, accurate, interpretable, and scalable framework to address high-dimensional, nonlinear problems. To address these gaps, we proposed a Mixed Integer Quadratically Constrained Programming (MIQCP) framework using KART, incorporating two formulations: SOS2 [5] and Big-M [6] with additional strengthening constraints. To explore further scalable implementation, we extend the KART framework into machine learning by designing a custom multilayer perceptron (MLP) inspired by KAN. The key strength of our sparse neural network lies in the computational simplicity and efficiency of the ReLU functions, and the interpretability achieved through piecewise linear decomposition without relying on symbolic regression used in KAN. We tested our models on a high-dimensional thermodynamics problem and validated their performance through comparative runtime and mean-squared error analysis against the equation of state and Artificial Neural Networks (ANN), highlighting the reduction in computational cost, improved approximation accuracy, and enhanced scalability.

This paper is organised as follows: Section 1 provides the background of KART, the development of KAN, and its utilisation in energy optimisation. Section 2.1 outlines the MIQCP formulations. Section 2.2 illustrates our implementation of the formulation into machine learning. The results of our computational testing are detailed and discussed in Section 3. The proposed formulation is then applied to a thermodynamics case example. Finally, Section 4 presents our discussion and future research focus and opportunities.

2. Methodology

2.1 MIQCP formulation

The proposed formulation combines the strengths of mathematical programming and the KART to approximate complex, nonlinear functions. The method begins by utilising a piecewise linear (PWL) approximation to represent the real function in the context of the KART. The real function is approximated as a series of linear segments, each defined over specific intervals of the input space as shown in figures 1. In this approach, the inner layers are formulated as Mixed Integer Linear Programming problems. This structure is extended to the outer layers, resulting in a more complex Mixed Integer Quadratically Constrained Program (MIQCP) formulation due to the combination of both layers. This method is designed to capture highly nonlinear functions, such as trigonometric and exponential functions, which are challenging to model with logical constraints. By using a non-parametric approach, it avoids the need for prior knowledge of the function's structure. Unlike traditional ML, our approach guarantees global optimality under certain conditions. The size of the problem is mainly

controlled by the number of breakpoints in each PWL and the amount of data. Figure 1 provides a visualized example of a single-variable function, illustrating how the piecewise linear approximation of the ϕ functions transform the input variable. The outputs of the ϕ function are summed with respect to the variable index p. The result of the summation becomes the input of the outer function Φ.

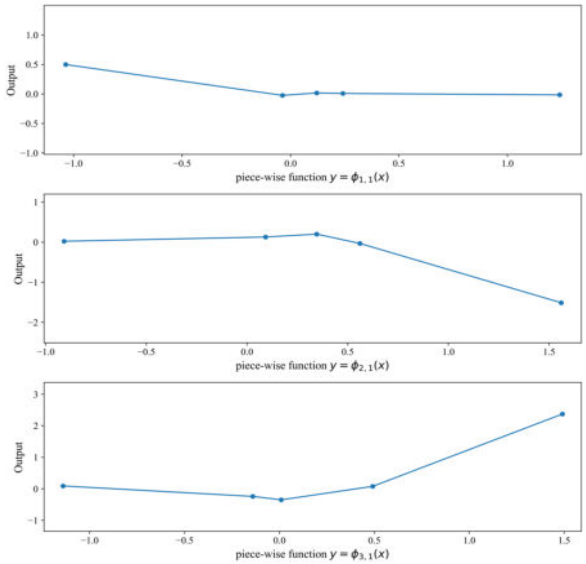


Figure 1: *Example of ϕ layer for 1D function*

2.1.1 SOS2 formulation

The SOS2 [5] formulation is useful in the Mixed-Integer Programming models to enforce piecewise linear approximations of nonlinear functions while maintaining continuity between the segments. The formulation divides the domain of x into n intervals defined by breakpoints $\{x_1, x_2, \dots, x_n\}$. λ and Λ are the interpolation coefficients (weights) of the adjacent breakpoints in the convex combination of the line segments for the first and second layers respectively. The SOS2 constraints in ensure that only adjacent λ and Λ can be non-zero. Gurobi[7] inherently recognises and manages SOS2 constraints internally, eliminating the need for manual constraint formulation.

Table 1: Indices

Description	*Index*	*Range*
Set of Datapoints	$d \in D$	$D = \{1, \dots, \lvert D \rvert\}$
Set of Intervals	$j \in J$	$J = \{1, \dots, \lvert J \rvert\}$
Input Variables	$p \in P$	$P = \{1, \dots, \lvert P \rvert\}$
Outer Layer	$q \in Q$	$Q = \{1, \dots, \lvert Q \rvert\}$

Table 2: Variables

Description	*Variable*	*Type*
Weights for piecewise input interval	λ	Continuous
Weights for piecewise output interval	Λ	Continuous
Anchor points for input intervals	m	Continuous
Anchor points for output intervals	n	Continuous
Univariate inner function output for input variable mapping	ϕ	Continuous
Values of the inner function ϕ outputs at interval boundaries	w	Continuous
Summed contributions of inner functions for each dimension q	Ψ	Continuous
Values of the outer function Φ outputs at interval boundaries	W	Continuous
Outer univariate function output	Φ	Continuous
Final predicted target value	$\hat{y}$	Continuous

$$\min \sum_{d=1}^{D} (y_d - \hat{y}_d)^2 \tag{5}$$

s.t.

$$\sum_j \lambda_{d,j,p} = 1 \quad \forall\, d \in D, p \in P \tag{6}$$

$$\sum_j \Lambda_{d,j,p} = 1 \quad \forall\, d \in D, p \in P \tag{7}$$

$$\Lambda_{d,j,p}, \lambda_{d,j,p} \geq 0 \quad \forall\, d \in D, j \in J, p \in P \tag{8}$$

$$x_{d,p} = \sum_j \lambda_{d,j,p,q} \cdot m_{j,p} \quad \forall\, d \in D, p \in P, q \in Q \tag{9}$$

$$\phi_{d,p,q} = \sum_j \lambda_{d,j,p} \cdot w_{d,j,p,q} \quad \forall\, d \in D, j \in J, p \in P \tag{10}$$

$$\Psi_{d,q} = \sum_p \phi_{d,p,q} \quad \forall\, d \in D, q \in Q \tag{11}$$

$$\Psi_{d,q} = \sum_j \Lambda_{d,p,q} \cdot n_{d,q} \quad \forall\, d \in D, p \in P \tag{12}$$

$$\Phi_{d,q} = \sum_j \Lambda_{d,p,q} \cdot W_{d,j,q} \quad \forall\, d \in D, p \in P \tag{13}$$

$$\hat{y}_d = \sum_{q \in Q} \Phi_{d,q} \quad \forall d \in D \tag{14}$$

The equation (5) is the objective function. The constraints provided in equation (6) to equation (8) implement interpolation coefficients λ and Λ to ensure continuity between segments. Equations (9) and (12) ensure that the input values $x_{d,p}$ and $\Psi_{d,q}$ are expressed as a weighted sum of interval boundaries with the weights $\lambda_{d,j,p,q}$ and $\Lambda_{d,p,q}$ respectively. Equation (11) converts the outputs of inner function $\phi_{d,p,q}$ to the inputs of the outer function $\Phi_{d,q}$. The transformed values $\phi_{d,p,q}$ and $\Phi_{d,q}$ are expressed as the weighted sum of the values at the boundaries by equations (10) and (13) respectively. Equation (14) sums all the outputs of $\Phi_{d,q}$ to give the final prediction.

2.1.2 Big-M formulation

Table 3: Additional Parameters for Big M formulation

Description	*Parameter*
Minimum spacing between breakpoints	δ
A large constant for Big-M constraints	M

Table 4: Additional variables for Big M formulation

Description	*Variable*	*Type*
Binary activation variable for inner, outer function segments correspondingly	z, Z	Binary

Slope of the piecewise linear approximation for the ϕ, Φ function	k/K	Continuous
Intercept of the piecewise linear approximation for the ϕ, Φ function	b, B	Continuous

The Big-M method divides the domain of a univariate function into intervals using breakpoints. The Big-M constraints ensure that only one segment is activated at a time through binary variable z. These constraints are applied to both the inner function $\phi_{q,p}(x_p)$ and the outer function $\Phi_{\mathrm{q}}(z_q)$. Given a set of data, for data points $d = 1, \dots, D$ and input parameters $p = 1, \dots, N, (x_1, \dots, x_n)$, we proposed the following MIQCP formulation to perform the approximation:

s.t.

$$m_{j,p} \le m_{j+1,p} \quad \forall j \in J, p \in P \tag{15}$$

$$x_{d,p} \le m_{j+1,p} + \left(1 - z_{d,j,p,q}\right) \cdot M \tag{16}$$
$$x_{d,p} \ge m_{j,p} - \left(1 - z_{d,j,p,q}\right) \cdot M \tag{17}$$
$$\forall \ d \in D, j \in J, p \in P, q \in Q$$

$$\sum_{j=1}^{J} z_{d,j,p,q} = 1 \qquad \forall\, d \in D, p \in P, q \in Q \tag{18}$$

$$k_{j,p} \cdot x_{d,p} + b_{j,p} = w_{d,j,p} \quad \forall d \in D, j \in J, p \in P \tag{19}$$

$$\phi_{d,p,q} \le w_{d,j,p,q} + \left(1 - z_{d,j,p,q}\right) \cdot M \tag{20}$$
$$\phi_{d,p,q} \ge w_{d,j,p,q} - \left(1 - z_{d,j,p,q}\right) \cdot M \tag{21}$$
$$\forall \quad j \in J, p \in P, q \in Q$$

$$k_{j,p} \cdot m_{j+1,p} + b_{j,p} = k_{j+1,p} \cdot m_{j+1,p} + b_{j+1,p} \tag{22}$$
$$\forall d \in D, p \in P, j = 1, \dots, |J| - 1$$

The above equations show the formulation of PWL functions in the inner functions. Equation (15) arranges the breakpoints in ascending order. Equations (16) and (17) show how each input $x_{d,p}$ is located by a binary variable $z_{d,j,p,q}$. If $x_{d,p}$ belongs to the interval, $z_{d,j,p,q}$ is activated, satisfying both constraints. If $x_{d,p}$ does not belong to the interval, $z_{d,j,p,q}$ is inactivated and the constraints are relaxed by the Big-M term. The equation (18) ensures only one interval is activated. The equation (19) shows the linear transformation by the slope and intercept of that interval. Equations (20) and (21) have a similar logic to equations (16) and (17), ensuring that $\phi_{d,p,q}$ corresponds to the value of $w_{d,j,p,q}$ within the activated interval. The last equation (22) ensures the continuity between intervals.

$$n_{j,q} \le n_{j+1,p} \qquad \forall j \in J, q \in Q \tag{23}$$

$$\Psi_{d,q} \ge n_{j,q} - \left(1 - Z_{d,j,q}\right) \cdot M \tag{24}$$
$$\Psi_{d,q} \le n_{j+1,q} + \left(1 - Z_{d,j,q}\right) \cdot M \tag{25}$$
$$\forall \ d \in D, j \in J, q \in Q$$

$$\sum_{j=1}^{J} Z_{dj,q} = 1 \qquad \forall d \in D, q \in Q \tag{26}$$

$$K_{j,q} \cdot \Psi_{d,q} + B_{j,q} = W_{d,j,q} \ \forall d \in D, j \in J, q \in Q \tag{27}$$

$$\Phi_{d,q} \le W_{d,j,q} + \left(1 - Z_{d,j,q}\right) \cdot M \tag{28}$$
$$\Phi_{d,q} \ge W_{d,j,q} - \left(1 - Z_{d,j,q}\right) \cdot M \tag{29}$$
$$\forall\, d \in D, j \in J, q \in Q$$

$$K_{j,q} \cdot n_{j+1,q} + B_{j,q} = K_{j+1,q} \cdot n_{j+1,q} + B_{j+1,q}$$
$$\forall d \in D, q \in Qj = 1, \dots, |J| - 1 \tag{30}$$

The above equations show the formulation of the outer $\Phi_{d,q}$ functions which follow the same logic as the first layer. Equation (23) arranges the breakpoints of each outer function in ascending order. Equation (11) is also implemented in the model to obtain the inputs of the outer functions. Equations (24) and (25) specify the binary variable $Z_{d,j,q}$ is activated when each $\Psi_{d,q}$ belongs to that corresponding interval. Equation (26) guarantees only one interval is activated. Equation (27) provide the linear mapping of $\Psi_{d,q}$ within each interval. Equation (28) and (29) provide the output value of $\Phi_{d,q}$ corresponds to the value of $W_{d,j,q}$ at the activated interval. Equation (30) ensures the continuity between intervals. Finally, the predicted output is obtained by summing the values of $\Phi_{d,q}$ in equation (14)

2.1.3 Strengthening the formulation

$$X_{d,q} \ge X_{d,q+1} \ \forall\, d \in D, q = 1, \dots, |Q| - 1 \tag{31}$$

The constraint stated in equation (31) prevents multiple equivalent solutions that differ only in the arrangement of variables. The KA representation[3] exhibits symmetry because the slope and intercept terms have interchangeable indices across different q within the same p. This symmetry can cause the solver to waste time exploring equivalent solutions. By introducing this constraint, redundancy in the search space is avoided, allowing the solver to focus on a representative subset of solutions.

$$n_{j,q} + \delta \le n_{j+1,p} \ \forall j \in J, q \in Q \tag{32}$$
$$m_{j,p} + \delta \le m_{j+1,p} \ \forall j \in J, p \in P \tag{33}$$
$$k_{j,p}{}^2 \ge \delta \ \forall j \in J, p \in P \tag{34}$$
$$K_{j,q}{}^2 \ge \delta \ \forall j \in J, q \in Q \tag{35}$$
$$b_{j,p}{}^2 \ge \delta \ \forall j \in J, p \in P \tag{36}$$
$$B_{j,q}{}^2 \ge \delta \ \forall j \in J, q \in Q \tag{37}$$

In the original approach, equations (15) and (23) failed due to overlapping intervals when multiple m and n are identical. This issue prevented the formation of a valid PWL function, but rather multiple points on the same x-axis. In our scenario, all input data values were first mapped to the same x-coordinate, 0, collapsing the data to a single point. As a result, all the slopes k and intercepts b in the first layer are 0. The second layer then creates multiple overlapping intervals at that same x-value. The prediction becomes trivial because the output at that x-value can always be expressed as a combination of the overlapping points. These intervals act as a direct mapping mechanism, where any inputs are reduced to their exact match of target values, bypassing the need to capture the overall structure. To resolve this, a perturbation term δ is introduced in equations (32) and (33) to prevent overlapping intervals but still allow for sharp changes within intervals. Additional constraints in equations (34) to (37) are also implemented to avoid slopes and intercepts taking 0 values.

$$f(X) = f(x_1, \dots, x_n) = \sum_{q=1}^{2n+1} \Phi\left(\sum_{p=1}^{n} \phi_{q,p}(x_p)\right) \quad (38)$$

One modification of the model is to adapt a variant of the KA representation [8] as shown in equation (37), which reduces the number of outer functions. By replacing multiple outer functions with a single function, this approach also reduces the number of variables and simplifies the problem size. Despite these simplifications, the variant is theoretically equivalent to the standard representation. Therefore, the model should retain the same accuracy in terms of its theoretical capabilities.

Another approach to reducing the problem size is to predefine the breakpoints of the inner functions, which can be evenly distributed within the specified bounds. While this method sacrifices some flexibility and accuracy of the model, it remains acceptable when the function is not overly complex. The PWL functions are then constructed based on these predefined breakpoints within the given intervals. By fixing the breakpoints, the search space is significantly reduced because the optimisation no longer involves finding the optimal values for the breakpoints. This reduction eliminates constraints associated with breakpoint positioning, such as ensuring proper ordering or non-overlapping intervals, thereby simplifying the overall formulation.

2.2 Machine Learning Approach

A neural network can be formulated to represent the above MIQCP formulation by using multiple ReLU [9] activation functions to model the PWL function. In this framework, the training parameters correspond to the shifts and scalings of the ReLU functions, enabling the PWL structure to be learned and optimised through training. This architecture differs from a standard neural network, as shown in figure 2, is that not all nodes are fully connected to every node in the subsequent layer, due to the difference in the underlying theoretical foundation of the two networks. While traditional neural networks rely on the Universal Approximation Theorem, our approach is derived from the KART. This representation has sparse connectivity, reduces the number of parameters, which is worth to explore in comparison to traditional neural networks.

The motivation for using this structure lies in its scalability. The ML model does not require the use of binary variables, which means all variables are continuous and can be trained by gradient-based optimisation methods. These methods ensure progress towards an optimum at every step, leading to faster convergence. In contrast, the branch and bound method employed to solve MIQCP relies on logical partitioning and pruning, which may not consistently reduce the search space and objective values effectively. However, this approach sacrifices the guarantees of global optimality provided by MIQCP. For problems where global optimality is crucial, MIQCP is still an invaluable tool. On the other hand, for applications that prioritise scalability, or where near-optimal solutions are adequate, the neural network-based approach offers a highly efficient option.

2.2.1 Modified Kolmogorov-Arnold Network

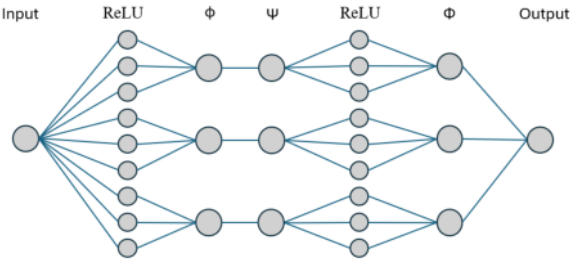


***Figure 2**: Our Multi-Layer Perceptron Network*

$$\phi(x_p) = k_1 \max(0, x + b_1) + \cdots + k_n \max(0, x + b_n) \quad (39)$$

The network illustrated in figure 2 demonstrates the construction of the KA representation using ReLU activation functions with 3 intervals in each PWL function. The network consists of a total of 5 hidden layers arranged in the following order: Linear, ReLU, Linear, Linear, and ReLU, followed by a final linear layer.

Starting from a single input, the value is first passed through nine linear, non-trainable functions with weights fixed at 1. This ensures that each node in the second layer receives equal input values. In the third step, ReLU activation functions are used. Each ReLU unit corresponds to an interval of the PWL function, defined by its activation $\max(0, x + b_i)$ as shown in equation (39). The output of each ReLU activation contributes to a segment of the PWL function, with the slope of each segment determined by the associated weight k_i. The weighted outputs are then summed to construct the first-level PWL function ϕ as shown in equation (39) [9].

For multivariable functions, multiple ϕ functions are summed over all variables in Q to form Ψ. The formulation of PWL function is then repeated by passing Ψ through a linear, non-trainable transformation to redistribute the outputs of the Ψ nodes. ReLU activations are applied again to form the PWL representation, then these outputs are summed into second-level functions Φ, and ultimately combining all Φ functions to predict the final output $\hat{y}$. This stepwise construction is conceptually aligned with how MIQCP models approximate solutions through piecewise linear approximations in each stage of the optimisation process.

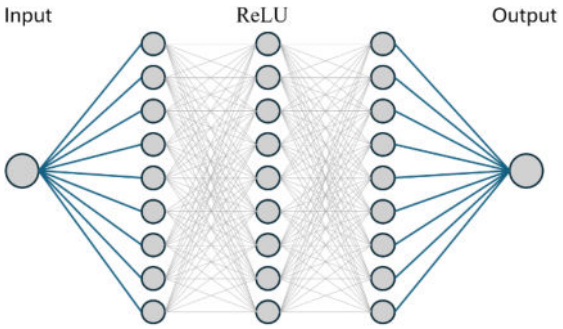


***Figure 3**: Standard Artificial Neural Network (ANN)*

In a standard Artificial Neural Network, each node is connected to every node in the subsequent layer. The configuration of this dense structure is based on the Universal Approximation Theorem which states that any continuous function can be approximated by a feedforward neural network with at least one hidden layer. While this approach ensures maximum information transmission, it simultaneously introduces significant computational complexity, requiring extensive weight parameters and increasing the risk of overfitting. [10] Sparse MLP, on the other hand, reduces the number of parameters which potentially enhances computational efficiency.

2.3 Testing

All models discussed were tested using a 12th Gen Intel® Core™ i7-12650H CPU@ 2.3GHz with 10 physical cores and 16 logical processors. The formulations were implemented using Pyomo version 6.8.2 [11]. The MIQCP models, SOS2, and Big-M formulations were solved with Gurobi version 12.0.0 [7]. The ANN and our ML model were implemented using PyTorch (torch) version 2.5.1 [12], with training performed using the Adam optimiser [13]. Our ML model was evaluated using 3 intervals in each PWL function and has a total of 90 trainable parameters. The ANN model architecture consisted of 2 input nodes, 10 hidden nodes in the first layer, 5 hidden nodes in the second layer, and 1 output node, connected with ReLU activation in the hidden layers and a linear activation in the output layer. There are 91 trainable parameters for the ANN model. We use this architecture because the two models have similar trainable parameters.

For testing our MIQCP model, the variant of KART as shown in equation (38) was implemented. The breakpoints were not fixed and the constraints from (31) to (37) were all implemented. For the MIQCP analysis and MIQCP vs ML comparison, we tested these models with datapoints generated from 2D functions. These functions ranged from simple to complex with increasing computational demand. These functions are as follows:

$$f(x) = x_1 x_2 \quad (40)$$

$$f(x) = x_1 \cos(\pi x_2) \quad (41)$$

$$f(x) = e^{x_1} \cos(\pi x_2) \quad (42)$$

$$f(x) = x_1^3 e^{x_2} \quad (43)$$

Each datapoint in the figures below represents an average value of the evaluated metric. The runtime measures both the efficiency and scalability while the mean squared error measures the accuracy of the predicted output. The data for testing was primarily generated uniformly within the range $[0,1)$. For each of the four functions, each test was run three times, and the minimum result was recorded. However, for problems requiring more than one hour to solve, only one test was run due to time constraints. The reported time is the average of the minimum times from all four equations.

3. Results and Discussion

This section evaluates the computational efficiency, accuracy and scalability of our framework through a series of experiments. The experiments are structured into 4 phases: testing MIQCP formulations, comparing MIQCP with ML. and finally testing the models in thermodynamics case study.

MIQCP runtime analysis (Big-M vs SOS2)

The experiments measured the runtimes required for the models to converge to a solution below our absolute error of 5e-3. The results show that the Big-M formulation consistently outperforms SOS2 in efficiency. Both models experience challenges with scalability as datapoints increases.

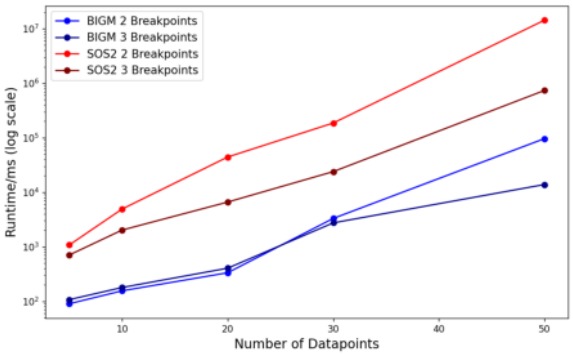


***Figure 4**: Runtime comparison - Big M and SOS2*

Figure 4 shows the runtime of our Big M and SOS2 model with the increasing number of data points. Even on a logarithmic scale (base 10), the lines exhibit an increasing gradient, indicating a very rapid growth in runtime as the number of data points increases. Big M consistently requires less time to achieve the set threshold of absolute error of 5e-3 compared to SOS2. In both cases, configurations with 2 breakpoints require more time than those with 3 breakpoints. This is likely because more line segments provide greater flexibility to capture the shape of complex functions and predict values more accurately. Notably, for processing 50 data points with 2 breakpoints, SOS2 requires approximately 4.5 hours, while Big M completes the task in roughly 1.5 minutes, underscoring the substantial computational efficiency of Big M. The potential reasons for this limitation include the inefficient reformulation of the SOS2 constraints and the poor selection of the big-M constants by the Gurobi solver [7]. This can lead to numerical instability and challenges in finding the solution space.

MIQCP runtime analysis across datapoints (Big M)

Since Big-M formulation performed better, we decided to proceed with this formulation to further analyse the

effect of the number of breakpoints on runtime across datasets ranging from 20-250 points. The models with single breakpoint performed poorly while multiple breakpoints improved performance but introduced scalability for larger datasets.

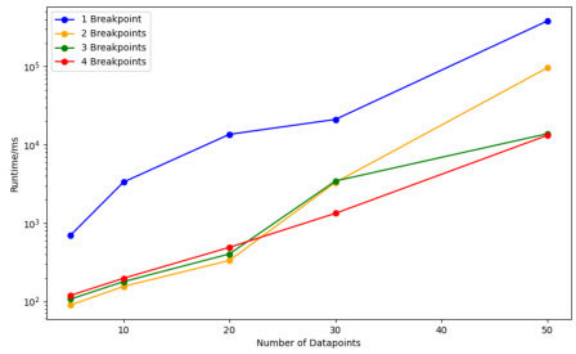


***Figure 5**: Runtime vs Breakpoints Across Datapoints*

Figure 5 compares the runtime of different breakpoints across a range of datapoints. The model with 1 breakpoint requires significantly more runtime to reach the minimum error threshold. This is due to the limits in flexibility, as it is constrained to only 2 linear line segments, making it hard to approximate all the output values and capture the shape of complex functions, e.g., high-frequency oscillating functions. On the other hand, models with multiple breakpoints perform similarly for small datasets and they almost instantaneously reach the minimum threshold. Nonetheless, as the number of datapoints increases, the runtime requirement grows exponentially. For datasets exceeding 100 datapoints, their runtimes reach our time limit of 30 minutes while the mean squared error plateaus, showing the model's limitations in scalability.

MIQCP MSE analysis across datapoints (Big M)

The experiments analysed the relationship between the number of breakpoints and the negative logarithm of the mean squared error (-log(MSE)) across datasets of varying sizes. Results show a diminishing return in accuracy for larger datasets with increasing breakpoints.

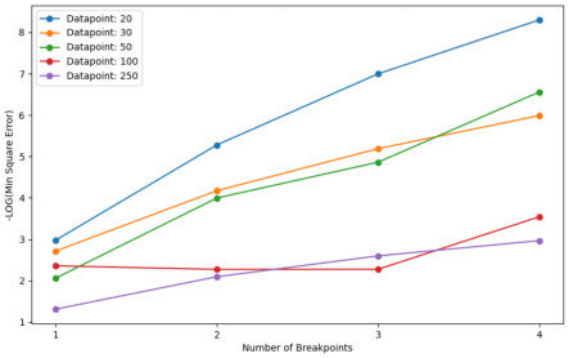


***Figure 6:** -log(MSE)vs Breakpoints Across Datasets*

For smaller datasets (e.g., 20 and 30 datapoints), increasing the number of breakpoints significantly improves the error metric. However, this improvement diminishes for larger datasets, such as those with 250 datapoints, where the curves plateau beyond three breakpoints. This indicates that additional model complexity yields diminishing returns as the dataset size increases. Furthermore, the sharp improvement observed in smaller datasets raises concerns about overfitting, where the model has too many intervals for this limited amount of data. For larger datasets, the limited performance gains beyond a certain number of breakpoints suggest scalability challenges, as the added computational cost may not justify the marginal error reduction.

MIQCP vs MLP Runtime Analysis

To address the scalability challenge of MIQCP, we implemented our piecewise linear framework into a customised sparse multilayer perceptron network (MLP). Figure 7 compares the runtime between our MLP model and the Big-M MIQCP formulation, evaluated with a fixed absolute error tolerance of $3 \cdot 10^{-3}$. We tested our models across [20, 30, 50, 100, 150, 200, 300, 500] datapoints.

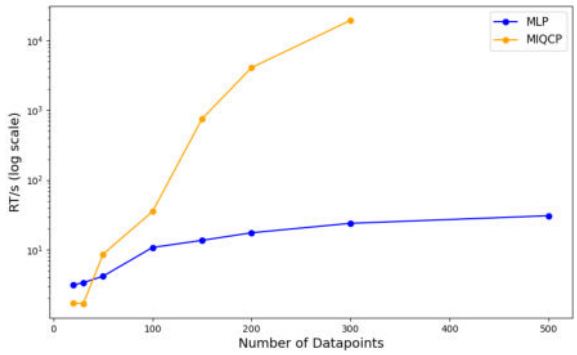


***Figure 7:** Runtime Comparison of MLP and MIQCP Models*

Initially, the MIQCP model demonstrates near-instantaneous convergence to a solution below the error threshold, outperforming the MLP in runtime for datasets with 20 and 30 datapoints. However, beyond 50 datapoints, the MIQCP's runtime increases exponentially with increasing data size and exceeds our 25,000 seconds time limit for 500 datapoints. Meanwhile, the MLP model exhibits consistent linear increase in runtime as it remains at around 30 seconds even at 500 datapoints with a fixed batch size of 16.

This behaviour can be attributed to their fundamental mechanisms. We implemented the Adaptive Moment Estimation (Adam) optimiser [13] in our MLP model. Adam is a stochastic gradient descent variant that updates network weights during training. It operates in linear time, hence with double the data size, we expect a proportionally doubling of the total training time. On the other hand, the Gurobi solver [7] in the MIQCP model is a branch-and-bound solver where the runtime depends on the solver's ability to heuristically find feasible regions and prune the search space. The efficiency of branch-and-bound relies on effective relaxation techniques to approximate feasible solutions and the depth of the branching tree.

MLP vs ANN Accuracy Analysis

The experiments compared the mean squared error of our MLP model and a standard ANN, with increasing epochs under a controlled batch size of 16 datapoints and a constant learning rate of 0.05. To ensure a fair comparison, both models were configured with an equal

number of parameters. We tested the models using several parameter configurations and averaged the results to produce the final graph.

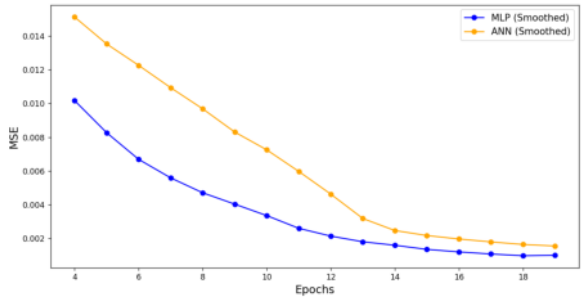


Figure 8: *Mean squared error of modified MLP and ANN Across Epochs*

Across all epochs, the MLP consistently achieves lower mean squared error than the ANN. Additionally, the MLP converges more quickly to an optimal solution, hence proving its higher efficiency and performance. There are two potential causes for this result. In comparing the performance of a standard the ANN and our approach, it is essential to consider the relationship between the number of parameters and predictive accuracy. For the MLP, increasing the number of nodes or parameters generally improves predictive performance, as the network gains a greater capacity to approximate complex functions. In contrast, the inherent structure of KA representations enables them to achieve a more compact representation of the underlying relationships within the data. This structural advantage allows the MLP to fully capture the necessary interactions with fewer parameters than the ANN. As a result, when both architectures are configured with a similar number of parameters, the MLP often exhibits superior performance due to its efficient parameter utilisation.

Secondly, the optimisation landscape of our MLP is less sensitive to flat regions or chaotic training behaviour. Since we implemented the Adam optimisation algorithm [13] in both MLP and ANN models, MLP's sparser connections and fewer highly interdependent weights allow gradient descent algorithms like Adam to converge more consistently with fewer oscillations. The reduced weight interdependence in MLP simplifies gradient calculations during backpropagation, promoting smoother and more consistent convergence compared to ANN, while avoiding chaotic updates that can destabilise training in ANN.

3.1 Chemical Engineering Application

This session evaluates the applicability of the MIQCP method and the machine learning approach in solving chemical engineering problems. Our proposed method is applicable in areas such as dynamical systems and thermodynamics, where the relationships are often nonlinear in nature. There are some existing methods such as sparse regression and symbolic genetic algorithm. Our work presents an alternative framework that offers a new approach to these challenges. One case study is discussed, with comparisons made to a well-established equation.

3.1.1 Thermodynamic Properties of gas

After testing with synthetic data, we applied our model to a real-life example, focusing on the thermodynamic properties of methanol. Gas properties are often evaluated using equations of state such as the Van der Waals equation [14] , which is expressed as follows (Equation 44). However, this approach can be limited by the assumptions made in the model and requires a lot of domain knowledge. Some equations of state are also highly specific to a system and have limited generalization. In contrast, the data-driven approach proposed in this study offers significant advantages by eliminating the need for predefined equations and domain knowledge. It provides a more adaptive framework for modelling thermodynamic properties. Additionally, due to the more interpretable property from KART, there is also a potential for exploring previously undiscovered trends. In this example, we compare the performance of our data driven approach against the Van der Waals equation.

$$P = \frac{RT}{V-b} - \frac{a}{V^2} \tag{44}$$

Data from the NIST Chemistry WebBook [15] was obtained, which contains real experimental measurements. A total of 150 data points were collected across different temperatures, pressures, and molar volumes. For this analysis, temperature and molar volume were used as inputs to predict pressure. The constants a and b are calculated from the critical pressure and temperature from NIST Chemistry WebBook [15].

The conditions for data processing and model evaluation were consistent with the previous study. To evaluate the performance of both methods, the dataset was split into training (80%) and testing (20%) subsets. 4 intervals are used in each PWL function. For the MIQCP, breakpoints of the first layers are fixed, and an absolute tolerance of 5e-5 is employed. For the neural network, we utilised a batch size of 16, a learning rate of 0.01, and trained the model for 500 epochs.
The data was first log-transformed to account for magnitude differences and then normalized using a min-max transformation to scale all values to the interval $[0,1]$.

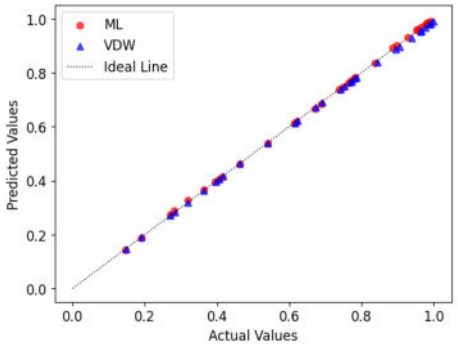


Figure 9*: ML vs Van der Waals equation*

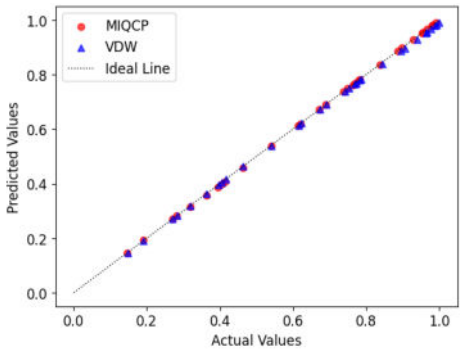


***Figure 10**: MIQCP vs Van der Waals equation*

The predictions of both models demonstrates high accuracies with an average MSE of approximately 1e-5 with model training completed in under 5 minutes. The model generates better predictions than the Van der Waals equaiton within the range of data. This can be attributed to two reasons. One is due to the numerical precision error in the critical pressure and temperature obtained from NIST, as the predicitons from the VdW eqaution are also very accurate. Another potential reason is the assumptions made by the VdW equation, which neglect some features of real gas behaviour and losing some of the accuracies, which is common in equations of states. Our model has shown the ability to capture real-life nonlinear dynamics in a short period of time and has the potential to perform better than some well-established equations. It is confident that our models can be extended to other chemical engineering applications such as biological systems and transfer processes.

Despite its advantages, there is a limitation to our proposed methods. The intervals of the PWL functions are determined based on the training data set and they are normally bounded in the range of the training set. Consequently, the model's predictive capabilities are constrained to the domain of the training data, making it highly effective at interpolation. However, the model is less effective in extrapolation, reducing its applications in chemical engineering scenarios that require predicting system behaviour beyond the current experimental conditions. This is common in non-parametric modelling where the model's intervals are not trained to capture the dynamics in the unexplored regions.

4. Conclusion

This study introduced and evaluated two formulations of MIQCP (Big-M and SOS2) based on the Kolmogorov- Arnold Representation Theorem (KART) to perform regression, with the primary goal of handling complex multivariate functions and achieving global optimality. We also developed a sparse neural network-based approach that mirrors our MIQCP Piecewise Linear Approximation model, providing an alternative for faster and more scalable solutions.

The experimental results highlighted the distinct strengths and trade-offs of the proposed approaches. The MIQCP formulations successfully captured the nonlinear behaviour of systems and demonstrated the ability to reach globally optimal solutions, with the Big-M method achieving significantly faster performance than SOS2. However, both methods faced scalability challenges as dataset size increased. As for our machine learning approach, our sparse MLP model proved highly efficient and scalable with runtimes growing linearly with dataset size. The machine learning approach exhibited much faster training speeds compared to the MIQCP formulations, though it lacks theoretical insight into its proximity to global optimality.

The proposed models were successfully applied to chemical engineering applications, modelling the thermodynamic properties of fluids. Results demonstrated that both the MIQCP and machine learning models offered predictive accuracies surpassing traditional equations of state, such as the Van der Waals equation, highlighting the applicability of data-driven methods to generalise nonlinear relationships while reducing reliance on domain-specific assumptions and experimental coefficients.

Although there many advantages to our methods, there is a limitation in our proposed method. The intervals of the PWL functions are determined based on the training data set and the bounds are set within the range of the training set. Consequently, the model's predictive capabilities are limited to the domain of the training data, making it less effective for extrapolation. Our model is unable to predict system behaviour beyond the untrained experimental conditions that deviate from the existing trend. This work bridges optimisation and machine learning to provide an accurate, efficient, scalable and reasonably interpretable framework for high-dimensional problems. The findings affirm the applicability of the proposed methods in chemical engineering while laying the foundation for future improvements in scalability and interpretability.

5. Future Works

A direction for future work involves exploring the reduction of initial layers for simpler functions. In practice, this adjustment may be feasible for certain cases and could significantly reduce computational costs, especially when the systems being modelled are not overly complex. Researchers could begin with this simplified structure first and revert to the full representation if the results are unsatisfactory. However, this approach raises theoretical concerns, as it deviates from the guarantees provided by KART. Currently, there is limited theoretical foundation on whether this simplification is consistently feasible or how it affects the performance. It is uncertain whether achieving global optimality in the reduced network truly represents the best solution or merely a local optimum. However, exploring this approach would provide invaluable insights for chemical engineering applications through the development of efficient and adaptable computational tools.

Another area that could potentially improve the model is to use predefined functions, such as trigonometric,

logarithmic, exponential or low-order polynomials to replace piecewise linear functions. In this case, the number of integer variables are determined by the possible combinations of the predefined functions rather than the number of datapoints. This could potentially improve scalability as the growth of integer variables would be more controlled and independent of the dataset size, thus limiting the number of possible combinations that the model needs to evaluate.

The current neural network demonstrates a much faster training runtime compared to the MIQCP approach. However, it remains slower than standard ANNs for low-dimensional problems due to its more complex structure. Future work should focus on optimising the implementation techniques to improve training efficiency. Additionally, researchers could apply the model to high-dimensional datasets. In such cases, the relatively small number of hidden layers in our model could lead to training times that are closer to those of traditional ANNs, which typically require deeper architectures to accommodate increased complexity.

Finally, the modified MIQCP and machine learning approaches could be tested against more state-of-the-art numerical methods using a variety of benchmark datasets in the future.

Acknowledgement
The authors would like to express their gratitude to Prof. Benoit Chachuat and Dr Lifeng Zhang for their invaluable guidance and support throughout this project. A special thanks to Dr Lifeng Zhang for providing foundational programming knowledge to facilitate development and implementation of this work.

References

[1] Zhang, T. (2005). Learning Bounds for Kernel Regression Using Effective Data Dimensionality. Neural Computation, 17(9), pp.2077–2098. doi:https://doi.org/10.1162/0899766054323008.

[2] Belyaev, M., Burnaev, E. and Kapushev, Y. (2014). Exact Inference for Gaussian Process Regression in case of Big Data with the Cartesian Product Structure. [online] arXiv.org. Available at: https://arxiv.org/abs/1403.6573 [Accessed 12 Dec. 2024].

[3] Braun, J. and Griebel, M. (2009). On a Constructive Proof of Kolmogorov's Superposition Theorem. *Constructive Approximation*, 30(3), pp.653–675. doi:https://doi.org/10.1007/s00365-009-9054-2.

[4] Ansar, T. and Ashraf, W.M. (2024). Kolmogorov–Arnold Network And Multi-Layer Perceptron for Modelling and Optimisation of Energy Systems. [online] doi:https://doi.org/10.2139/ssrn.4998362.

[5] Beale, E.M.L. and Forrest, J.J.H. (1976). Global optimization using special ordered sets. Mathematical Programming, 10(1), pp.52–69. doi:https://doi.org/10.1007/bf01580653.

[6] Crema, A. (2014). Mathematical programming approach to tighten a Big-M formulation. [online] Available at: https://optimization-online.org/wp-content/uploads/2014/08/4483.pdf [Accessed 12 Dec. 2024].

[7] Gurobi Optimization, LLC (2024) Gurobi Optimizer Reference Manual, version 12.0.0 https://www.gurobi.com

[8] Lorentz, G.G. (1962). Metric Entropy, Widths, and Superpositions of Functions. American Mathematical Monthly, 69(6), pp.469–485. doi:https://doi.org/10.1080/00029890.1962.11989915.

[9] Zhu, Z., Zhou, Y., Dong, Y. and Zhong, Z. (2023). PWLU: Learning Specialized Activation Functions with the Piecewise Linear Unit. IEEE Transactions on Pattern Analysis and Machine Intelligence, [online] 45(10), pp.12269–12286. doi:https://doi.org/10.1109/tpami.2023.3286109.

[10] Voigtlaender, F. (2023). The universal approximation theorem for complex-valued neural networks. Applied and Computational Harmonic Analysis, [online] 64, pp.33–61. doi:https://doi.org/10.1016/j.acha.2022.12.002.

[11] Bynum, Michael L., Gabriel A. Hackebeil, William E. Hart, Carl D. Laird, Bethany L. Nicholson, John D. Siirola, Jean-Paul Watson, and David L. Woodruff. Pyomo - Optimization Modeling in Python. Third Edition Vol. 67. Springer, 2021.

[12] Paszke, A., Gross, S., Soumith Chintala, Chanan, G., Yang, E., DeVito, Z., Lin, Z., Alban Desmaison, Luca Antiga and Lerer, A. (2017). Automatic differentiation in PyTorch. [online] Openreview.net. Available at: https://openreview.net/forum?id=BJJsrmfCZ.

[13]Kingma, D.P. and Ba, J. (2014). Adam: A Method for Stochastic Optimization. [online] arXiv.org. Available at: https://arxiv.org/abs/1412.6980.

[14] Berberan-Santos, M.N., Bodunov, E.N. and Pogliani, L. (2007). The van der Waals equation: analytical and approximate solutions. Journal of Mathematical Chemistry, [online] 43(4), pp.1437–1457. doi:https://doi.org/10.1007/s10910-007-9272-4.

[15] National Institute of Standards and Technology (2018). Thermophysical Properties of Fluid Systems. [online] Nist.gov. Available at: https://webbook.nist.gov/chemistry/fluid/

Expanding the applicability of the SAFT-γ Mie group-contribution approach to model thermodynamic properties of pure acetonitrile

Jake Ng and Zain Azeem
Department of Chemical Engineering, Imperial College London, U.K.

Abstract Predicting the thermodynamic properties of acetonitrile is important and of interest to many industries due to the chemical's wide use as a solvent. However, developing a model with reliable predictive capacity can be challenging when the compound of interest is highly polar or hydrogen bonding, like acetonitrile. In this work, we extend the applicability of the SAFT-γ Mie group-contribution method to model acetonitrile by presenting a new CH3CN molecular group, composed of two identical Mie segments. The parameters characterising group interactions are obtained using literature data for vapour pressure, saturation density and speed of sound, and used, without further fitting, for predictive calculations of the thermophysical properties (vapour pressure, saturation density, vaporisation enthalpy and second-order thermodynamic derivative properties) of pure acetonitrile. The comparison between the theoretical calculations and the experimental data highlighted the strong predictive capability of the presented model for the fluid-phase behaviour of pure acetonitrile, including accurate predictions for properties which have not been used to fit the model parameters.

1. Introduction

Acetonitrile (CAS number: 75-05-8), whose chemical formula is CH_3CN, is a solvent used widely across industries for its ability to dissolve a range of both polar and nonpolar compounds. It is commonly used as a solvent in the extraction of butadiene and its aqueous solution is also a common choice of mobile phase in reversed-phase liquid chromatography [1]. It is also a common starting material for the synthesis of many chemicals, pharmaceuticals and pesticides [1]. However, these are only some of the compound's applications. Notably, acetonitrile containing electrolyte is attracting increasing interest following a report it enhances lithium-ion battery's low-temperature performance and high-temperature durability, two longstanding challenges of the technology [2].

The knowledge of thermodynamic properties of pure acetonitrile and its mixtures is relevant for the design and continuous improvement of industrial processes like those aforementioned. However, making accurate predictions of such properties can be challenging for highly polar and hydrogen bonding molecules. Acetonitrile is known to be a highly polar molecule with a dipole moment of 3.92 D [3], but it has also been reported to be capable of forming hydrogen bonds, both as a pure substance and in mixtures [4–6].

The statistical associating fluid theory (SAFT) is an equation of state (EoS) [7,8], with many variants, that boasts a strong predictive ability [9]. Molecules are modelled as a chain of spherical monomers with the directionality of intermolecular interactions introduced via the short-range square-well association sites that mediate hydrogen bonding and other strong directional interactions. These SAFT approaches have been shown to provide accurate representation of both simple and complex fluids, including polymers and electrolytes [10,11]. The SAFT-γ Mie group-contribution method [12,13] considered in our current work is a successor to the SAFT-VR Mie EoS [14] and it was developed with a focus on its predictive capabilities for mixture thermodynamic properties. The behaviour of acetonitrile-containing systems has been studied with several SAFT EoS, including soft-SAFT [15] and variants of PC-SAFT [16,17], as well as non-SAFT approaches, such as COSMO-RS [18,19] and COSMO-SAC [20].

The aim of this current work is to propose a model capable of reliable predictions of thermophysical properties of pure acetonitrile using the SAFT-γ Mie EoS. A brief overview of the SAFT-γ Mie group-contribution model is presented in Section 2, followed by the methodology of model development.

2. SAFT-γ Mie theory and model development

The SAFT-γ Mie group-contribution method [12,13] models molecules of fluid as associating heteronuclear chains of spherical segments. It is also based on the group-contribution concept that the properties of a given molecule, or a mixture, can be estimated by considering the contributions from the groups that make up the molecule. This allows for parameters characterising any given group to be used for any molecules and mixtures containing that same group.

A given group k comprises of v_k^* identical spherical segments with a shape factor S_k ($0 \leq S_k \leq 1$), describing the extent each segment contributes to the molecule's overall thermodynamic properties. The interaction between two groups k and l is described via a Mie potential, which is a function of the intersegment distance, r_{kl} [21]:

$$\Phi_{kl}^{\text{Mie}}(r_{kl}) = C_{kl}\varepsilon_{kl}\left[\left(\frac{\sigma_{kl}}{r_{kl}}\right)^{\lambda_{kl}^r} - \left(\frac{\sigma_{kl}}{r_{kl}}\right)^{\lambda_{kl}^a}\right] \qquad (1)$$

which gives rise to parameters ε_{kl}, the depth of the potential well; σ_{kl}, the segment diameter; and λ_{kl}^r and λ_{kl}^a, which are the repulsive and attractive exponents, respectively. The prefactor C_{kl} is a function of these exponents and is there to ensure the minimum value of such an interaction corresponds to $-\varepsilon_{kl}$:

$$C_{kl} = \frac{\lambda_{kl}^{r}}{\lambda_{kl}^{r} - \lambda_{kl}^{a}} \left(\frac{\lambda_{kl}^{r}}{\lambda_{kl}^{a}}\right)^{\lambda_{kl}^{a}/(\lambda_{kl}^{r} - \lambda_{kl}^{a})} \quad (2)$$

Off-centre short-range square-well association sites are added to these segments to represent hydrogen bonding and strong polar interactions. $N_{ST,k}$ is the number of different types of association sites present in group k and $n_{k,a}$ is the number of association sites of a given type, where $a = 1, \ldots, N_{ST,k}$. The square-well interaction between a site of type a on segment k and a site of type b on segment l is characterised by the association energy, $\varepsilon_{kl,ab}^{HB}$, and the bonding volume, $K_{kl,ab}$, a parameter representing the strength and the range of the association. It is given by:

$$\Phi_{kl,ab}^{\mathrm{HB}}(r_{kl,ab}) = \begin{cases} -\varepsilon_{kl,ab}^{\mathrm{HB}}, & \text{if } r_{kl,ab} \leq r_{kl,ab}^{c} \\ 0, & \text{if } r_{kl,ab} > r_{kl,ab}^{c} \end{cases} \quad (3)$$

where $r_{kl,ab}^{c}$ is the cutoff range of the interaction. A SAFT-γ Mie representation of acetonitrile, considered in this work, is given in Figure 1 as an example.

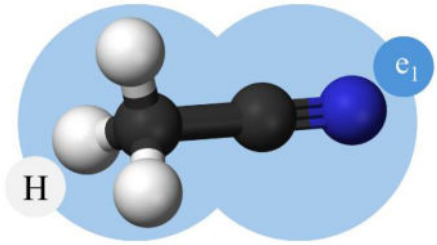


Figure 1. SAFT-γ Mie molecular model of acetonitrile including association sites. The model for acetonitrile comprises two identical segments decorated with two sites for like association – one of type H and another of type e_1. The model is further discussed in Section 3.

After determining their values, these parameters are used to calculate the total Helmholtz free energy of the fluid, A, which is expressed as a sum of four contributions for non-ionic species [12]:

$$A = A^{\mathrm{ideal}} + A^{\mathrm{monomer}} + A^{\mathrm{chain}} + A^{\mathrm{assoc.}} \quad (4)$$

where A^{ideal} refers to the free energy of the ideal gas mixture, A^{monomer} to the contribution due to dispersive interactions between monomeric segments, A^{chain} to the contribution due to the formation of fused monomer chains, and $A^{\mathrm{assoc.}}$ to the contribution due to association. The reader is referred to previous papers [12,13] for detailed expressions of each term and for the explanations on the Born and ion terms that appear in the full SAFT-γ Mie expression.

Once the functional form of the total Helmholtz free energy of the given compound or mixture is established, then its other single-phase and equilibrium properties can be determined using standard thermodynamic relationships [22]. In particular, the key relationships are those for pressure,

$$P = -\left(\frac{\partial A}{\partial V}\right)_{N,T} \quad (5)$$

and the chemical potential of component i,

$$\mu_i = \left(\frac{\partial A}{\partial N_i}\right)_{V,T,N_{j\neq i}} \quad (6)$$

along with an expression for the Gibbs free energy in terms of the Helmholtz free energy, given as:

$$\begin{aligned} G &= A + pV \\ &= A - V\left(\frac{\partial A}{\partial V}\right)_{N,T} \end{aligned} \quad (7)$$

These expressions are used to obtain fluid-phase equilibrium properties by solving equalities while ensuring that minimum free energy is achieved.

2.1. Parameter estimation process

The parameters for group–group interactions are determined by minimising the deviations between experimental and calculated data. These interactions are essential for accurately modelling bulk properties of fluids. The parameter estimation minimises a least-square objective function, which evaluates the overall differences between experimental and calculated values for each property and data point. The general form of F_{obj} is given by:

$$F_{obj} = \sum_{z} w_z \frac{1}{N_z} \sum_{i=1}^{N_z} \left[\frac{Z_i^{exp} - Z_i^{calc}}{Z_i^{exp}}\right]^2 \quad (8)$$

where Z represents the different properties considered in the estimation, which, in this current work, are saturated vapour pressure, saturated liquid density and liquid speed of sound; N_Z is the total number of experimental data points for each property Z; w_z is the weight applied to each property Z, which is used to adjust the influence of each property on the overall objective function; and Z_i^{exp} and Z_i^{calc} are the experimental and calculated values of property Z at a given data point i, respectively.

The parameter estimation process is carried out using the gPROMS® platform, which performs the optimisation using numerical solvers. Experimental data are sourced from DETHERM, a reliable thermodynamic database. The estimation procedure aims to provide the best possible match between the calculated and experimental values for the thermodynamic properties used in the estimation, based on local optimisation from a set of Sobol points within defined parameter space boundaries. The numerical solvers identify the best result within the explored parameter space, though this does not guarantee a global optimum.

The quality of the SAFT-γ Mie group contribution method in describing the thermodynamic properties of a fluid is evaluated using the percentage average absolute

deviation (%AAD) and the average absolute deviation (AAD). These metrics provide quantitative measures of how well the theoretical model replicates the experimental data. For each property Z, the %AAD is defined as:

$$\%AAD_Z = \frac{100}{N_z}\sum_{i=1}^{N_z}\left|\frac{Z_i^{calc} - Z_i^{exp}}{Z_i^{exp}}\right| \quad (9)$$

Similarly, the AAD is defined as:

$$AAD_Z = \frac{1}{N_z}\sum_{i=1}^{N_z}\left|Z_i^{calc} - Z_i^{exp}\right| \quad (10)$$

These measures are applied to properties of interest, such as enthalpy of vaporisation and saturated gas density, to assess the accuracy of the theoretical approach.

3. Results and discussion

In this section, we present the SAFT-γ Mie model developed for pure acetonitrile and evaluate its accuracy in describing the compound's thermodynamic properties. Accurate modelling of the pure-component properties of acetonitrile is essential to ensure a reliable description of its behaviour in mixtures can be achieved.

The ability of the SAFT-γ Mie model to predict the thermodynamic properties of acetonitrile is evaluated, focusing on key properties including vapour pressure, saturated liquid density and heat capacity. The descriptions of the thermodynamic properties of other compounds, such as water, alcohols and ketones, have been addressed in previous studies. We compare the parameter values developed for acetonitrile with those for similar molecules of comparable molecular characteristics to ensure consistency and reliability of the group contribution parameters (see Table 1).

3.1. SAFT-γ Mie model for acetonitrile

As described in Figure 1, acetonitrile is modelled as a molecular group consisting of $\nu_k^* = 2$ identical Mie segments in this current work. The group has two association sites, one each of types H and e_1, in order to account for the like-group association interactions taking place in pure acetonitrile fluid.

Available property data for saturated liquid density [23] and vapour pressure [24] in range 270–520 K and speed of sound data in range 260–315 K [25] are obtained from DETHERM for use in parameter regression. The SAFT-γ Mie parameters for the acetonitrile group presented in Table 1 are developed using only these data. Once estimated, these parameters enable the description of the vapour–liquid equilibrium and other thermodynamic properties of pure acetonitrile.

Figures 2 and 3 present the SAFT-γ Mie model calculations alongside the corresponding experimental property data. The results demonstrate that the proposed model predicts accurately the vapour-liquid equilibrium, with deviations (%AAD) of 0.98% for vapour pressure and 0.30% for saturated liquid density. Presented alongside these are the predictive calculations by a non-associating model of acetonitrile. The comparison emphasises the improvements achieved by taking into account the like-group association in pure acetonitrile fluid. In particular, it is shown that incorporating association sites into the model suppresses significantly the overestimation of the critical temperature. Significant error in the speed of sound predictions can also be observed in the non-associating model. For completeness, SAFT-γ Mie group parameters for the non-associating model are $\nu_k^* = 2$, $S_k = 0.70127$, $\lambda_{kk}^r = 18.426$, $\lambda_{kk}^a = 6.0000$, $\sigma_{kk} = 3.9127$ Å and $\varepsilon_{kk}/k_b = 507.09$ K. This model, unlike the associating model proposed in our current work, was not fitted to the experimental speed of sound data. An associating model regressed without speed of sound data displays better agreement with other properties, however, significant discrepancies with experimental speed of sound data remain. The finalised model includes the speed of sound data in the parameter estimation process to ensure that a more accurate overall description of the system is achieved.

Table 1
SAFT-γ Mie like-group interaction parameters. The significance of each parameter is provided in Section 2.

k	Group k	ν_k^*	S_k	λ_{kk}^r	λ_{kk}^a	σ_{kk} (Å)	(ε_{kk}/k_B) (K)	$N_{ST,k}$	$n_{k,H}$	n_{k,e_1}	n_{k,e_2}	Ref.
1	CH_3CN	2	0.80934	16.004	6.0000	3.5389	270.34	2	1	1	0	*
2	H_2O	1	1.00000	17.020	6.0000	3.0063	266.68	2	2	2	0	[26]
3	CH_3OH	2	0.83517	19.235	6.0000	3.2462	307.69	2	1	2	0	[26]
4	CH_3COCH_3	3	0.72135	17.433	6.0000	3.5980	286.02	3	1	1	1	[27]

Note: The asterisk * in the Ref. column indicates that the CH_3CN group is characterised in the current work.

Table 2
SAFT-γ Mie group association energy ($\varepsilon_{kl,ab}^{HB}$) and bonding volume ($K_{kl,ab}$) parameters.

k	l	Group k	Site a of group k	Group l	Site b of group l	$(\varepsilon_{kl,ab}^{HB}/k_B)$ (K)	$K_{kl,ab}$ (Å^3)	Ref.
1	1	CH_3CN	H	CH_3CN	e_1	1811.4	2042.9	*
2	2	H_2O	H	H_2O	e_1	1985.4	101.69	[26]
3	3	CH_3OH	H	CH_3OH	e_1	2062.1	106.57	[26]
4	4	CH_3COCH_3	H	CH_3COCH_3	e_1	980.20	2865.2	[27]

Note: The asterisk * in the Ref. column indicates that the CH_3CN group is characterised in the current work.

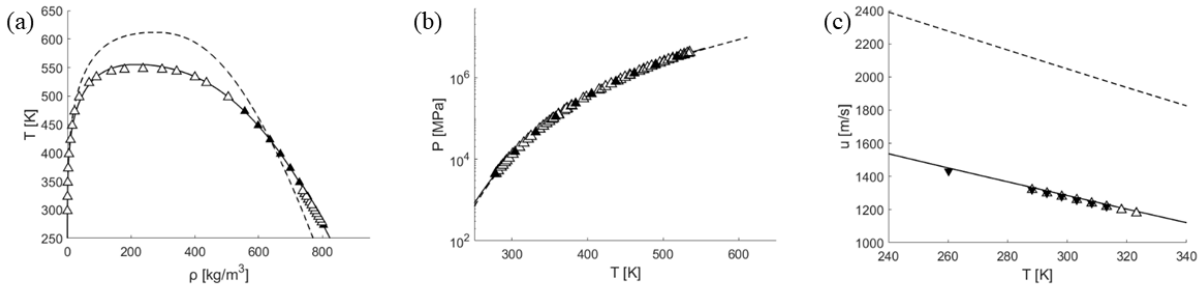


Figure 2. SAFT-γ Mie calculations for pure acetonitrile alongside data used in parameter estimation: (a) vapour-liquid coexistence density ρ_{sat} (b) vapour pressure P_{vap} (c) liquid speed of sound u. The open symbols correspond to experimental data not used in parameter estimation, sourced from [23] for saturated vapour density, [23,28–30] for saturated liquid density, [24,31] for vapour pressure and [25,32,33] for speed of sound. The filled upwards pointing symbols correspond to experimental data used in parameter estimation for both the associating and non-associating models, sourced from [23] for saturated liquid density and [24] for vapour pressure. The filled downward pointing symbols correspond to experimental speed of sound data used in fitting the associating model only, sourced from [25]. The dashed curves represent SAFT-γ Mie calculations with the non-associating model and the continuous curves represent those with the associating model.

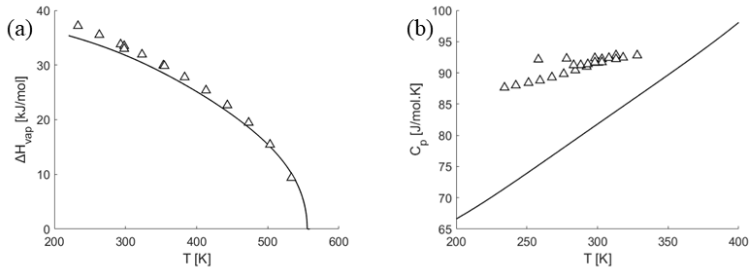


Figure 3. SAFT-γ Mie predictive calculations: (a) vaporisation enthalpy ΔH_{vap} (b) isobaric liquid heat capacity c_P. The symbols correspond to experimental data, sourced from [34,35] for vaporisation enthalpy and [36–41] for isobaric heat capacity. The continuous curves represent the SAFT-γ Mie calculations.

Table 3

Overview of the accuracy of SAFT-γ Mie in the calculation of pure compound properties: critical temperature T_c, saturated vapour density ρ_{sat}^{vap}, saturated liquid density ρ_{sat}^{liq}, vapour pressure P_{vap}, liquid speed of sound u, vaporisation enthalpy ΔH_{vap} and isobaric heat capacity c_P.

Critical temperature							
Compound	T range (K)	P	$N_{s,p}^{est}$	$N_{s,p}^{tot}$	%AAD	AAD (K)	Ref.
Acetonitrile	545.41	-	0	1	1.84	10.06	*
Cyclohexanol	647.1	-	0	1	0.6336	4.100	[42]
Menthol	637	-	0	1	1.110	7.700	[42]
Saturated vapour density							
Compound	T range (K)	P	$N_{s,p}^{est}$	$N_{s,p}^{tot}$	%AAD	AAD (kg/m^3)	Ref.
Acetonitrile	297-533	saturation	0	11	7.65	1.77	*
Cyclohexanol	519-606	saturation	11	11	2.807	1.059	[42]
Menthol	511-600	saturation	11	16	10.34	4.276	[42]
Saturated liquid density							
Compound	T range (K)	P	$N_{s,p}^{est}$	$N_{s,p}^{tot}$	%AAD	AAD (kg/m^3)	Ref.
Acetonitrile	268-533	saturation	10	34	0.30	2.16	*
Water	280-640	saturation	-	37	1.95	-	[43]
Methanol	180-510	saturation	-	34	0.72	-	[43]
Acetone	180-460	saturation	-	29	0.25	-	[27]
Vapour pressure							
Compound	T range (K)	P	$N_{s,p}^{est}$	$N_{s,p}^{tot}$	%AAD	AAD (kPa)	Ref.
Acetonitrile	277-528	-	10	63	0.98	13.62	*
Water	280-640	-	-	37	1.15	-	[43]
Methanol	180-510	-	-	34	0.43	-	[43]
Acetone	180-460	-	-	29	0.19	-	[27]
Liquid speed of sound							
Compound	T range (K)	P	$N_{s,p}^{est}$	$N_{s,p}^{tot}$	%AAD	AAD (m/s)	Ref.
Acetonitrile	260-328	1	7	34	1.05	13.23	*
Menthol	307-413	1.013	0	6	2.652	29.86	[42]
Vaporisation enthalpy							
Compound	T range (K)	P	$N_{s,p}^{est}$	$N_{s,p}^{tot}$	%AAD	AAD (kJ/mol)	Ref.
Acetonitrile	263-533	saturation	0	14	5.86	1.63	*
Cyclohexanol	300-600	saturation	14	14	5.883	2.210	[42]
Menthol	298-623	saturation	0	12	10.59	5.556	[42]

Isobaric heat capacity Compound	T range (K)	P	$N_{s,p}^{est}$	$N_{s,p}^{tot}$	%AAD	AAD (J/mol K)	Ref.
Acetonitrile	234-328	1.01325	0	30	12.18	11.07	*
Menthol	333-633	saturation	0	11	11.45	46.99	[42]

Note: The asterisk * in the Ref. column indicates that the CH_3CN group is characterised in the current work.

Table 4

Percentage average absolute deviations for saturated liquid density ρ_{sat}^{liq} and vapour pressure P_{vap} for modelling pure acetonitrile with SAFT variants

EoS	T range (K)	%AAD ρ_{sat}^{liq}	T range (K)	%AAD P_{vap}	Ref.
SAFT-γ Mie	268-533	0.30	277-528	0.98	*
soft-SAFT	280-528	0.70	280-528	2.03	[15]
PC-SAFT	292-529	6.14	292-529	1.69	[16]
PCP-SAFT	292-529	2.32	292-529	1.92	[16]
PCIP-SAFT	292-529	3.31	292-529	0.66	[16]
mg-SAFT	230-545	1.79	230-545	2.58	[17]
mg-SAFT	230-545	5.16	230-545	4.50	[17]
mg-SAFT	230-545	5.20	230-545	1.84	[17]

Note: The asterisk * in the Ref. column indicates that the model is characterised in the current work.

3.2. Assessment of thermodynamic property predictions

3.2.1. Critical temperature

The SAFT-γ Mie model overestimates the experimental critical temperature by 10.06 K, corresponding to a %AAD of 1.84%. This slight overestimation is expected, as accurately describing the near-critical region presents significant challenges due to the sensitivity of the thermodynamic properties near the critical point. An overprediction is generally preferable to an underprediction, as overestimation ensures that the model describes thermodynamic behaviour more conservatively, avoiding properties above the critical temperature being incorrectly deemed infeasible. This trend aligns with previously developed SAFT-γ Mie models, such as the model for menthol [42], which has a reported percentage deviation of 1.110%.

3.2.2. Saturated vapour density

The model yields a %AAD of 7.65% for the saturated vapour density of acetonitrile. While this value is relatively high in percentage terms, the AAD is much smaller at 1.77 kg/m^3, which is comparable to the AAD of 2.16 kg/m^3 for saturated liquid density. This difference in percentage deviation can be attributed to the fact that the saturated vapour density values are closer to zero, meaning that small absolute deviations can result in a larger percentage deviation. For comparison, SAFT-γ Mie models for menthol and cyclohexanol [42] report %AAD values of 2.807% and 10.34% respectively. Both models include vapour density data in their parameter estimation, which would be expected to reduce deviations. Despite this, our model's %AAD falls within this range, demonstrating that it achieves an acceptable level of accuracy.

3.2.3. Saturated liquid density

The model provides an excellent description of the vapour-liquid equilibrium of acetonitrile, with a %AAD of 0.30% for saturated liquid density. This value is comparable to other modelled systems: earlier studies reported %AADs of 1.95% for water, 0.72% for methanol [43] and 0.25% for acetone [27] for their respective saturated liquid densities.

3.2.4. Vapour pressure

The %AAD for vapour pressure in this study is 0.98%, demonstrating the reliability of the proposed associating model for acetonitrile. This level of accuracy is comparable to, or better than, predictions for other compounds using the SAFT-γ Mie approach. For comparison, methanol, which is also a hydrogen bonding compound, has a %AAD of 0.43%, while water shows a higher %AAD of 1.15%, as reported in an earlier study on their respective model developments [43].

3.2.5. Speed of sound

Excellent agreement is observed with the experimental speed of sound data for acetonitrile, with a %AAD of 1.05%. Including the speed of sound data in parameter estimation was crucial, as significant error was observed in preliminary models that excluded it.

3.2.6. Enthalpy of vaporisation

The model provides a good description for the vaporisation enthalpy of acetonitrile, as shown in Figure 3, with a %AAD of 5.86%. This performance is comparable to, or better than, other SAFT-γ Mie models developed for different compounds, such as cyclohexanol and menthol, which report %AAD values of 5.883% and 10.59%, respectively [42]. These results demonstrate that the model provides a reliable description of vaporisation enthalpy, consistent with the performance of other SAFT-γ Mie models for different compounds.

3.2.7. Isobaric heat capacity

The model exhibits a %AAD of 12.18% for the isobaric liquid heat capacity of acetonitrile. While this level of deviation is higher compared to that of other properties, it remains comparable with the performance of SAFT-γ Mie models for some of the other compounds. For example, menthol [42] is reported to have a %AAD of 11.45% for this property, with an absolute deviation of 46.99 J/mol K, compared to our AAD of 11.07 J/mol K. The set of parameters presented in our current work delivers the best achievable balance for accurately describing the overall properties of acetonitrile.

Improved predictions of heat capacity may be obtained with additional experimental data to refine the model. However, to the best of our knowledge, such data are not currently available.

3.3 Comparison with other SAFT approaches

Table 4 summarises the temperature ranges and %AADs of various SAFT-based approaches for predicting the saturated liquid density and vapour pressure of pure acetonitrile fluid. The SAFT-γ Mie model, presented in our current work, demonstrates a %AAD of 0.30% for saturated liquid density and 0.98% for vapour pressure. For both properties, this is an improvement upon the accuracy of previously reported SAFT models concerning pure acetonitrile. The only outperformer to this current work is the PCIP-SAFT model [16] for vapour pressure, with a reported %AAD of 0.66% against our 0.98%. However, the PCIP-SAFT model has a higher %AAD for saturated liquid density compared to the SAFT-γ Mie model, suggesting the latter provides a more accurate overall description of the vapour-liquid equilibrium.

Conclusions

We have successfully extended the SAFT-γ Mie framework by introducing a reliable model for the acetonitrile molecular group, addressing the challenges posed by its strong dipole and hydrogen bonding capabilities. Three different thermophysical data for pure acetonitrile have been used for parameter regression. The quality of the proposed model has been presented visually by comparing the calculated fluid behaviour with the experimental data. It was also assessed quantitatively through %AADs and AADs.

The inclusion of association sites has enhanced significantly the model's overall reliability, improving its description of vapour-liquid equilibrium and other key thermodynamic properties. As a result, an improvement over previous SAFT-based works for modelling pure acetonitrile has been achieved, once again demonstrating the strong predictive capability of the SAFT-γ Mie approach for describing the thermodynamic behaviour of associating fluids.

The model establishes a solid foundation for extending the SAFT-γ Mie approach to acetonitrile containing mixtures. Future work should aim to build upon this model by characterising unlike interactions. This may involve introducing the so-called unlike induced association site. Developing a full SAFT-γ Mie model for acetonitrile capable of describing multicomponent systems will further expand the applicability of SAFT-γ Mie to industrial processes involving acetonitrile.

Acknowledgements

The authors thank Dr. Thomas Bernet, Shubhani Paliwal and all other members of the MSE group for the support and guidance they have provided over the course of this research.

References

[1] European Commission: Joint Research Centre. European Union risk assessment report : acetonitrile : CAS75-05-8 No: 200-835-2. Volume 18. Publications Office; 2002.

[2] Matsuoka N, Kamine H, Natsume Y, Yoshino A. Moderately Concentrated Acetonitrile-containing Electrolytes with High Ionic Conductivity for Durability-oriented Lithium-Ion Batteries. ChemElectroChem 2021;8:3095–104. https://doi.org/10.1002/celc.202100927.

[3] Alston Steiner P, Gordy W. Precision measurement of dipole moments and other spectral constants of normal and deuterated methyl fluoride and methyl cyanide. J Mol Spectrosc 1966;21:291–301. https://doi.org/10.1016/0022-2852(66)90152-4.

[4] Kharat B, Deshmukh V, Chaudhari A. Hydrogen-bonding interactions in acetonitrile oligomers using density functional theory method. Struct Chem 2012;23:637–44. https://doi.org/10.1007/s11224-011-9909-6.

[5] Gopi R, Ramanathan N, Sundararajan K. Acetonitrile–water hydrogen-bonded interaction: Matrix-isolation infrared and ab initio computation. J Mol Struct 2015;1094:118–29. https://doi.org/10.1016/J.MOLSTRUC.2015.03.046.

[6] Kryachko ES, Nguyen MT. Hydrogen bonding between phenol and acetonitrile. Journal of Physical Chemistry A 2002;106:4267–71. https://doi.org/10.1021/jp020423i.

[7] Chapman WG, Gubbins KE, Jackson G, Radosz M. SAFT: Equation-of-state solution model for associating fluids. Fluid Phase Equilib 1989;52:31–8. https://doi.org/10.1016/0378-3812(89)80308-5.

[8] Chapman WG, Gubbins KE, Jackson G, Radosz M. New reference equation of state for associating liquids. Ind Eng Chem Res 1990;29:1709–21. https://doi.org/10.1021/ie00104a021.

[9] Wilhelmsen Ø, Aasen A, Skaugen G, Aursand P, Austegard A, Aursand E, et al. Thermodynamic Modeling with Equations of State: Present Challenges with Established Methods. Ind Eng Chem Res 2017;56:3503–15. https://doi.org/10.1021/acs.iecr.7b00317.

[10] Müller EA, Gubbins KE. Molecular-Based Equations of State for Associating Fluids: A Review of SAFT and Related Approaches. Ind Eng Chem Res 2001;40:2193–211. https://doi.org/10.1021/ie000773w.

[11] Economou IG. Statistical Associating Fluid Theory: A Successful Model for the Calculation of Thermodynamic and Phase Equilibrium Properties of Complex Fluid Mixtures. Ind Eng Chem Res 2002;41:953–62. https://doi.org/10.1021/ie0102201.

[12] Papaioannou V, Lafitte T, Avendaño C, Adjiman CS, Jackson G, Müller EA, et al. Group contribution methodology based on the statistical associating fluid theory for heteronuclear molecules formed from Mie segments. J Chem Phys 2014;140. https://doi.org/10.1063/1.4851455.

[13] Haslam AJ, González-Pérez A, Di Lecce S, Khalit SH, Perdomo FA, Kournopoulos S, et al. Expanding the Applications of the SAFT-γ Mie Group-Contribution Equation of State: Prediction of Thermodynamic Properties and Phase Behavior of Mixtures. J Chem Eng Data 2020;65:5862–90. https://doi.org/10.1021/acs.jced.0c00746.

[14] Lafitte T, Apostolakou A, Avendaño C, Galindo A, Adjiman CS, Müller EA, et al. Accurate statistical associating fluid theory for chain molecules formed from Mie segments. J Chem Phys 2013;139. https://doi.org/10.1063/1.4819786.

[15] Belkadi A, Hadj-Kali MK, Gerbaud V, Joulia X, Llovell F, Vega LF. Modeling the phase equilibria of nitriles by the soft-SAFT Equation of State. Computer Aided Chemical Engineering, vol. 25, 2008, p. 739–44. https://doi.org/10.1016/S1570-7946(08)80129-0.

[16] Kleiner M, Gross J. An equation of state contribution for polar components: Polarizable dipoles. AIChE Journal 2006;52:1951–61. https://doi.org/10.1002/aic.10791.

[17] NguyenHuynh D, Nguyen TTX. Application of the modified group-contribution PC-SAFT to nitrile and their mixtures. Fluid Phase Equilib 2017;450:112–25. https://doi.org/10.1016/j.fluid.2017.07.017.

[18] Eckert F, Leito I, Kaljurand I, Kütt A, Klamt A, Diedenhofen M. Prediction of acidity in acetonitrile solution with COSMO-RS. J Comput Chem 2009;30:799–810. https://doi.org/10.1002/jcc.21103.

[19] Zhang Z, Lu R, Zhang Q, Chen J, Li W. COSMO-RS based ionic liquid screening for the separation of acetonitrile and ethanol azeotropic mixture. Journal of Chemical Technology & Biotechnology 2020;95:1191–202. https://doi.org/10.1002/jctb.6305.

[20] Chen W-L, Hsieh C-M, Yang L, Hsu C-C, Lin S-T. A Critical Evaluation on the Performance of COSMO-SAC Models for Vapor–Liquid and Liquid–Liquid Equilibrium Predictions Based on Different Quantum Chemical Calculations. Ind Eng Chem Res 2016;55:9312–22. https://doi.org/10.1021/acs.iecr.6b02345.

[21] Mie G. Zur kinetischen Theorie der einatomigen Körper. Ann Phys 1903;316:657–97. https://doi.org/10.1002/andp.19033160802.

[22] Bett KE, Rowlinson JS, Saville G. Thermodynamics for Chemical Engineers. The MIT Press; 2003. https://doi.org/10.7551/mitpress/6793.001.0001.

[23] Francesconi AZ, Franck EU, Lentz H. Die PVT-Daten des Acetonitrils bis 450 °C und 2500 bar. Berichte Der Bunsengesellschaft Für Physikalische Chemie 1975;79:897–901. https://doi.org/10.1002/bbpc.19750791012.

[24] Ewing MB, Ochoa JCS. Vapor Pressures of Acetonitrile Determined by Comparative Ebulliometry. J Chem Eng Data 2004;49:486–91. https://doi.org/10.1021/je034144f.

[25] Niepmann R. Thermodynamic properties of acetonitrile 1. Speeds of sound between 240 and 475 K and up to 60 MPa. J Chem Thermodyn 1984;16:779–85. https://doi.org/10.1016/0021-9614(84)90062-4.

[26] Dufal S, Papaioannou V, Sadeqzadeh M, Pogiatzis T, Chremos A, Adjiman CS, et al. Prediction of thermodynamic properties and phase behavior of fluids and mixtures with the SAFT-γ mie group-contribution equation of state. J Chem Eng Data 2014;59:3272–88. https://doi.org/10.1021/je500248h.

[27] Sadeqzadeh M, Papaioannou V, Dufal S, Adjiman CS, Jackson G, Galindo A. The development of unlike induced association-site models to study the phase behaviour of aqueous mixtures comprising acetone, alkanes and alkyl carboxylic acids with the SAFT-γ Mie group contribution methodology. Fluid Phase Equilib 2016;407:39–57. https://doi.org/10.1016/j.fluid.2015.07.047.

[28] Chen F, Yang Z, Chen Z, Hu J, Chen C, Cai J. Density, viscosity, speed of sound, excess property and bulk modulus of binary mixtures of γ-butyrolactone with acetonitrile, dimethyl carbonate, and tetrahydrofuran at temperatures (293.15 to 333.15) K. J Mol Liq 2015;209:683–92. https://doi.org/10.1016/j.molliq.2015.06.041.

[29] Iulian O, Ciocirlan O. Volumetric Properties of Binary Mixtures of Two 1-Alkyl-3-Methylimidazolium Tetrafluoroborate Ionic Liquids with Molecular Solvents. J Chem Eng Data 2012;57:2640–6. https://doi.org/10.1021/je300316a.

[30] Ikeuchi H, Kanakubo M, Okuno S, Sato R, Fujita K, Hamada M, et al. Densities and Viscosities of Tris(acetylacetonato)cobalt(III) Complex Solutions in Various Solvents. J Solution Chem 2010;39:1428–53. https://doi.org/10.1007/s10953-010-9598-6.

[31] Shakhmuradov ShG, Guseinov SO. Investigation of thermodynamic characteristics of saturated nitriles in a broad temperature and pressure range. Izv Vyssh Uchebn Zaved, Neft Gaz 1984:65–9.

[32] Varfolomeev MA, Rakipov IT, Solomonov BN, Marczak W. Speed of Sound, Density, and

Related Thermodynamic Excess Properties of Binary Mixtures of 2-Pyrrolidone and *N* -Methyl-2-pyrrolidone with Acetonitrile and Chloroform. J Chem Eng Data 2016;61:1032–46. https://doi.org/10.1021/acs.jced.5b00474.

[33] Grande M del C, Juliá JÁ, Barrero CR, Marschoff CM. Sound velocity measurements in the water + acetonitrile system at temperatures from 293.15 to 323.15 K and its implications on thermodynamic data processing. Phys Chem Liquids 2013;51:457–68. https://doi.org/10.1080/00319104.2012.737790.

[34] Liessmann G, Schmidt W, Reiffarth S. Recommended thermophysical data. Data Compilation of the Saechsische Olefinwerke Boehlen Germany 1995;1.

[35] Antosik M, Gałka M, Malanowski SK. Vapor−Liquid Equilibrium for Acetonitrile + Propanenitrile and 1-Pentanamine + 1-Methoxy-2-propanol. J Chem Eng Data 2004;49:11–7. https://doi.org/10.1021/je025660t.

[36] Putnam WE, McEachern DM, Kilpatrick JE. Entropy and Related Thermodynamic Properties of Acetonitrile (Methyl Cyanide). J Chem Phys 1965;42:749–55. https://doi.org/10.1063/1.1696002.

[37] Al'per GA, Peshekhodov PB, Nikiforov MYu, Krestov GA. Mutual Association of Structure-Specialities and of the Character of Transition of the Excess Heat Capacity in Binary Solutions with Chloroform. Zh Fiz Khim 1990;64:652–5.

[38] Kolker AM, Kulikov MV, Krestov AlG. Volumes and heat capacities of binary non-aqueous mixtures. Part 1. The systems acetonitrile—formamide and acetonitrile—N-methylformamide. Thermochim Acta 1992;211:61–72. https://doi.org/10.1016/0040-6031(92)87007-W.

[39] Checoni RF, Francesconi AZ. Measurements of the Molar Heat Capacities and Excess Molar Heat Capacities for Acetonitrile + Diethylamine or sec-Butylamine Mixtures at Various Temperatures and Atmospheric Pressure. J Solution Chem 2007;36:913–22. https://doi.org/10.1007/s10953-007-9155-0.

[40] Kolker AM, Safonova LP. Molar heat capacities of the (water+acetonitrile) mixtures at T=(283.15, 298.15, 313.15, and 328.15)K. J Chem Thermodyn 2010;42:1209–12. https://doi.org/10.1016/j.jct.2010.04.019.

[41] Anouti M, Jacquemin J, Lemordant D. Volumetric Properties, Viscosities, and Isobaric Heat Capacities of Imidazolium Octanoate Protic Ionic Liquid in Molecular Solvents. J Chem Eng Data 2010;55:5719–28. https://doi.org/10.1021/je100671v.

[42] Bernet T, Paliwal S, Alyazidi A, Standish R, Haslam AJ, Adjiman CS, et al. Modeling the thermodynamic properties of cyclic alcohols with the SAFT-γ Mie approach: application to cyclohexanol and menthol systems. Mol Phys 2024. https://doi.org/10.1080/00268976.2024.2394132.

[43] Papaioannou V, Calado F, Lafitte T, Dufal S, Sadeqzadeh M, Jackson G, et al. Application of the SAFT-γ Mie group contribution equation of state to fluids of relevance to the oil and gas industry. Fluid Phase Equilib 2016;416:104–19. https://doi.org/10.1016/j.fluid.2015.12.041.

Spontaneous Non-Planar Material Assembly via Frontal Photopolymerisation

Isobel Bentley

Department of Chemical Engineering, Imperial College London, SW7 2AZ, UK

Abstract

This study investigates the potential of frontal photopolymerisation (FPP) for the spontaneous self-assembly of complex, three-dimensional structures from two-dimensional scaffolds. FPP creates polymerisation conversion gradients due to the exponential attenuation of light through a sample. These gradients impact the extent of shrinkage throughout the sample causing structural stresses, compensated for by spontaneous sample bending. Parameters affecting the process, including photoinitiator concentration, UV dosages and hinge dimensions, were systemically tuned to successfully fabricate a 6-sided, self-assembling cube. Due to the precise alignment of the faces and exact 90 degrees angles, a cube serves as a robust test case for the notion that FPP offers huge potential in the realm of complex shape manufacturing. This approach holds promise for various applications including microfluidics, microelectronics and soft robotics. Further research will focus on expanding the scope of self-assembling geometries as well as material detachment improvements with the use of substrates.

1 Introduction

Complex, non-planar geometries with high functionality have applications across various fields from microfluidics [1] and electronics [2] to photonics [3] and soft-robotics [4].

As an example, microfluidic technologies often require the fabrication of microchannels with intricate 3D structures to manipulate fluids at the submillimetre scale [1]. Similarly, microelectronics and chip fabrication demand highly precise 3D structures to ensure optimal chip functionality. Given the industry's ever-changing performance goals, emphasis must be placed on developing processes, like FPP, capable of producing a wide range of geometries [5].

Precision is not the sole challenge of material assembly. Other industries, such as 3D soft robotics, instead require a broader range of tolerances to adapt effectively to a wider variety of tasks. The assembly of such intricate structures requires controlled manufacturing techniques to accommodate their objectives [6]. Continuous innovation and advancement in fabrication methods are essential to meet the growing demand for more complex devices.

Frontal photopolymerisation uses UV light to polymerise a resin. The intensity of the light reduces as it travels through the material due to scattering and absorption [7]. This phenomenon is known as optical attenuation and causes the extent of polymerisation deeper in the sample to be less than that near the exposed surface [8]. Lambert-Beer's Law (see Equation 1) is used to describe the exponential attenuation of light intensity with depth in the resin. As a result of the polymerisation gradient (see Equation 4) a front is formed where the critical polymer conversion threshold, ϕ_c, is met. Polymer beneath this front remains liquid [8].

Lambert-Beer's Law:

$$I(z) = I_0 e^{-\mu z} \tag{1}$$

where I is the UV intensity at depth z ($\mathrm{mW/cm^2}$), I_0 is the UV intensity at $z = 0$ ($\mathrm{mW/cm^2}$), μ is the optical attenuation coefficient ($\mathrm{mm^{-1}}$) and z is the material depth (mm).

As resin converts to polymer, it shrinks; the extent of this shrinkage depends on the polymerisation conversion in that region. The difference in shrinkage with the polymerisation gradient generates internal stresses which result in sample bending [9]. By controlling the exposure of UV light to the liquid polymer, it is possible to induce controlled bending in specific regions of a sample, enabling the fabrication of hinges and other 3D structures. Materials that spontaneously bend could provide a solution to challenges in fabricating intricate geometries. Self-assembly will eliminate intricate assebley steps in manufacturing.

The goal of this paper is to investigate parameters that may affect the bending and demonstrate that, through adequate tuning, frontal photopolymerisation can be used to fabricate intricate, self-assembling, 3D shapes from 2D scaffolds. The formation of a cube serves as a test case for this concept as six faces must fold simultaneously and align at exact 90-degree angles. The parameters investigated include resin solu-

tions, UV dosages, hinge dimensions and the development procedure.

2 Background

Origami, the ancient art of folding paper, transforms two-dimensional sheets into intricate three-dimensional structures [10]. This concept has proven its versatility across various scales and serves as inspiration for this project and many other engineering solutions. At the largest scale, deployable devices exploit the efficiency and compactness of folded materials for storage and portability [11]. On the nanoscale, DNA origami enables the folding of a single long strand of DNA into complex three-dimensional shapes [12]. Both applications use principles of origami to enhance functionality of a minimised space.

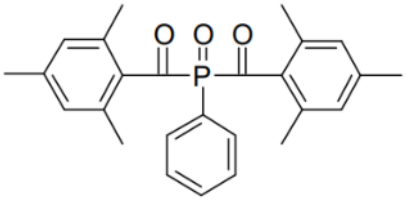


Figure 1: Molecular structure of Irgacure 819, the photoinitiator used in experiments [13]. It induces a high polymerisation initiation rate due to the generation of highly reactive phosphinoyl radicals upon UV exposure, which then go on to react with the acrylate double bonds in PEGDA [14].

Poly(ethylene glycol) diacrylate (PEGDA), mixed with the photoinitiator Irgacure 819 (see Figure 1) and exposed to UV light, has also been proven effective for origami-inspired applications and will serve as the resin in the following report [15]. The two acrylate groups in PEGDA (see Figure 4) enable free radical polymerisation. In this process photoinitiator is cleaved and the resulting free radicals attack the double bonds in the acrylate units, which go on to further react with other oligomers in a chain reaction, creating polymer chains. Cross-linking of polymer chains leads to the formation of complex polymer networks [16].

While origami has been successfully applied to PEGDA systems, current material assembly requires some manual manipulation after exposure. Spontaneous bending in these materials remains an area that has yet to be thoroughly explored.

3 Materials and Methods

3.1 Solution Preparation

Depending on the objective of each experiment, solutions of a specific PEDGA molecular weight, photoinitiator (Irgacure 819) concentration and SUDAN I concentration were prepared. SUDAN I is a dye added to a resin to help control the light penetration and therefore limit the cured depth of the resin after UV exposure [17]. The addition of SUDAN I is necessary to alter optical attenuation without changing other parameters.

Each sample was prepared in a 30ml vial, and the required weight of each component was calculated according to the necessary concentration. Mixtures were then set on a hotplate stirrer for 1 hour or until solid photoinitiator was no longer visible in the vial. To isolate samples from light exposure, solutions were wrapped in aluminium foil.

3.2 Standard Fabrication Procedure

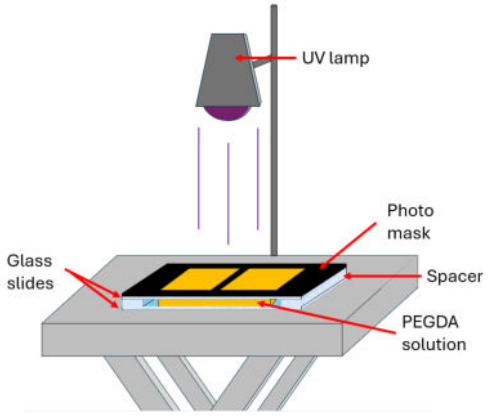


Figure 2: Experimental setup for the FPP standard fabrication procedure. The setup includes glass slides separated by spacers, with PEGDA solution forming a capillary in between the slides. The photomask controls regions of UV light exposure.

A standard method was used to create all polymer samples via frontal photopolymerisation. Using a clamp and clamp stand, a 200 W UV lamp (Omnicure s1500) was set up above an adjustable stand. The height and position of the stand were adjusted to ensure an approximately consistent UV exposure of 2.5 mW/cm^2 across the surface of the stand. Two glass slides, separated by spacers at either end, were then placed in the centre. Using a plastic dropper, the required PEGDA solution, prepared following instructions summarised in Section 3.1, was placed into the gap between the glass slides to create a capillary. The experimental set up needed to ensure that that the resin filled an area large enough that, when the photomask was placed on top of the glass slide, the exposed areas were filled with solution. The photomask, with cutouts specific to each experiment, remained on top of the slide whilst the required UV dosage was applied to the exposed resin (see Supplementary Information Section 1.2 for photomask designs).

The dosage is calculated using the following equation:

$$d = I \times t \tag{2}$$

where d is the UV dosage (mJ/cm^2), I is the UV light intensity (mW/cm^2) and t is time (s).

After the necessary dosage had been applied to the sample, it was rinsed with ethanol, to remove residue

liquid polymer, and dried with an air gun. This rinsing procedure was repeated twice.

3.3 Data Collection

3.3.1 Effect of Photoinitiator Concentration

Initial experiments tested the effect of photoinitiator (PI) concentration on the depth of the polymerisation front (measured by sample thickness), z_f, of resin exposed to various light dosages. A series of experiments were conducted with mixtures at 1:88, 1:107 and 1:248 PI:PEDGA mol/mol ratios for PEGDA 250, 575 and 700.

The standard procedure for polymer sample fabrication described in Section 3.2 was used for photoinitiator concentration experiments. Two samples were prepared for each light dosage at intervals between 10-700 mJ/cm^2. The height of each sample was then measured, using a digital caliper, and recorded.

3.3.2 Development Procedure Experiments

Two polymer samples were prepared as described in Section 3.2, each exposed to the constant irradiation dosage of 35 mJ/cm^2. However, one sample was rinsed with ethanol and left to bend without air drying. The other was developed using the standard method: rinsing with ethanol and drying with an air gun. Images were taken at regular time intervals for each sample and the curvature at each time stamp was recorded. Curvature is defined as the reciprocal of the radius of a circle that fits the curve of the sample.

3.3.3 Optimisation of Light Dosage for Hinge Fabrication

Here, a resin of PEGDA 700 with 1:88 mol/mol ratio PI:PEGDA is used. To control the monomer-to-polymer gradient of conversion, we added 0.05% w/w of the photoabsorber SUDAN I. A series of experiments were conducted to determine the optimum UV dosages at which a hinge of 0.1 cm width and square face length of 0.5 cm would be printed. The standard method of sample fabrication described in Section 3.2 was used with a two-step photomask to ensure correct UV exposure to each region of the hinge (see Supplementary Information Section 1.2 for photomask designs). The optimal dosage would produce a 90° right angle. For this to occur, the follow conditions had to be met:

- A hinge stiff enough to hold the weight of a folded side without collapsing, yet thin enough to still bend
- A face thin enough to minimise the weight and gravitational effect on the face, yet thick enough to remain flat

The faces of the hinges were exposed to a constant dosage of 90 mJ/cm^2. At high dosages the polymerisation conversion gradient diminishes and sample is more uniform, eliminating bending. An initial dosage of 19 mJ/cm^2 was applied to the hinge. This value was adjusted until the above conditions were met.

3.3.4 Effect of Dimensional Variability on Origami of Polymer

Using the optimal hinge dosage (29 mJ/cm^2) determined in Section 4.3, hinges of different dimensions were fabricated using the standard procedure (Section 3.2). Two layers of photomask were used; one of which was removed to expose the hinge once the face thickness had grown sufficiently (see Supplementary Information Section 1.2 for photomask designs). Hinges of 0.1 cm, 0.2 cm and 0.3 cm were tested with small, medium and large face sizes, of square side length 0.3 cm, 0.5 cm and 1 cm respectively. Images were taken of each sample after a controlled amount time (4 hours) and the spontaneous hinge bending angle for each sample was observed and plotted.

To validate that the optimal hinge width result held true for a box formation, identical experiments were then conducted for unlidded cubes (i.e. 5 faces) of hinge size 0.1 cm, 0.2 cm and 0.3 cm at the middle face size, 0.5 cm.

3.3.5 Fabrication of a Self-Assembled Non-Planar Material

A 6-sided cube is fabricated as a prototype to demonstrate a spontaneously self-assembled, non-planar material. To create a 6-sided cube at the optimal hinge width, 0.1 cm (see Section 4.4), and middle face size, 0.5 cm, the UV dosage at the 2-step folding hinge was further tuned as it held the weight of 2 faces. The 5 remaining hinges were exposed to the optimal dosage found in Section 4.3 (29 mJ/cm^2).

With knowledge from previous experiments, it was determined that the 2-step folding hinge required a greater UV dosage so it was stiffer and could withstand the weight of both faces without collapsing. Photomasks with varied opacity at that specific hinge were created and tested using the standard procedure (see Section 3.2) until a self-assembling, 6-sided box was produced. For photomask design methods, refer to Supplementary Information Section 1.2.1.

4 Results and Discussions

4.1 Effect of Photoinitiator Concentration

Experimental data shows that for any photoinitiator (PI) to PEDGA molar ratio, regardless of the PEGDA molecular weight, the thickness of the polymer sample produced increases linearly with irradiation dose in a log scale. This relation between dosage and sample thickness, z_f, is already well defined by the following equation:

$$z_f = \frac{1}{\mu}\left[\ln d + \ln K - \ln\left(\ln\left(\frac{1}{1-\phi_c}\right)\right)\right] \quad (3)$$

where μ is the attenuation coefficient (mm^{-1}), K is the conversion rate constant (cm^{2}mW^{-1}s^{-1}) and ϕ_c is the critical polymerisation conversion [9].

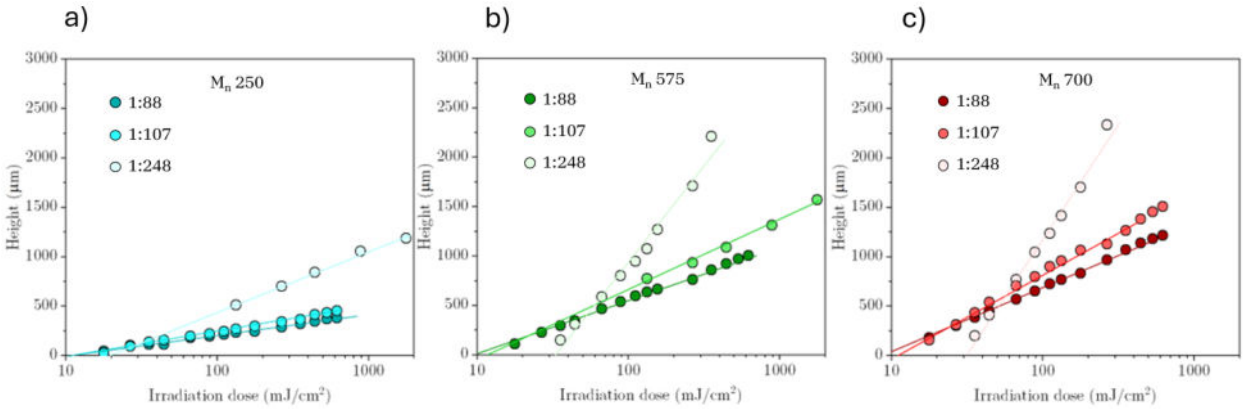


Figure 3: Polymerized height (mm) as a function of irradiation dose (mJ/cm^2) for three molecular weights (M_n): (a) M_n=250 g/mol (b) M_n=575 g/mol (c) M_n=700 g/mol. Data points are further categorized by photoinitiator (PI) concentration ratios (1:88, 1:107, 1:248), with lighter shades representing lower PI concentrations. Solid lines indicate linear fits. The plots show that lower PI concentration produce greater rates of change in height with UV dosage. However, lower PI concentrations are slower to start polymerisation and therefore a cross-over point is seen.

Moreover, for any singular PI mol/mol ratio, the greater the PEDGA molecular weight, the steeper the change in thickness is per unit increase in irradiation dose (see Figure 5). This may be explained by the greater effective size of the polymer network formed by the longer polymer chains found in a higher molecular weight liquid polymer solution. Polyethylene glycol diacrylate (PEGDA) consists of 2 acrylate groups and a number, n, of ethylene glycol units. The number, n, determines the length and molecular weight of the polymer. As n increases, the mesh size of the polymer network gets larger as bridging length between polymer chains increases [18]. This may result in physical expansion of the polymerised region in comparison to that of a shorter chain PEGDA solution and hence a greater measured polymer sample thickness [19]. In addition to this, the effect of molar weight also comes into play. Although the mol/mol ratio of PI:PEGDA in each solution is the same, due to the varying molecular weights of PEGDA, their mass ratios are not. The PEGDA 700 mixture has a greater mass of PI compared to the PEGDA 250 mixture, for example. The effects of a greater mass ratio of PI are discussed later (see Figure 6).

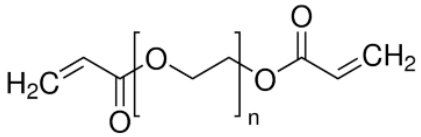


Figure 4: Molecular strucutre of PEGDA where n changes with molecular weight. PEGDA is susceptible to free radical polymerisation due to double bonds in acrylate groups at either end of molecule [20].

Figure 3 highlights the changes in the thickness of the polymer samples, z_f, for changing PI concentrations. Note that a higher PI:PEGDA ratio is a reduced concentration of PI. The gradients of the plots increase as the concentration is reduced. At the highest dosages, the sample with the greatest thickness has the lowest initial PI concentration. It appears that less PI is conducive to a greater polymer conversion rate at high dosages which seems counter-intuitive. It is suggested that by increasing the PI concentration, the absorption of light at the top of the sample is increased and therefore leaves less light to penetrate deeper into the sample. As a result, the critical polymerisation conversion, ϕ_c, is met at a shallower depth and the final thickness, z_f, is lower.

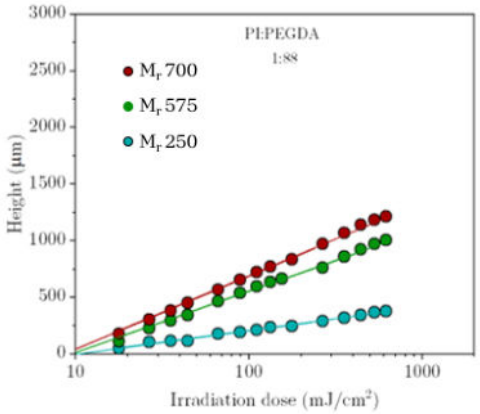


Figure 5: Variation of polymerised height as a function of irradiation dose on a log scale for 1:88 PI:PEGDA mol/mol ratio. Solid lines indicate linear fits to the data. The plot shows a logarithmic increase in height with higher irradiation doses. The rate of this of change of height with dosage increases with PEGDA molecular weight. For comparison of M$_n$ solution solutions at other PI concentrations see Supplementary Information Section 2.1.

To prove that UV absorption is greater when initial PI concentration is high, optical attenuation coefficients, μ, for each experimental setup (1:88, 1:107 and 1:248 PI:PEDGDA mol/mol ratios for PEGDA 250, 575 and 700) were calculated in Table 1. This was done using the well-defined relation between sample thickness, z_f, and dosage (see Equation 3) and gradients of the height vs. irradiation dose in a log scale (see Figure 3). Typically, at high dosages, the value of μ changes, however for these calculations,

it is taken to be constant across the UV dosage range tested.

See Supplementary Information Section 1.1 for experiments conducted to verify that the UV dosage range used do not result in significant bleaching. μ remains approximately constant for a single solution regardless of the irradiation dose within the experimental range.

Table 1: μ values corresponding to different PI:PEGDA ratios for varying M_n of PEGDA. μ is taken as a constant for each set of conditions.

	PI:PEGDA Ratio		
M_n of PEGDA	1:88	1:107	1:248
250	10.93	8.79	3.72
575	4.37	3.23	1.18
700	3.41	2.69	0.99

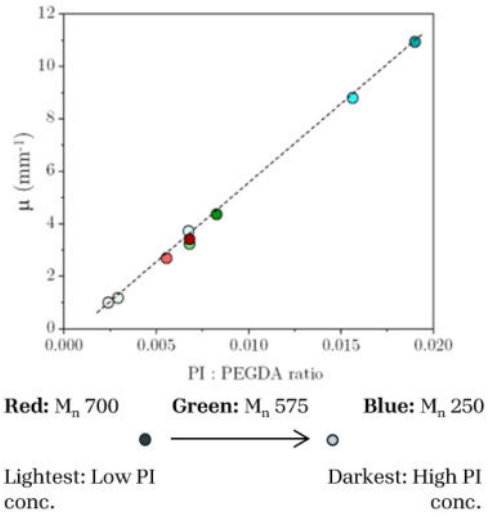


Figure 6: Optical attenuation coefficient (μ) as a function of PI:PEGDA mass ratio. Data points are colour-coded according to molecular weight: red (M_n=700), green (M_n=575) and blue (M_n=250). Marker darkness indicates the PI concentration, with darker shades representing higher PI concentrations. A linear trend is seen; as PI mass concentration increases,the light attenuation increases linearly.

μ values are then plotted against the PI:PEGDA mass ratio for each solution (see Figure 6). This differs from the typical ratios used in Figure 3 which are on a molar basis. The plot shows a positive correlation: as PI:PEGDA mass ratio increases, the optical attenuation coefficient, μ, increases linearly. For any single PI:PEGDA mass ratio, μ is constant. The optical attenuation coefficient quantifies the extent of attenuation or weakening of light when passing through a medium [21]. Increasing PI increases the extent of UV absorption as there is more PI to be cleaved. Therefore, changing PI concentration directly affects μ, as seen in Figure 6. This provides evidence for the theory that the high PI concentration solutions do not produce the thickest polymer samples because the extent of light absorption is so great, that the depth of light penetration is reduced.

From the original height vs. dose plots for changing PI concentration (3), it is also observed that the low concentration solutions are initially slower. The dosage at which critical conversion is first met, and the liquid polymer solidifies, is greater. As a result, a cross over point is seen. It is inferred from this data that there are competing effects of increasing PI concentration. At low dosages, resins with low PI concentration perform poorly (that is, sample thickness is low), yet as the dosage is increased the thickness grows at a greater rate when compared to solutions with a high PI concentration. To further explore this, the polymerisation rate constant, K, for each resin solution is investigated.

Using intercept values for each plot in Figure 3, Equation 3 and the respective μ value, the conversion rate constant, K, is calculated. Molecular weight is plotted against conversion rate constant in Figure 7.

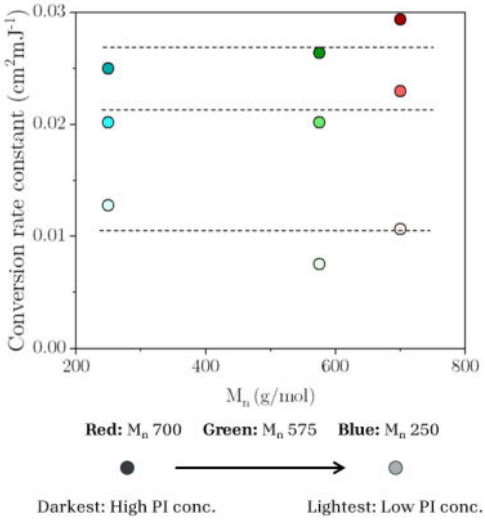


Figure 7: Effect of molecular weight (M_n) and photoinitiator concentration on the conversion rate constant, K (cm^2mJ^{-1}). Data points are colour-coded: red (M_n=700), green (M_n=575) and blue (M_n=250). Darkness of markers corresponds to PI concentration, with darker shades indicating high PI concentrations. The plot shows that changing the PI concentration has a greater effect on K than changing the M_n of PEGDA. K increases as PI concentration increases.

Each PEDGA solution shows an increase in K with increased PI concentration. This outcome is explained by the underlying free radical photopolymerisation reaction mechanism. Higher PI concentration results in generation of more radicals per unit of light intensity. Hence, propagation occurs faster and the polymer conversion rate increases.

In summary, increasing the PI concentration increases the conversion rate constant, K. It also increases the optical attenuation coefficient, μ, which means the extent of UV light penetration is reduced. These two effects are competing. At low UV doses, K,

is dominant and therefore, solutions with high PI concentration produce thicker samples. However, as the dosage is increased, the effect of a higher μ value in high PI concentration solutions dominates. At these doses, a low PI concentration leads to thicker polymer samples; greater z_f. Samples with a larger z_f value indicate that a greater proportion of the initial liquid polymer now has polymerisation conversion greater than critical conversion, ϕ_c. It is likely that the conversion profiles throughout samples of different initial PI concentration would vary largely. At a high dosage, it is likely that the conversion, ϕ, at the top of sample (where z = 0), is greater if the initial PI concentration of the resin is greater. However, the thickness, z_f, would be smaller than a sample produced with a solution of a low PI concentration. The polymerisation conversion, at conditions were the optical attenuation coefficient μ does not change, can be defined by Equation 4.

$$\phi(z,t) = 1 - exp[-KI_0 exp(-\mu z)t] \tag{4}$$

While it is evident that PI concentration affects the conversion profile in a polymer sample, this topic is beyond the scope of this report.

A final observation can be seen from Figure 7. As there is is no difference between solutions of the same PI concentration and varied molecular weights, it can be concluded that K is highly dependent on the PI:PEGDA mol ratio in the initial solution, but less so on the PEGDA molecular weight. For any singular PI:PEGDA mol ratio, the conversion rate constant, K, is approximately constant. This helps to explain the trend seen earlier in Figure 5; increasing M_n of a PEGDA mixture increases the samples thickess. Since the PI:PEGDA mol ratio between solutions is kept constant, K is not changing. However, because the mass ratio is changing, μ is also changing, but now there is no competing affect of K. μ increases as the mass concentration of PI increases. This explains why PEGDA 700 demonstrated the thickest final sample when compared to other mixtures of different molecular weight at the same PI:PEGDA mol/mol ratio; the mass of PI was the lowest and therefore so was the optical attenuation coefficient, μ. This allowed for deeper UV penetration and a thicker sample.

It is decided that the subsequent non-planar material assembly is to be conducted with the PEGDA 700 liquid polymer. Firstly, this solution has the greatest viscosity, making the insertion of the fluid between the glass slides easier. Also, the irradiation time required to create the necessary sample thickness is lower for the higher molecular weight PEGDA solution. Therefore, choosing PEGDA 700 improves time efficiency.

Solutions for material assembly will be prepared in a 1:88 molar ratio of PI:PEGDA. The highest PI concentration is selected because the thickness will not change significantly if the dosage is slightly off. This is seen by shallow gradients in Figure 3 for high concentration solutions. Moreover, since the PEGDA 700 solution is selected as the liquid polymer, which tends to fabricate thicker sample per unit time, the high PI concentration ratio is used to ensure that samples remain thin enough to bend.

4.2 Development Procedure Experiments

Images taken at identical timestamps are compared for a sample solely rinsed with ethanol and a sample rinsed with ethanol and then dried with air gun. From the images (see Supplementary Information Section 2.2), it is observed that the addition of air drying results in a faster rate and a greater degree of sample bending. As for the change in quantified curvature with time for different development procedures, Figure 8 shows that ethanol with airdrying has a greater initial gradient (i.e. greater curvature per unit time). It also produces a larger final value of curvature, where curvature is defined as the reciprocal of the radius of a circle that fits into the sample bending.

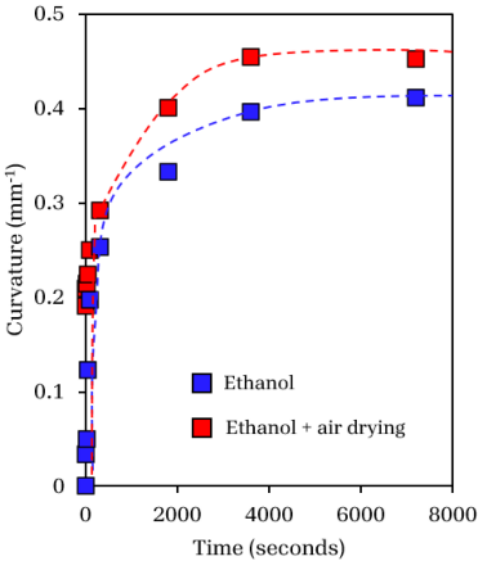


Figure 8: Curvature (mm^{-1}) of polymer sample with changing development procedure. The addition air drying increases the rate of curvature and the extent of curvature after a 7200 second period (2 hours)

Is it proposed that when the sample is rinsed, the unreacted, residue liquid polymer particles are deabsorbed and replaced with ethanol molecules. This occurs mainly at the polymerisation front of the sample due to the gradient of absorbing capacity as a result of the polymerisation gradient. At the polymerisation front, the degree of polymerisation is lowest due to the lower density crosslinked structure and hence larger pores that can accommodate ethanol [22]. It is suggested that when these ethanol particles then evaporate. The sample compensates by experiencing compressive stress, especially at the front. This phenomenon causes the polymer to bend. The extent and

rate of this evaporation is increased when the sample is dried with an air gun and therefore the extent and rate of bending increases also.

4.3 Optimisation of Light Dosage for Hinge Fabrication

Figure 9 shows the variation in the hinge bending as the hinge irradiation dosage is tuned. The hinge size is maintained at 0.1 cm with face size of 0.5 cm and the face dosage is also constant at 90 mJ/cm^2. See Supplementary Information Section 2.3 for images of samples at varied hinges dosages.

An optical dosage is the value at which a 90° right-angle is reached. This is found at a hinge dosage of approximately 29 mJ/cm^2. Either side of this optimum, the hinge flattens. From this data it is inferred than some equilibrium has been met. There is a balance between the effect of gravity felt on the face and the torsional resistance of the hinge. At lower dosages, the hinge is not stiff enough to hold the weight of the face, at higher dosages the hinge becomes too stiff to adequately bend.

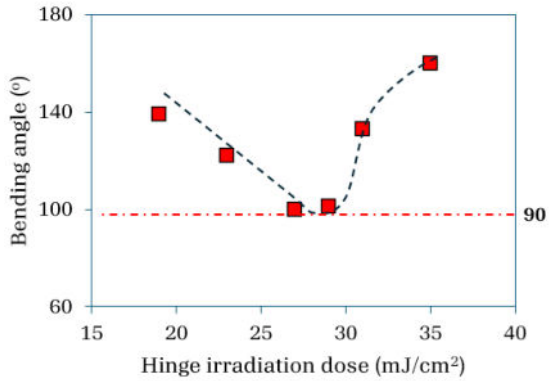


Figure 9: Variation in bending angle of hinge (°) with hinge dosage (mJ/cm^2). The dosage applied to the face is kept constant at 90 mJ/cm^2. A 90° angle is met at 29 mJ/cm^2 and this dosage is taken as the optimum. Either side of this the hinge flattens, due to dominant effect of gravity.

4.4 Effect of Dimensional Variability on Origami of Polymer

Figure 10a shows a grid of each square side length with the tested hinge widths. UV dosages were maintained at 90 mJ/cm^2 for the faces and 29 mJ/cm^2 for the hinge. Bending angles for each sample were measured and plotted in Figure 10b.

0.1 cm and 0.2 cm hinges show little variation in their bending angles, regardless of the square face size. So, it is concluded that the extent of bending has less dependence on the square side length and is more dependent on the hinge width. For these hinge dimensions, the bending angle remains approximately constant for each side size. This does not hold true for the 0.3 cm hinge at a 29 mJ/cm^2 dosage; it showed great variation in bending angle with square side length.

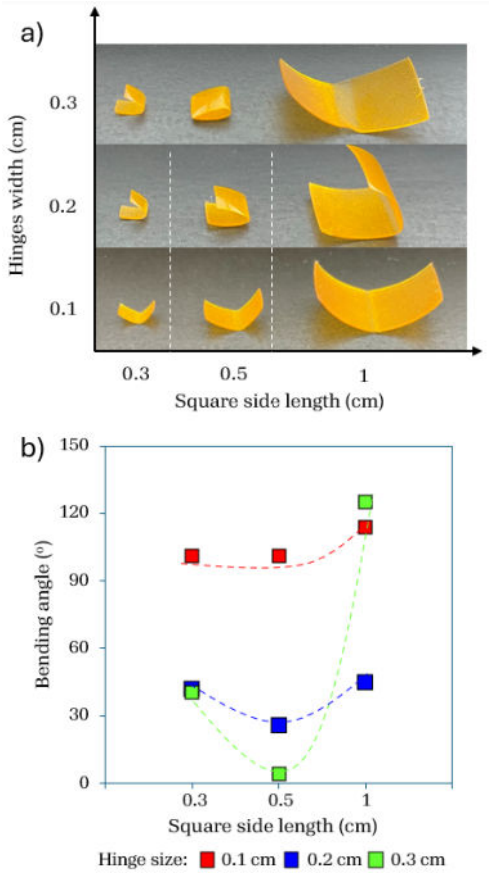


Figure 10: (a) Grid of hinge images with varied hinge and face dimensions. 0.1, 0.2 and 0.3 cm hinges with 0.3, 0.5 and 1 cm square side length faces are tested. Different bending regimes are observed: controlled, gravity-dominant and gravity failure (b) Bending angles of hinges seen in Figure 10a (°) with square side length (cm). Data points are colour-coded according to hinge size: red (0.1 cm), blue (0.2 cm), green (0.3). 0.1 and 0.2 cm hinges show little variation in bending, 0.3 has no observable trend.

It is worth noting that the intervals at which the square side lengths were taken were not even; there is a significantly smaller difference between 0.3 cm and 0.5 cm compared to 0.5 cm and 1 cm. For the 2 smaller face sizes, 0.3 and 0.5 cm, the hinge bending angle increases with decreasing hinge width. For the largest face size, this correlation is not observed. It appears that the face is too heavy for the hinge to bend and a threshold for failure has been met.

The 0.1 cm hinge consistently reached an angle closest to a 90° right-angle and is deemed optimal. Although it is the smallest, this hinge is able to withstand gravity, regardless of face sizes and weights tested. It may then be possible to categorise the hinges depending on their behaviour at the constant dosage. The following regimes are proposed.

A hinge is in a **controlled regime** if the bending angle hardly changes with face size. Within a particular face size range, there is a single intrinsic bending

angle that a hinge of a specific size reaches at a defined dosage. This regime is observed for the 0.2 cm hinge widths as there is no considerable change in bending angle throughout the samples. Should the hinge bend and then be dominated by gravity, causing it to produce an angle that is not the intrinsic angle, the hinge is said to be **gravity dominated**. This is seen for the 0.1 cm hinge with a square side length of 1 cm. If the effect of gravity greatly outweighs the bending torsion and the hinge either does not bend or collapses completely, it said to be a **gravity failure**, as seen for the 0.3 cm hinge with 1 cm face size. For this dosage the intrinsic bending angle of 0.1 cm hinge is the closest to 90°.

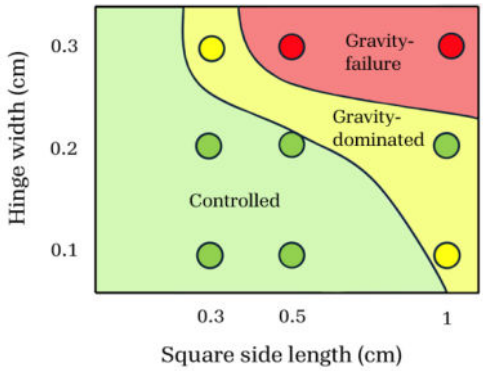


Figure 11: Morphology map representing regions of different bending regimes for hinges of varying dimensions. Circles represent experimental data points (see Figure 10a): green is the controlled regime, yellow is the gravity dominated regime, red is the gravity failure regime. The background colours represented proposed thresholds for each regime.

In Figure 11, circles represent experimental data points where their colour shows the regime that was experimentally observed. The background colours show estimates of regions of different regimes. It is proposed that each hinge width has an intrinsic angle for a specific dosage. If the data point falls within the controlled regime region (the green region), the hinge will demonstrate the intrinsic angle for that hinge size at that dosage. Yellow regions are gravity dominated and therefore do not produce the bending angle intrinsic to the hinge width at that dosage. Red represents gravity failure, whereby samples do not bend or completely collapse due to gravity dominance. The existence of these regions is purely theoretical and has not been conclusively proven. The 1 cm side size with a 0.2 cm hinge width indicates that there may be exceptions to the trends. It demonstrated a controlled regime yet falls within the gravity dominated region. To validate this theory, many more experiments would have to be conducted with varying hinge sizes, face sizes and dosages.

5-sided box experiments showed similar trends (see Figure 12). At the optimum dosage from Section 4.3, the 2D scaffold with 0.1 cm hinge width spontaneously produces a lidless box. This further verifies that the

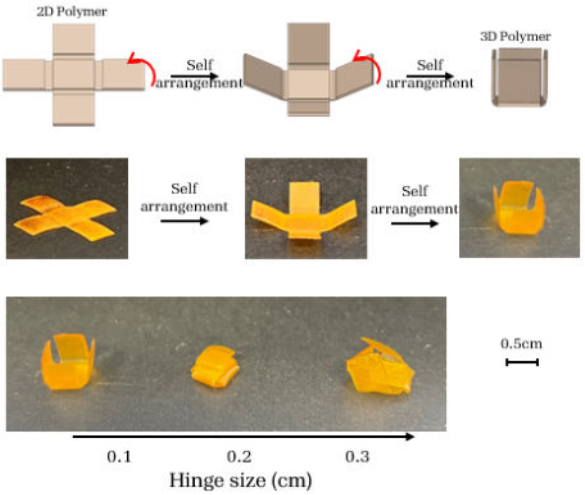


Figure 12: Self-assembly process of a 2D polymer into a 3D lidless box structure. The top panel illustrates the expected transformation from a flat 2D polymer sheet into a 3D cube structure via self-arrangement. The middle panel shows the actual sequential stages of self-arrangement for a 0.5 cm square side length and 0.1 cm hinge width structure. The bottom panel displays the resulting 3D structures at hinge sizes of 0.1 cm, 0.2 cm, and 0.3 cm, demonstrating the influence of hinge dimensions on the folding outcome. Scale bar: 0.5 cm.

0.1 cm hinge produces an angle closest to 90° when using a dosage of 29 mJ/cm^2. At 0.2 cm and 0.3 cm hinges, the boxes collapsed as expected from Figure 10. At these dimensions, the effect of gravity on the faces of the boxes outweighs the stiffness of the hinges.

4.5 Fabrication of a Self-Assembled Non-Planar Material

The tuning of opacity of the 2-step folding hinge portion of the photomask is conducted using educated trial-and-error attempts. Due to the success of the 0.1 cm hinge width in Section 4.4, this dimension, with a middle face size of 0.5 cm, is chosen for the 6-sided mesoscale box. The five hinges holding a single face are irradiated with a 29 mJ/cm^2 dosage, and the faces all had 90 mJ/cm^2 dosages. It is found that at a dosage of 35 mJ/cm^2 on the two-step folding hinge, a box with a lid spontaneously self-assembles (see Figure 13).

Increasing the UV dosage increases the thickness of the polymer in this region. Therefore, the stiffness of the hinge also goes up. At 35 mJ/cm^2 the 2-step folding hinge it able to hold the weight of 2 faces, whilst still being able to bend. An equilibrium has been reached between the bending torsion at the hinge and the effects of gravity on the faces, which is now greater as there are 2 involved.

5 Conclusion

The conclusions of this study can be drawn on two levels: those specific to the fabrication of a 6-sided mesoscale cube from a liquid PEGDA solution and

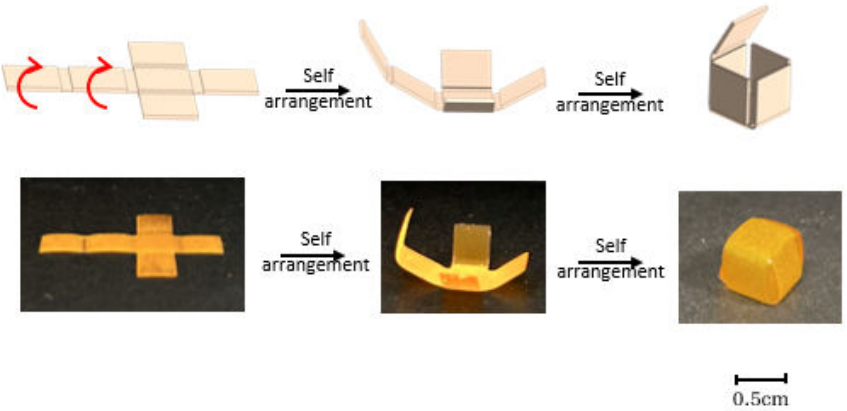


Figure 13: Self-assembly process of a 2D polymer into a 3D cube structure. The top panel illustrates the expected transformation from a flat 2D polymer sheet into a 3D enclosed cube structure via self-arrangement. The middle panel shows the actual sequential stages of self-arrangement for a 0.5 cm square side length and 0.1 cm hinge width structure. Scale bar: 0.5 cm.

those more broadly applicable to spontaneous, non-planar material assembly.

By systematically optimising hinge widths, cube face length and UV irradiation dosages, we demonstrate the successful fabrication of a self-assembling six-sided cube. An intriguing correlation was observed between photoinitiator (PI) concentration and the resulting sample fabrication thickness. While a higher PI concentration increases the number of free radicals and enhances the conversion rate conversion constant, K, it also introduces a competing effect due to increased optical attenuation, μ. Higher μ values are observed as PI concentration is increased which means the extent of UV light weakening through the sample is greater. Penetration deeper into the sample is reduced as a result. At low UV dosages, the effect of a greater K value dominates and thicker samples (i.e. greater z_f values) are produced with higher initial PI concentration solutions. As the UV dosage increases, the effect of a greater μ value in high PI concentration samples dominates, and thicker samples are produced if the concentration of PI is lower in the original resin.

Optimal hinge performance was achieved with a UV dosage of 29 mJ/cm^2 for single-face hinges and 35 mJ/cm^2 for the two-step folding hinge. These dosages ensured sufficient stiffness to bear the weight of faces while maintaining the flexibility required for folding. At 29 mJ/cm^2 narrower hinges, specifically those of 0.1 cm width, consistently produced 90° bending angles spontaneously, facilitating proper self-assembly. Wider hinges were less effective, particularly under increased face weights, where gravitational forces outweighed the bending torsion.

Hinge bending was separated into 3 regimes:

- Controlled regime: the hinge folds to the intrinsic bending angle dependent on the hinge width and irradiation dosage, yet independent of face size.
- Gravity-failure regime: the hinge is unable to fold or collapses due to the dominance of gravitational forces.
- Gravity-dominated regime: the hinge bends to an angle that is not its intrinsic angle as it has been dominated by the effect of gravity.

In theory, a complex shape can be scaled up or down whilst maintaining a constant hinge size, provided it remains within the controlled regime. This study concludes that, from a design perspective, scaling considerations must account for not only UV dose, optical attenuation and other operating parameters but also the dimensions of the structure and bending regimes to ensure reliable self-assembly.

More generally, this study demonstrates the feasibility of spontaneous origami through frontal photopolymerisation, proving that self-assembling three-dimensional structures can emerge directly from planar polymer precursors. The successful fabrication of a six-sided box serves as a robust test case for this concept, as its construction requires precise alignment of edges and hinges to form a fully enclosed structure. By demonstrating the self-assembled fabrication of a box, we validate the methodology and highlight the potential of this approach for creating complex geometries with high accuracy spontaneously.

Looking forward, exploring a suitable coating for the glass slides could significantly enhance the fabrication process. While the polymer sample bends spontaneously once detached, the current method requires scraping with a scalpel to overcome adhesion to the glass surface. Developing a non-stick coating or alternative substrate could reduce damage to the polymer samples and improve overall efficiency.

Further research should focus on the effects of scaling up and down the system. While, at certain dimensions the hinge bending angles appear to remain

unaffected by the weight they lift, the limits of this weight independence need to be investigated thoroughly. Greater understanding of this range will enable more precise control over hinge and bending behaviour and broaden the scope of applications.

Acknowledgment

The author of this report would like to extend their utmost gratitude to Ghifari Ridwan for his efforts and invaluable support throughout this project.

References

[1] L. C. Duarte, F. Figueredo, C. L. S. Chagas, E. Cortón, and W. K. T. Coltro, "A review of the recent achievements and future trends on 3D printed microfluidic devices for bioanalytical applications," *Analytica Chimica Acta*, vol. 1299, p. 342429, Apr. 2024.

[2] D. Karnaushenko, T. Kang, V. K. Bandari, F. Zhu, and O. G. Schmidt, "3D Self-Assembled Microelectronic Devices: Concepts, Materials, Applications," *Advanced Materials*, vol. 32, no. 15, p. 1902994, 2020. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/adma.201902994.

[3] P. Lalanne, W. Yan, K. Vynck, C. Sauvan, and J.-P. Hugonin, "Light Interaction with Photonic and Plasmonic Resonances," *Laser & Photonics Reviews*, vol. 12, no. 5, p. 1700113, 2018. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/lpor.201700113.

[4] F. Iida and C. Laschi, "Soft Robotics: Challenges and Perspectives," *Procedia Computer Science*, vol. 7, pp. 99–102, Jan. 2011.

[5] M. Datta, "Electrochemical processing technologies in chip fabrication: challenges and opportunities," *Electrochimica Acta*, vol. 48, pp. 2975–2985, Sept. 2003.

[6] S. Coyle, C. Majidi, P. LeDuc, and K. J. Hsia, "Bio-inspired soft robotics: Material selection, actuation, and design," *Extreme Mechanics Letters*, vol. 22, pp. 51–59, July 2018.

[7] E. T. Baker and J. W. Lavelle, "The effect of particle size on the light attenuation coefficient of natural suspensions," *Journal of Geophysical Research: Oceans*, vol. 89, no. C5, pp. 8197–8203, 1984. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/JC089iC05p08197.

[8] J. T. Cabral, S. D. Hudson, C. Harrison, and J. F. Douglas, "Frontal Photopolymerization for Microfluidic Applications," *Langmuir*, vol. 20, pp. 10020–10029, Nov. 2004. Publisher: American Chemical Society.

[9] J. Wang, B. Zhang, P. Lei, and N. Dai, "Photo-induced Bending and Buckling of Polymer Sheets," *IOP Conference Series: Materials Science and Engineering*, vol. 493, p. 012029, Mar. 2019. Publisher: IOP Publishing.

[10] M. Meloni, J. Cai, Q. Zhang, D. Sang-Hoon Lee, M. Li, R. Ma, T. E. Parashkevov, and J. Feng, "Engineering Origami: A Comprehensive Review of Recent Applications, Design Methods, and Tools," *Advanced Science*, vol. 8, no. 13, p. 2000636, 2021. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/advs.202000636.

[11] P. P. Pratapa and A. Bellamkonda, "Thick panel origami for load-bearing deployable structures," *Mechanics Research Communications*, vol. 124, p. 103937, Sept. 2022.

[12] B. Saccà and C. M. Niemeyer, "DNA Origami: The Art of Folding DNA," *Angewandte Chemie International Edition*, vol. 51, no. 1, pp. 58–66, 2012. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/anie.201105846.

[13] C. S. C. , "Ciba IRGACURE 819 Photoinitiator," technical Data Sheet, Basle, Switzerland, Aug. 2001.

[14] R. Tripathy, J. V. Crivello, and R. Faust, "Photoinitiated polymerization of acrylate, methacrylate, and vinyl ether end-functional polyisobutylene macromonomers," *Journal of Polymer Science Part A: Polymer Chemistry*, vol. 51, pp. 305–317, Jan. 2013. Publisher: John Wiley & Sons, Ltd.

[15] Z. Zhao, J. Wu, X. Mu, H. Chen, H. J. Qi, and D. Fang, "Origami by frontal photopolymerization," *Science Advances*, vol. 3, p. e1602326, Apr. 2017. Publisher: American Association for the Advancement of Science.

[16] M. Levin, Y. Tang, C. D. Eisenbach, M. T. Valentine, and N. Cohen, "Understanding the Response of Poly(ethylene glycol) diacrylate (PEGDA) Hydrogel Networks: A Statistical Mechanics-Based Framework," *Macromolecules*, vol. 57, pp. 7074–7086, Aug. 2024. Publisher: American Chemical Society.

[17] S. Pal and A. S. K, "Biodegradable, Solvent-Free Photocrosslinkable PLLA Resin Formulations for Additive Manufacturing," *Macromolecular Chemistry and Physics*, vol. 223, p. 2200139, Oct. 2022.

[18] A. E. Ross, M. Y. Tang, and R. A. Gemeinhart, "Effects of Molecular Weight and Loading on Matrix Metalloproteinase-2 Mediated Release from Poly(Ethylene Glycol) Diacrylate Hydrogels," *The AAPS Journal*, vol. 14, pp. 482–490, Apr. 2012.

[19] C. W. H. Rajawasam, O. J. Dodo, M. A. S. N. Weerasinghe, I. O. Raji, S. V. Wanasinghe, D. Konkolewicz, and N. D. A. Watuthanthrige, "Educational series: characterizing crosslinked polymer networks," *Polymer Chemistry*, vol. 15, no. 4, pp. 219–247, 2024. Publisher: Royal Society of Chemistry.

[20] F. Ayhan and H. Ayhan, "Functional PEG Macromers for Biomedical Applications," *JOURNAL OF COMPOSITES AND BIODEGRADABLE POLYMERS*, vol. 5, pp. 48–61, Jan. 2019.

[21] S. Chang and A. K. Bowden, "Review of methods and applications of attenuation coefficient measurements with optical coherence tomography," *Journal of Biomedical Optics*, vol. 24, p. 090901, Sept. 2019.

[22] R. Nigmatullin, M. Bencsik, and F. Gao, "Influence of polymerisation conditions on the properties of polymer/clay nanocomposite hydrogels," *Soft Matter*, vol. 10, no. 12, p. 2035, 2014.

Optimisation of hard carbon-based cathodes for sodium-ion batteries through varying binder content

Yujie Zou and Gergely Hajdu
Department of Chemical Engineering, Imperial College London, U.K.

0. **Abstract** Sodium-ion batteries are a promising emergent technology, offering an alternative to lithium-ion batteries due to their lower cost, toxicity and abundance of starting materials. Instead of traditional binders, we used green binders which are soluble in water. In this work, the effects of varying the ratios of binder and active material for cathodes were investigated by building and characterising coin cells. Mixtures of the active material (hard carbon, HC) and binders (CMC and SBR) were prepared as an aqueous slurry, with compositions ranging from 80-95% HC, 1-5% CMC and 0-2% SBR. The cathode slurry was applied on an aluminium current collector with a coating thickness of 200 μm. Coin cells, with elemental sodium anode and NaPF6 electrolyte, were assembled and tested for specific current density, voltage, current, and internal coulombic efficiency with respect to time and charging rates. It was found that a cathode with a binder content of 5% CMC and a HC: Binder ratio of 90:10 possesses the highest overall specific capacity.

1. Introduction

Efficient energy storage is vital in the modern aspects of life, with modern battery technology becoming widespread and ubiquitous, powering nearly every portable device used today [1]. It is also important in utilising renewable energy, helping to balance out its intermittent nature. The most common battery chemistry being used today is lithium-ion technology, offering high energy density, low weight and a long lifespan [2]. However, the disadvantages are the need for rare metals, fire hazards and toxic components [1] [2]. Sodium-ion batteries offer lower cost, more abundant starting materials, and a more environmentally friendly production [2].

In this research, sodium-ion coin cells were made and tested, with a battery chemistry consisting of elemental sodium anode and hard carbon cathode. This differs from most other battery chemistries, where carbon acts as an anode, as in this case the energy is provided via metal plating/stripping, while in other cases it happens via intercalation/de-intercalation

Batteries are commonly characterised by graphs of voltage-current, capacity-voltage and cycle efficiency-specific capacity. The objective of the research was to find the cathode composition that offered the best overall performance based on these indicators.

2. Background

2.1 Sodium-ion Batteries (SIBs)

Sodium-ion batteries (SIBs) have gained considerable attention as a viable alternative to lithium-ion batteries (LIBs) due to the abundance, low cost and high availability of sodium compared to lithium. Although SIBs offer a promising solution for large-scale energy storage, they suffer from limitations such as lower energy density and shorter cycle life than LIBs. The primary challenge lies in optimizing the materials and components to achieve performance that meets or exceeds that of LIBs. Sodium ions, being larger and heavier than lithium ions, can cause more significant volume changes during cycling, leading to capacity fading and lower overall efficiency [3].

One of the most crucial factors in improving the performance of SIBs is the choice of electrode materials. Hard carbon has emerged as a leading candidate for the anode material in SIBs due to its excellent cycling stability, high capacity, and low cost. KT2-009 hard carbon, in particular, has been shown to exhibit good electrochemical properties, making it a suitable anode material for sodium-ion systems. Unlike graphite, which is commonly used in LIBs, hard carbon has better structural stability when intercalating sodium ions, making it a favourable option for SIBs [4].

2.2 Binders and Their Role in Electrodes

Binders are essential components in the fabrication of battery electrodes. They help bind the active material to the current collector and maintain the structural integrity of the electrode during cycling. The binder material also influences the electrode's mechanical properties, flexibility, and conductivity. Traditional binders such as polyvinylidene fluoride (PVDF) are often used in LIBs but are less environmentally friendly and require toxic organic solvents. As a result, there has been growing interest in aqueous-based binders, such as carboxymethyl cellulose (CMC), which is a water-soluble polymer that provides good adhesion, while offering environmental and safety benefits [5].

To further improve the mechanical properties and adhesion of the binder, styrene-butadiene rubber (SBR) is often combined with CMC. SBR improves the flexibility of the binder and enhances the overall performance of the electrode by ensuring better adhesion and reducing brittleness [6], particularly during charge/discharge cycles where the electrode material experiences volume changes. According to research by Müllner et al. [7], incorporating SBR into CMC-based binders improved the adhesion properties and cycling stability of electrodes in lithium-ion batteries.

2.3 Previous Studies on Binder Compositions

Several studies have investigated the impact of binder compositions on the performance of various battery systems, focusing on the balance between CMC and SBR. Zhang et al. [6] conducted a comprehensive study comparing different ratios of SBR and CMC in lithium-ion batteries, particularly in anodes made from $ZnFe_2O_4$. Their results indicated that a balanced combination of CMC and SBR improved both the mechanical strength and electrochemical properties of the electrode, enhancing its long-term cycling stability and rate performance. Furthermore, they found that the binder's ratio affected the ionic conductivity and the formation of a stable solid electrolyte interphase (SEI) layer, which is critical for preventing capacity degradation over time.

The combination of CMC and SBR has also been studied in LIBs with anodes made from silicon materials.

Research showed [6] that a 1:1 ratio of CMC and SBR resulted in the highest discharge capacity and the best cycling stability, demonstrating the potential of this binder system to improve battery efficiency. Additionally, the mechanical properties of the binder significantly influence the structural integrity of the electrode, reducing electrode cracking and capacity fading during prolonged cycling.

2.4 Gap in Research and Study Objectives

While much has been learned from the use of CMC and SBR binders in lithium-ion batteries, there is still a limited understanding of how different compositions of these binders affect the performance of sodium-ion batteries, especially when paired with hard carbon anodes such as KT2-009. Much of the existing research focuses on binder systems used in LIBs, leaving a gap in knowledge regarding their applications in SIBs. This research aims to address this gap by exploring the impact of various CMC/SBR binder ratios on the electrochemical performance of sodium-ion batteries using KT2-009 as the anode material.

By investigating different binder compositions, this study seeks to optimize the performance of SIBs, contributing to the development of more sustainable and cost-effective energy storage systems for future applications in electric vehicles and large-scale energy storage.

3. Methodology

The electrode preparation process involves several key steps to ensure the proper assembly and performance of the coin cells. Initially, the required binder composition is prepared and mixed with hard carbon to form a uniform slurry. This slurry is then coated onto the current collector, dried, and subsequently cut into appropriately sized squares to fit inside a coin cell. The following sections provide a detailed explanation of each step in this process.

3.1 Materials Preparation: Binder and Slurry

3.1.1 Binder preparation

Two types of binder were used in the experiment: one was a pure 5 wt% Carboxymethyl Cellulose (CMC) of a molecular weight of 250 000, sodium salt of a degree of substitution of 1.2 solution, and the other was a mixture of CMC and Styrene-Butadiene Rubber (SBR) solution. To prepare the mixed binder, aqueous solutions of 5 wt% CMC and 48 wt% SBR were used. The required amounts of

CMC solution, SBR solution, and deionised (DI) water were weighed into glass vials to make approximately 20 g of binder mixture. A magnetic stir bar was added to each vial, and the contents were mixed overnight using a magnetic stirrer to achieve a homogeneous solution.

3.1.2 Slurry preparation

To prepare the slurry, approximately 1 g of KT2-009 hard carbon and the corresponding type and amount of binder solution were weighed into a container. The components were then mixed in a centrifugal planetary mixer for 5 minutes at 2000 rpm. Care was taken to keep the hard carbon in a sealed container, as it readily adsorbs moisture from the air. The finished slurry mixture was visually inspected to confirm uniform mixing. The ratios of hard carbon to binder with different binder compositions are demonstrated in Tables 3-1 and 3-2 below.

Table 3-1: Slurry with Varying HC-Binder Ratios, CMC only

HC: Binder (wt%)	Binder Ratio - CMC only (wt%)
95:5	5%
90:10	
80:20	

Table 3-2: Slurry with Varying HC-Binder Ratios and Varying CMC & SBR Ratios

HC: Binder (wt%)	95:5	90:10	85:15
Binder Ratio - CMC: SBR (wt%)	4%: 1%		
	3%: 2%		
	2.5%: 2.5%		

3.2 Coating the Slurry

The current collector used was a 12 µm thick aluminium foil. Initially, the original carbon coating on the foil was removed by rubbing it with a wet paper towel. The cleaned aluminium foil was placed on a glass plate, and the prepared slurry was applied onto one side using a four-sided slurry applicator, achieving a layer thickness of 200 µm. The coated foil was left to dry in the air initially. Once sufficiently dry, the foil was sandwiched between wax paper and placed with the coating facing down onto a glass plate.

3.3 Electrode Cutting and Preparation

The dried coated foil was covered in weighing paper on both sides to protect it and was cut into 8mm squares using a scalpel and a metallic guide grid with the coating facing down, ensuring that the cutting process did not cause flaking of the electrode layer. It was then placed in a folded pouch. This ensured uniformity in electrode size, which is critical for consistent cell performance.

3.4 Drying and Measuring Electrodes

The pouch containing the electrodes was placed in a drying oven and dried at a constant temperature for one day. Following the drying process, the electrodes were removed from the oven and allowed to cool to room temperature. Subsequently, the weight of each electrode was measured using a precise analytical balance. For quality control and consistency, three electrodes from each slurry composition were selected and weighed, ensuring the verification of their mass before cell assembly.

3.5 Cells Assembly

Coin cells were assembled within a controlled glove box environment, which was continuously purged with Argon gas to maintain an inert atmosphere. Before doing the assembly, the Argon gas level was checked to verify that it was within the acceptable range. More importantly, the water and oxygen levels in the glove box were also monitored and confirmed to be below 0.5 ppm and 25 ppm, respectively, to ensure an optimal environment for handling sensitive materials such as sodium. To further ensure the purity of the atmosphere, the glove box was flushed three times, for five minutes each time, with Argon gas before use.

The electrodes, which had been prepared earlier, were placed into a sample box and transferred into the glove box through the small chamber, maintaining a sealed environment throughout the process. The assembly process began by placing the top casing onto the work surface. Next, a piece of sodium metal, verified to be shiny and free from oxidation, was ground to a flat surface to ensure uniform contact and minimise any irregularities. Any dust on the sodium metal was removed using a brush. A separator was then placed on top of the sodium metal to prevent direct contact between the positive and negative components.

An electrode, with the hard carbon facing down, was then placed on the separator. By using a micropipette, 100 µL of electrolyte was carefully

transferred and dispensed around the electrode to ensure complete wetting. A stainless steel spacer was then positioned over the electrode, followed by the insertion of a spring. Finally, the bottom casing was added to complete the assembly of the coin cell. The fully assembled cell was then placed into an automatic crimping machine, which sealed the cell and ensured proper contact between the internal components.

Once the assembly process was complete, all tools and materials were returned to their designated locations, and the assembled cells were transferred from the glove box to the external. The oxygen and water levels, as well as the Argon gas level, were rechecked to ensure they remained within a proper limit before the glove box was closed. In general, three cells of each specific hard carbon to binder ratio were made based on the different binder ratios.

3.5 Characterization of Cells

The characterisation of the assembled coin cells was carried out using the LAND Battery Testing System, CT 2001A. The cells were placed onto electronic channels within the testing system, and specific testing protocols were set before initiating the system. The testing protocol began with a rest period for the batteries, followed by setting the discharge and charge rates (DC and CC) sequentially at 0.1C, 0.2C, 0.5C, 1C, 2C and back to 0.1C. Each rate was looped five times, resulting in a total of 30 cycles for each cell.

Throughout the testing process, the system recorded various data points, including cell capacity (CapC/mAh), specific cell capacity (SpeCapC/mAh/g), discharge capacity (CapD/mAh), specific discharge capacity (SpeCapD/mAh/g), and efficiency. These metrics were displayed in real-time on the command window, providing a comprehensive overview of the cell performance.

Upon completion of the testing cycles, the data was analysed using three sets of graphs, which are the Time - Voltage & Current graph, the Capacity - Voltage graph and the Cycle - Efficiency & Specific Capacity Graphs, allowing for a detailed evaluation of the cell's performance characteristics. A full analysis is provided in section 4.

4. Results and Discussion

4.1 Specific Capacity vs. Voltage

Figure 4.1-1 illustrates the specific capacity vs. voltage profiles for sodium-ion batteries with varying binder compositions and hard carbon-to-binder ratios from Cycle 1 to 2. Initially, there is a rapid voltage drop from nearly 2.6V, which is related to the electrochemical potential of sodium (Na/Na+) around -2.71V vs. the standard hydrogen electrode (SHE). This initial high voltage represents the open-circuit potential of the fully charged cell. As the cell begins to discharge, sodium ions flow from the sodium metal anode, through the electrolyte, and into the hard carbon cathode, contributing to the voltage drop. This process, known as sodiation, involves the insertion of sodium ions into hard carbon during the initial discharge. According to Dose and Johnson, 2022 [8], the main purpose of this initial sodiation is to compensate for active ion loss (AIL) in the first cycle, thus increasing the reversible capacity. After sodiation, the voltage increases steadily as sodium ions are extracted from the hard carbon cathode (desodiation) during charging. This desodiation process effectively "empties" the cathode of sodium ions, making it ready to accept more during the next discharge cycle [9]. Interestingly, there is a gap between graphs of Cycle 1 and 2, which is due to the formation of the sodium electrolyte interphase (SEI) on the hard carbon cathode surface. This SEI layer, while consuming some sodium ions in its formation, plays a crucial role in preventing further electrolyte and electrode degradation by acting as an electronic insulator, which enables long-term cycling [10].

As shown in Figure 4.1-1, increasing the CMC content in the binder generally leads to higher specific capacity, particularly at higher hard carbon ratios. This trend is likely attributed to enhanced adhesion and ionic conductivity within the electrode facilitated by CMC. Within each binder composition, increasing the hard carbon-to-binder ratio from 85: 15 to 95: 5 also results in a noticeable increase in specific capacity because of increased active material content. The 5% CMC binder with a 95:5 hard carbon-to-binder ratio exhibits the highest specific capacity, reaching approximately 260 mAh/g at 2.5 V over Cycle 2.

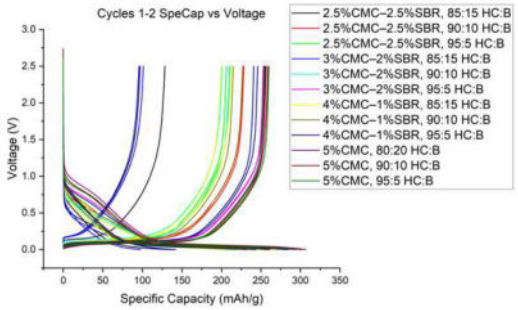


Figure 4.1-1: Specific Capacity vs. Voltage Graph for Cells of Different Composition from Cycle 1 to 2

Figures 4.1-2 to 4.1-7 compare the specific capacity and voltage profiles of sodium-ion batteries with varying binder compositions and hard carbon-to-binder ratios across multiple cycles. As observed in the figures, the voltage profiles show similar trends across all cells, with specific capacity generally increasing as the SBR content in the binder decreases. This suggests higher CMC in the binder leads to improved performance. For instance, cells with 5% CMC consistently demonstrates a higher specific capacity at the same voltage compared to those with lower CMC content, indicating that the addition of SBR to the binder may negatively impact capacity.

Furthermore, the figures suggest that a hard carbon-to-binder ratio of 80:20 generally results in better performance, which is opposite to the conclusion drew in Figure 4.1-1 as discussed previously. Therefore, further investigation is needed to fully understand the mechanism behind different hard carbon-to-binder ratios to determine the optimal compositions for maximizing the performance of sodium-ion batteries.

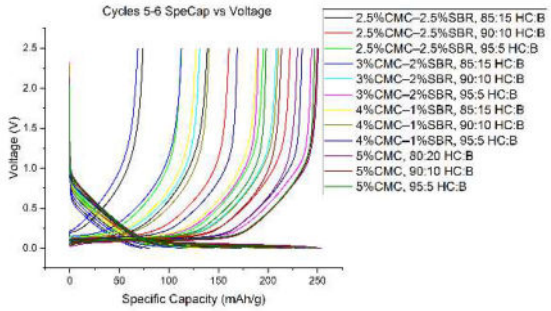


Figure 4.1-2: Specific Capacity vs. Voltage Graph for Cells of Different Composition from Cycle 5-6

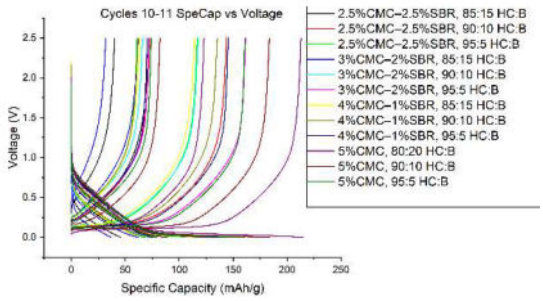


Figure 4.1-3: Specific Capacity vs. Voltage Graph for Cells of Different Composition from Cycle 10-11

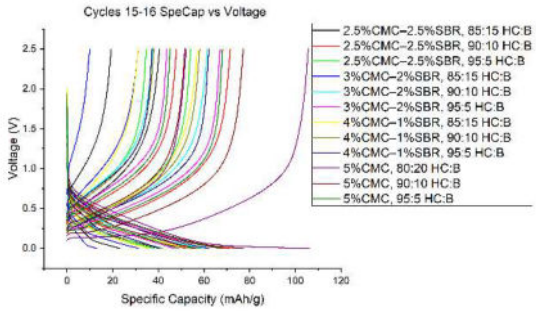


Figure 4.1-4: Specific Capacity vs. Voltage Graph for Cells of Different Composition from Cycle 15-16

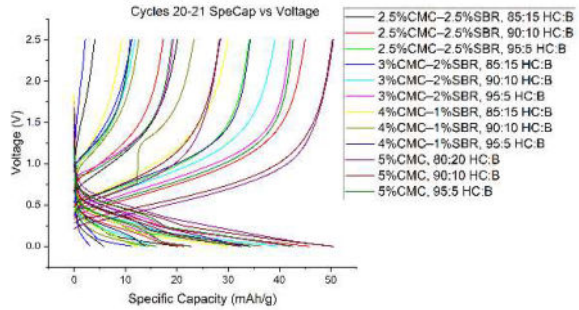


Figure 4.1-5: Specific Capacity vs. Voltage Graph for Cells of Different Composition from Cycle 20-21

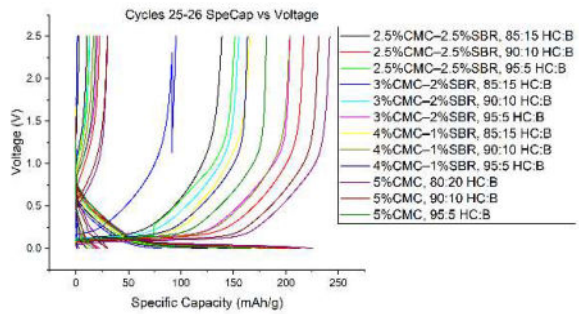


Figure 4.1-6: Specific Capacity vs. Voltage Graph for Cells of Different Composition from Cycle 25-26

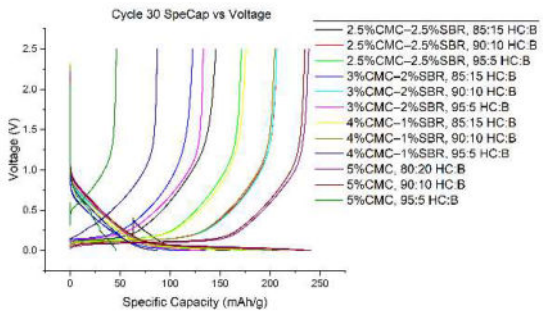


Figure 4.1-7: Specific Capacity vs. Voltage Graph for Cells of Different Composition from Cycle 30

4.2 Time vs. Specific Capacity and Efficiency

To understand the effects of cathode composition, the assembled half-cells were characterized in a special battery testing apparatus. Data from the best-performing cells from each composition were selected. Specific capacity during charge and discharge, with respect to the testing cycles were measured and plotted, along with their ratios which is internal coulombic efficiency.

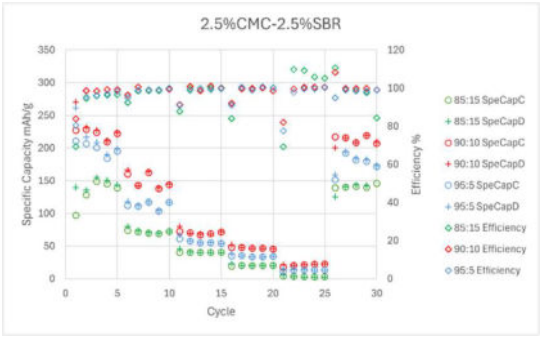


Figure 4.2-1: Cycle vs. Specific Capacity & Efficiency Graph for 2.5% CMC: 2.5% SBR

Results of the 2.5% CMC-2.5% SBR are shown in Figure 4.2-1. Here, the 85:15 hard carbon-to-binder ratio performs noticeably worse than ratios 90:10 and 95:5, of which the former performs the best. Internal coulombic efficiencies of the 90:10 HC: Binder ratios are approximately a percentage higher than the other compositions.

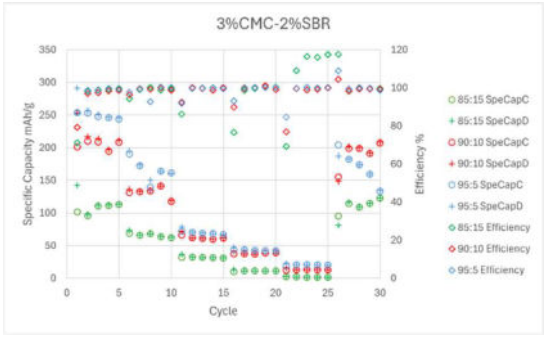


Figure 4.2-2: Cycle vs. Specific Capacity & Efficiency Graph for 3% CMC: 2% SBR

Results are similar for 3% CMC-2% SBR composition, as seen in Figure 4.2-1. The 85:15 hard carbon-to-binder ratio has significantly lower specific capacity than ratios 90:10 and 95:5, of which the latter performs better, with internal coulombic efficiencies being approximately equal.

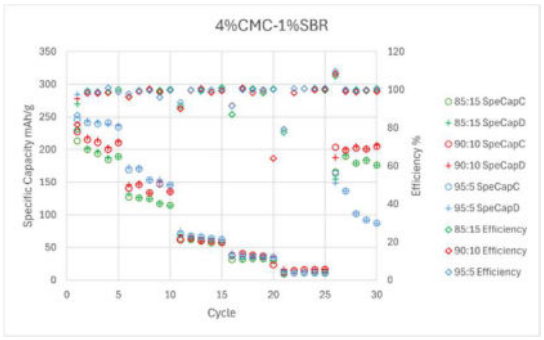


Figure 4.2-3: Cycle vs. Specific Capacity & Efficiency Graph for 4% CMC: 1% SBR

As SBR content is further lowered, as seen in composition 4% CMC-1% SBR in Figure 4.2-3, the difference in specific capacities exhibit more similar values for different hard carbon and binder ratios. During cycles 10-25, with higher charging/discharging ratios of 0.5C, 1C and 2C, the effects of differing HC: Binder ratios on specific capacities seem to matter very little, with values being approximately the same. During the last 5 cycles with 0.1C charging/discharging rate, the electrode with HC: Binder ratio 95:5 seemed to fail with performance falling off rapidly, while composition 90:10 performed the best.

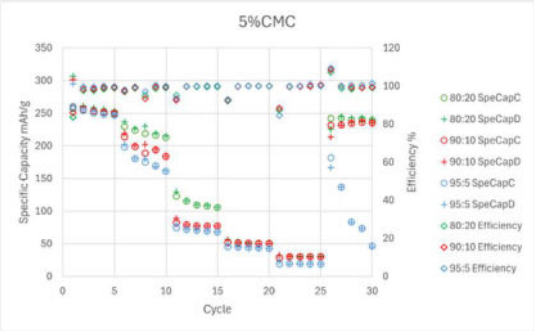


Figure 4.2-4: Cycle vs. Specific Capacity & Efficiency Graph for 5% CMC

As SBR content is eliminated, with a pure CMC binder of 5% CMC being used in Figure 4.2-3, the difference in specific capacities exhibits similar values for different hard carbon and binder ratios. During cycles 15-25, with higher charging/discharging ratios of 1C and 2C, the effects of differing HC: Binder ratios on specific capacities seem to matter very little, with values being approximately the same. Like the previous graph, during the last 5 cycles with 0.1C charging/discharging rate, the electrode with HC: Binder ratio 95:5 seemed to fail with performance falling off dramatically, with compositions 80:20 and 90:10 performing similarly well.

To summarize, the trends in Figures 4.2-1 to 4.2-4 indicate that for higher charging-discharging ratios of around 1C and 2C, adding SBR to the binder composition hinders performance, however for high HC: Binder ratios of 95:5, SBR seems to slow down rapid performance reduction during the last 5 cycles. The reason for that could be that CMC on its own is too fragile, and without the elasticity of SBR, the mechanical integrity can be easily compromised during the last 5 cycles. This is supported by the literature [11], and the fact that 95:5 HC: Binder ratios for 5% CMC binder had to be handled very delicately during assembly, as the active layer formed cracks and flaked off very easily.

As seen in Figure 4.2-5, during most cycles, cells with 5% CMC. 90:10 HC: Binder cathode composition possesses the highest specific capacity, with efficiencies being roughly equal.

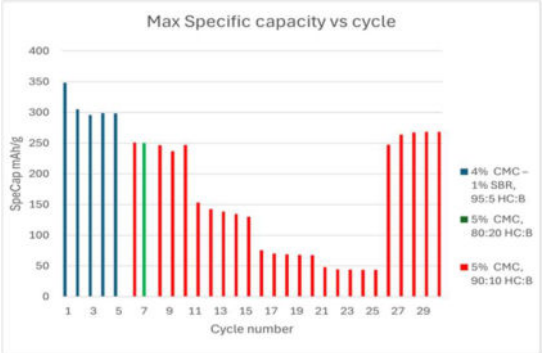


Figure 4.2-5: Cycle Number vs. Specific Capacity Graph

The absence of SBR causing higher performance could be caused by SBR introducing larger tortuosity due to its different binding properties, hindering ion conduction, because its flexibility prevents the formation of ion-conducting channels. Another factor could be that CMC is a better conductor than SBR [12].

5. Conclusion and Outlook

5.1 Conclusion

To conclude, this study demonstrates the critical role of binder composition and hard carbon-to-binder ratio in determining the performance of sodium-ion batteries. Higher CMC content in the binder generally leads to improved specific capacity due to enhanced adhesion and ionic conductivity. On the other hand, the addition of SBR, while potentially beneficial for the mechanical integrity of the binder, appears to negatively impact the cell performance. Optimising the hard carbon-to-binder ratio is also crucial for maximising capacity, but further research is needed to determine the optimal ratio for achieving peak performance. The capacity fade remains another significant challenge which impacts long-term cycling stability, therefore, further investigations should focus on the mechanisms behind the capacity fade, as well as exploring and modifying the structure of CMC, SBR and hard carbon for better cell performance.

5.2 Outlook and Future Work

5.2.1 Investigating the Structure of CMC, SBR and Hard Carbon

To understand the impact of SBR on CMC binder performance, further investigation of the interactions between these components and hard

carbon is needed. Since CMC has numerous hydroxyl (-OH) and carboxyl (-COOH) groups, these groups form strong hydrogen bonds with hydroxyl groups on the surface of hard carbon particles, creating adhesion [12]. However, SBR is a hydrocarbon-based polymer with limited polar groups, and it primarily interacts with hard carbon through weaker van der Waals forces. Therefore, introducing SBR into the CMC matrix may disrupt the extensive hydrogen bonding network between CMC and hard carbon, resulting in a negative effect on cell performance. However, since CMC itself is an extremely stiff and brittle polymer, the addition of SBR could strengthen the structure of the binder, thus enhancing the flexibility of the electrode [12].

Several techniques can be used to visualise the distribution of hard carbon particles, binders, and pores in the electrode. For example, X-ray Diffraction (XRD) can be used to determine if there are any changes in the crystallinity of the hard carbon or the formation of any new crystalline phases due to binder interactions, as well as analysing crystallite size and orientation, for instance, if the binder affects the size or orientation of hard carbon crystallites, which further influence electrochemical performance. In addition, the use of X-ray Photoelectron Spectroscopy (XPS) can be used to investigate the chemical bonding between the binder and hard carbon to understand how SBR might disrupt these interactions.

5.2.2 From Lab-Based Research to Real-Life Manufacture

While this study highlights the potential drawback of adding SBR to CMC binders for lab-scale sodium-ion batteries, further investigation can be done to assess its impact on large-scale manufacturing. Scalability, cost, and processing conditions should be carefully considered. SBR's elastomeric properties could offer advantages in manufacturing, such as improved flexibility and handling of electrodes. However, its potential negative effects on adhesion, conductivity, and cycling stability need to be addressed. For instance, a comprehensive analysis of optimised SBR/CMC ratios needs to be conducted to determine if a specific ratio can balance the benefits and drawbacks of each polymer, as well as evaluate the cost-effectiveness of incorporating SBR in large-scale production.

6 References

[1] Nayak, P.K., Yang, L., Brehm, W. and Adelhelm, P. (2018). From Lithium-Ion to Sodium-Ion Batteries: Advantages, Challenges, and Surprises. Angewandte Chemie (International ed. in English), [online] 57(1), pp.102–120. doi:https://doi.org/10.1002/anie.201703772.

[2] Kazuki Ohishi, Igarashi, D., Tatara, R., Nishimura, S., Koda, A., Shinichi Komaba and Sugiyama, J. (2021). Na Diffusion in Hard Carbon Studied with Positive Muon Spin Rotation and Relaxation. ACS Physical Chemistry Au, 2(2), pp.98–107. doi:https://doi.org/10.1021/acsphyschemau.1c00036.

[3] Liu, T., Zhang, Y., Jiang, Z., Zeng, X., Ji, J., Li, Z., Gao, X., Sun, M., Lin, Z., Ling, M., Zheng, J. and Liang, C. (2019). Exploring competitive features of stationary sodium ion batteries for electrochemical energy storage. Energy & Environmental Science, 12(5), pp.1512–1533. doi:https://doi.org/10.1039/c8ee03727b.

[4] Xie, F., Xu, Z., Guo, Z. and Titirici, M.-M. (2020). Hard carbons for sodium-ion batteries and beyond. Progress in Energy, 2(4), p.042002. doi:https://doi.org/10.1088/2516-1083/aba5f5.

[5] Li, R.-R., Yang, Z., He, X.-X., Liu, X.-H., Zhang, H., Gao, Y., Qiao, Y., Li, L. and Chou, S.-L. (2021). Binders for sodium-ion batteries: progress, challenges and strategies. Chemical Communications, [online] 57(93), pp.12406–12416. doi:https://doi.org/10.1039/d1cc04563f.

[6] Zhang, R., Yang, X., Zhang, D., Qiu, H., Fu, Q., Na, H., Guo, Z., Du, F., Chen, G. and Wei, Y. (2015). Water soluble styrene butadiene rubber and sodium carboxyl methyl cellulose binder for ZnFe2O4 anode electrodes in lithium ion batteries. Journal of Power Sources, 285, pp.227–234. doi:https://doi.org/10.1016/j.jpowsour.2015.03.100.

[7] Müllner, S., Michlik, T., Reichel, M., Held, T., Moos, R. and Roth, C. (2023). Effect of Water-Soluble CMC/SBR Binder Ratios on Si-rGO Composites Using µm- and nm-Sized Silicon as Anode Materials for Lithium-Ion Batteries. Batteries, 9(5), pp.248–248. doi:https://doi.org/10.3390/batteries9050248.

[8] Dose, W.M. and Johnson, C.S. (2022). Cathode pre-lithiation/sodiation for next-generation batteries. Current Opinion in Electrochemistry, 31, p.100827. doi:https://doi.org/10.1016/j.coelec.2021.100827.

[9] Wu, C.-M., Pan, Ping-I., Cheng, Y.-W., Liu, C.-P., Chang, C.-C., Avdeev, M. and Lin, S. (2017). The mechanism of the sodiation and desodiation in Super P carbon electrode for sodium-ion battery. Journal of Power Sources, 340, pp.14–21. doi:https://doi.org/10.1016/j.jpowsour.2016.11.048.

[10] Weadock, N., Nitinun Varongchayakul, Wan, J., Lee, S., Joonil Seog and Hu, L. (2013). Determination of mechanical properties of the SEI in sodium ion batteries via colloidal probe microscopy. Nano energy, 2(5), pp.713–719. doi:https://doi.org/10.1016/j.nanoen.2013.08.005.

[11] Hofmann, K., Hegde, A.D., Liu-Theato, X., Gordon, R., Smith, A. and Willenbacher, N. (2024). Effect of mechanical properties on processing behavior and electrochemical performance of aqueous processed graphite anodes for lithium-ion batteries. Journal of Power Sources, [online] 593, p.233996. doi:https://doi.org/10.1016/j.jpowsour.2023.233996.

[12] Müllner, S., Michlik, T., Reichel, M., Held, T., Moos, R. and Roth, C. (2023). Effect of Water-Soluble CMC/SBR Binder Ratios on Si-rGO Composites Using µm- and nm-Sized Silicon as Anode Materials for Lithium-Ion Batteries. Batteries, 9(5), pp.248–248. doi:https://doi.org/10.3390/batteries9050248.

A novel CoP_x@C cocatalyst for enhanced selective visible light photocatalytic C-C coupling of methanol to ethylene glycol over CdS

Jonathan Low Yi Wei, Kar Jan Chan, Humza Bhatti
Department of Chemical Engineering, Imperial College London, United Kingdom

Abstract

Ethylene glycol (EG) is a critical precursor in the synthesis of many polymers, with a global demand of over 32 million metric tons, and a sizeable carbon footprint in production. Due to its ability to reduce this footprint, photocatalytic conversion of C1 precursors, such as methanol to EG via C-C coupling, remains a dream reaction for the chemicals industry. However, it suffers from poor selectivity and formation rates. Noting this, this paper proposes a novel CoP_x@C cocatalyst loaded onto a CdS catalyst capable of enhancing the selectivity and EG formation rate of visible light photocatalytic coupling of methanol to EG. The cocatalyst is synthesised via phosphorylation of zeolite imadiazole framework (ZIF)-67, producing a highly active morphology with a high density of micropores. The synthesised cocatalyst at 2 w/t% enhances the EG formation rate over CdS by 51-fold to 1.6 mmol g_{cat}^{-1} h^{-1}, while increasing the selectivity towards EG from 42% to 78%, achieving the highest reported enhancement in selectivity and activity over the base catalyst. This work not only offers an interesting non-petroleum route to produce EG, but also investigates the effect of variable cobalt phosphide loadings, morphologies and synthesis techniques on the activity and selectivity for this reaction, opening potential for further performance enhancements.

Keywords

C-C Coupling, Photocatalysis, Cobalt Phosphide, Cadmium Sulfide, Ethylene Glycol, Methanol

1. Introduction

Ethylene glycol (EG, $HOCH_2CH_2OH$) is a colourless and odourless compound which is a critical precursor in the synthesis of many polymers, plastics as well as chemicals used in the automotive, chemical, paint and coating industries.[1] With a global demand of over 32 million metric tons annually that is projected to grow by over 4% annually over the next 10 years, sustainable methods of EG production are of great commercial interest.

Currently, commercial EG production largely occurs via a 2-step process, consisting of epoxidation of petroleum-derived ethylene to ethylene oxide (EO) and the subsequent hydrolysis of EO.[2] While mature, this multistep process suffers from low efficiency due to the low EO yield in the first step, and sizeable energy consumption and carbon footprint due to the harsh reaction conditions of over 300 °C and 30 bar.[3]

With increasing global demand for EG and a greater push towards more sustainably produced products, focus on developing non-petroleum based EG production routes has grown. C1 precursors are particularly attractive for this process due to the milder reaction conditions and high atom economy of C-C coupling reactions to produce EG. Methanol has emerged as a particular interest due to its relatively low cost, ease of procurement and potential for sustainable production from the oxidation of biomass or organic waste.[4]

A variety of methods to synthesise EG from methanol have previously been investigated to overcome the thermodynamic limitations of the reaction, including dielectric barrier discharge, electrocatalytic, biosynthetic and photocatalytic methods.[5] Amongst these alternatives, visible light photocatalysis remains most attractive due to higher yields and the abundance of visible light as an energy source for the reaction.

Several preliminary successes utilising hydrogen evolution reaction (HER) catalysts, such as CdS, ZnS, TiO_2 and Pt, have shown promising findings, achieving yields of up to 16% over 100 h when paired with certain co-catalysts under specific reaction conditions.[6] However, limited research has been done to create and optimise cocatalysts that improve the selectivity and activity of existing catalysts.

Noting the vast economic potential for enhancing this reaction through development of an optimised cocatalyst, this paper presents a novel, highly active CoP_x@C/CdS photocatalyst capable of producing ethylene glycol from methanol with a high selectivity.

2. Background

2.1. Reaction pathway

The methanol to ethylene glycol (MTEG) reaction (Eqn. 1) is a C-C coupling reaction proposed to consist of 3 steps,[6] as outlined in Fig. 1 below.

$$2CH_3OH \rightarrow HOCH_2CH_2OH + H_2 \quad (1)$$

First, a photon is absorbed by the photocatalyst, causing an electron to transition from the valence to the conduction band of the photocatalyst, photogenerating an electron-hole pair. Next, the electron transfers this energy to a methanol molecule adsorbed onto the photocatalyst surface, activating its C-H bond.

This can occur in 2 different ways, firstly via a stepwise proton transfer followed by an electron transfer. Proton transfer can occur via heterolytic cleavage of the bond, producing CH_2OH^- and a proton. An electron is then transferred from CH_2OH^- to the hole to fill it, producing a hydroxymethyl (CH_2OH) radical. Alternatively, the C-H activation can occur via homolytic cleavage of the C-H bond in a concerted

proton electron transfer (CPET) step. Finally, two of the evolved CH_2OH radicals combine, forming EG.

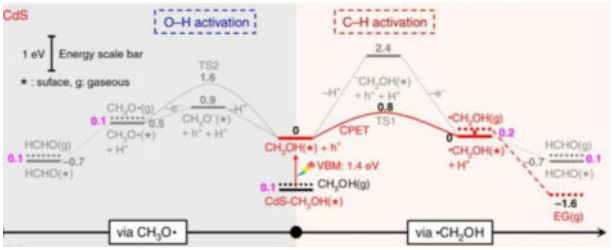


Fig. 1: MTEG reaction pathway over CdS proposed by Xie et al.[6]

As shown by Fig. 1, the C-H activation step competes with the activation of the O-H bond in methanol via proton transfer, leading to the formation of a CH_3O^- radical, then formaldehyde (FA) as a product via Eqn. 2:

$$CH_3OH \rightarrow HCHO + H_2 \quad (2)$$

In methanol, the O–H bond is generally weaker than the C-H bond, hence naturally more reactive.[7] This results in the selectivity of the reaction towards EG being naturally low. This selectivity towards EG instead of FA is thus principally dependent on the ability of the catalyst to enhance C-H activation over O-H activation. Noting the reaction pathway,[6] the activity of any catalyst is dependent on two factors: The degree of electron-hole pair separation, which determines the extent of charge recombination, and the band gap of the photocatalyst, which determines the ease of photoexcitation and the percentage of incident visible light photons that can trigger the reaction. Maximising EG yield and selectivity is paramount for efficient reagent utilisation and product isolation. Therefore, any catalytic system for this reaction must aim to improve these two parameters.

2.2. CdS as a catalyst

For a catalyst to successfully catalyse the MTEG reaction, the conduction band must be more negative than the reduction potential of methanol to EG (-0.38 V at pH 7), while the valence band must be more positive than the reduction potential of H^+/H_2.

Due to the proximity of this potential to that of hydrogen reduction, a variety of HER catalysts which meet this criterion have been investigated in Table 1 below. The band gaps of these catalysts are shown in Fig. 2.

Table 1: MTEG reaction of various catalysts under visible light.[6]

Catalyst	Reaction conditions	EG formation rate (mmol g^{-1} h^{-1})	Selectivity (%)	Yield (%)
ZnS	Xe lamp (420–780 nm), 300 W light intensity, 12 h reaction time	0	0	0
CdS nanorods		0.46	71	0.112 (12 h)
CdS nanoparticles		0.28	56	0.068 (12 h)
Cu_2O		0	0	0

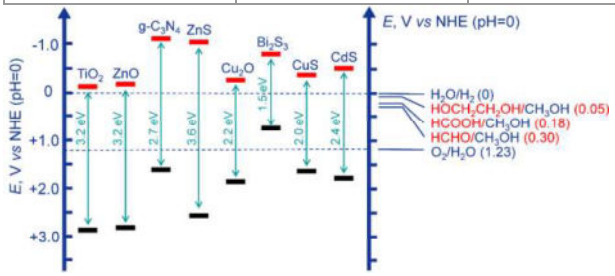


Fig. 2: Band gaps of commonly tested materials.[6]

As shown by the relative catalyst performance in Table 1 and Fig. 2 above, CdS is particularly attractive due to its relatively narrow band gap (2.4 eV) and good charge capacity, which can greatly enhance the seperation of electron–hole pairs to extend their lifetime.[8] The morphology of CdS is also relatively important, with CdS nanorods found to have higher activity and selectivity then other morphologies.[6]

2.3. CoP as a cocatalyst

To improve photocatalyst performance, semiconductors are typically modified through the addition of cocatalysts. This is done to improve electron-hole separation, suppress charge recombination, or modify the band gap of the system or binding energy to various intermediates in a favourable manner.[3]

Several typical HER cocatalysts such as Pt, Pd, NiO_x and MoS_2 have been investigated[6] to elevate the performance of CdS nanorods, with MoS_2 nanofoam reported to achieve the best results, enhancing the EG formation rate from 0.46 to 11 mmol g^{-1} h^{-1} and EG selectivity from 71% to 90%.[6]

Noting existing attempts, cobalt phosphide (CoP) is an exceptionally active, non-noble metal bifunctional alkaline HER and OER catalyst[9] which could be useful as a cocatalyst for this reaction. With low HER overpotentials (136 mV) and Tafel slope (56.2 mV dec^{-1}) in alkaline media, CoP could enhance hydrogen formation rates and the overall reaction. CoP has previously been loaded onto $Zn_2In_2S_5$ at nominal rates of 0.125 – 0.5 w/t%, and has been shown to enhance the rate of EG formation from 1.3 to 18.9 mmol g_{cat}^{-1} h^{-1} while increasing the selectivity of the reaction from 75 to 90%, with an optimised loading of 0.25 w/t%.[10] However, the paper fails to accurately suggest the role of CoP in enhancing the selectivity of the reaction or attempt to optimise the morphology of CoP loaded onto photocatalysts.

CoP is known for creating an "ensemble effect" where the contrasting electronegativity of the Co and P sites can enhance the electron withdrawing, and donating effects of each.[11] The Co and P atoms in the CoP crystalline structure may function as anion-acceptor and proton-acceptor sites, respectively.[11] This enables it to achieve better charge separation, and adsorption of both H^+ and other anions, allowing for more dynamic tuning of the binding energy of various transition species on the cocatalyst surface, affecting their rates of adsorption and desorption. Combined, this could improve H_2 formation rates and improve C-H activation, boosting catalytic performance.

A wide variety of CoP morphologies such as nanowires, nanosheets, nanotubes, hollow polyhedrons and nanocubes have been investigated[12] for improving the number of catalytically active sites and rate of gas emission. However, it is unclear how well these findings translate to similar reactions like C-C coupling, especially as a cocatalyst.

Noting this gap in literature, this paper proposes a novel CoP_x@C cocatalyst synthesised from zeolitic imadiazole framework (ZIF)-67 with a high surface area, and improved selectivity and activity in the MTEG reaction. This paper also investigates the effect of different precursors on the activity and selectivity of CoP_x as a cocatalyst. ZIF-67[13] and Co-Co-PBA[14] were selected as target precursors for CoP_x synthesis due to their highly porous structure and extremely high surface area to volume ratio. Additionally, $CoCl_2$, Co_3O_4 and $Co_3(PO_4)_2$ were used to explore a variety of conventional morphologies such as nanoparticles and nanosheets that have been previously reported in literature.

3. Methodology

3.1. Synthesis of CdS nanorods

CdS nanorods were synthesised through a solvothermal method.[13] In this procedure, 4.62 g of cadmium chloride hemipentahydrate (*VWR, >99%*) and 4.62 g of thiourea (*Sigma-Aldrich*) were dispersed in 60 ml of ethylenediamine (*VWR*). The mixture was then transferred to a Teflon-lined autoclave and was maintained at 160 °C for 24 h. After cooling down to room temperature, the yellow solid product was collected by centrifugation, washed with distilled water and ethanol three times, then dried at 60 °C.

3.2. Synthesis of co-catalysts

To investigate the effect of different CoP_x morphologies on the catalyst activity and selectivity towards EG, a variety of methods were investigated to synthesise CoP_x as a cocatalyst. CoP_x samples were mostly synthesised in a two-step process, consisting of a precursor preparation step, followed by a phosphorylation step in a tube furnace (*Carbolite Gero EZS 1200*) under inert atmosphere.

3.2.1. Synthesis of CoP_x from cobalt (II) phosphate

CoP_x was synthesised from cobalt phosphide in a two-step procedure,[14] consisting of a precipitation reaction to obtain cobalt phosphide, followed by reduction in a tube furnace to obtain the CoP_x. In a typical precipitation procedure, 1.6 g of cobalt chloride (*Alfa Aesar, 97%*) was dissolved in 50 ml deionised water. The mixture was stirred for 5 min, and 9 ml of 0.5 M diammonium phosphate solution (*Sigma-Aldrich, ≥98%*) was added dropwise under continuous stirring. After 1 h, the suspension was heated on a hot plate overnight at 110 °C. The solid pink product from the reaction was then collected by centrifugation, washed with deionised water and ethanol two times each, then dried in a vacuum oven at 40 °C for 24 h. The cobalt phosphide was then heated at 800 °C for 3 h in a tube furnace with a ramp rate of 5 °C min^{-1} in 5% H_2/N_2. The resultant black metal phosphides were then washed with deionised water and ethanol two times each and dried in a vacuum oven at 40 °C for 24 h.

3.2.2. Synthesis of ZIF-67

ZIF-67 was synthesised via a precipitation method.[15] In a typical procedure, 1.43 g of cobalt (II) nitrate hexahydrate *(Sigma Aldrich)* and 3.24g of 2-methylimidazole (2-MIM) *(Sigma Aldrich, 99%)* were separately dissolved in 50 ml methanol each *(VWR, >99.9%)*. The two solutions were then mixed under magnetic stirring for 5 min, and the obtained blue solution was left for 12 h without stirring at room temperature. The solid product from the reaction was then collected by centrifugation at 4500 rpm and washed with deionised water and ethanol two times each. The resultant purple product was then dried in a vacuum oven at 40 °C overnight to obtain ZIF-67.

3.2.3. Synthesis of Co-Co-PBA nanocubes (NCs)

Co-Co-PBA NCs ($Co_3[Co(CN)_6]_2$) were synthesised via a precipitation method.[16] In a typical precipitation procedure, 4.8 mmol of cobalt acetate tetrahydate *(Sigma-Aldrich, >98%),* 7.2 mmol sodium citrate *(Fisher Scientific, >99%),* and 3.2 mmol of potassium hexacycobaltate *(Thermo Scientific, 95%)* were dissolved in 320 ml of deionised water. The reactant mixture was then stirred at 700 rpm for 3 min, then left for 8 h without stirring at room temperature. The resultant products were collected via centrifugation at 4500 rpm and washed deionised water and ethanol twice each. The resultant pink Co-Co-PBA powder was then dried in a vacuum oven at 40 °C for 24 h.

3.2.4. CoP_x synthesis via precursor phosphorylation

CoP_x was synthesised from the above precursors (PBA NCs and ZIF-67), cobalt (II,III) oxide *(Alfa Aesar, 99%)* and cobalt chloride (*Alfa Aesar, 97%*) via phosphorylation in a tube furnace, using the heat to generate phosphine gas to reduce the precursor to CoP_x. [16] In this procedure, the precursor and sodium hypophosphite were placed at opposite ends of a porcelain crucible in a 1:10 mass ratio. The sample was then annealed under an inert nitrogen atmosphere in a tube furnace at 300 °C for 2 h, with a ramping rate of 2 °C min^{-1}.

3.3. Loading of cocatalysts onto catalyst

The CoP_x/CdS catalyst was prepared by an ultrasonication method.[10] In a typical procedure, 10 mg of CoP_x was first ultrasonically dispersed in 10 ml N,N-dimethylformamide (DMF, *(Fluorochem)*) in a centrifuge tube for 3 h at room temperature. To achieve desired loading rates, variable amounts of this suspension was added to a separate centrifuge tube with 50 mg of CdS and topped up to 5 ml with fresh DMF, following which this new mixture was dispersed ultrasonically for 2 h to achieve close contact between CoP_x and CdS. The loading amount of CoP_x onto CdS was varied from 0.5 w/t% to 4 w/t%, calculated as a nominal percentage of the mass of the CdS added. The obtained products were washed with deionised water and ethanol twice each, then dried in a vacuum oven at 40 °C for 24 h.

3.4. Photoreactor experiments

Photocatalytic reactions were carried out in an EvoluChem PhotoRedOx Box high throughput reactor, shown in Fig. 3 below, which can hold up to 4 reaction vials at a time. The reactor utilises a 450 nm, 30 W LED lamp and has a light intensity of 30 mW cm^{-2} on the front of the sample at Point A (Fig. 3) and an intensity of 21 mW cm^{-2} at Point B (Fig. 3) as measured by a an ILT 1400 radiometer photometer.

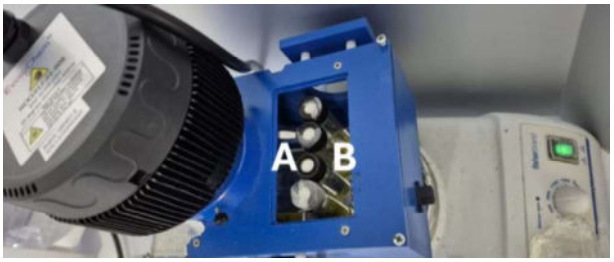


Fig. 3: *Photocatalytic reactor setup.*

To prepare each reaction vial, 10 mg of solid catalyst was ultrasonically dispersed for 10 min in 4.0 cm^3 methanol *(VWR, >99.9%)* and 1.0 cm^3 deionised water, with a magnetic stirrer added to each vial to ensure sufficient mass transfer. Each vial was then evacuated by purging the vial with N_2 gas using a Schlenk line, then loaded into the reactor (Fig. 3). Photocatalytic reactions were performed for 19 h under magnetic stirring at 700 rpm. Following the reaction, the liquid products were analysed using a high-performance liquid chromatograph (*HPLC, Shimazu LC-20A*) with refractive index and ultraviolet detectors. A Supelcogel C-610H column (30 cm by 7.8 mm) and a 5 mM sulfuric acid as the mobile phase were used.

3.5. Characterisation techniques

To characterise the material, X-ray diffraction (XRD) and scanning electron microscopy (SEM) were conducted to investigate the morphology of the catalyst. XRD patterns were measured on a Malvern Panalytical XRD X'Pert PRO system using, while field-emission gun SEM micrographs were obtained using a LEO Gemini 1525 FEG-SEM and an electron high tension of 5 kV.

The material's band gap was characterised via UV-Vis diffuse reflectance spectra using a Shimadzu UV-3000 equipped with a diffuse reflectance integrating sphere. Elemental composition was estimated via a combination of energy-dispersive X-ray spectroscopy (EDX) and X-ray fluorescence (XRF), with EDX conducted with a Zeiss Auriga FIB-SEM and XRF with an Epsilon 3 Pananalytical. Barrett-Joyner-Halenda (BJH) pore volume and distribution and Brunauer-Emmett-Teller (BET) surface area analysis were conducted to quantify surface area and pore size distribution with a Tristar 3000 via N_2 adsorption at 77 K.

4. Results and discussion

4.1. CoP_x/CdS photocatalytic reactions

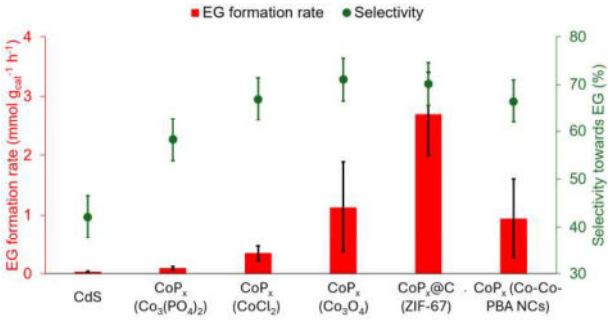


Fig. 4: *EG formation rates over CdS modified with various CoP_x cocatalysts. The compound in brackets represents the precursor used to synthesise the CoP_x. All cocatalysts were loaded at 3% nominal weight.*

A range of CoP_x morphologies was investigated for their ability to enhance the selectivity and activity of the MTEG. As shown by Fig. 4, loading of CoP_x onto CdS enhanced the selectivity of the MTEG reaction towards EG from 42.1% to 58.2-71.0% for CoP_x derived from all precursors. All CoP_x samples also enhanced EG formation rates by 3-77x over the base CdS (0.035 mmol g_{cat}^{-1} h^{-1}), with CoP_x@C (ZIF-67) presenting the highest EG formation rate of 2.7 mmol g_{cat}^{-1} h^{-1}. The large variation in sample performance likely stems from the uneven cocatalyst loading achieved from the ultrasonication procedure, which is relatively typical in photocatalysts.[16]

Aside from the CdS modified with CoP_x ($Co_3(PO_4)_2$), all other catalysts showed comparable selectivities of around 70%. This suggests that the enhancement in selectivity towards EG could be an intrinsic property of CoP_x itself, rather than the result of any specific morphology produced. However, changing the CoP_x precursor used in phosphorylation resulted in vastly different activities, with the most active CoP_x@C (ZIF-67) having an EG formation rate around 7.8 times that of the least active CoP_x ($CoCl_2$). This could suggest that the changes in precursor affect some combination of the electrochemical surface area, hydrogen formation rate and degree of charge separation, all of which affect the activity of the catalyst.

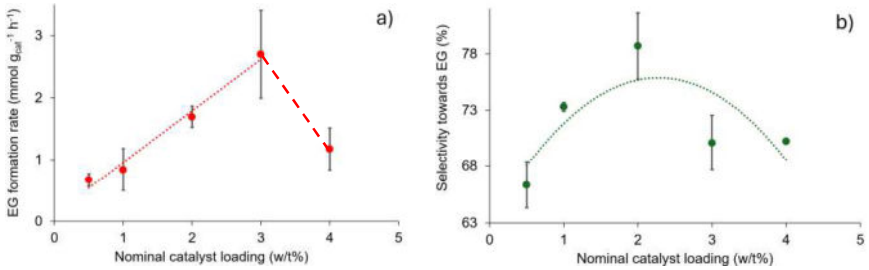


Fig 5a-b: *MTEG a) activity and b) selectivity of CoP_x@C (ZIF-67)/CdS catalysts with variable CoP_x@C loadings.*

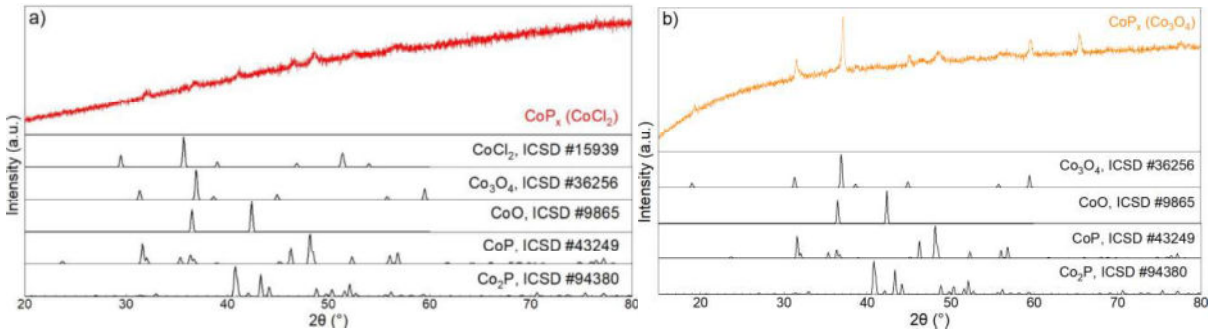
Table 2: *Comparison of optimised CoP_x@C with other cocatalysts tested with CdS. Samples highlighted in yellow use a much higher light intensity and baseline than in this paper, indicated in Table 1.*

Cocatalyst used with CdS, w/t% loading	EG formation rate (mmol g^{-1} h^{-1})	Selectivity (%)	AQY, 450 nm (%)
None	0.035	42.1	0.023
CoP_x@C (ZIF-67) 2%	1.694	78.7	1.772
CoO_x 3%	0.356	30.2	0.233
Pt 0.5%	0.043	35.5	0.028
NiCoP 0.5%	0.037	38.1	0.024
$Co_3(PO_4)_2$	0.009	12.8	0.006
MoS_2 7%	11	90	Unknown
CdS nanorods	0.46	71	Unknown

CoP_x@C (ZIF-67)/CdS was identified as the best catalyst due to its exceptionally high catalytic activity and good selectivity. Its performance was optimised by varying the cocatalyst nominal loading between 0.5-4 w/t%, as shown in Fig. 5a and 5b. The selectivity of all loaded catalysts tested followed a parabolic trend, with a peak at 2% selectivity. Activity was directly proportional to cocatalyst loading up to 3 w/t%, with an EG formation rate of 2.7 mmol g_{cat}^{-1} h^{-1}, before sharply declining above this point. The activity curve as a function of cocatalyst loading suggest that the reaction is first order with respect to the concentration of CoP_x@C at least up to a loading of 3 w/t%, suggesting that a key rate determining step in the reaction involves CoP_x@C. A higher selectivity is more challenging to attain than a higher activity since activity can be easily enhanced by increasing light intensity or adding scavengers.[6] Hence, the selectivity of the catalyst is prioritised, and the optimal loading of the cocatalyst is determined to be 2 w/t%.

The performance of the most effective CoP_x@C/CdS catalyst was compared against CdS loaded with other cocatalysts that are known to have good HER performance,[17] as shown in Table 2. This catalyst demonstrates a significantly higher activity and selectivity than all other catalysts tested in tests with the same light intensity. It also exhibits with the highest enhancement factor of the EG formation rate and selectivity compared to the literature benchmark (7 w/t% MoS_2 nanofoam), improving activity by a factor of 51 and selectivity by 36.6% compared to a 24-fold activity increase and 19% selectivity increase on MoS_2.

The significantly improved performance of CoP_x@C/CdS relative to these benchmarks suggests that the improvements in selectivity and activity do not purely come from improved H_2 formation rates. This is observed by noting how samples loaded with HER cocatalysts such as Pt and NiCoP, both of which are known for improving charge separation and having many HER active sites promoting H^+ dissociation,[17] showed an increase in activity, but poor selectivity towards EG formation. This could suggest, by elimination, that the enhancements in selectivity from adding CoP_x emerge from altering the binding energies of one or more reaction intermediates.



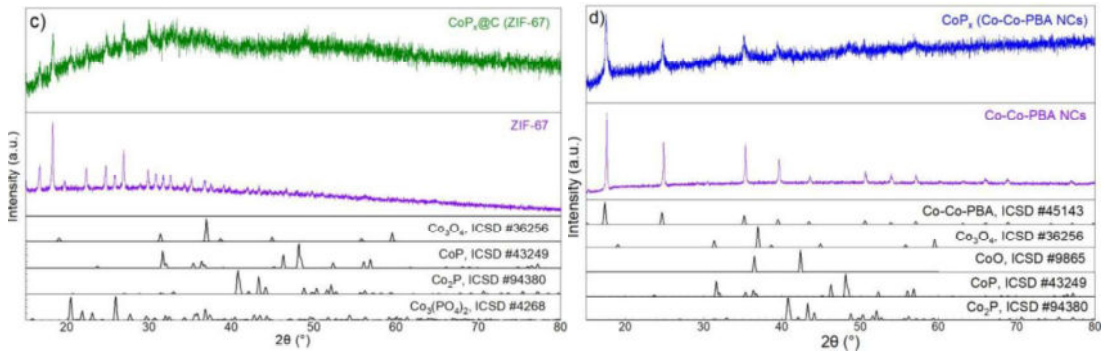


Fig. 6a-d: *XRD of a) CoP_x ($CoCl_2$), b) CoP_x (Co_3O_4), c) CoP_x@C (ZIF-67) and ZIF-67, d) CoP_x (Co-Co-PBA NCs) and Co-Co-PBA NCs.*

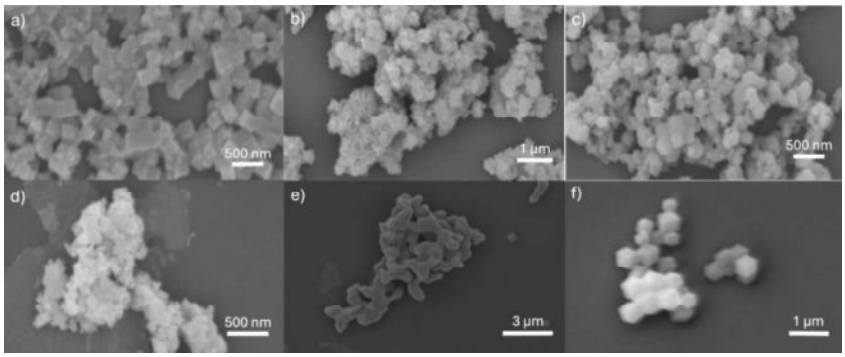


Fig. 7a-f: *SEM of a) CoP_x ($CoCl_2$), b) CoP_x (Co_3O_4), c) CoP_x@C (ZIF-67), d) CoP_x (Co-Co-PBA NCs), e) CoP_x ($Co_3(PO_4)_2$), f) ZIF-67.*

4.2. Characterisation of synthesised cocatalysts

4.2.1. XRD and SEM results

To gain better insight into the nature and performance of the synthesised CoP_x and CoP_x/CdS, a series of characterisation tests were performed.

As shown by Fig. 6a-6d, the CoP_x samples synthesised from carbon containing precursors (ZIF-67 in Fig. 6c and Co-Co-PBA NCs in Fig. 6d) show an elevated amount of sample noise, relative to samples not synthesised from carbon containing precursors ($CoCl_2$ in Fig. 6c and Co_3O_4 in Fig. 6d). This elevated noise level is possibly caused by the presence of a sizeable amount of amorphous carbon deposited on the CoP_x surface which scatters the X-rays at all angles. This carbon is evolved as a by-product of the breakdown of the precursor in the phosphorylation reaction. Hence, any small peaks in the sample are likely obscured by this elevated noise, leading to any compounds with such peaks, or in low concentrations, to be masked by the noise.

All samples phosphorylated using NaH_2PO_2 (Fig. 6a-6d) also exhibit numerous broad, poorly defined peaks. The appearance of such peaks could be possibly caused by poor crystallinity in the samples, which causes the broadening, and lower intensity observed. This poor definition is likely amplified by the noise created by presence of amorphous carbon in Fig. 6c and 6d. This poor crystallinity is supported by the SEM images of these materials (Fig. 6a-6d), all of which show a range of crystalline morphologies with varying degrees of deformation, indicating poor, but present crystallinity in these samples. The poor crystallinity likely emerges from the insufficient temperature (300 °C) in the phosphorylation step. It has been observed in literature that lower temperatures favour the formation of a less crystalline structure by providing inadequate driving force for the atoms to rearrange themselves into a more ordered, stable crystalline arrangement.[18]

Several of the observed XRD spectra (Fig. 6b, 6c and 6d) show the presence of peaks from the cobalt precursors (Co_3O_4 in Fig. 6b, ZIF-67 in Fig. 6c and Co-Co-PBA NCs in Fig. 6d) used for the reaction (as discussed below). This indicates the phosphorylation reaction did not go to completion, as there is still unreacted precursor. This could be due to insufficient dwelling time in the tube furnace (2 h), insufficient phosphorous precursor used (10:1 mass ratio), or insufficient contact area between the precursor and the evolved gas, causing the powder at bottom to be hindered from contacting the phosphine gas by those at the top.

As shown by Fig. 6a, the XRD spectra of CoP_x ($CoCl_2$) exhibits a series of poorly-defined peaks, corresponding to various facets orthorhombic CoP and Co_2P. 4 small peaks are observed, centred at approximately at 31.5°, 35.6°, 40.5° and 48.2°, corresponding to the (011) facet of CoP, (020) facet of CoP, (201) facet of Co_2P and (211) facet of CoP respectively. Notably, the first 2 peaks also overlap with the spectra

of Co_3O_4 shown in Fig 6a., indicating that may be possible that the mixture could contain various cobalt oxides. The poor crystallinity is supported by the SEM image of the same sample in Fig. 7a, which shows a mixture of orthorhombic and cubic particles with principal lengths between 200-300 nm with variable degrees of deformation, which would lead to the observed broad peaks in Fig. 6a.

As shown by Fig. 6b, the cocatalyst synthesised from Co_3O_4 exhibits 6 peaks at 31.3°, 37°, 45°, 48.2°, 59.5° and 65.2°. The peaks at 37°, 59.5° and 65.2° correspond to the (220), (511) and (440) facets of cubic Co_3O_4, the peak at 48.2° corresponds to the (211) facet of CoP, and the remaining 2 peaks could be attributed either to Co_3O_4 or CoP. The observed XRD spectra suggests that the cocatalyst may be a mixture of Co_3O_4 and CoP. This is supported by the varying particle morphologies in Fig. 7b, ranging from amorphous (bottom left) to flower-like (middle).

As shown by Fig. 6c, CoP_x@C (ZIF-67) exhibits 1 clear peak at 17.5°, and several poorly defined peaks at 22.5°, 26°, 28°, 31.5° and 48.2°. The peaks at 17.5° and 26° correspond to various facets of cubic ZIF-67, the peak at 48.2° corresponds to the (211) facet of CoP, while the peak at 22.5° corresponds to the (101) facet of $Co_3(PO_4)$. The peak at 26° could be attributable to either ZIF-67 or $Co_3(PO_4)$. The presence of peaks from different molecules indicates that this is catalyst is a mixture of compounds, notably ZIF-67, CoP and $Co_3(PO_4)_2$. This is supported by the SEM in Fig. 7c, which shows a mixture of deformed and nondeformed cube-like structures, possibly inherited from the ZIF-67 precursor with principal lengths between 300-500 nm. The similar cube-like morphology between ZIF-67 in Fig. 7f and the cocatalyst suggests that the morphology of the precursor is likely conserved during the phosphorylation process. The poor resolution of the images makes it challenging to determine an exact morphology, and hence, transmission electron microscopy (TEM) should be done to confirm the morphology of the synthesised CoP_x@C.

As shown by Fig. 6d, CoP_x (Co-Co-PBA NCs) exhibits several well-defined peaks at 17.4°, 24.7°, 35.2° and 39.5°, with 2 poorly defined peaks at 31.6° and 48.2°. The well-defined peaks at 17.5°, 26° and 28° correspond to various facets of cubic Co-Co-PBA NCs, while the peaks at 31.5° and 48.2° correspond to the (011) and (211) facets of CoP. The presence of peaks from different molecules indicates that this is catalyst is a mixture of compounds, notably Co-Co-PBA NCs and CoP_x. This is supported by the variety of morphologies SEM image of the same sample in Fig. 7d, which shows a several cube-like compounds, likely Co-Co-PBA NCs on the surface of a largely amorphous structure.

As shown by Fig. 7e, the cocatalyst synthesised from cobalt phosphate appears irregular and amorphous. The cocatalyst also appears to form large aggregates. The excessive clumping of particles suggests the active surface area of this cocatalyst is likely very low, which contributes to the significantly lower activity of CdS loaded with this catalyst observed.

4.2.2. EDX and XRF

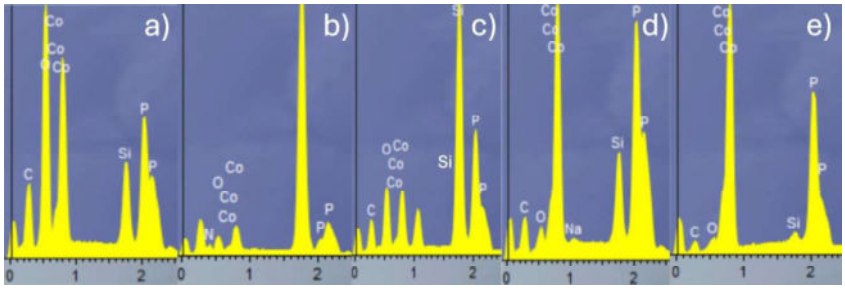


Fig. 8a-e: *EDX spectra of a) CoP_x ($CoCl_2$), b) CoP_x (Co_3O_4), c) CoP_x@C (ZIF-67), d) CoP_x (Co-Co-PBA NCs), e) CoP_x ($Co_3(PO_4)_2$)*

As shown above by Fig. 8a-e above, the cocatalysts synthesised from $CoCl_2$, Co_3O_4 and ZIF-67 contain a sizeable atomic weight percentage of cobalt, phosphorous and oxygen, while the cocatalysts synthesised from Co-Co-PBA NCs and $Co_3(PO_4)_2$ contain a sizeable atomic weight percentage of cobalt, phosphorous and only trace amounts of oxygen. The oxygen present in CoP_x (Co_3O_4) likely originates from unreacted Co_3O_4, as suggested from the XRD spectra in Fig. 6b. The large amounts of O in CoP_x@C (ZIF-67) likely stems from $Co_3(PO_4)_2$ as suggested by the XRD spectra in Fig. 6c.

The large amount of oxygen in the cocatalysts synthesised from $CoCl_2$ could come from several sources. First, it could result from oxidation of the precursor, or obtained products into cobalt oxide, either due to a small amount of oxygen being present in the tube during phosphorylation, or through long term exposure to air. Alternatively, amorphous cobalt phosphates or oxides may have formed from reactions between degradation products of the sodium hypophosphite decomposition, and the compounds in the mixture, creating oxides. Due to the unclear conclusions from the characterisation, further analysis via X-ray photoelectron spectroscopy

(XPS) is likely required to confirm the resultant species from this reaction and their compositions.

Interestingly, the cocatalysts containing sizeable amounts of oxygen show the highest activities for the reaction, while not affecting the selectivity of the reaction. This suggests that the oxygen-containing compound(s) present in the cocatalyst could play a key role in enhancing the charge separation, and potentially H_2 evolution rate over the catalyst system.

4.2.3. Combined discussion on cocatalysts

As highlighted from the above characterisation, the optimal cocatalyst synthesised from ZIF-67 is likely a mixture of CoP and $Co_3(PO_4)_2$ and unreacted ZIF-67. The deformations in the observed CoP_x@C (ZIF67) observed likely result from the compression of the hollow structure of ZIF-67 from the reaction, causing the original structure of the MOF to be deformed, leading to the observed compressions.

The elevated activity observed from the oxygen-rich cocatalysts possibly comes from the oxygen sites from $Co_3(PO_4)_2$, and potentially Co_3O_4 serving as additional electron rich sites.[20] This could modulate the local electron density on the surface of the CoP, leading to a sizeable increase in the surface area for hydrogen evolution, by creating a localised electron rich region that attracts H^+ ions to its surface. Moreover, the modulated electron density can result in improved charge separation by encouraging hole formation around the CoP_x sites.

4.3. Characterisation of CoP_x@C (ZIF-67)/CdS

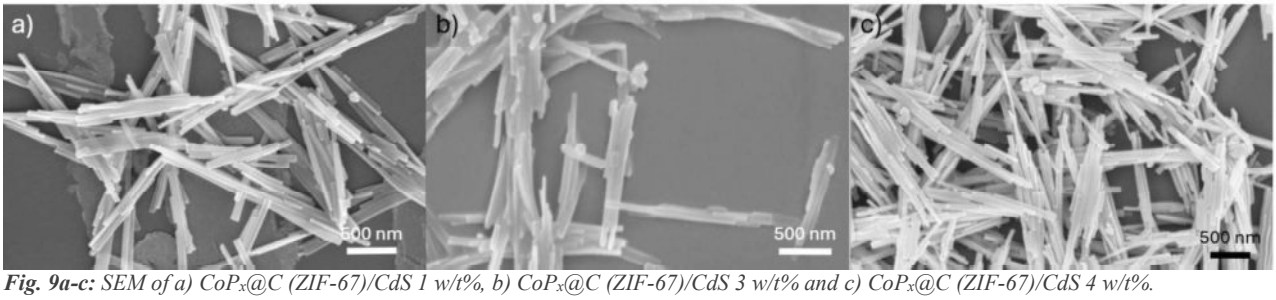


Fig. 9a-c: *SEM of a) CoP_x@C (ZIF-67)/CdS 1 w/t%, b) CoP_x@C (ZIF-67)/CdS 3 w/t% and c) CoP_x@C (ZIF-67)/CdS 4 w/t%.*

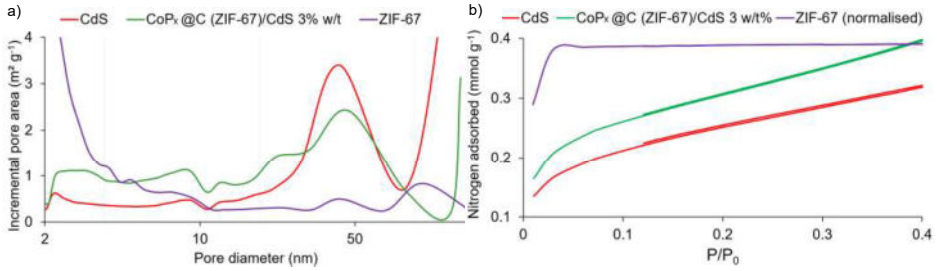


***Fig. 10a-b*:** a) *Differential area against logarithm of pore diameter plot* and *b) low pressure BET isotherms of CdS, CoP_x@C (ZIF-67)/CdS 3 w/t% and ZIF-67.*

As shown in Fig. 9a and b, CoP_x@C particles deposit regularly on the surface of the CdS nanorods at loadings beneath the optimal cocatalyst loading. Conversely, at catalyst loadings above the optimum cocatalyst loading (4 w/t%), CoP particles start to agglomerate with each other. This could possibly lead to a lower total contact area between CoP_x@C and CdS, hence leading to a lower activity.

The pore diameter distribution and BET surface area of the catalyst were also analysed to investigate the effect of cocatalyst loading on the active surface area of the catalyst. It was found that the addition of CoP_x@C to CdS leads to a 21% increase in the surface area of the catalyst from 20.46 $m^2 g^{-1}$ to 24.65 $m^2 g^{-1}$. The increase in area in demonstrated by an increase in the absorbance at low pressures (Fig. 10b), indicating that the addition of CoP_x@C (ZIF-67) to the CdS likely increases the volume of micropores in the catalyst. This is verified by Fig. 10a, which shows that the loaded CdS has a sizeable increase in pore area from pores with a diameter under 25 nm, resulting in a substantial decrease in the average pore diameter of the CdS from 34 to 13 nm. The presence of numerous micropores likely is inherited from the ZIF-67 metal organic framework utilised to synthesise the CoP cocatalyst, shown by the green line. This suggests that the porous structure of ZIF-67 is maintained through the loading and phosphorylation procedures. The microporous structure observed possibly provides a substantially more surface area, excellent surface area to volume ratio and high catalytic site density. Combined, these factors likely lead to increased catalytic activity, as observed in Table 2.

5. Discussion

The results from Table 2 and Fig. 5b suggest that the optimised cocatalyst is CoP_x@C synthesised from ZIF-67, and loaded onto CdS with a nominal loading of 2 w/t%. This co-catalyst on CdS shows as an impressive 51-fold

enhancement over CdS in EG formation rate, and a 36% selectivity shift towards EG. The characterisations, shown in Fig. 6c, 7c and 8c, suggest that this CoP_x@C contains a significant amount of cobalt, oxygen and phosphorous, and is possibly a mixture of cobalt (II) phosphate, and cobalt phosphides, and unreacted ZIF-67 precursor. The cocatalyst is likely covered in a layer of amorphous carbon, masking the true morphology of the cocatalyst synthesised, which is likely rhombic dodecahedral. The exact ratio of these components in the sample is unclear from the existing characterisation, and XPS should be conducted to confirm this.

The increased activity of CdS loaded with CoP_x likely stems from improved charge separation and electron density modulation, boosting activity.[10] CoP_x is known for its ability to modulate electron density around a surface, with the positively charged Co ions and negatively charged P anions in the crystalline structure serving as electron/hydroxide acceptor and H^+ acceptor sites respectively,[9,14] improving charge separation, and hence, hydrogen formation rates. This translates into an increase in catalytic activity. Noting the possibility that $Co_3(PO_4)_2$ may be present in this sample, it is possible that the $Co_3(PO_4)_2$ could synergistically improve the performance of the CoP in the sample, by contributing additional negatively charged O anions as H^+ acceptor sites. This is supported by comparing the activity of $Co_3(PO_4)_2$/CdS in Table 2 (0.009 mmol $g_{cat}^{-1}h^{-1}$) and the activity of the CoP/CdS ($Co_3(PO_4)_2$) (0.097 mmol $g_{cat}^{-1}h^{-1}$) from Fig. 4 with the optimized catalyst. The significant increase in activity when both compounds are present, as compared to individually supports a potential for this synergistic effect. Further characterization with XPS[6], and TEM should be conducted to confirm this phenomenon, along with appropriate measurements of the gaseous products of this reaction.

The enhanced activity of CdS loaded with CoP_x@C (ZIF-67) compared to other sources likely emerges from the high density of micropores present in ZIF-67 compared to other sample (Fig. S3), shown in the BJH pore area plot (Fig. 10a). The macrostructure of ZIF-67 is relatively hollow, giving it a very large internal surface area to volume/weight ratio, and hence, potential to have a high density of catalyst active sites. As shown in Fig. 10b, these morphological benefits are inherited by the final catalyst when CoP synthesized from ZIF-67 is loaded on it, leading to improved performance.[19]

The enhanced selectivity of the CdS loaded with CoP_x possibly emerges from how the contact between CdS and CoP_x affects the binding energy of key intermediates in the C-H activation step, improving selectivity. The presence of CoP_x in contact with CdS could modulate the local electron density of the surface. This modulation could create electron rich and poor regions on the catalyst surface, as the cobalt phosphide pulls more electrons away from the photogenerated CdS surface. The modified electron density can alter the binding energy of methanol, and the various radical and anionic intermediates involved in the reaction, particularly CH_3O^-, CH_2OH^- and their associated radicals.[6] Notably, as radicals are electron deficient species, while the anions are electron rich species, changes to the local electron density affect the binding energy of these species in opposite directions. This could give rise to the parabolic selectivity trend observed, with the desorption or adsorption of different intermediates serving as the rate determining step under different cocatalyst loadings.

Literature has proposed that desorption of CH_2OH radicals from the catalyst into the bulk phase could be a limiting step in the reaction pathway, [6] and hence, the addition of CoP_x may make this more energetically favorable by causing the catalyst surface to be more locally electron poor. However, an excessive withdrawal of electron density may make it challenging for electron-deficient species to adsorb onto the surface or stay adsorbed onto it. This could hence lead to the parabolic correlation between selectivity and CoP_x loading, where above the optimal loading, the rate determining step changes from radical desorption to something else. Further characterisation via in-situ Fourier transform infrared spectroscopy (FTIR), Raman spectroscopy, XPS and electron plasmonic resonance (EPR),[6,10] along with the necessary density functional theory (DFT)[10] calculations should be carried out to develop and justify this pathway.

6. Conclusion

This study successfully fabricates a unique CoP_x@C cocatalyst with an excellent EG formation rate enhancement and selectivity when loaded on CdS. The optimal loading of this cocatalyst was found to be 2 w/t%. The paper presents a unique approach within C-C coupling for synthesising CoP_x@C, utilising ZIF-67 as a precursor to produce CoP_x@C via phosphorylation with sodium hypophosphite. While still unoptimised, the synthesised catalyst produces a 51-fold increase in EG formation rates and a 36% selectivity shift to EG over pristine CdS nanorods. This is the highest activity and selectivity shift enhancement reported for any cocatalyst in literature. This performance is highly promising, and future work should look towards investigating the efficacy of this catalyst in similar C-C coupling reactions, such as production of 2,3 butandiol from ethanol, and the synthesis of hydrobenzoin from benzyl alcohol, which follow similar mechanistic pathways. It should also be tested in a photoelectrochemical cell, which typically has higher and more scalable production rates.

The characterisation performed on the catalyst reveals poor crystallinity in the sample, as well as the presence of sizeable oxygen in the cocatalyst, possibly attributable to cobalt phosphate. This makes it challenging to determine the exact composition of the cocatalyst synthesised, making a thorough mechanistic analysis challenging. Hence, more comprehensive characterisation of the system should be carried out to confirm the co-catalyst composition and identify the features causing the increase in catalytic activity.

In view of the limited understanding of the MTEG reaction mechanism, further investigation should be carried out to better understand how selectivity and yield can be improved. EPR spectroscopy allows for the identification of radical intermediates[6,10], while the efficiencies of light harvesting and charge separation can be investigated through measuring photoluminescence intensity, photocurrent density under periodic irradiation, and electrochemical impedance.[10] Finally, H_2 evolution should be measured using a gas chromatograph reactor as an additional indicator of catalytic activity, and to determine the ratio of holes to electrons in the system.

XPS should also be carried out to investigate the bonds present, and hence compounds present in the system, and their relative abundance. This would quantify the mixture's composition and enable a better understanding of the interactions between CoP_x, other compounds in the cocatalyst and CdS to glean insight into their roles in the reaction. TEM imaging should be conducted to observe the contact between CdS nanorods and CoP and determine if this is a key driver of catalytic actvity.[6] Inductively coupled plasma mass spectrometry (ICPMS) could also be employed to quantify cobalt loading, allowing for further optimisation of cocatalyst loading. Finally, catalyst stability, crucial for prolonged functionality, should also be analysed by monitoring the reaction rates over time in a kinetically controlled reaction.

There is also significant room for optimisation of the existing procedure to modify the mixture of compounds obtained, and the Co:P ratio in the compound. Particularly, the dwelling temperature can be increased in the phosphorylation step to obtain a more crystalline structure and more extensive phosphorylation. The amount or type of phosphorous precursor can similarly be changed to change the Co:P ratio, which can yield a variety of CoP, Co_2P and CoP_2 mixtures, [20] each with differing physiochemical properties. Given the existing performance of the material, and room for further optimisation, this cocatalyst shows great potential for uses in larger scale production in the future.

Acknowledgements

The authors would like to thank Humza Bhatti for assistance with CdS synthesis, and conducting the SEM and EDX, as well as Dr Salvador Eslava, Woojin Mun, Bradley Lewis and Mengya Wang for their advice in the production of this paper.

References

[1] Yue H, Zhao Y, Ma X, Gong J. Ethylene glycol: properties, synthesis, and applications. *Chemical Society Reviews*. 2012;41(11):4218–44. https://doi.org/10.1039/c2cs15359a.

[2] Fortune Business Insights. *Ethylene Glycol Market Size, Share | Global Report [2024-2032]*. https://www.fortunebusinessinsights.com/ethylene-glycol-market-110172 [Accessed 1 December 2024].

[3] Fan Y, Bao J, Shi L, Li S, Lu Y, Liu H, et al. Photocatalytic Coupling of Methanol and Formaldehyde into Ethylene Glycol with High Atomic Efficiency. *Catalysis Letters*. 2018;148(8):2274–82. https://doi.org/ 10.1007/s10562-018-2465-6.

[4] Deka TJ, Osman AI, Baruah DC, Rooney DW. Methanol fuel production, utilization, and techno-economy: a review. *Environmental Chemistry Letters*. 2022;20. https://doi.org/ 10.1007/s10311-022-01485-y.

[5] Ayodele OB. Plasma-Assisted One-Step Direct Methanol Conversion to Ethylene Glycol and Hydrogen: Process Intensification. *Energies*. 2024;17(13):3216. https://doi.org/10.3390/en17133216.

[6] Xie S, Shen Z, Deng J, Guo P, Zhang Q, Zhang H, et al. Visible light-driven C–H activation and C–C coupling of methanol into ethylene glycol. *Nature Communications*. 2018;9(1). https://doi.org/ 10.1038/s41467-018-03543-y.

[7] Cheng JK, Loh TP. Copper- and Cobalt-Catalyzed Direct Coupling of sp^3 α-Carbon of Alcohols with Alkenes and Hydroperoxides. *Journal of the American Chemical Society*. 2014;137(1):42–5. https://doi.org/ 10.1021/ja510635k.

[8] Zheng X, Liu Y, Yang Y, Song Y, Deng P, Li J, et al. Recent Advances in Cadmium Sulfide-Based Photocatalysts for Photocatalytic Hydrogen Evolution. *Renewables*. 2023;1(1):39–56. https://doi.org/10.31635/renewables.022.202200001.

[9] Wang J, Liu Z, Zheng Y, Cui L, Yang W, Liu J. Recent advances in cobalt phosphide-based materials for energy-related applications. *Journal of Materials Chemistry*. 2017;5(44):22913–32. https://doi.org/10.1039/c7ta08386f.

[10] Zhang H, Xie S, Hu J, Wu X, Zhang Q, Cheng J, et al. C–H activations of methanol and ethanol and C–C couplings into diols by zinc–indium–sulfide under visible light. *Chemical Communications*. 2020;56(12):1776–9. https://doi.org/ 10.1039/c9cc09205f.

[11] Deng R, Guo M, Wang C, Zhang Q. Recent advances in cobalt phosphide-based materials for electrocatalytic water splitting: From catalytic mechanism and synthesis method to optimization design. *Nano Materials Science*. 2022; https://doi.org/10.1016/j.nanoms.2022.04.003.

[12] Du H, Liu Q, Cheng N, Asiri AM, Sun X, Chang Ming Li. Template-assisted synthesis of CoP nanotubes to efficiently catalyze hydrogen-evolving reaction. *Journal of materials chemistry*. 2014;2(36):14812–6. https://doi.org/10.1039/c4ta02368d.

[13] Jiang D, Sun Z, Jia H, Lu D, Du P. A cocatalyst-free CdS nanorod/ZnS nanoparticle composite for high-performance visible-light-driven hydrogen production from water. *Journal of Materials Chemistry*. 2016;4(2):675–83. https://doi.org/10.1039/c5ta07420g.

[14] Zhang C, Huang Y, Yu Y, Zhang J, Zhuo S, Zhang B. Sub-1.1 nm ultrathin porous CoP nanosheets with dominant reactive {200} facets: a high mass activity and efficient electrocatalyst for the hydrogen evolution reaction. *Chemical Science*. 2017;8(4):2769–75. https://doi.org/10.1039/C6SC05687C

[15] Weng B, Wei W, Yiliguma Yiliguma, Wu H, Alenizi AM, Zheng G. Bifunctional CoP and CoN porous nanocatalysts derived from ZIF-67 in situ grown on nanowire photoelectrodes for efficient photoelectrochemical water splitting and CO_2 reduction. *Journal of Materials Chemistry*. 2016;4(40):15353–60. https://doi.org/ 10.1039/c6ta06841c.

[16] Ji L, Wang JY, Teng X, Meyer TJ, Chen Z. CoP Nanoframes as Bifunctional Electrocatalysts for Efficient Overall Water Splitting. *ACS Catalysis*. 2020;10(1):412–9. https://doi.org/10.1021/acscatal.9b03623

[17] Mahmood N, Yao Y, Zhang JW, Pan L, Zhang X, Zou JJ. Electrocatalysts for Hydrogen Evolution in Alkaline Electrolytes: Mechanisms, Challenges, and Prospective Solutions. *Advanced Science*. 2017;5(2):1700464. https://doi.org/10.1002/advs.201700464.

[18] Huang G, Huang Y, Ali A, Chen Z, Shen PK, Ni B, et al. Phase-controllable cobalt phosphide heterostructure for efficient electrocatalytic hydrogen evolution in water and seawater. *Electron*. 2024;2(3). https://doi.org/10.1002/elt2.58.

[19] Huang B, Wang F, Li H, Li Y, Li K. Synergistic role of Co and CoP on highly dispersed cobalt nanoparticles embeded in porous carbon for efficient PH3 decomposition. *Separation and Purification Technology*. 2024;348:127689. https://doi.org/10.1016/j.seppur.2024.127689.

[20] Fang L, Xie Y, Yang Y, Zhu B, Wang Y, Liu M, et al. Interfacial Electronic Modulation of Multishelled CoP Hollow Spheres via Surface Reconstruction for High-Efficient Hydrogen Evolution Reaction. *ACS Applied Energy Materials*. 2019;3(1):309–18. https://doi.org/ 10.1021/acsaem.9b01562.

Computational fluid dynamics modeling of Onyx liquid embolic system

Soumya Jonnalagadda[1] and Yi Jin[1]

[1] *Department of Chemical Engineering, Imperial College London, South Kensington Campus, SW7 2AZ, UK*

Abstract

Arteriovenous malformations (AVMs) are abnormal tangles of blood vessels that disrupt normal blood flow and pose significant treatment challenges. Onyx, a liquid embolic agent, is frequently used to occlude the AVM nidus by precipitating and solidifying upon injection into the blood. In this study, we utilized computational fluid dynamics (CFD) modeling in ANSYS ®Fluent to simulate the propagation and solidification of Onyx. The Volume-of-Fluid (VoF) method was employed to capture the interphase dynamics between Onyx and blood, while mass transfer was modeled to represent the diffusion of DMSO into blood, leading to the viscosity changes responsible for Onyx solidification. Sensitivity analyses were performed on mesh size, time step, and viscosity models to evaluate their influence on simulation accuracy and computational efficiency. A fine mesh and optimized time step captured sharp interphase boundaries and temporal characteristics critical to Onyx behavior, while incorporating a dynamic viscosity model improved the representation of the solidification process. These parameters enabled the identification of the most reliable properties for simulating embolization, including interfacial mass transport, viscosity evolution, and flow dynamics under different conditions. Although this model has not yet reached the stage of clinical validation, it provides a valuable framework for investigating the influence of injection parameters, such as flow rate and viscosity, on Onyx penetration and solidification in blood vessels. This research lays the groundwork for future studies aimed at identifying key factors responsible for embolization outcomes and potential post-embolization complications, ultimately contributing to safer and more effective preoperative planning for AVM treatment.

Keywords: Arteriovenous Malformation (AVM), embolization, Onyx, viscosity, nidus, hemodynamics, simulation, two-fluid model, scalar transport, Computational Fluid Dynamics (CFD), Volume-of-Fluid Method (VoF Method), surface tension.

1 Introduction

While numerous embolic liquid materials exist, each employing various strategies to occlude vessels, many are hindered by significant drawbacks that limit their effectiveness. A prime example of this is the category of embolic agents known as sclerosants, which achieve vessel occlusion by irreversibly damaging the endothelial lining of vessel walls, leading to fibrosis and eventual closure [1]. Although effective in some cases, sclerosants present notable limitations. These materials are inherently non-targeting, often dispersing beyond the intended treatment area. To mitigate the risk of drainage into surrounding tissues, the use of external occlusion balloon catheters is typically required [2]. However, the introduction of additional devices can increase complexity and the potential for complications, particularly in patients undergoing treatment for fistulas, which are commonly associated with arteriovenous malformations (AVMs) [3].

Onyx, a non-adhesive liquid embolic agent, addresses many of these limitations and offers distinct advantages over traditional materials. Unlike adhesive agents that can bond uncontrollably with vessel walls and surrounding tissues, Onyx is designed to remain cohesive yet non-sticky, allowing for precise control during embolization procedures. Onyx consists of an ethylene-vinyl alcohol (EVOH) copolymer dissolved in dimethyl sulfoxide (DMSO), with added trace amounts of tantalum powder for radio opacity, which enables real-time visualization under fluoroscopy. The overarching mechanism is defined by EVOH. When Onyx comes into contact with blood, the DMSO diffuses away, causing the EVOH to precipitate and solidify. This solidification process effectively occludes the targeted blood vessels by forming a cohesive mass that blocks blood flow.

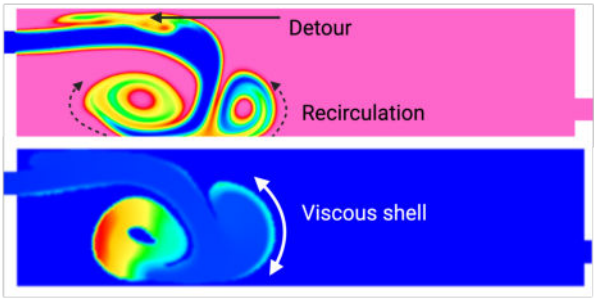


Figure 1: Illustration of detour and recirculation effects

Onyx is currently available for treatment as Onyx 18, 20, and 34—each with varying ratios of DMSO and EVOH, as per Table 1. Its ability to be produced at varying viscosities enables a tailored approach to treatment, allowing for controlled injections that reduce the likelihood of non-target embolization. The material's unique properties facilitate the gradual delivery of Onyx, enabling precise occlusion of target vessels with a reduced risk of unintended dispersion. The implications of this for a multiphase computational fluid dynamics (CFD) model are significant, as the different formulations result in variations in viscosity, propagation, and solidification behaviors.

A key feature of Onyx is its lack of fragmentation during injection [4], a property that distinguishes it from other embolic agents. This behavior, often described as exhibiting a "lava-like" flow, ensures a cohesive and predictable material trajectory during administration. By maintaining structural integrity, Onyx minimizes the risk of embolic material mi-

gration, which can otherwise result in complications such as tissue ischemia, damage to surrounding healthy vessels, or incomplete occlusion.

Onyx formulation	EVOH (%)	DMSO (%)
Onyx 18	6.0	94.0
Onyx 20	6.5	93.5
Onyx 34	8.0	92.0

Table 1: Onyx formulations in centipoise (cP) and their corresponding compositions.

The controlled and cohesive delivery of Onyx is further enhanced by its viscous consistency and non-adhesive nature, which allow for slow and deliberate treatment of vascular malformations. These characteristics are particularly beneficial in treating AVMs, where precision is critical to achieving successful outcomes while minimizing collateral damage. In addition, the unique rheological properties of Onyx make it suitable for a wide range of vascular conditions, offering versatility and adaptability that are often lacking in other embolic agents.

2 Background

Research to date has primarily examined the impact of embolization on pressure, concluding that increased embolization correlates with heightened pressure [5]. Other studies focus on determining safe embolization scenarios [6] by using a one-dimensional two-phase filtration model and real patient clinical data - observing the process as an optimal control problem and utilizing algorithms to find the best control functions that achieve the desired therapeutic outcomes. Jain., *et al*, [7] propose graphics processing unit (GPU) computation to replicate more complex nidus, demonstrating the potential of advanced computational methods in accurately modeling the hemodynamics of arteriovenous malformations (AVMs).

A key effect that defines the embolization process is the development of the viscous shell. This effect is visualized in Fig. 1, whereby the expansion of this shell forms the first blocking wall for the nidus and ensures that the embolic material stays localized within the target region. This viscous shell was first observed and described by Orlowski., *et al*, 2012 [8], who validated the described computational findings with clinical imaging techniques. The clinical imaging provided robust confirmation of the model's accuracy and applicability in real-world scenarios. Orlowski., *et al* also observed the detour and recirculation effect, also described in Fig. 1. Whilst we intend on using this as a point of validation for our simulations, we intend on expanding upon this work by analyzing time-step and mesh size to optimize embolization in both exactly and nearly static nidus, as well as moving blood scenarios.

In this paper, we simulate the propagation and solidification of Onyx after injection through the use of CFD. By leveraging its non-adhesive, highly viscous, and cohesive properties, Onyx allows for precise and effective occlusion with minimal complications. This paper aims to model these effects using a simplified vascular geometry.

3 Methodology

3.1 Simulation Properties

All simulations discussed below use a simplified rectangular geometry of size 25 cm $\times$ 6 cm, with tubular inlet and outlet diameter of 1cm. This geometry allows us to focus on key phenomena and clarity in result interpretation [8]. Simplifications are made to reduce computational cost while maintaining the essential physics of the embolization process.

The viscosity and density of blood are maintained at 0.004 kg/m/s and 1063 kg/m^3, respectively, under standard conditions. These values are consistent with physiological parameters and ensure the accurate representation of blood flow dynamics. The simulation assumes a steady flow regime, with no pulsatile effects, to simplify the analysis and focus on the embolization process. There is also an additional transient flow model considered whereby blood is introduced with an inflow velocity of 0.05 m/s, simulating a more realistic clinical scenario.

The embolization of vessels or niduses of interest with Onyx is the result of precipitation of EVOH and tantalum at the interface where glue and blood interpenetrate into each other. Blood, acting as the anti-solvent, has the affinity to dissolve DMSO only and leads to the formation of tiny solid particles of solutes suspended. This phenomenon is reflected by the increasing viscosity at the phase boundary. To effectively represent the propagation and solidification of the glue material in blood vessel, a multiphase model using Volume of Fluid (VoF) method in ANSYS ®Fluent is proposed, where both phases are treated as incompressible [9]. The intrinsic equations governing the phases are the continuity (Eq. 1) and conservation of momentum equation (Eq. 2) for the mixture:

$$\frac{1}{\rho_q}\left[\frac{\partial}{\partial t}\left(\alpha_q\rho_q\right)+\nabla\cdot\left(\alpha_q\rho_q\vec{v}_q\right)\right]=S_{\alpha_q}+\sum_{p=1}^{n}\left(\dot{m}_{pq}-\dot{m}_{qp}\right) \quad (1)$$

$$\frac{\partial}{\partial t}(\rho\vec{v})+\nabla\cdot(\rho\vec{v}\vec{v})=-\nabla p+\nabla\cdot\left[\mu\left(\nabla\vec{v}+\nabla\vec{v}^{T}\right)\right]+\rho\vec{g}+\vec{F} \quad (2)$$

With volume fraction α following the sum:

$$\sum_{q=1}^{n}\alpha_q=1 \quad (3)$$

Where subscripts q and p correspond to either glue and blood respectively, α is the volume fraction, ρ is the density, v is the velocity, $S\alpha$ is generation rate, m_{pq} and m_{qp} are the mass transfer from phase p to q and reversed respectively, μ is the dynamic viscosity and F represents the additional forces exerted on the mixture. Here, both source and mass transfer terms are set to zero. Additionally, the flow neglects gravitational body forces, as their influence is negligible compared to the driving pressure gradients in this scenario.

Beyond this, the continuity equation is reformulated in the implicit discrete form to fit the finite volume approach

used in ANSYS ®Fluent:

$$\frac{\alpha_q^{n+1}\rho_q^{n+1} - \alpha_q^n \rho_q^n}{\Delta t} V + \sum_f \left(\rho_q^{n+1} U_f^{n+1} \alpha_{q,f}^{n+1}\right) = \left[S_{\alpha_q} + \sum_{p=1}^{n} (\dot{m}_{pq} - \dot{m}_{qp})\right] V \tag{4}$$

The volume fraction cutoff in equation above is 10^{-6}. The interface is modeled as dispersed so that transitional concentrations will appear at the boundary, demonstrating the interpenetration between two fluids. All the solidification processes are assumed under fixed body temperature, so no energy equation is activated.

The dynamic viscosities for the blood and glue are set to be 0.004 and 0.018 $kg/m/s$ consistently, whereas in the transient model, the viscosity of the mixture fluid was assumed to follow a linear relationship with the concentration of a passive scalar that was created to represent the solvent, namely, DMSO :

$$\mu(x,t) = \mu_{\text{onyx}} \left(1 + \beta \left(1 - \gamma(x,t)\right)\right) \tag{5}$$

Where the value of scalar, $\gamma(x,t)$, is normalized between 0 and 1, μ_{onyx} is the Onyx viscosity, and β is the enlargement factor chosen as 100, as opposed to 500 proposed by Orlowski., *et al*, 2012 [8]. This is simply to ensure that the viscosity for these simulations remain within a realistic region. This demonstrates that the viscosity of the glue corresponds to its initial value when the local scalar value is 1 and increases as the scalar value decreases. The scalar, initialized with a value of 1 and injected alongside the glue from the feed vessel, is used to simulate the concentration profile of the solvent (DMSO) as it undergoes convection and diffusion between the two fluids. This process is implemented using the User-Defined Scalar (UDS) function in Fluent, which solves the advection equation as follows:

$$\frac{\partial \alpha_p \rho_p \gamma_p^k}{\partial t} + \nabla \cdot \left(\alpha_p \rho_p \vec{v}_p \gamma_p^k - \alpha_p \Gamma_p^k \nabla \gamma_p^k\right) = S_p^k \tag{6}$$

Where α, ρ and v are the the volume fraction, density, and velocity of phase p as above. The UDS index is given as k, and Γ_p^k is the diffusion coefficient for the scalar in phase p, with units of $kg/m/s$. The source term is given as S_p^k, which we have taken to be zero.

Two scenarios are tested - one is for a nearly static bloodstream, and one for flowing blood. In the sensitivity analysis discussed in Section 3.2, the glue is injected with a constant velocity of 0.15 m/s into static blood. In the varying viscosity simulations for both scenarios in Section 4.1 and Section 4.2, the injection lasts for 2.5 s. This is because in real surgery, the clinician may adopt stage wise operations while monitoring the patient's cerebral conditions. The outlet vessel is set at a gauge pressure of 0 Pa, and the operating pressure is maintained at atmospheric conditions over the simulation. Due to the low inlet velocity which gives a maximum Reynolds number of only 88, only laminar flow is used. A piecewise function dependent on the volume fractions is employed to guarantee the scalar motion is convection dominated in glue-rich region and diffusion dominated in the mixing area. It is expected that a large viscosity will be observed at the interface where the scalar diffuses from glue to blood.

As for solution settings, a default convergence criteria of 10^{-3} is used. The solver is based on the coupled pressure-velocity scheme [10] and second order accuracy methods are implemented for spatial and temporal discretization. PRESTO! [11] approach is applied, particularly for the pressure term in favor of convergence. The time-step is chosen to be 0.01 s as suggested by [8], discussed further in Section 3.2. In each time-step, at most 20 iterations are solved to accelerate computing.

3.2 Sensitivity Analysis

Prior to implementing variable viscosity models, sensitivity analyses were conducted on different mesh sizes and time-steps to ensure accurate capture and efficient representation of key flow pattern characteristics. The results from these initial simulations served as a reference point for the subsequent implementation of variable viscosity models in Section 4.1.

3.2.1 Mesh Size

Given the lack of literature on mesh size, three mesh sizes of 2 mm, 1 mm, and 0.5 mm were chosen, providing minimum cells per gap of 5, 10, and 20, and total numbers of 9,142, 36,294, and 144,650. It was expected that the time required for simulations would increase in proportion to the total number of cells. All three models were run with a time-step of 0.01 s, as suggested by Orlowski et al., [8].

The results are shown in Fig. 2, which represent the concentration map for each mesh size at different times after injection. Note that color bar is scaled using the volume fraction. As illustrated in the color bar, the region with warmer colors on the right indicates a higher concentration of blood, while colors closer to the blue side signify more glue in that space. A similar pattern is observed across all three sets of profiles: as the glue progresses forward, two side streams split from the frontier and recirculate backward. This may be explained by the greater viscosity of the glue compared to blood. As more glue enters the nidus, the existing glue acts as an effective blocker, forcing the subsequent flows to detour toward the blood on the two sides, where lower viscosity provides less resistance. Since the inlet and outlet vessels are not centered, the velocity field in the nidus is asymmetric, causing the glue flow to bend. Together with the effect of viscosity, the glue can turn at a right angle and settle in the first half of the nidus.

At first glance, the most obvious and significant difference between the different mesh sizes is that finer meshes produce a clearer phase boundary, whereas in coarser meshes, the interface appears more vague with more medium concentrations in the remixing region (recirculation). To quantify the influence of mesh sizes on accuracy, pressure values were extracted from a vertical line 3 cm from the left nidus wall at 2.5 s post-injection (the last frame). In Fig. 3, the discrete pressures are shown in a scatterplot, with values on the left vertical axis. The relative errors from the model with the finest mesh are also plotted as a dotted line and superimposed in the same graph, with values on the right vertical axis.
From Fig. 3, it can be concluded that both models with mesh sizes of 0.5 mm and 1 mm converge to the same result, while the model with the coarsest mesh diverges significantly near

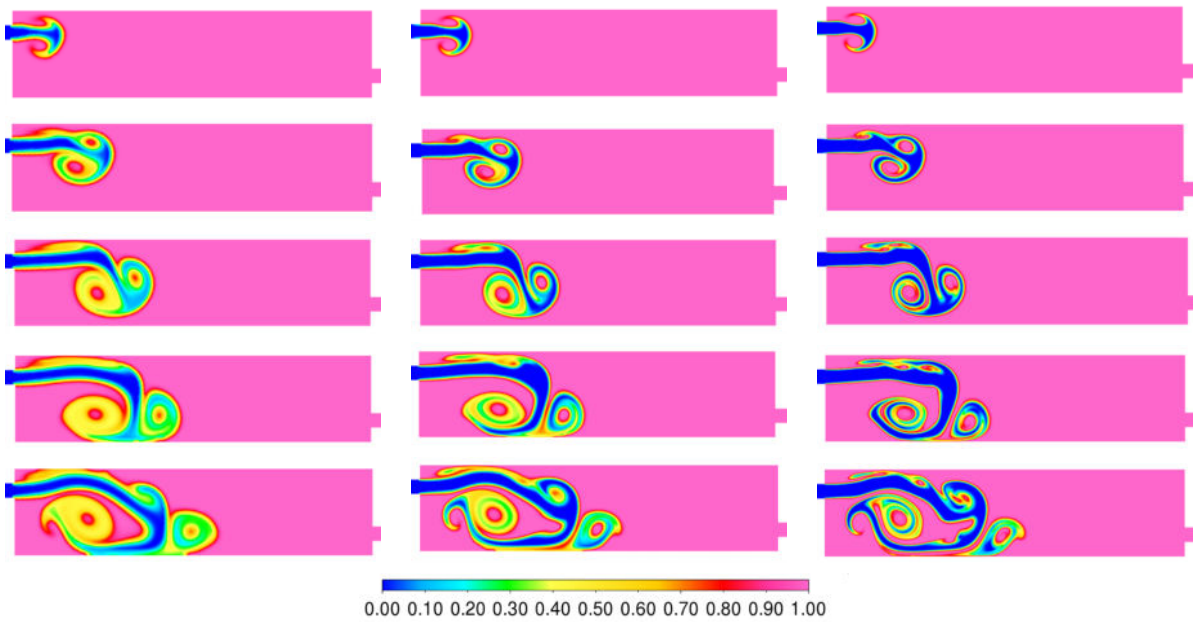


Figure 2: Concentration contours for course mesh size of 2 mm (left column), medium mesh size of 1 mm (central column), and fine mesh size of 0.5 mm (right column).

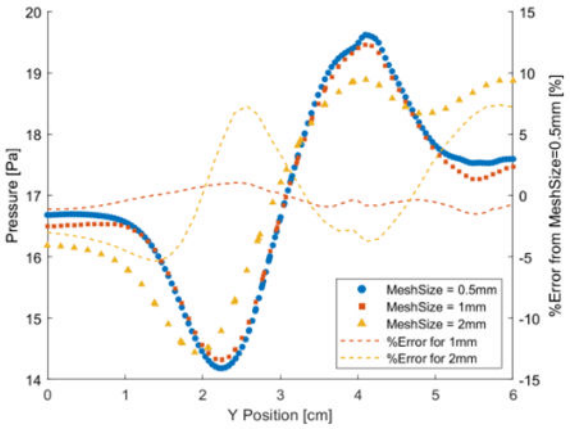


Figure 3: Pressure relationship graph with mesh size and associated error

the upper nidal wall ($y = 6\,\text{cm}$). There also appears to be a phase difference in the region near the bottom ($y = 0\,\text{cm}$). The peak percentage error for the coarse-mesh model reaches approximately 8% compared to the fine-mesh model, whereas the medium-mesh model maintains its relative error mostly below 1%. However, transition from the medium mesh to the fine mesh would increase the computational time by 4 times, given the shift from the coarse to the medium one only doubles the time. Considering the trade off between mesh precision and computational time, a mesh size of 1 mm was chosen for further simulations.

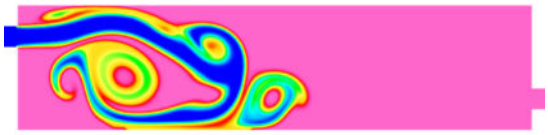
3.2.2 Time-step

To ensure the solution converges during the embolization process, testing on time-step selection can be conducted, with the current time-step being 0.01 s. Two smaller time-steps, 0.005 and 0.0025 s were taken for cohesive validation. This procedure follows that of the mesh size sensitivity analysis. Fig. 4 depict the volume fractions at a flow time of 2.5 s for the specified time-steps. The results indicate that any observed differences between the simulations are negligible. This conclusion is further supported by the pressure profiles at X=3cm shown in Fig. 5, where the results for time-steps of 0.01 s and 0.005 s closely overlap. Notably, the model with the shortest time-step produces slightly higher amplitude values. Additionally, the relative errors for the 0.01 s and 0.005 s time-steps remain below 3%.

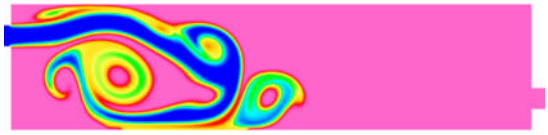
These findings confirm that a time-step of 0.01 s is sufficiently small to capture all critical flow features accurately. Further refinement, such as halving or quartering the time-step, would provide only marginal improvements in accuracy while significantly increasing computational costs—requiring at least four times the computing time for each reduction.

4 Results / Discussion

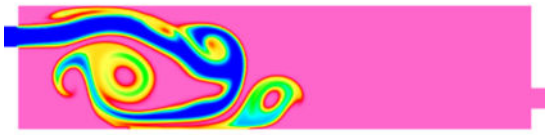
After determining the appropriate mesh size and time-step, the next logical steps lie in examining mixture viscosity to succinctly understand the effects of the sensitivity analysis parameters on the solidification and propagation of the glue. This model relies on the boundary conditions and solution settings described at the beginning of Section 3, and Eq. 5 is used to model the scalar-viscosity relationship.



(a) Time-step of 0.01 s.



(b) Time-step of 0.005 s.



(c) Time-step of 0.0025 s.

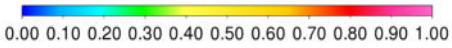


Figure 4: Concentration maps taken at a flow time of 2.5 s for different time-steps. The colorbar below illustrates the scale used in the analysis.

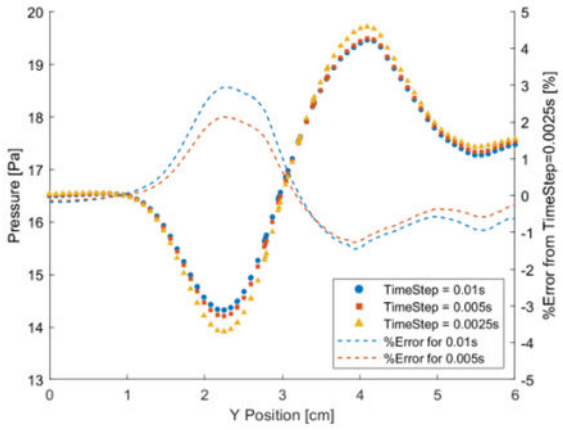


Figure 5: Pressure relationship graph with time-step and associated error

4.1 Static blood flow

The above selections were first applied to a static blood flow. As expected, a similar 'umbrella-like' flow pattern is obtained in the concentration profiles. As for the scalar motion shown in Fig. 8, it follows a very similar trend as the concentration, but spreads over and forms a vaguer boundary. This effectively reflects the possible solvent transport among the glue and blood phases. High viscosities are witnessed as expected at the interface where the scalar advances faster than Onyx and displays a negative gradient. Over the whole operation, the maximum viscosity is found to be around 0.4 kg/m/s, which is 100 times larger than the blood viscosity.

Compared to simulations with fixed viscosities, there seems to be greater resistance against recirculation as the stream redirects vertically downwards at earlier flow times. This may be explained by the viscous shell in Fig 7, which bounds the head of flow and restricts splitting. Over the sampled time, no specific dissociation of side streams from the main body is observed before the glue gets solidified. This can be attributed to the viscous shield. Both changes are positive, as they avoid any unpredictable pressure variation inside the vessel due to the flow pattern. Meanwhile, the operation is considered safer because any escaping glue would not bypass the nidus needed to undergo embolization and thereby cause problems associated with control and supervision of the process.

After 2.5 s (the fifth image in each set of figures), the injection is terminated. At this point, Onyx finishes settling and hardens evenly due to the transport of the scalar from the center. This happens very quickly, with a noticeable increase in viscosity to 0.3 kg/m/s for most regions within the later second. The final structure is promising, as the solidified glue blocks the entire nidus successfully.

4.2 Transient blood flow

Thus far, the models have demonstrated scenarios for embolization in exactly and nearly static nidus. This subsection introduces a scenario whereby a blood inflow velocity of 0.05 m/s is introduced from the left nidal wall - this is 1/3rd of the glue injection rate. The outlet on the bottom right of the geometry remains as the only exit. The results are observed in Fig. 9, 10 and 11.

Compared to Section 4.1, a thoroughly different flow pattern is achieved this time. Since the inlet injection flow of Onyx is surrounded by moving blood flow, a slim pathway is formed with a much smaller recirculation and detour initially, as observed clearly in Fig. 9 - highlighting the effects of advection. In spite of that, a viscous shell is again formed during the simulation. Although the glue fails to completely settle, precipitation is observed quickly once the Onyx inflow injection is terminated, as per Fig. 10. This effect occurs due to a considerable amount of scalar adverted away via the moving blood, leading the visocity of glue to grow dramatically to around 0.35 kg/m/s. The front of the 'umbrella' is exposed when it approaches the outflow vessel, and it occupies this region entirely - indicating accomplished embolization. It must be noted that is not the case in a clinical setting, as the pressure would drive flow instead, causing glue to slow down as its viscosity increases and obstruct the outlet.

5 Conclusion & Outlook

In this study, we developed a model to simulate the interphase movement and solidification of Onyx when injected into blood, following the methodology of Orlowski *et al.* [8]. Using this replicated model, we conducted sensitivity analyses to identify the optimal mesh size and time-steps, balancing computational cost and accuracy. We then investigated two scenarios with varying viscosity via a static and moving blood flow by implementing a simple linear relationship with a passive scalar to represent the transport of DMSO from Onyx into the blood.

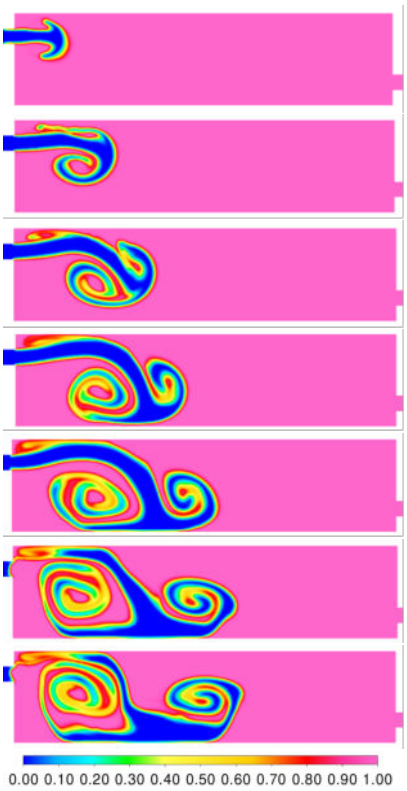


Fig. 6: Concentration maps for static blood varying with time for a 1 mm mesh size and 0.01 s time-step. The scalar motion reflects possible solvent transport among glue and blood phases.

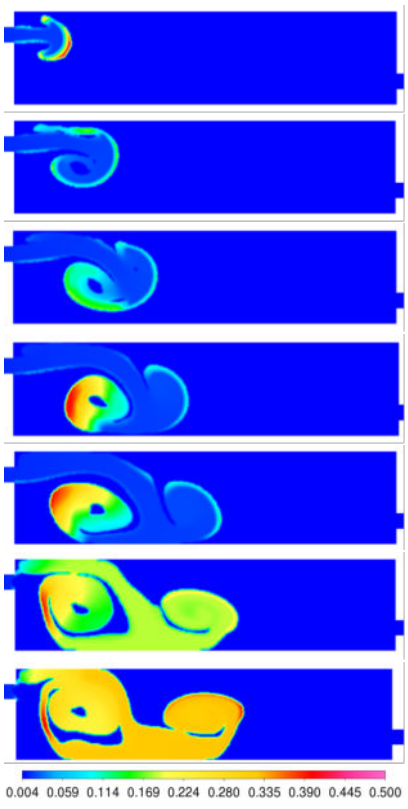


Fig. 7: Viscosity contour maps for static blood in units of $[kg/m/s]$ varying with time for a 1 mm mesh size and 0.01 s time-step. The Onyx settles and hardens evenly after 2.5 s.

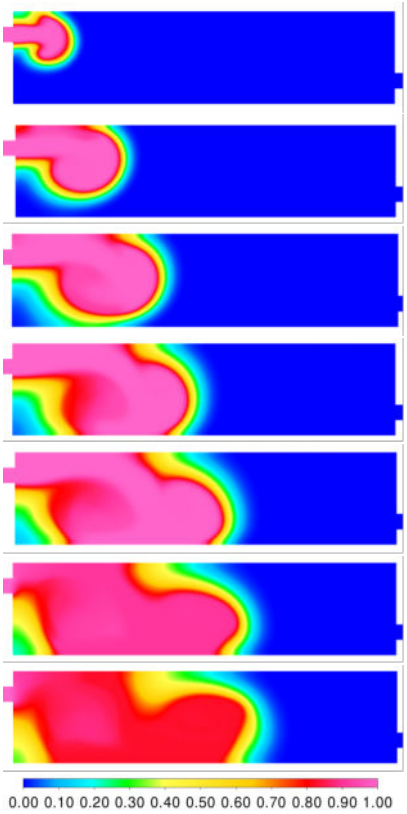


Fig. 8: Scalar maps for static blood varying with time for a 1 mm mesh size and 0.01 s time-step. As for scalar motion, it advances faster than glue in early stages, leading to high viscosities at the interface.

Our simulations revealed that the formation of a viscous shell in stationary or nearly still blood had positive effects in mitigating recirculation and detour phenomena observed in fixed-viscosity simulations. The increasing viscosity also reduced the likelihood of Onyx dissociation, enhancing stability and reducing uncertainty in potential clinical applications. In simulations with flowing blood, the model demonstrated successful embolization, although the behavior differed. The Onyx solidifies more rapidly in moving blood due to the convective transport of the scalar, ultimately causing blockage at the outlet.These findings highlight the importance of viscosity variations and the formation of a viscous shell in improving the effectiveness and stability of Onyx as an embolic agent. The model's ability to predict these behaviors offers valuable insights for clinical applications, supporting the potential for more precise and controlled embolization treatments.

Future work should focus on testing additional boundary conditions, such as constant pressure drops at the inlet, to better replicate clinical scenarios. It is also worth including interphase mass transfer as DMSO transfers from Onyx to the blood. Moreover, employing exponential viscosity-scalar relationships could better reflect the dynamics of EVOH precipitation, especially in the supersaturation zone required for nucleation and growth. Considering the complex structure of fistulous and plexiform compartments in real AVMs, integrating porous medium models would also enhance the simulation's applicability to real-life conditions.

Furthermore, transitioning to platforms for advanced multiphase-multicomponent modeling with higher capability for freedom would allow for sharper and more biologically accurate boundaries, bridging the gap between computational modeling and clinical practice. This step would help refine the representation of Onyx behavior during embolization and its interaction with blood, ultimately contributing to the development of optimized treatment protocols for AVMs and other vascular abnormalities.

This study focused on modeling the interphase movement and solidification of Onyx using the Volume-of-Fluid (VoF) methodology. While the VoF approach effectively captured critical dynamics, incorporating the Compressive Continuous Species Transfer (C-CST) method could significantly enhance the accuracy of species transfer simulations. The C-CST approach, as proposed by Maes et al. (2018) [12], is specifically designed to reduce numerical diffusion by introducing a compressive term in the species conservation equation. This feature is particularly valuable for high Péclet number scenarios, such as blood flow, where convection dominates diffusion. The

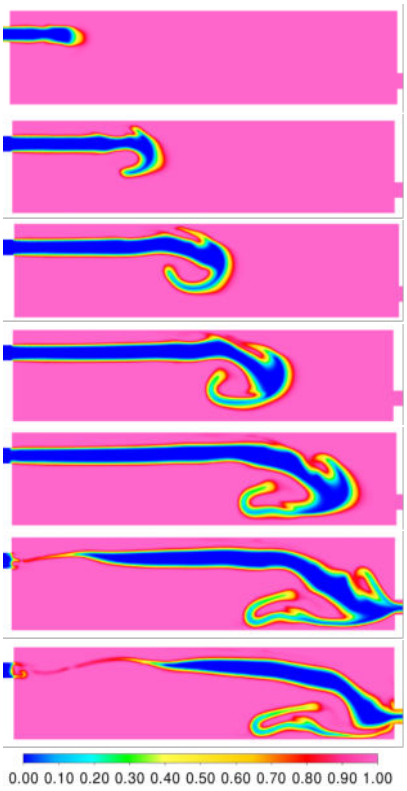


Fig. 9: Concentration maps for moving blood flow varying with time for a 1 mm mesh size and 0.01 s time-step.

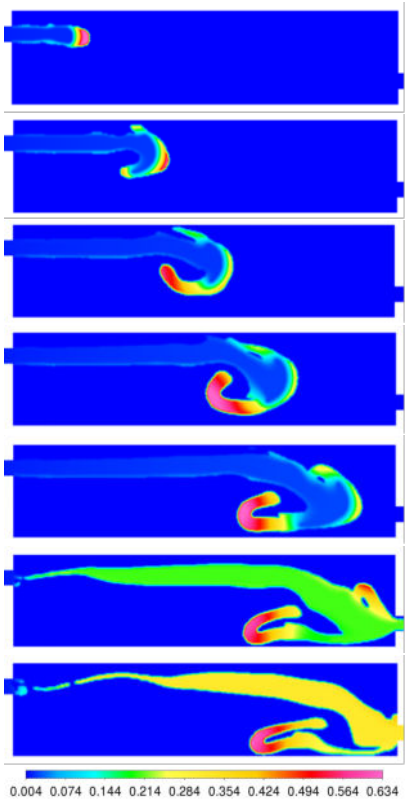


Fig. 10: Viscosity contour maps for transient blood flow in units of $[kg/m/s]$ varying with time for a 1 mm mesh size and 0.01 s time-step.

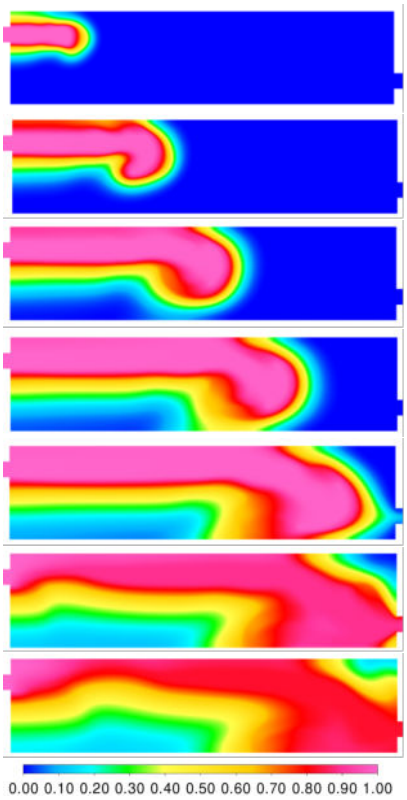


Fig. 11: Scalar maps for transient blood varying with time for a 1 mm mesh size and 0.01 s time-step.

addition of advection flux adjustments and sharp interface preservation terms within the C-CST framework ensures a more accurate representation of species transfer across fluid interfaces. The use of C-CST could also address some of the limitations observed in the VoF method, making the simulations more cohesive.

6 Acknowledgments

We extend our heartfelt gratitude to Dr. Yu Zhu for his invaluable insights, guidance, and unwavering support throughout the course of this study.

References

[1] Pal A, Blanzy J, Gómez KJR, Preul MC, Vernon BL. Liquid embolic agents for endovascular embolization: A review. Gels. 2023;9(5):378.

[2] Shen Y, Wang D, Fan X, Zheng L, Su L, Yang X. Ethanol embolization of arteriovenous malformations in the buttock: ten-year experiences in diagnoses and treatment options. Orphanet Journal of Rare Diseases. 2024;19(1):195.

[3] Tomlinson FH, Rüfenacht DA, Sundt TM, Nichols DA, Fode NC. Arteriovenous fistulas of the brain and the spinal cord. Journal of neurosurgery. 1993;79(1):16-27.

[4] Senturk C. Mechanical removal of migrated Onyx due to microcatheter rupture during AVM embolization: a technical case report. Cardiovascular and interventional radiology. 2015;38:1654-7.

[5] Henkes H, Gotwald T, Brew S, Kaemmerer F, Miloslavski E, Kuehne D. Pressure measurements in arterial feeders of brain arteriovenous malformations before and after endovascular embolization. Neuroradiology. 2004;46:673-7.

[6] Sharifullina T, Cherevko A, Ostapenko V. Optimal control problem arising in mathematical modeling of cerebral vascular pathology embolization. Scientific Reports. 2022;12(1):1302.

[7] Jain MS, Do HM, Wintermark M, Massoud TF. Large-scale ensemble simulations of biomathematical brain arteriovenous malformation models using graphics processing unit computation. Computers in Biology and Medicine. 2019;113:103416.

[8] Orlowski P, Summers P, Noble JA, Byrne J, Ventikos Y. Computational modelling for the embolization of brain arteriovenous malformations. Medical engineering & physics. 2012;34(7):873-81.

[9] ANSYS Inc . ANSYS Fluent User's Guide. Canonsburg, PA, USA; 2023. Version 2023 R2. Available from: https://www.ansys.com.

[10] Latimer B, Pollard A. Comparison of pressure-velocity coupling solution algorithms. Numerical Heat Transfer. 1985;8(6):635-52.

[11] Bershad BN, Lazowska ED, Levy HM. Presto: A system for object-oriented parallel programming. Software: Practice and Experience. 1988;18(8):713-32.

[12] Maes J, Soulaine C. A new compressive scheme to simulate species transfer across fluid interfaces using the volume-of-fluid method. Chemical Engineering Science. 2018;190:405-18.

Techno-Economic Analysis and Life Cycle Assessment of Sustainable Aviation Fuel: Comparing Direct Air Capture and Point-Source CO_2 Capture Technologies

Isabelle Beddows (02025680), Ashley Lim (02040695)
Department of Chemical Engineering, Imperial College London, U.K.

Abstract

Global warming has been exacerbated by the rapid growth of air travel and its rising carbon emissions. To address this, sustainable aviation fuels (SAF) have emerged as a promising alternative. SAFs are commonly produced via the Fischer-Tropsch process, with the carbon feedstock being the critical point of variation that determines the SAF eligibility. Recently, the European Union has recognised recycled carbon aviation fuels as an eligible form of SAF. This report evaluates the implications of this policy change and examines the feasiblility of incorporating point-source carbon capture (PSC) as a viable pathway for SAF production to that of direct air capture pathways. The mean costs of SAF production were calculated as €14/kg_{SAF} for direct air capture and €10/kg_{SAF} for point source capture, while mean emissions were found to be 40 gCO_2/MJ for direct air capture and 44 gCO_2/MJ for point source capture. The EU's decision to recognise recycled aviation fuels as SAFs represents a significant and positive step forward. This policy has the potential to drive technological advancements and accelerate the adoption of SAF across EU.

1. Introduction

Technological advancements in the aviation industry have made air travel more affordable and accessible, fuelling rapid growth in the aviation market. The aviation sector is responsible for 2.5% of global CO_2 emissions and contributes 4% to climate change[1]. As the demand for air travel surges, the industry faces an urgent challenge to reduce their carbon footprint and mitigate its impact on global warming.

Sustainable aviation fuels (SAF) present a promising solution to decarbonise aviation by replacing conventional aviation fuels (CAF). The UK mandate declares that SAF must meet a strict sustainability criterion with a minimum 50% emission reduction as compared to CAF[2]. This translates to a maximum allowable emission of 44.5 gCO_2/MJ to be eligible as SAF. As a leading player in sustainability, the European Union (EU) has recognised the importance of transitioning to SAFs. Previously, only direct air capture was allowed for valid SAF production in the EU. However, the EU recently passed a policy that makes recycled carbon aviation fuels eligible as SAF[3]. The EU has also set ambitious targets for SAF adoption, aiming to incorporate 2% SAF in 2025, 6% in 2030 and 70% in 2050.

Fischer-Tropsch (FT) process is one of the commonly used pathways for SAF production. Traditionally, biofuel have been utilised as the primary source of carbon dioxide (CO_2) in the feedstock. This approach involves biomass gasification to produce syngas, for conversion into liquid hydrocarbons. While biofuels are often favoured as a sustainable alternative to fossil fuels, they pose several drawbacks to society. Several review papers have identified key issues such as direct and indirect land-use changes causing significant GHG emissions and a loss of biodiversity. Resource competition for first-generation feedstocks could drive up food insecurity. Biofuels also create environment impacts like increased water footprint, acidification and eutrophication[4]. Hence, there has been shifting sentiments to use Carbon Capture and Storage (CCS) technologies to provide the CO_2 feedstock in the FT process.

The water-gas shift reaction is responsible for adjusting the hydrogen (H_2) to carbon monoxide (CO) ratio in the syngas for optimal FT synthesis. The main synthesis reaction then converts the syngas, a mixture of CO and H_2 into predominantly linear hydrocarbons.

$$CO_2 + H_2 \rightleftharpoons CO + H_2O \quad (1)$$

$$(2n+1)H_2 + nCO \leftrightharpoons C_nH(2n+2) + nH_2O \quad (2)$$

Two approaches to carbon capture were considered for SAF production. Direct Air Capture (DAC) involves the removal of CO_2 directly from the atmosphere for storage or utilisation. Point-Source Capture (PSC) of industrial emissions was proposed as it falls under the eligibility criteria for SAF production. PSC targets concentrated CO_2 emissions from industrial facilities before they are released into the atmosphere.

To comprehensively review the environmental impacts of SAFs, a Life-Cycle Assessment (LCA) would be essential. LCA is a tool used to assess the environmental impacts associated with all stages of a product over the entire period of its life, from acquiring raw material, production, usage and disposal[5]. Hence, providing a framework to assess the potential of SAFs to decarbonise aviation.

2. Background

2.1 Current Studies and Knowledge Gaps

Due to the recent EU policy change[3] of allowing SAF to be made from "recycled CO_2", there is a need to evaluate what impact this would have on the net reduction of CO_2 emissions. Previous studies that look at SAF production via the FT process mainly focus on using CO_2 from DAC as a feedstock,[6 7] so it is important to compare this scenario with CO_2 from PSC. They also rarely calculate emissions and costs for different EU targets, which means there is a lack of clarity on what the actual impact of implementation would be.

Furthermore, there is rarely a breakdown on which technologies contribute most to total cost and CO_2 emissions. This prevents clarity on assessing which technologies need the most cost and emission reduction and can make it more difficult for decisions to be made

on what future research is most important to decrease these two variables.

Finally, literature studies do not always compare environmental impacts to SAF emission reduction targets, which prevents the opportunity to assess whether extra carbon removal is needed and does not provide an accurate final cost.

2.2 Project Objective

This study aims to evaluate the economic viability and environmental implications of the EU's recent policy change that permits PSC for SAF production. By referencing the UK mandate of 50% emission reductions for SAF eligibility as a baseline, a direct comparison between DAC and PSC for SAF production will be provided. The objective is to determine the feasibility of scaling up the carbon capture technologies for SAF production, to achieve the EU's SAF adoption targets by 2050.

2.3 Fischer-Tropsch Pathway to SAF

In addition to CO2 capture, there are several other technologies and resources needed to complete synthesis of SAF via FT. A simplified block diagram for the CO2 feedstock DAC route showing the technologies involved and the inputs and outputs is displayed below in Figure 1, and for PSC in Figure 2.

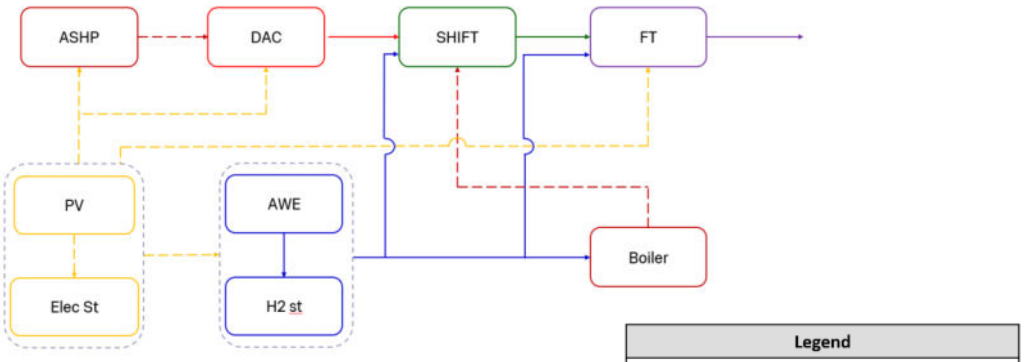


Figure 1 – Block Diagram for DAC route

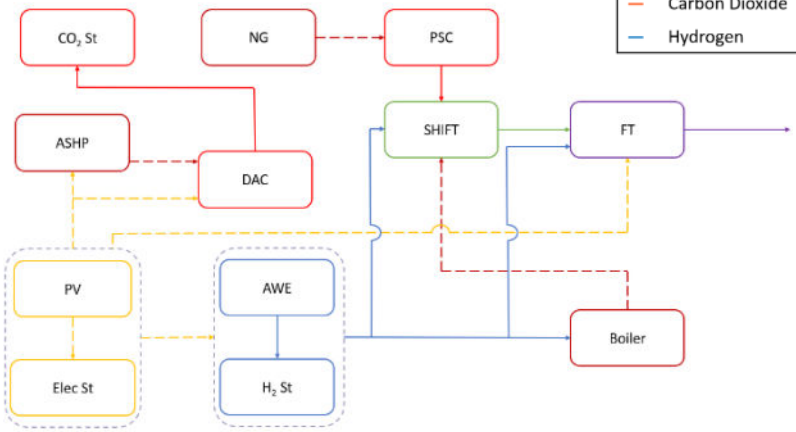


Figure 2 – Block Diagram for PSC route

The technologies needed for both scenarios are: alkaline water electrolysis (AWE) for H2 production, water gas shift reactor (SHIFT) for CO production, Fischer-Tropsch (FT) reactor for SAF production, hydrogen boiler for thermal heat generation and solar photovoltaics (PV) for electricity generation. For the DAC route, a heat pump was used to supply thermal energy, and for PSC natural gas was chosen. Justification of all technologies are detailed in the following Section 2.4.

2.4 Justification

Electricity: To minimise energy related CO_2 emissions, electricity was chosen to be sourced from solar photovoltaics (PV). Although solar has a lower capacity factor[8] than other renewable energy sources, it has less geographical limitations compared to HEP, geothermal and offshore wind energy. Onshore wind output is less predictable[9], which would complicate energy storage planning so was not chosen. PV was selected over solar CSP, as it is a more mature[10] technology with lower upfront costs. Due to solar intermittency, energy storage would need to be installed to compensate, and PV would need to be oversized.

Electricity Storage: Lithium-ion batteries were chosen due to their high energy density[11] and low maintenance costs. Although they have higher upfront costs than lead-acid batteries, they are more cost effective for long term applications, which is ideal for continuous SAF production.

Heating: Air source heat pump (ASHP) was chosen for low temperature heating such as in DAC, where electricity is converted into thermal energy. An air source heat pump was chosen over ground source heat pump, as the latter has more restrictive installation[12] requirements. Hydrogen boiler was selected for high

temperature SHIFT reaction, as it has lower carbon emissions than fossil fuels and leverages existing hydrogen infrastructure given that H2 is a reactant in the process. Natural gas was used for PSC. Although this analysis did not include capturing natural gas combustion emissions, they could be theoretically captured in the existing PSC system[13] which would reduce CO_2 emissions and wouldn't require new infrastructure for say hydrogen boilers. Electric heaters were considered but not chosen as they are currently unproven in industrial applications. Off-gas burning[14] from the FT process was used to satisfy heat requirements for distillation columns. The CO_2 emissions were then captured by point source and recycled back as feed. This system helps minimise emissions, whilst maximising resource utilisation.

Hydrogen: Alkaline water electrolysis was selected over PEM electrolysis for H2 production, as it is a more mature and scalable technology which has lower capital costs. Despite electrolysis producing less than 1%[15] of global hydrogen, fossil fuels were avoided as a hydrogen source due to their high emissions. For intermediate storage, compressed hydrogen gas storage[16] and liquified hydrogen storage[17] were chosen over salt caverns. Although salt caverns are cheaper, they have more geographical limitations.

Water gas shift reaction: The SHIFT reactor is needed as it converts H2 and CO_2 at 800C into CO, which is the feedstock for the FT reaction. Purification of reactants in final mixture is disregarded in this analysis.

Fischer-Tropsch: Converts H2 and CO reactants into linear hydrocarbons, which are then processed in distillation columns and crackers to produce SAF and gasoline.

Carbon capture and storage: The pathway for CCS assumed was DAC, compression of CO_2, transportation along pipelines and storage. Storage options such as depleted oil and gas fields (DOGF) and deep saline aquifers (SA) were explored, and both considered in this analysis due to each having their own benefits. DOGF offer cheaper costs, due to existing infrastructure and less monitoring needed as it has proven containment of hydrocarbons. However, SA were also included as they have a larger storage capacity and are more widespread. For this study, monitoring costs were not included, as there is a large uncertainty on how expensive this would be.

3. Methodology

3.1 Goal and Scope Definition

The goal of the LCA study in this report was to evaluate the environmental and economic performance of SAF which is produced via the FT process, and compare the two sources of CO_2; PSC and DAC. The functional unit for the enviornemntal assessment was defined as gCO2/MJ, allowing for a direct comparison to the 50% emission reduction target. The economic evaluation was was calculated per kg of SAF, as this unit is the most used in literature, which allows for straightforward comparison with existing studies.

The system boundaries used were cradle to grave, as this includes the whole life cycle of the fuel from raw material extraction to combustion. This approach is the same as that done for the 44.5gCO2/MJ reduction target[18], which makes it easier to compare final results with this benchmark. Geographical data such as for PV was collected from EU data sources, and plant cost data was scaled to 2023 using the Chemical Engineering Plant Costs Index[19] (CEPCI). Emissions were assumed to only come from solar PV, energy storage and process heat emissions, due to PV and battery manufacture being energy and resource intensive. CO2 emissions from construction of other infrastructure was assumed to be negligible, as it would be ammortised over the technology's lifetime.Transportation of materials between unit operations pre-CCS was also excluded from the assessment, as it was assumed all linked technologies were located at the same site as each other for ease of calculations. The environmental impact assessed was chosen as CO_2 emissions, as unlike other SAF such as biofuels, factors like water usage and competition of land for food crops is not as significant[20]. The overview of the SAF emission pathway can be seen in Figure 3., and FT was assumed to have constant and continuous production.

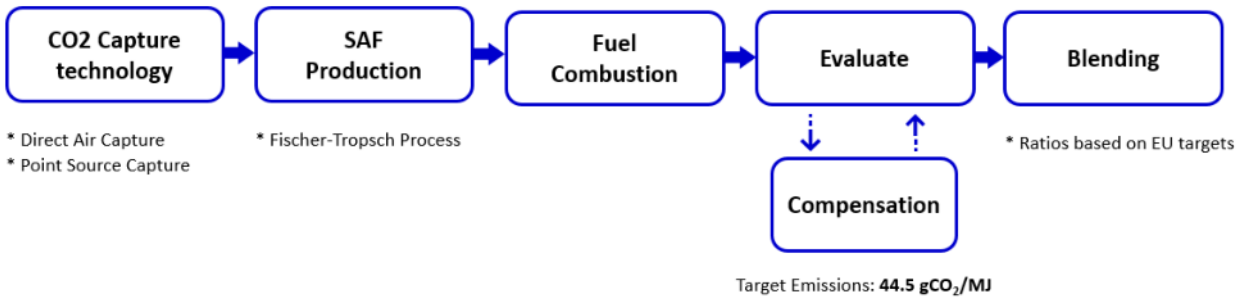


Figure 3 – Overview of SAF emission pathway

3.2 Inventory Analysis – Data Collection

This study used data from LCA and economic literature reports, as well as data from industry reports for specialised technologies, such as from the DAC company Climeworks. The inputs needed to be found for each technology were the: material requirements, energy requirements, CAPEX costs and emission factors (where applicable) for all technologies shown in Figure. 1 and 2.

3.3 Impact Assessment – Excel Model

An Excel model was developed to evaluate the environmental and economic impacts of the two

pathways for SAF in terms of the functional unit. This was split up into three main steps. The first one was finding CAPEX costs for each technology and annualising them using the capital recovery factor CRF. This is calculated using Equation 3 which utilises the interest rate (i) which was averaged for the EU, and the number of operational years (n) of the analysed technology was found from literature.

$$CRF = \frac{r(1+r)^n}{(1+r)^n - 1} \quad (3)$$

The next step was finding the energy requirements of each technology. For electricity requirements for the ASHP for thermal regenerartion of DAC adsorbents, a coefficient of performance (COP) was used to convert the thermal energy requirments into electricity needed. The COP used for this study was was 2.51[21], and conversion was calculated using Equation 4.

$$Elec._{ASHP} = \frac{Heat_{DAC}}{COP} \quad (4)$$

To calculate the amount of electricity storage needed per day, capacity factor and the number of hours a day with zero electricity generation were found from a solar database[22] and plotted against each other to get a correlation. The fraction of electricity storage needed per day could then be calculated for any CF, and then multiplied by the corresponding total energy requirement to find the total amount of electricity storage needed per day.

The final step was finding the energy costs and associated emissions. The levelized cost of solar energy was calculated using Equation 4, where CF of solar was taken as the average of all EU countries.

$$LCOE = \frac{(Total\ installed\ cost \times CRF) + Fixed\ O\&M\ Cost}{8760 \times CF_{solar}} \quad (4)[23]$$

The emission factor of solar was found from literature, and averaged for all EU countries. The energy requirements could then be multiplied by these values to get costs and emissions. Costs and emissions were then summed up to get final nominal values for each CO2 capture route.

3.4 Interpretation – Python Model

Initially, input parameters across all unit operations were systematically extracted from an Excel dataset. Each parameter was characterized using a triangular probability distribution based on its minimum, median, and maximum values to account for inherent uncertainties.

Monte Carlo simulation was implemented to generate 10,000 sets of random input parameter values. To validate calculations and ensure consistency of outputs, the Excel-based computational model was translated into Python.

A target emissions level of 44.5 gCO_2/MJ was established to align with sustainability benchmarks. For the PSC route, an iterative adjustment mechanism was implemented using a while-loop structure to compensate for emissions exceeding the target level. Secondary emissions resulting from additional DAC operations, as well as CO_2 storage and transportation, were incorporated into this adjustment mechanism.

Final values for costs and emissions for each route were obtained, and the results were visualized in a histogram, as seen in Figure 4. Comprehensive breakdowns of costs and emissions were generated, grouped by unit operations to facilitate easy visualization and comparison.

The solar capacity factor was identified as the key variable influencing cost and emission outputs. A relationship between the solar capacity factor and hydrogen storage hours was established by plotting data from multiple locations with varying capacity factors. A sensitivity analysis, presented in Figure 10, was then conducted using linear regression to explore the relationship between capacity factor and cost/emission outputs.

Blend ratios were then integrated into the model to examine the projected costs, emission savings, and capacity requires. Finally, the projected carbon tax required to bridge the cost and emission gap between SAFs and CAFs was calculated. This analysis provided valuable insights into the feasibility of the proposed EU policy changes.

4. Results and Discussion

4.1 Key Summary Values

Table 1 – Mean output data from Python Model

Component	Units	DAC	PSC
Emissions	gCO_2/MJ	40	44
Compensation	gCO_2/MJ	-	8.0
SAF Cost	€/kg_{SAF}	14	10
CO_2 Capture Cost	€/kg_{CO2}	0.98	0.13
H_2 Production Cost	€/kg_{H2}	13	13

The following insights were extracted from Table 1. DAC route emissions were 40 gCO_2/MJ, well under the target emissions of 44.5 gCO_2/MJ. PSC route emissions originally reached 52 gCO_2/MJ which exceeded the targeted emissions, hence requiring further compensation. The resultant extra contribution to PSC route costs was accounted for in the model. Post compensation emissions came out to 44 gCO_2/MJ for PSC with 8.0 gCO_2/MJ of emissions compensated via a new unit operation of direct air capture.

The cost of SAF for PSC route (€10/kg_{SAF}) was lower as compared to DAC route (€14/kg_{SAF}). Thus, PSC could be considered as the more economically viable route of the two. However, the higher carbon intensity of SAF via PSC route should be kept in consideration.

The cost of carbon capture for DAC route (€0.98/kg_{CO2}) was approximately seven times larger than PSC route (€0.13/kg_{CO2}). DAC technology targets atmospheric CO_2 of very low concentrations while PSC

technology targets industrial emissions of CO_2 with high concentrations of CO_2. This results in a huge difference in energy requirements to capture the same amount of carbon. The difference could also be explained by high CAPEX and FOC values for DAC, making DAC route the more expensive choice.

The cost of hydrogen production was calculated to be €13/kg_{H2} for both DAC and PSC route. This uniform cost stems from similar hydrogen requirements for both pathways. In comparison to literature values, this hydrogen cost lies on the upper end of the reported range. The elevated cost is attributed to the conservative assumptions employed during the modelling process. In the calculations for the Levelised Cost of Hydrogen (LCOH), the additional cost of electricity storage systems was included. This is unlike the standard LCOH assessments found in literature. Integration of electricity storage was crucial in ensuring continuous hydrogen production throughout periods of intermittent renewable electricity supply.

4.2 Sensitivity Analysis

Through the Monte Carlo Simulations, each output generated a set of values from the sampled input parameters. The output values were then plotted on a histogram to show the distribution and range of values produced.

From Figure 4, the kernel density equation (KDE) curves revealed that PSC's cost distribution is more concentrated, while DAC's cost distribution is more spread out. This implies that the PSC route has less variability and uncertainty in costs.

It was found that the mean cost of SAF in DAC and SAF route were €14.33/kg_{SAF} and €10.31/kg_{SAF} respectively. The plots of 5% and 95% boundary lines were used to determine the 90% confidence interval of the distribution. The SAF costs for DAC route lies between €11.33/kg_{SAF} and €17.21/kg_{SAF} with 90% confidence. Similarly, SAF costs for PSC route lies between €8.21/kg_{SAF} and €12.40/kg_{SAF} with 90% confidence.

The mean emissions of SAF through DAC and SAF route were 39.95 gCO_2/MJ and 44.10 gCO_2/MJ respectively. These values ranged from 3.64 gCO_2/MJ to 76.27 gCO_2/MJ for DAC route, with 90% confidence. For PSC route, they ranged from 11.00 gCO_2/MJ to 77.21 gCO_2/MJ, with 90% confidence.

The target emissions of 44.5 gCO_2/MJ is represented by the red dashed line. A while-loop mechanism was used to iteratively reduce the mean emissions via PSC below the target emissions.

The KDE curve for the emission distribution plot is significantly more spread out than the cost distribution plots. This suggests that the emission outputs have a very high uncertainty as compared to cost ouputs.

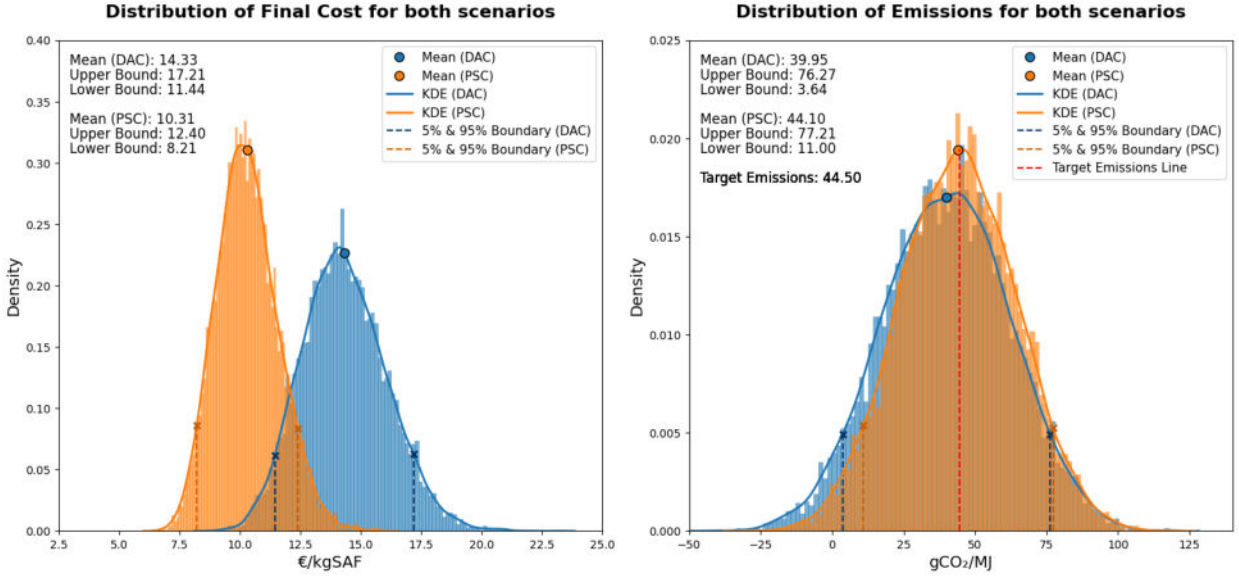


Figure 4 – Distribution plot of final cost and emissions for each route

4.3 Breakdown of Cost Components

Figure 5 breaks down the cost components for both DAC and PSC routes to SAF. From the breakdown, it can be seen that AWE makes up a significant portion of the costs for both pathways, as they both require the same amount of hydrogen. The largest contributor from AWE, is the electrcicty storage cost. This could be attributed to the high electricity requirement of this technology, as well as the assumption of constant and continuous FT operation during the whole day, including night.

The main differentiator for cost between DAC and PSC is the CAPEX +FOC of DAC, which results in an almost 50% increase. These high costs could be due to the use of solid sorbent technology in DAC, which is relatively newer so less developed with higher expenses[24]. Additionally, the CAPEX of DAC includes the ASHP, further increasing costs. In contrast, PSC uses a more mature technology of liquid sorbent MEA, and captures higher concentrations of CO2, which makes the process inherently cheaper[25]. These findings indicate that with current technology, PSC is more economically attractive than DAC, and for DAC to be on par with PSC there needs to be significant

technology improvements and cost reduction. However, both pathways still have high overall costs, mainly due to the significant expense of AWE. To achieve

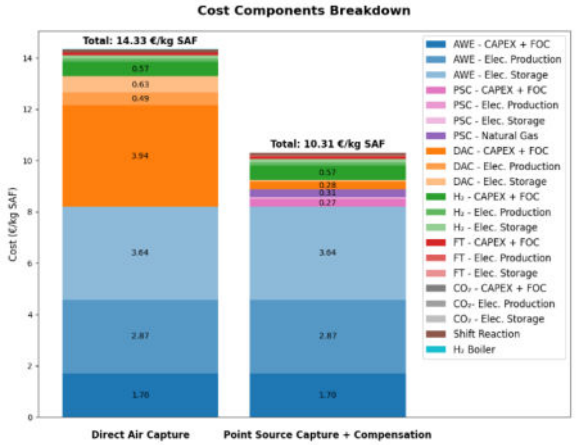


Figure 5 – Breakdown of cost components for both routes

4.4 Breakdown of Emission Components

The breakdown of emission components for both DAC and PSC routes, shown in Figure 6, display that their emissions are relatively similar after DAC compensation for the PSC route. This is due to the DAC route already being close to the 44.5gCO2/MJ SAF target, while PSC requires extra DAC compensation from DAC to reach this target. Similar to costs, most of the emissions are from AWE electricity production and storage, which has energy requirements substantially higher than any other technology. This can be seen in Figure 7 and 8, where AWE consumes 46kWh per kg SAF, and the next highest contributor is DAC electricity requirements, which consumes 7.21kWh per kg. The PSC route has higher overall emissions, primarly due to the use of natural gas for heating. As seen on Figures 6 and 7, natural gas has a considerable higher emission factor of 0.191 kgCO2/kWh, compared to 0.05 kgCO2/kWh for PV. However, there is the potential that emission for PSC could be greatly reduced if thermal

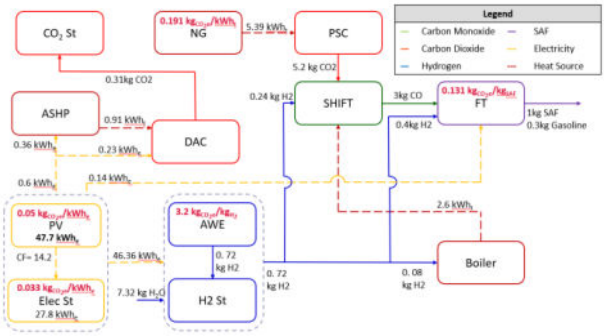


Figure 7 – Block Diagram for PSC route

4.6 Comparison with Literature Values

This section compares this reports SAF costs and emissions with values reported in literature. The selected studies were all from recent years to ensure relevance to current technology, and to have the most up to date costs and emissions data to provide a good benchmark for evaluating the differences in values. The discrepancies in values come from differences in

economic viability for SAF production therfore, efforts would have to be made to either reduce the electricity or storage requirements of the system

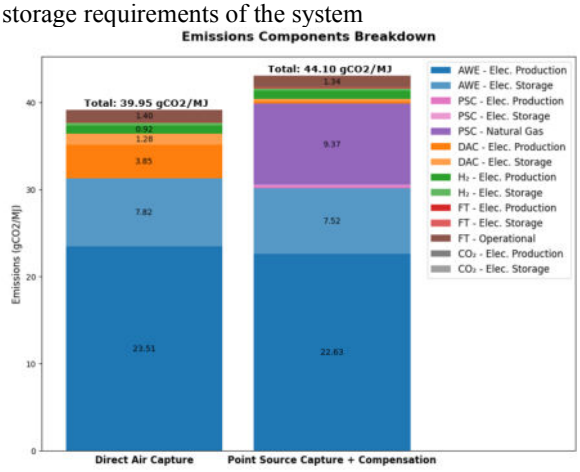


Figure 6 – Breakdown of emission components for both routes

energy was supplied via an ASHP. Figures 7 and 8 indicate that the thermal requirements for PSC of 5.39kWh are lower than the DAC thermal requirements of 10.95kWh. This suggests that if an ASHP was used, PSC route could result in lower emissions than the DAC route. Overall, these results show that PSC has the potential for lower emisisons than DAC and combined with its already lower cost, presents a promising pathway for SAF production.

4.5 Block Diagrams

Figure 7 and Figure 8 show a summary of all the energy and material balances needed per kg of jet fuel produced, for PSC and DAC SAF routes respectively. As mentioned in earlier results, these figures show that AWE is the most energy intensive part of the process, and natural gas combustion has the highest emission fator. These are two key areas where resource requirements influence SAF values of costs and emissions considerably, and any reduction in this would be favourable.

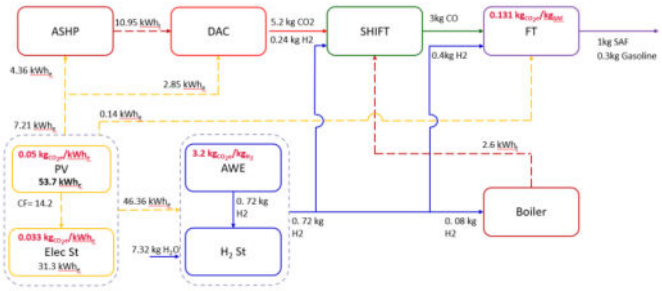


Figure 8 – Block Diagram for DAC route

assumptions and methodologies of assessments. The findings chosen were from Gonzales, Bernadi and Rojas, who all did economic and environmental assessments of SAF made from DAC.

SAF Cost: A Comparison of costs from this study to literature values as seen in Figure 9 , shows that the SAF production cost in this study is significantly higher. This is due to multiple reasons. The study from Gonzales et al. assumed that FT production could be

flexible, where 85% of demand could be met during the day and only 15% at night which reduced electricity storage needs and costs. They also did not include DAC CAOEX or energy costs. In comparison, this study assumed constant operation with no fluctuation resulting in a much higher electricity storage need. CAPEX DAC costs were also included, and were a significant contributor, as seen in Figure 5 As a result, the cost from Gonzales et al. was significantly lower. Rojas et al. used offshore wind energy which had a capacity factor of 36%, and assumed that electricity storage was not needed, and instead used the grid as a virtual storage system. This approach resulted in a much lower overall cost compared to the analysis in this paper, where electricity related costs were substantial, as seen Figure 5. Similar to this study, hydrogen production dominated costs, highlighting how this technology is a significant cost driver. Bernadi et al, used wind energy for hydrogen production, and assumed a general LCOH for wind powered AWE of H2 of 4.2$/kg[26], which is around 3 times less than the LCOH calculated in this analysis of 12.46€/kg. The cost differences for hydrogen in all these studies, highlight how there is the potential for significant cost savings for SAF if wind energy is to be used instead of solar.

SAF Emissions: In comparison to cost, the emission values of SAF in literature are more spread out. The study from Gonzales had the highest emissions, due to in depth modelling of fuel distribution, which was not a factor considered in this analysis. This highlights the importance of considering these emissions for a more complete LCA. Bernadi had lower emissions than this study, due to higher capacity factors from wind, and from selling surplus wind energy for carbon credits which could make a significant difference to offsetting emissions and provide financial incentive. The lowest emissions were from Rojas, which was achieved by using offshore wind and avoiding storage emissions by utilising the grid as a virtual storage system. This kind of assumption is not common practise for an LCA and can lead to underestimates of the carbon intensity of SAF. Overall, these studies show the importance of how the source of electricity can significantly influence costs and emissions, and how assumptions and the level of detail in modelling can result in either over or underestimates.

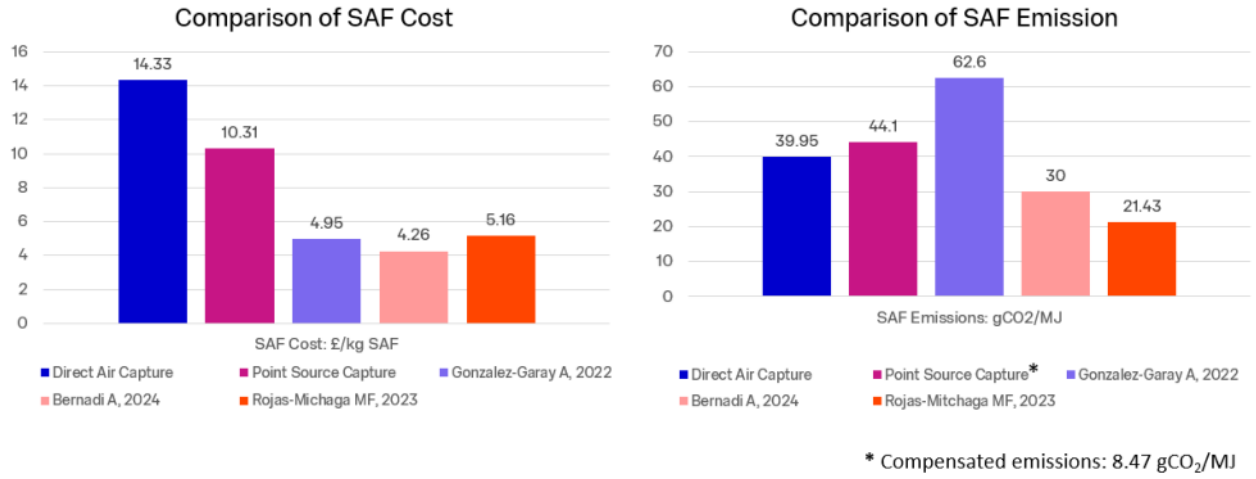


Figure 9 – Comparison of cost and emission values between modelled outputs and literature values

4.7 Solar Capacity Factor

Given that both electricity generation and storage are influenced by solar CF with significant associated costs and emissions as shown in Figures 5 and 6, a sensitivity analysis was performed to assess the impact of varying CF on these factors and plotted on Figure 10. This analysis shows that as solar CF increases, both the mean cost and emissions of SAF production linearly decrease. A higher CF indicates a higher annual energy output, which leads to reduced PV needing to be installed, resulting in lower emissions and costs. Higher realistic CF would be around 25% and would be common in countries like Saudi Arabia or Austrialia. At lower CF factors, e.g. around 10%, PV needs to be oversized, and more electricity storage installed to ensure continuous SAF production and not run the risk of having power outages. Further improvements in CF, therefore solar output could be achieved by efficiency improvements in solar technology. For this model, a system energy loss of 14% was assumed, but if this was decreased then cost and emission reductions could be reached without the need to change geographical location. Furthermore, alternative technologies like CSP or wind energy in combination with PV could reduce costs and emissions. CSP can generate some electricity at night and provide storage which would reduce reliance on batteries, while wind energy has a higher CF, so would improve efficiency of energy generation. Regions that have a higher solar CF have a natural advantage, so FT SAF production which utilises PV as energy would be recommended most for these countries to get fuel with lower costs and emissions. For regions where solar output is less efficient, different energy sources or a diversified mix would be more beneficial.

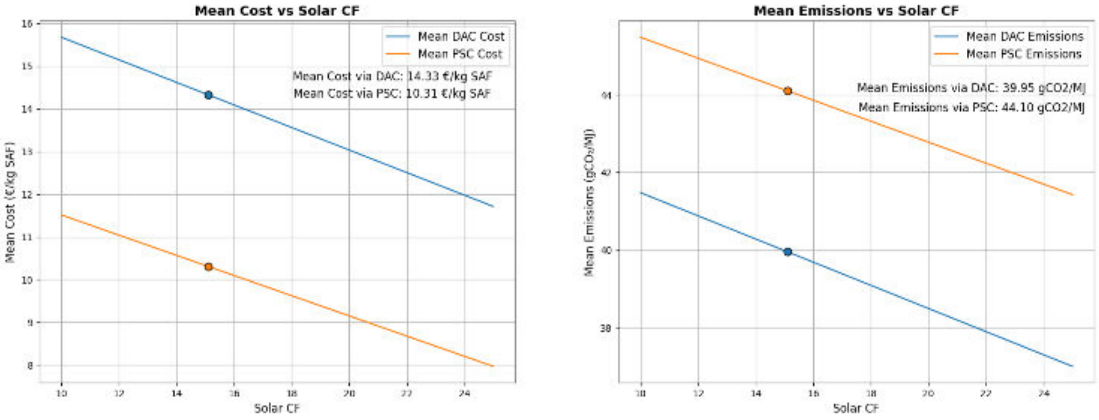


Figure 10 – Sensitivity analysis of solar capacity factor against cost and emission outputs via python model

4.8 Projected Capacities

Further analysis was conducted on the 70% blend ratio target for 2050, projecting to meet the EU aviation demand. Table 2 contains the annual cost, emission savings, installed capacities for both DAC and PSC routes.

Annual cost provides rough gauge of the amount that needs be invested by the EU to implement their 70% blend ratio target. Annual emission savings were calculated based on the amount of CO_2 emissions avoided by using the 70% SAF blend ratio as compared to CAF. While the DAC route requires a higher annual cost than the PSC route, modelled calculations found that the DAC route saves an additional 5MtCO_2/yr as compared to the PSC route.

The EU currently has 320 TWh/yr installed solar capacity. To meet the production needs for both DAC and PSC route given by the annual solar electricity capacity, an additional 369% and 306% increase in solar installations would be needed respectively, assuming all energy requirements are to be provided by solar.

To support the annual energy storage capacity required by the DAC route (760 TWh/yr), deployment would need to be larger than the PSC route (670 TWh/yr). The energy storage capacities required exceed both current EU and global available capacities, indicating the need for significant infrastructure development in this area. Alternatively, pursuing energy sources with higher capacity factors could help mitigate electricity storage demands. Nuclear energy offers consistent power supply and minimises storage needs, but face challenges like waste management and public resistance in some countries.

Table 2 – Sumamary of annual parameters for 70% blend ratio

Component	Units	DAC	PSC	Current
Annual Cost	B€/yr	410	300	-
Annual Emission Savings	MtCO_2/yr	59	54	-
Annual Solar Electricity Capacity	TWh/yr	1500	1300	EU – 320[27]
Annual Energy Storage Capacity	TWh/yr	760	670	EU – 0.036[28] Global – 8.5
Annual Carbon Capture Capacity	MtCO_2/yr	140	160	Global – 43 (DAC)
Annual AWE Capacity	MtH_2/yr	20	20	EU – 0.021

4.9 Carbon Tax

Table 3 – Potential carbon tax for 70% SAF blend ratio

Component	Units	DAC	PSC	Current
Carbon Tax	€/kg_{CO2}	6.3	4.8	0.068
Carbon Tax	€/t_{CO2}	6300	4800	68

In future attempts to encourage producers to shift to SAF production, SAF costs must align with current production costs in CAF. This study explored the use of carbon tax as a means to levelise the playing field with CAF.

Table 3 presents the carbon tax calculated for a 70% SAF blend ratio to meet projected EU aviation demand. The carbon tax for DAC and PSC routes are €6.3/kg_{CO2} and €4.8/kg_{CO2} respectively, which are substantially higher compared to ther current carbon tax of €0.068/kg_{CO2}.

For the EU to meet their SAF blending targets in the future, more effort needs to be done to reduce the production cost of SAF. With current technologies and infrastructure, substituting CAF with SAF on a large scale is not yet economically viable.

For the DAC route to achieve full carbon feedstock requirements, it would necessitate the installation of approximately 23 million DAC or equivalent, assuming an annual capture capacity of 140 MtCO_2/yr. This is based on the maximum reported capture capacity of 6000 tCO_2/yr per Climeworks facility. However, DAC system total cost currently makes up 27% of the SAF cost, equivalent to €3.94/kg_{SAF}. Therefore, significant technological and cost reductions in DAC systems

would be crucial to advancing the adoption of this technology on a larger scale.

The required AWE annual production capacity required to meet the hydrogen demand for feedstock is 20 MtH_2/yr, which is around 100 times larger than the current EU AWE capacity. This substantial gap highlights the magnitude of infrastructure expansion that would be needed to achieve this. However, AWE technology is currently expensive so getting to this level of deployment would need substantial additional funding. Current EU funding options, such as from Innovation Fund calls and EU Hydrogen Bank are limited and complex. As a result, acquiring this necessary funding could be a challenge and might impact the ability to meet future green hydrogen demands. Hence, both advancmenets in technology and funding would be required to achieve this.

5. Conclusions and Outlook

From an environmental standpoint, DAC route offers a more sustainable pathway of SAF production due to its lower mean emissions as compared to PSC route. However, substituting natural gas consumption in the PSC process with an electric heat pump could potentially reduce PSC route emissions below those of the DAC route.. Additionally, the high uncertainty of input parameters led to significant uncertainty bounds in the output emissions, despite the 90% confidence level. This underscores the need for further research and the development of more extensive models to strengthen conclusions and provide robust evidence.

From an economic perspective, PSC stands out as the cheaper pathway for SAF production. The bulk of cost components for both routes were attributed to alkaline water electrolysis and direct air capture technology. Future research and developments on the technology could potentially drive costs down, enabling increased rates of SAF adoption.

The price of CAF, SAF from DAC and SAF from PSC stand at €0.80/kg_{fuel}, €14.31/kg_{fuel} and €10.31/kg_{fuel} respectively, highlighted that SAF is not yet economically viable on a large scale.

The EU's decision to recognise recycled carbon aviation fuels as Sustainable Aviation Fuels is a significant step forward. Life cycle emissions of SAF produced via point-source carbon capture were reasonably able to stay below the target threshold of 44.5 gCO_2/MJ through direct air capture to compensate the surplus emissions. As a more cost-effective alternative to direct air capture, this policy has potential to attract investments, encourage producers and drive advancements in SAF production technologies.

To address existing challenges and enhance research efforts, future efforts should explore the alternative high-capacity energy sources like wind and nuclear. Renewable energy modelling could be refined by integrating spatio-temporal modelling of solar output to accurately obtain electricity storage capacity. Building-related emissions are a potential component to incorporate into lifecycle emissions modeling, contingent on the availability of published data in the future. Additionally, a more comprehensive review of transportation emissions within the broader context of the system could be undertaken.

Acknowledgements

We would like to express our sincere gratitude to Research Fellow, David Danaci for his invaluable advice and support throughout this project. His encouragement has made this research journey both productive and enjoyable.

References

[1] Ritchie, H. (2024). What share of global CO2 emissions come from aviation?. Our World in Data. [online] Available at: https://ourworldindata.org/global-aviation-emissions.

[2] Department for Transport (2021). Mandating the use of sustainable aviation fuels in the UK. [online] GOV.UK. Available at: https://www.gov.uk/government/consultations/mandating-the-use-of-sustainable-aviation-fuels-in-the-uk.

[3] Council of the EU (2023). RefuelEU aviation initiative: Council adopts new law to decarbonise the aviation sector. [online] Consilium. Available at: https://www.consilium.europa.eu/en/press/press-releases/2023/10/09/refueleu-aviation-initiative-council-adopts-new-law-to-decarbonise-the-aviation-sector/.

[4] Jeswani, H.K., Chilvers, A. and Azapagic, A. (2020). Environmental Sustainability of biofuels: a Review. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, [online] 476(2243), p.20200351. doi:https://doi.org/10.1098/rspa.2020.0351.

[5] Finnveden, G. and Potting, J. (2014). Life Cycle Assessment. [online] ScienceDirect. Available at: https://www.sciencedirect.com/science/article/abs/pii/B9780123864543006278.

[6] Marchese, M., Buffo, G., Santarelli, M. and Lanzini, A. (2021). CO2 from direct air capture as carbon feedstock for Fischer-Tropsch chemicals and fuels: Energy and economic analysis. Journal of CO2 Utilization, 46, p.101487. doi:https://doi.org/10.1016/j.jcou.2021.101487.

[7] Terlouw, T., Treyer, K., Bauer, C. and Mazzotti, M. (2021). Life Cycle Assessment of Direct Air Carbon Capture and Storage with Low-Carbon Energy Sources. Environmental Science & Technology, 55(16), pp.11397–11411. doi:https://doi.org/10.1021/acs.est.1c03263.

[8] Tan, S.H. (2011). Solar Intermittency: How Big is the Problem? [online] Renewable Energy World. Available at: https://www.renewableenergyworld.com/storage/solar-intermittency-how-big-is-the-problem/.

[9] IEA (2019). Offshore Wind Outlook 2019 World Energy Outlook Special Report. [online] Available at: https://iea.blob.core.windows.net/assets/495ab264-4ddf-4b68-b9c0-514295ff40a7/Offshore_Wind_Outlook_2019.pdf.

[10] Alless, R. (2021). Concentrated Solar Power (CSP) Vs Photovoltaic (PV): An In-depth Comparison. [online] SF Magazine. Available at: https://www.solarfeeds.com/mag/csp-and-pv-differences-comparison/.

[11] cmxbattery and cmxbattery (2024). Lithium Batteries vs Lead Acid Batteries: A Comprehensive Comparison. [online] Custom LiFePo4& rechargeable li ion battery pack Manufacturer supplier. Available at: https://cmxbattery.com/lithium-ion-batteries/lithium-batteries-vs-lead-acid-batteries-a-comprehensive-comparison/ [Accessed 12 Dec. 2024].

[12] Energy Saving Trust (2018). Air source vs ground source heat pumps. [online] Energy Saving Trust. Available at: https://energysavingtrust.org.uk/air-source-heat-pumps-vs-ground-source-heat-pumps/.

[13] NETL (n.d.). Point Source Carbon Capture from Power Generation Sources. [online] netl.doe.gov. Available at: https://www.netl.doe.gov/carbon-capture/power-generation.

[14] Gonzalez-Garay, A., Heuberger-Austin, C., Fu, X., Klokkenburg, M., Zhang, D., Made, A. van der and Shah, N. (2022). Unravelling the potential of sustainable aviation fuels to decarbonise the aviation sector. Energy & Environmental Science, [online] 15(8), pp.3291–3309. doi:https://doi.org/10.1039/D1EE03437E.

[15] Hydrogen production – Global Hydrogen Review 2024 – Analysis - IEA (2024). Hydrogen production – Global Hydrogen Review 2024 – Analysis - IEA. [online] IEA. Available at: https://www.iea.org/reports/global-hydrogen-review-2024/hydrogen-production.

[16] Irena (2022). 5 Compressed hydrogen storage. [online] Irena.org. Available at: https://www.irena.org/Innovation-landscape-for-smart-electrification/Power-to-hydrogen/5-Compressed-hydrogen-storage [Accessed 12 Dec. 2024].

[17] Naquash, A., Agarwal, N. and Lee, M. (2024). A Review on Liquid Hydrogen Storage: Current Status, Challenges and Future Directions. Sustainability, [online] 16(18), pp.8270–8270. doi:https://doi.org/10.3390/su16188270.

[18] The Royal Society (2023). Net zero aviation fuels: resource requirements and environmental impacts POLICY BRIEFING. [online] Available at: https://royalsociety.org/-/media/policy/projects/net-zero-aviation/net-zero-aviation-fuels-policy-briefing.pdf.

[19] CEPCI (2014). The Chemical Engineering Plant Cost Index - Chemical Engineering. [online] Chemengonline.com. Available at: https://www.chemengonline.com/pci-home.

[20] EASA Eco (2022). How sustainable are SAF? | EASA Eco. [online] EASA Eco. Available at: https://www.easa.europa.eu/eco/eaer/topics/sustainable-aviation-fuels/how-sustainable-are-saf#additional-effects-from-use-of-saf.

[21] Deutz, S. and Bardow, A. (2021). Life-cycle assessment of an industrial direct air capture process based on temperature–vacuum swing adsorption. Nature Energy, [online] 6(2), pp.203–213. doi:https://doi.org/10.1038/s41560-020-00771-9.

[22] Renewables Ninja (2019). Renewables.ninja. [online] Renewables.ninja. Available at: https://www.renewables.ninja/.

[23] NREL (2010). Simple Levelized Cost of Energy (LCOE) Calculator Documentation | Energy Analysis | NREL. [online] Nrel.gov. Available at: https://www.nrel.gov/analysis/tech-lcoe-documentation.html.

[24] Lebling, K., McQueen, N., Pisciotta, M. and Wilcox, J. (2021). Direct Air Capture: Resource Considerations and Costs for Carbon Removal. World Resources Institute. [online] Available at: https://www.wri.org/insights/direct-air-capture-resource-considerations-and-costs-carbon-removal.

[25] Jiang, Y., Mathias, P.M., Zheng, R.F., Freeman, C.J., Barpaga, D., Malhotra, D., Koech, P.K., Zwoster, A. and Heldebrant, D.J. (2023). Energy-effective and low-cost carbon capture from point-sources enabled by water-lean solvents. Journal of Cleaner Production, [online] 388, p.135696. doi:https://doi.org/10.1016/j.jclepro.2022.135696.

[26] https://www.irena.org/publications/2019/Sep/Hydrogen-A-renewable-energy-perspective

[27] European Commission (2023). Solar Energy. [online] energy.ec.europa.eu. Available at: https://energy.ec.europa.eu/topics/renewable-energy/solar-energy_en.

[28] Solar Power Europe (2023). New analysis reveals European solar battery storage market increased by 94% in 2023 - SolarPower Europe. [online] Solarpowereurope.org. Available at: https://www.solarpowereurope.org/press-releases/new-analysis-reveals-european-solar-battery-storage-market-increased-by-94-in-2023.

Process Flexibility Analysis of a Robotic Parallel Mini Bioreactor Platform for High-Throughput Bioprocess Development

Bhargavi Bavaharan and Joshua Samuel
Department of Chemical Engineering, Imperial College London, U.K.

Abstract

The demand for scalable and flexible bioprocesses has driven the development of high-throughput (HT) mini-bioreactor (MBR) platforms. This study investigates the process flexibility of the KIWI-biolab platform, leveraging a computational Quality by Digital Design (QbDD) framework to identify robust Design Spaces (DSps) and assess the impact of Key Performance Indicator (KPI) constraints on process flexibility with reliability. Using an emulator, we identified a unified DSp at a tolerance level of 23.91%, balancing flexibility with reliability. A sensitivity analysis revealed trade-offs between protein productivity, biomass productivity and dissolved oxygen tension, highlighting the need for tailored design decisions (DDs). By analysing Nominal Operating Points (NOPs) we demonstrated the influence of design decisions on process robustness and flexibility. This work advances scalable bioprocess development through model-based insights into operational flexibility.
Keywords: Design Space, Flexibility, Key Performance Indicators, Nominal Operating Point, Sensitivity

1. Introduction

The growing demand for biological products in pharmaceuticals, biofuels and food production underscores the need for efficient bioprocess scale-up [1,2]. Bioprocesses contribute to meeting global biologics demand while improving sustainability and cost efficiency [3]. However, scaling up from laboratory to industrial scales is challenging due to inhomogeneities in substrate concentration, temperature, pH and CO_2 gradients [4]. These gradients create heterogeneous conditions that complicate scale-up [4].

High-throughput (HT) mini-bioreactor (MBR) platforms address these challenges by enabling parallel millilitre-scale cultivations under well controlled conditions, with online monitoring and cultivation strategies that can mimic industrial processes [5]. Such platforms enhance experimental throughput, optimise feeding strategies and reduce scale-up risks [5]. Despite these advancements, a gap persists between bioprocess design and scale-up.

A key concept in bridging this gap is the Design Space (DSp) defined in the Quality by Design (QbD) framework [6]. A DSp represents the multidimensional range of operating conditions ensuring that Critical Quality Attributes (CQAs) are consistently met. Identifying a DSp enables flexible operation compared to single-point processes [7].

While HT MBR platforms focus on optimising experimental designs, they often provide limited insight into how Key Performance Indicators (KPIs) vary across ranges of operating conditions. They do not consider the influence of KPI-based constraints or how DSps are affected by these constraints [5,8-10]. For example, Janakiraman et al. identified a DSp for a HT MBR platform but did not quantify process flexibility [10].

To address these issues, we apply a Quality by Digital Design (QbDD) framework, which integrates computational tools into QbD [7]. We employ a model-based approach developed by Sachio et al. [7], to assess process flexibility and evaluate how KPI constraints shape the DSp [7,11]. Specifically, we examine an HT MBR platform by (1) identifying DSps, (2) analysing flexibility and (3) investigating how KPI constraints influence DSp size.

These insights contribute to more robust operating regions, accelerating bioprocess development while supporting QbD principles and better bridging the gap between design and scale-up.

2. Background

2.1. High-Throughput (HT) Mini-Bioreactor Platforms

HT MBR platforms are scaled-down bioreactors designed for high experimental throughput and comprehensive monitoring and control. With up to 48 parallel MBRs, they can generate large datasets. Optical methods measure parameters such as biomass, dissolved oxygen (DO), pH and fluorescence, while liquid handling robots automate procedures for improved reproducibility [12].

These platforms are instrumental in strain and media optimisation, enhancing productivity and stability [12-14]. For instance, Rohe et al. used an automated HT MBR platform with online monitoring and precise sampling, demonstrating scalability from millilitre to litre scales. However, their design was static, without

dynamic experimental re-design [9].

Janakiraman et al. incorporated Design of Experiments (DoE) and scale-down models across 15 mL, 5 L and 15,000 L scales. They identified Critical Process Parameters (CPPs) that influenced Critical Quality Attributes (CQAs), accelerating process characterisation [10]. However, their study did not explicitly address the quantification of process flexibility or the incorporation of KPI based constraints into the DSp.

Cruz et al. and Haby et al. advanced HT MBR platforms by integrating dynamic, real-time control, data integration and online optimal experimental redesign strategies [5,8]. These improvements address static limitations, but the extent of DSp flexibility remains unclear.

This study focuses on the KIWI-biolab platform at TU Berlin, described by Cruz et al. and Haby et al. [5,8]. Equipped with an emulator, it allows for model-based approaches to examine process flexibility and identify robust operating regimes.

2.2. The KIWI-biolab platform

The KIWI-biolab platform integrates two robotic units to fully automate bioprocess workflows. The Tecan Freedom EVO 200 liquid-handling station (LHS) (Tecan Group Ltd, Männedorf, Switzerland) manages operations involving the MBR system, while the Microlab Star LHS (Hamilton Company, Bonaduz, Switzerland) focuses on at-line analytical measurements [5]. These units are connected by a transfer unit enabling continuous and automated cultivation. The main part of the system is 24 continuously stirred MBRs, each with a working volume of 10 mL, filled with a glucose supplemented medium and inoculated with an *E. coli* strain cultivated for protein production [5].

The process starts with a batch phase, during this initial phase, all the glucose is provided at the beginning of the cultivation and no additional glucose is added, this phase continues until the microorganisms consume the initial nutrients. This transitions into a exponential batch phase, where glucose pulses are delivered to each MBR to match the exponential growth of the microorganisms, keeping microorganisms growing at a controlled rate. Finally, an induction phase begins with the addition of an inducer, activating the production of a fluorescent protein and there is constant feeding to maintain steady growth and production [5]. Figure 1 summarises this flow and Figure 2 shows an example feeding profile.

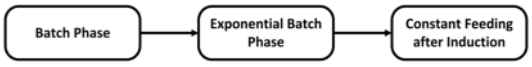


Figure 1: KIWI-biolab feeding strategy, the process begins with a Batch Phase (no feeding), transitions to an Exponential Batch Phase (pulsed glucose to match growth) and ends with constant feeding during the Induction Phase for protein production.

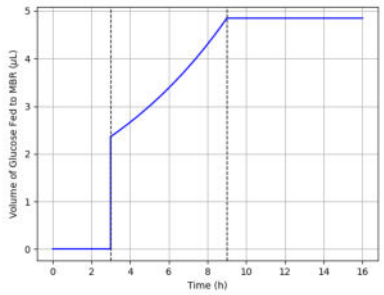


Figure 2: The first dashed line marks the start of the Exponential Batch Phase, while the second indicates the transition to constant feeding after induction. Glucose volume fed remains at zero initially, rises exponentially and is at a constant level post-induction.

Sampling is performed systematically, with columns sampled every 20 minutes and each reactor sampled once per hour. Collected samples are transferred into cooled 96-well plates containing sodium hydroxide to stop microbial activity, after which the plates are transported by the robotic arm to the analysis platform. Diluted samples undergo measurements for optical density (OD) and fluorescence to estimate biomass and protein production levels, while additional samples are analysed for glucose and acetate concentrations using enzymatic assays [5]. The system integrates all online sensor data, dissolved oxygen tension (DOT), along with at-line analytical results, into an SQL database for real-time monitoring and data management, as shown in Figure 3 [5].

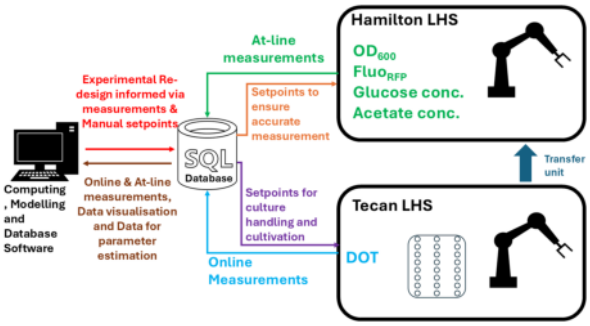


Figure 3: The KIWI-biolab platform integrates two robotic units: the Tecan LHS for managing MBR operations and the Microlab LHS for at-line analytical measurements. Connected by a transfer unit, the system automates microbial cultivation workflows, with online data integration into an SQL database for process monitoring and experimental redesign. Key measurements include optical density (OD), fluorescence, glucose concentration and DOT, adapted from [5].

2.3. KIWI-biolab Emulator

The KIWI-biolab emulator is a computational tool that replicates the physical KIWI-biolab platform, enabling *in silico* simulation of cultivation experiments. It allows users to customise experimental parameters, including initial glucose concentration for the MBR (S_0), batch time (t_{batch}), induction time (t_{ind}) and

growth rate setpoint (μ_{set}), which represent target values for the specific growth rate of microorganisms. Growth rate setpoints are critical in bioprocessing as they influence biomass production and protein yields.

These parameters determine the feeding profile for MBRs during both the exponential batch phase and post-induction phases, feeding stops at the end of experiment which can specified also but is kept at 16 hours for this study:

$$F_p = 1e6 \cdot \frac{(X_0+S_0Y)(V\Delta t)}{S_{in}Y}\mu_{set}e^{\mu_{set}(t-t_{batch})} \; if\, t < t_{ind} \tag{1}$$

$$F_p = 1e6 \cdot \frac{(X_0+S_0Y)(V\Delta t)}{S_{in}Y}\mu_{set}e^{\mu_{set}(t_{ind}-t_{batch})} \; if\, t \geq t_{ind} \tag{2}$$

Here, X_0 is the initial biomass, Y is the biomass yield, V is the MBR working volume, Δt is the feed pulse for every 10 minutes, S_{in} is the feed glucose concentration, and the $1e6$ scales the feeding profile volume from L to μL, Table 1 shows the relevant values.

Table 1: Feeding profile constant values.

Constant	Value
X_0 $(g_{biomass}L^{-1})$	0.18
Y $(-)$	0.50
V (L)	0.01
Δt (h)	0.17
S_{in} $(g_{glucose}L^{-1})$	200.00

The emulator integrates with Apache Airflow and Docker to automate workflows efficiently. Apache Airflow schedules and monitors task execution, ensuring seamless coordination of emulator components, while Docker packages the emulator and its dependencies into isolated environments, enabling reproducibility and scalability. Results are stored in an SQL database for efficient analysis and the emulator supports data integration for generating feeding profiles and collecting time-series data.

The emulator is available as open-source software at: https://git.tu-berlin.de/bvt-htbd/kiwi/tf2/emulator2/-/tree/one_shot.

2.4. Design Spaces

To understand DSps, it's important to first know about QbD and its computational extension, QbDD. QbD differs from traditional Quality by Testing (QbT) by building quality into products and processes rather than relying on post-production checks [7]. It ensures that CQAs, CPPs and Critical Material Attributes (CMAs) consistently meet specifications, forming the foundation for defining DSps.

QbDD extends QbD by using computational tools to accelerate development via simulations [7]. This approach helps identify optimal operating conditions, map DSps and reduce reliance on costly experiments.

Within this framework, Knowledge Spaces (KSps) are created by varying Design Decisions (DDs) within their feasible ranges, as defined for the process [7]. Each combination of DDs has a specific KPI value with it. This variation spans all possible conditions under which the process could potentially operate, represented geometrically in a multidimensional space. When we impose KPI constraints on the KSp, we extract a subset of the KSp that satisfies these constraints, referred to as the DSp, expressed as:

$$\text{DSp} = \{\boldsymbol{DD} \in \text{KSp} \mid \boldsymbol{DD} \text{ satisfies KPI constraints}\}, \tag{3}$$

where $\boldsymbol{DD}$ is the vector of DDs. It follows that:

$$\text{DSp} \subseteq \text{KSp}. \tag{4}$$

Once the DSp is identified, a Nominal Operating Point (NOP) is specified. After selecting the NOP, an Acceptable Operating Region (AOR) is defined. The AOR, as per Sachio et al. [7], is the region surrounding the NOP within which operations are guaranteed to remain within the DSp. This ensures that the process operates consistently and meets specifications even under permissible variations [7].

To define the AOR, Sachio et al. introduce the Multivariate Proven Acceptable Range (MPAR), which is the acceptable variation range for each input parameter while maintaining operation within specification [7]. The MPAR constrains the DDs, ensuring that the AOR is a subset of the DSp. Expressed as:

$$\text{AOR} = \{\boldsymbol{DD} \in \text{DSp} \mid \boldsymbol{DD} \text{ satisfies MPAR constraints}\}, \tag{5}$$

where:

$$\text{AOR} \subseteq \text{DSp}. \tag{6}$$

Essentially, the AOR is a subset of the DSp around a specified NOP, guaranteeing process reliability within acceptable DD variations defined by the MPAR. The hierarchical relationship between the KSp, DSp and AOR is shown in Figure 4.

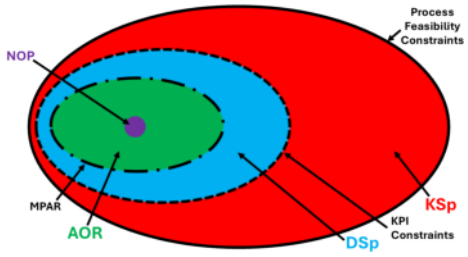


Figure 4: Venn diagram showing the hierarchical relationships between the KSp, DSp and AOR. The red KSp is created by varying DDs within feasible ranges, bounded by process constraints given by the solid black boundary around the KSp. The blue DSp, a subset of the KSp, is formed by applying KPI constraints given by the black dashed boundary around the DSp. The green AOR, centred around a specified NOP in purple, is bounded by the MPAR given by the black dashed-dotted boundary around the AOR.

2.5. Research Objectives

The primary aim of this study is to leverage a computational framework to support QbDD in assessing the flexibility of the KIWI-biolab platform. To achieve this, the study focuses on several key objectives: (1) identifying DSps from experimental DD ranges using data generated by an emulator for the KIWI-biolab platform, (2) quantifying the size of the DSp and AOR to evaluate their contributions to process flexibility, where larger DSps indicate more flexible processes and larger AORs signify greater flexibility around the NOP and (3) investigating the effect of varying KPI constraints on DSp size. Furthermore, the study seeks to identify operating points with high KPI values, which represent optimal conditions for robust process performance.

To achieve these goals, we employ a computational framework developed by Sachio et al. [7,11] to obtain and quantify DSp and AOR size. These methods are applied to data generated by an emulator that replicates the operation of the KIWI-biolab platform. The following section outlines the methodologies employed, detailing how the computational framework was adapted and applied in this study.

3. Methodology

3.1. Computational Framework

This study employs an adapted model-based computational framework adapted from Sachio et al. [7], aimed at identifying and analysing the DSp for the KIWI-biolab platform. This framework quantifies the flexibility of the bioprocess by linking DDs, KPIs and process performance metrics. The workflow involves the following steps, as illustrated in Figure 5:

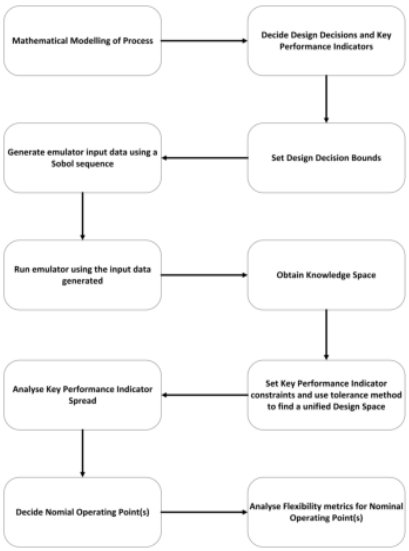


Figure 5: Computational framework for DSp identification and analysis, adapted from [7]. The workflow begins with mathematical modelling of the process, proceeds through selecting DDs and KPIs and, finishes in identifying robust and flexible operating points.

3.2. Mathematical Modelling of Process

In the initial step, a validated mathematical model of the KIWI-biolab platform, developed by Luna et al. [15], forms the foundation for the computational experiments in this study. The model has been verified against experimental data, ensuring that the emulator accurately replicates realistic process behaviour [15]. The emulator leverages this mathematical model to simulate the platform's operations.

3.3. Selection of Design Decisions and Key Performance Indicators

In this step we choose our DDs and KPIs, our DDs are the growth rate setpoint, induction time and initial glucose concentration in the MBR and the batch time. These are represented as:

$$\boldsymbol{DD} = [\mu_{set}, t_{ind}, S_0] \tag{7}$$

As a rule of thumb, the batch time is determined by the initial glucose concentration to ensure that microorganisms have sufficient time to consume the glucose before feeding starts, without unnecessary idle time. This relationship is given by:

$$t_{batch} = S_0 + 1 \tag{8}$$

This gives the microorganisms time to consume the glucose before feeding start, without having much idle time. The KPIs considered are the: Minimum Dissolved Oxygen Tension (DOT), Biomass Productivity and the $Fluo_{RFP}$ Productivity. Which are calculated as follows:

$$\text{Minimum DOT}\,[\%a.s.] = min(DOT) \tag{9}$$

$$\text{Biomass Productivity}[g_{biomass}L^{-1}h^{-1}] = \frac{X_{final} - 0.18 g_{biomass} L^{-1}}{t_{final}} \tag{10}$$

$$Fluo_{RFP}\ \text{Productivity}[RFUh^{-1}] = \frac{Fluo_{RFP,final}}{t_{final}} \tag{11}$$

Where X_{final} is the final biomass concentration for the MBR, biomass concentration is obtained by multiplying the OD measurement 2.7027, $Fluo_{RFP,final}$ is final fluorescence measurement for that MBR and t_{final} is the final measurement time for that MBR.

Dissolved Oxygen Tension (DOT) is the concentration of oxygen dissolved in the culture medium, essential for aerobic microorganisms as oxygen availability influences their growth and metabolism. Biomass productivity measures the rate at which microorganisms convert nutrients into cellular mass, indicating the overall efficiency of growth under given operating conditions. Fluorescence RFP productivity refers to the rate at which the recombinant fluorescent protein (RFP) is produced, providing a direct indication of the platform's ability to generate target proteins. Tracking these KPIs is crucial because they reflect

the health and performance of the culture, guiding process optimization and ensuring that the cultivation conditions are maintained to achieve desired product quality and process consistency.

3.4. Dataset Generation

After defining the DDs and KPIs, we established initial DD bounds based on expert guidance from the KIWI-biolab team, ensuring that the KSp reflects feasible and practical bioprocess conditions which are shown in Table 2.

Table 2: Initial Design Decision bounds.

Design Decision	Lower Bound	Upper Bound
S_0 $(g_{glucose}L^{-1})$	2.00	6.00
t_{ind} (h)	9.00	14.00
μ_{set} (h^{-1})	0.10	0.35

To generate the KSps for this study, we utilised a Sobol sequence, a quasi-random sampling method that ensures uniform coverage across the multidimensional design space [16]. This method reduces clustering, enhances sampling efficiency and guarantees equal representation of all DD combinations by generating 2^n unique inputs, where n can be specified.

Given the limitation of 24 MBRs per run, the generated designs were divided into groups of 24 input combinations, with each group processed in an emulation run. Each MBR was treated as an independent experiment and the emulator produced a detailed dataset for each condition, including time-series data for biomass concentrations, DOT and fluorescence measurements.

To streamline the data handling process, a Python script was developed to automate the extraction and curation of KPI values from the emulator outputs. The script identifies relevant KPIs and appends these to their corresponding input parameters, ensuring efficient and accurate preparation for downstream analysis.

The generated dataset was then used as input for the DSp identification, which leverages the open-source Python package called "*dside*" developed by Sachio et al. [7]. This package is available at https://github.com/stvsach/dside and was integral to the subsequent identification and analysis of the DSp.

3.5. Initial Design Space Identification and Bounds Refinement

Initially 512 (2^9) points are inputted into the emulator with the initial DD bounds specified, and we set initial KPI constraints to be the following, based off expert guidance from the KIWI-biolab team:

$$\text{Minimum DOT} \geq 20\ \%a.s. \tag{12}$$

$$\text{Biomass Productivity} \geq 0.4375\ g_{biomass}L^{-1}h^{-1} \tag{13}$$

$$Fluo_{RFP}\ \text{Productivity} \geq 62.5\ RFUh^{-1} \tag{14}$$

To define the boundaries of the DSp, points are classified as either "satisfied" or "violated" based on whether they meet the specified constraints. At 0% tolerance, only points that strictly satisfy all constraints are included within the DSp boundaries; these are referred to as "satisfied" points and are represented in orange in Figure 6. As the tolerance is gradually increased (e.g., to x%), points that fall within ±x% of the constraint values are also incorporated into the DSp. For instance, with a 25% tolerance, a constraint such as Min DOT $\geq$ 20%a.s. is relaxed to include points with Min DOT $\geq$ 15% a.s. (since 25% of 20 is 5, and 20 - 5 = 15). This relaxation applies to all constraints uniformly.

The process begins with a 0% tolerance, which is incrementally increased until a single, unified DSp region is obtained, this is referred to as the tolerance method by Sachio et al. [7].

Table 3: Initial Design Decision bounds.

DSp Size (-)	Number of Regions (-)	Tolerance (%)
0.026	4.00	0
0.046	1.00	3.20

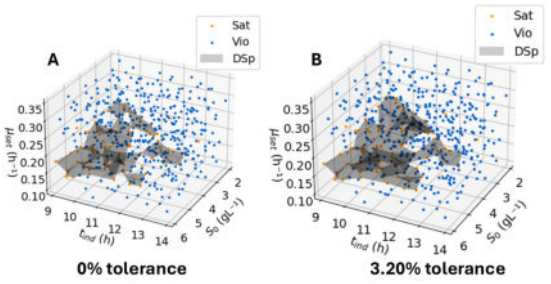


Figure 6: Initial DSps with 512 points used as input into DSp identification program, A: DSp obtained with 0% tolerance this results in 4 disjointed regions, B: DSp obtained with 3.20% tolerance this results in 1 unified region.

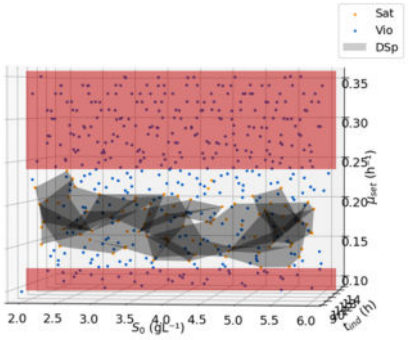


Figure 7: Refinement of initial KSp bounds based on the distribution of satisfied (Sat) and violated (Vio) points. The DSp identified at 3.2% tolerance shows most satisfied points within $\mu_{set} \in [0.12, 0.24]h^{-1}$, guiding the refinement of μ_{set}*bounds* while keeping S_0 and t_{ind} unchanged. The grey region represents the DSp, with orange and blue points indicating satisfied and violated conditions.

Following on from the initial DSp identification we looked at where the most of our satisfied points lie to refine our bounds for when we increase are our dataset size when we run the emulator for more points. Whilst from Figure 6B we observe the satisfied points within our DSp lie evenly across the S_0 axis and t_{ind}, however when we look at a different orientation of the initial DSp with 3.2% tolerance in Figure 7 across the μ_{set} axis the points lie between $[0.12,\ 0.24]h^{-1}$. So we refine the μ_{set} bounds to be as $\mu_{set} \in [0.12, 0.24]\, h^{-1}$ and keep the bounds the same for the other DDs, so we can reduce the amount of violated points in the DSp, which helps reduce the tolerance required to obtain a single DSp. It is noted that a DSp size of 1 means all of your points satisfy the KPI constraints that are imposed.

3.6. Design Space Identification

With refined bounds and a larger dataset (e.g., 4096 points), we repeat the procedure. We adjust tolerance until we identify a single DSp that meets the KPI constraints.

3.7. Flexibility Analysis

Finally, we select one or more NOPs and determine their AORs which the dside package calculates. By quantifying these AORs and their MPARs, we assess how flexible each operating point is, providing insights into process robustness and stability.

3.8. KPI Constraint Sensitivity Analysis

After identifying a baseline DSp and establishing the relationships between DDs and KPIs, a sensitivity analysis was conducted to investigate how varying KPI constraints affects the DSp size. This analysis is essential for understanding the trade-offs between different performance targets and for identifying operational regimes that maintain feasibility under more stringent or relaxed performance criteria.

To systematically investigate the impact of modifying KPI constraints, a Sobol sequence is utilised once again. However, in this case, the approach is applied not to the DDs but instead to generate a diverse sampling of new combinations of KPI constraints. This method enables the exploration of the KPI constraints. By varying, for instance, the minimum acceptable levels of biomass productivity and $Fluo_{RFP}$ productivity across their observed experimental ranges, multiple scenarios are systematically generated. Each of these scenarios corresponds to a unique and specific set of KPI constraints that must be satisfied simultaneously. The resulting set of scenarios provides a comprehensive framework for analysing trade-offs and interdependencies among the different KPIs under varying conditions. The distribution and visualization of these distinct scenarios are illustrated below in Figure 8.

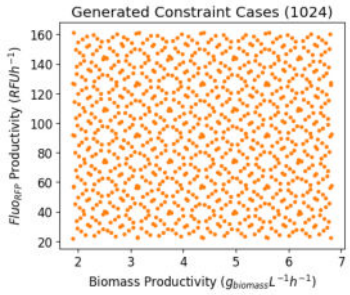


Figure 8: Generation of 1,024 constrained cases using Sobol sequencing, spanning the minimum to maximum ranges of $Fluo_{RFP}$ Productivity and Biomass Productivity. These ranges were derived from the dataset produced by the emulator, ensuring uniform and efficient sampling within the defined bounds.

For each scenario, the DSp identification is repeated using the updated KPI constraints. Points are re-screened, producing a scenario-specific DSp. If constraints are stricter, the DSp may shrink, fragment, or vanish; if they are relaxed, it may expand or merge.

To gauge these effects, the DSp size is recorded for each scenario, enabling direct comparisons. Visualisations such as contour or scatter plots then show how the DSp responds to varying KPI requirements.

4. Results and Discussion

4.1. Design Space Identification

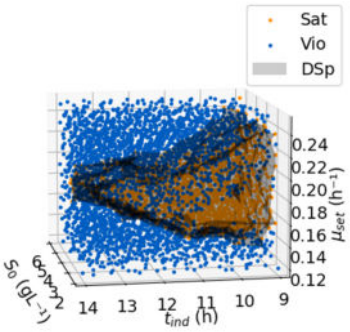


Figure 9: DSp obtained with 4096 points obtained from emulator with a tolerance of 23.91% to obtain a single DSp, with satisfied (Sat) points in orange, violated (Vio) points in blue and the DSp shown in grey.

To assess the flexibility of the KIWI-biolab platform, a DSp was identified using 4096 data points generated by the emulator. The DSp was defined by applying KPI constraints, specifically: (Minimum DOT $\geq$ 20% a.s.), Biomass Productivity ($\geq$ 0.4375 $g_{biomass}L^{-1}h^{-1}$), and Fluorescence Protein ($Fluo_{RFP}$) productivity ($\geq$ 62.5 $RFUh^{-1}$). The tolerance-based method was employed to account for variability within the constraints, achieving a unified DSp with a tolerance level of 23.91%.

The resulting DSp in Figure 9 highlights a trade-off between flexibility and constraint stringency. While the relaxed tolerance enabled a contiguous DSp with a

size of 0.37, questions arise regarding the practical implications of including violated points. Although these violations were deemed negligible, their presence could compromise the consistency of CQAs under industrial conditions, where even minor deviations may lead to scale-up failures. This reliance on relaxed thresholds suggests an inherent limitation in the approach, as it assumes that marginal violations will not significantly impact downstream processes in the MBRrs - a potentially risky assumption without rigorous experimental validation.

The DSp's central regions offer greater operational flexibility, with multiple viable combinations of DDs. However, the narrowing of flexibility near the DSp boundaries underscores a critical limitation: processes operating near these regions are more vulnerable to disturbances or variability, potentially leading to KPI violations. This raises concerns about robustness, especially in high-throughput settings where variability is often amplified.

Additionally, the computational cost of DSp generation was non-trivial, with 135 minutes required to generate 4096 points. This inefficiency limited the resolution of DSp mapping, potentially excluding finer insights into the operational space. While the tolerance-based approach is practical, adopting more computationally efficient frameworks, such as machine learning or surrogate models, could mitigate these challenges and enhance precision.

While the tolerance-based DSp provides a broad operational range and enhances process flexibility, its reliance on subjective tolerances, computational inefficiencies, and risks at the DSp boundaries warrant careful consideration for real-world applications.

4.2. Spread of KPIs in Design Space

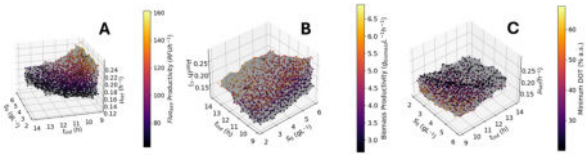


Figure 10: KPI distributions.

The distribution of KPIs within the DSp revealed distinct relationships between design decisions (DDs) and performance outcomes. As can be seen in Figure 10A, high protein ($Fluo_{RFP}$) productivity points were concentrated at shorter induction times ($t_{ind} \approx 9h$) and evenly distributed across initial glucose concentrations (S_0). As t_{ind} increases, protein productivity declined due to metabolic shifts. Early induction ensures microorganisms are in the early exponential growth phase, maximising resources for efficient protein synthesis [17]. Inducing production later on reduces $Fluo_{RFP}$ productivity as metabolic activity and protein synthesis capacity decline due to resources becoming more limited [18]. This trend emphasises the importance of precise timing in bioprocess optimisation.

In contrast, Figure 10B shows higher biomass productivity observed at longer t_{ind} ($\geq$ 13 h), though fewer points exhibit this trend compared to Figure 10A. With decreasing t_{ind} comes points decreasing in biomass productivity, although there are points exhibiting values between 5.5 and 6 $g_{biomass}L^{-1}h^{-1}$ spread across all t_{ind} values in the DSp. This is expected of the biomass productivity because the later the induction time, the more time the microorganisms have to actively cultivate in the bioreactor [19].

High biomass productivity greater than 6.5 $g_{biomass}L^{-1}h^{-1}$ are also observed at low μ_{set} of less than 0.15 h^{-1} as the microorganisms prioritise growth by allocating resources to build cellular components rather than protein production at these set points [20]. Additionally, low growth rate setpoints are associated with energy resources being used more efficiently for growth rather than secondary metabolite production or stress responses. This differs to protein productivity in Figure 10A, where there are values with high $Fluo_{RFP}$ productivity found at high growth rate setpoints of 0.20 h^{-1}, as microorganisms are often in a state of rapid metabolism and high resource turnover when the growth rate setpoint is high, which facilitates the higher productivity of $Fluo_{RFP}$ [21]. These conditions may lead to increased flux through the metabolic pathways that produce energy and precursors for protein synthesis [22]. This illustrates the metabolic trade off that exists between growth and protein synthesis.

Contrary to both Figures 10A and 10B, Figure 10C shows the spread of minimum dissolved oxygen (DOT) only at a growth rate setpoint of 0.15 h^{-1}, reflecting the reduced oxygen demand at this setting [23]. Similar to Figure 10A, high values of minimum DOT are found at early t_{ind} of 9 to 10 hours and with increasing t_{ind} the lower the minimum DOT. At low t_{ind}, protein production is induced early so less biomass has accumulated by the time induction begins resulting in more oxygen present in the media [23]. At later t_{ind} however, higher biomass would be accumulated leading to higher oxygen demand thus lower minimum DOT values.

Notably, all KPIs are independent of initial glucose concentration, as their distributions remain consistent across the DSp. This highlights the insignificance of this DD on the KPIs compared to induction time and growth rate setpoint. The trends highlight a critical challenge in bioprocess design: optimising multiple KPIs simultaneously is constrained by inherent trade-offs. For instance, conditions favourable for $Fluo_{RFP}$ production (high μ_{set} and early t_{ind}) are detrimental to biomass productivity and vice versa, underscoring the need for careful prioritisation of KPIs based on specific process goals.

4.3. Nominal Operating Points

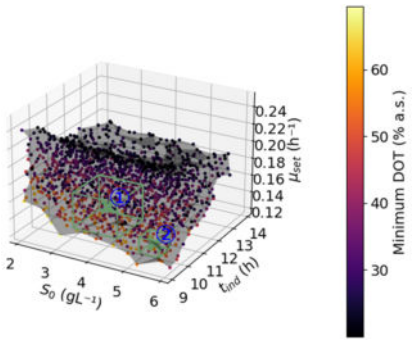


Figure 11: Nominal operating points within the design space.

Table 4: Comparison of Nominal operating points in the design space (percentages refer to the change from NOP1 to NOP2).

	NOP1	NOP2
MPAR: S_0 ($g_{glucose}L^{-1}$)	4.00 ± 0.62	5.50 ± 0.26 (-58%)
MPAR $t_{ind}(h)$	10.00 ± 0.77	9.50 ± 0.33 (-57%)
MPAR: $\mu_{set}(h^{-1})$	0.16 ± 0.02	0.14 ± 0.01 (-50%)
AOR size(-)	0.029109	0.002291 (-92%)

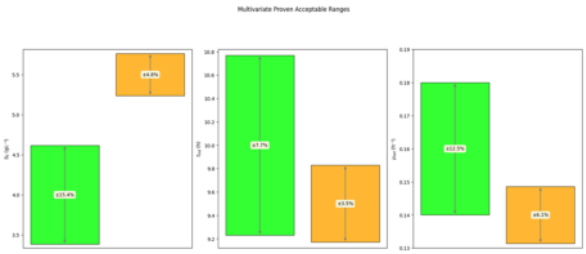


Figure 12: MPAR for each AOR found within the design space

Two arbitrary nominal operating points within the DSp identified in 4.1 were chosen to assess the feasibility and robustness of the process in the KIWI emulator (Figure 11). Table 1 quantifies the NOPs as well as the sizes of the corresponding acceptable operating regions. As can be seen in Figure 11 and table 1, NOP1 is found closer to the centre of the DSp compared to NOP2, which suggests the greater flexibility of operating the MBRs at NOP1 compared to NOP2. This is further highlighted by the AOR for NOP1 being significantly larger than the AOR for NOP2 by 92%. The difference in flexibility is also illustrated in Figure 12 (green area = NOP1 and orange area = NOP2), where the multivariate proven acceptable ranges quantified in Table 4, for the three DDs for NOP1 are significantly larger than those for NOP2. It should also be noted that the MPARs of the initial glucose concentration for both NOPs do not fall in range with each other however, as shown in section 4.3, this DD has little effect on all of the KPIs, therefore this is not a significant observation. Flexibility is measured by the intensity of disturbances the process can handle within the DD [7]. Operating the MBRs under the conditions of NOP2 would mean around 50 – 60% less variation in the DDs would be handled compared to the conditions of NOP1.

Table 5: Comparison of the KPI distributions for NOP1 and NOP2 (percentages refer to the change from NOP1 to NOP2).

	NOP1	NOP2
Max $Fluo_{RFP}$ Productivity ($RFUh^{-1}$)	96.80	78.30
Avg $Fluo_{RFP}$ Productivity ($RFUh^{-1}$)	79.90	70.45
Min $Fluo_{RFP}$ Productivity ($RFUh^{-1}$)	64.32	65.30
Max Biomass Productivity ($g_{biomass}L^{-1}h^{-1}$)	4.63	3.96
Avg Biomass Productivity ($g_{biomass}L^{-1}h^{-1}$)	3.76	3.47
Min Biomass Productivity ($g_{biomass}L^{-1}h^{-1}$)	3.08	3.35
Max Minimum DOT (% a.s.)	69.46	59.70
Avg Minimum DOT (% a.s.)	34.39	40.45
Min Minimum DOT (% a.s.)	20.00	21.32

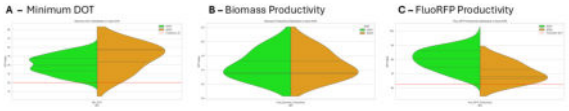


Figure 13: KPI distribution violin plots.

Figure 13 exhibits the KPI distributions for both AOR1 and AOR2 (green = AOR1, orange = AOR2), with Table 5 quantifying the distributions. It can be seen that NOP2 operation results in higher minimum DOT compared to NOP1 operation (Figure 13A), with an average increase of 17.62%. This can be attributed to the lower induction time and growth rate setpoint of NOP2 compared to NOP1 (Table 4) as higher values of minimum DOT occur at low induction times and a low growth rate setpoint as shown in Figure 10C. Additionally, the resulting minimum DOT within AOR2 has a wider range of variation compared to the one within AOR1, suggesting that operation at NOP2 would allow more flexibility with minimum DOT compared to operation at NOP1.

Figure 13B shows even variation of biomass productivity for both NOP operations, which is also quantified in Table 5 where the percentage difference for the average biomass productivity is considerably smaller than those for the averages of the other KPIs. As shown in Table 4, NOP2 has a slightly lower growth rate set point and induction time than NOP1; both DDs being lower does not have a great impact on the biomass productivity hence the similarity of the KPI's distribution.

The distribution for $Fluo_{RFP}$ productivity, however, exhibits a different trend; we would expect to see this distribution be somewhat similar to biomass productivity given the trade-off between the two KPIs however, from Figure 13C, NOP1 achieved a higher productivity than NOP2. This could be possibly attributed to the slightly higher growth rate setpoint of NOP1 than NOP2 by a difference of 0.02 h-1 de-

spite NOP1 having a slightly higher induction time than NOP2 by 0.5 hours (Table 4). This could suggest that growth rate setpoint has a stronger influence on protein productivity compared to induction time as a DD.

From the KPI distributions and the MPARs for the AORs, one can say that while NOP1 offers greater process flexibility and robustness, its higher induction time and growth rate setpoint may introduce inefficiencies in resource utilisation. Conversely, NOP2 sacrifices flexibility for potentially improved oxygen management, albeit with reduced protein yields. This highlights the importance of tailoring NOP selection to the specific priorities of the bioprocess [7].

4.4. KPI sensitivity analysis

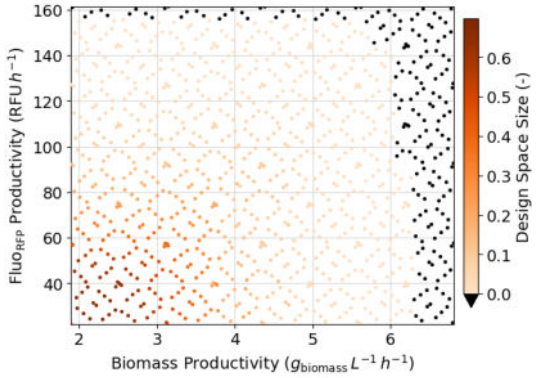


Figure 14: KPI sensitivity analysis.

To further investigate the trade-off between the productivity KPIs introduced in Section 4.2, we conducted a sensitivity analysis by relaxing the constraints on $Fluo_{RFP}$ and biomass productivities whilst keeping the minimum DOT constraint active above 20% a.s. (Figure 7).

A key observation is that at low productivity extremes for both $Fluo_{RFP}$ and biomass, large design spaces exceeding 0.60 occur. This indicates that suboptimal conditions, such as low t_{ind} and low μ_{set}, are easier to achieve and dominate the DSp. These conditions, however, are characterised by inefficient usage of metabolic resources, with the microorganisms operating in an unstressed and unoptimised environment. Such operating points are undesirable for the MBRs; hence the importance of experimental design is emphasised here to avoid these suboptimal regions.

As the productivities of biomass and $Fluo_{RFP}$ simultaneously increase, the size of the corresponding DSp narrows significantly. Achieving high productivities for both KPIs requires optimised trade-offs in metabolic resource allocation, which depends on precise operating conditions. Specifically:

- High $Fluo_{RFP}$ productivity demands a high μ_{set} and early t_{ind} which directs resources towards protein synthesis
- High biomass productivity requires a low μ_{set} and late t_{ind}, favouring resource allocation for cellular growth and maintenance

At the extremes of both KPIs, the design spaces become unfeasible, highlighted by the black points in Figure 7. This is reflective of the inherent difficulty in optimising both KPIs simultaneously due to the differing operating conditions, emphasising the need for targeted trade-offs. Overall, the sensitivity analysis emphasised the inverse relationship between the $Fluo_{RFP}$ and biomass productivities. The DSp size highlights feasible operational trade-offs within the DSp, exploring deeper into the trends shown in the spreads of the productivity KPIs across the DSp.

5. Conclusions

This novel study demonstrated the utility of a computational framework to assess and optimise the process flexibility of the KIWI-biolab platform within a Quality by Digital Design (QbDD) framework. By identifying a contiguous design space at a tolerance level of 23.91%, the research highlighted the platform's capacity for robust operation under constrained Key Performance Indicators (KPIs). The findings underscored the critical trade-offs between protein productivity, biomass productivity, and dissolved oxygen tension, emphasising the need for careful prioritisation of KPIs based on specific bioprocessing objectives.

Flexibility analysis revealed that operating closer to the design space's centre enhances robustness and stability, while operations near its boundaries may be susceptible to disturbances. Furthermore, the exploration of nominal operating points (NOPs) emphasised the importance of balancing flexibility with resource efficiency, as design decisions such as induction time and growth rate setpoints significantly influence process outcomes. Sensitivity analysis confirmed the inherent challenges of optimising multiple KPIs simultaneously, requiring deliberate trade-offs to achieve desired performance metrics.

These insights provide a foundation for scalable and adaptable bioprocess development, addressing key challenges in high-throughput bioprocessing platforms.

6. Outlook

Looking forward we could use a different method to identify the design space for the mini bioreactor platform. We used a method which allowed violations of the KPI constraints in the design space of a highly complex process however, by training a neural network instead, the multivariate data obtained from the KIWI emulator could be learned directly by the ANN rather than rely on predefined tolerances. This would require a higher computational cost yet it would be a lot more effective in the design space identification.

Other KPIs such as biomass yield and protein product yield could be assessed to see how the design

decisions affect their outputs and thus further refine the identified design space for a flexible process in order to enhance scalability to an industrial level.

Additionally, the constraints on the KPIs could be further relaxed to allow more points within the design space and accommodate for more disturbances and variability in the design decisions that can arise. This would further enhance the process flexibility however, relaxing the constraints should be balanced with optimising the mini bioreactor performance overall, which is to be further studied upon.

Acknowledgements

We wish to express our sincere gratitude to Dr.-Ing. M. Nicolas Cruz Bournazou (TU Berlin), Dr. Martin Luna and Mr. Federico Mione (CONICET Santa Fe), as well as Dr. Steven Sachio, Dr. Andrea Galeazzi and Dr. Maria Papathanasiou (Imperial College London), for their invaluable guidance and support throughout this project.

References

[1] Shanley A. Bioprocess Scale-Up: Getting Smarter, Sooner. BIOPHARM INTERNATIONAL. 2020 Nov 1;33(11):54-6.

[2] de Mello AF, de Souza Vandenberghe LP, Herrmann LW, Letti LA, Burgos WJ, Scapini T, Manzoki MC, de Oliveira PZ, Soccol CR. Strategies and engineering aspects on the scale-up of bioreactors for different bioprocesses. Systems Microbiology and Biomanufacturing. 2024 Apr;4(2):365-85.

[3] Pandey K, Pandey M, Kumar V, Aggarwal U, Singhal B. Bioprocessing 4.0 in biomanufacturing: paving the way for sustainable bioeconomy. Systems Microbiology and Biomanufacturing. 2024 Apr;4(2):407-24.

[4] Neubauer P, Anane E, Junne S, Cruz Bournazou MN. Potential of integrating model-based design of experiments approaches and process analytical technologies for bioprocess scale-down. Digital Twins: Applications to the Design and Optimization of Bioprocesses. 2021:1-28.

[5] Haby B, Hans S, Anane E, Sawatzki A, Krausch N, Neubauer P, Cruz Bournazou MN. Integrated robotic mini bioreactor platform for automated, parallel microbial cultivation with online data handling and process control. SLAS TECHNOLOGY: Translating Life Sciences Innovation. 2019 Dec;24(6):569-82.

[6] Mu'azzam K, da Silva FV, Murtagh J, Gallagher MS. A roadmap for model-based bioprocess development. Biotechnology Advances. 2024 May 15:108378.

[7] Sachio S, Kontoravdi C, Papathanasiou MM. A model-based approach towards accelerated process development: A case study on chromatography. Chemical Engineering Research and Design. 2023 Sep 1;197:800-20.

[8] Cruz Bournazou MN, Barz T, Nickel DB, Lopez Cárdenas DC, Glauche F, Knepper A, Neubauer P. Online optimal experimental re-design in robotic parallel fed-batch cultivation facilities. Biotechnology and bioengineering. 2017 Mar;114(3):610-9.

[9] Rohe P, Venkanna D, Kleine B, Freudl R, Oldiges M. An automated workflow for enhancing microbial bioprocess optimization on a novel microbioreactor platform. Microbial Cell Factories. 2012 Dec;11:1-4.

[10] Janakiraman V, Kwiatkowski C, Kshirsagar R, Ryll T, Huang YM. Application of high-throughput mini-bioreactor system for systematic scale-down modeling, process characterization, and control strategy development. Biotechnology progress. 2015 Nov;31(6):1623-32.

[11] Sachio S, Likozar B, Kontoravdi C, Papathanasiou MM. Computer-aided design space identification for screening of protein A affinity chromatography resins. Journal of Chromatography A. 2024 May 10;1722:464890.

[12] Hemmerich J, Noack S, Wiechert W, Oldiges M. Microbioreactor systems for accelerated bioprocess development. Biotechnology journal. 2018 Apr;13(4):1700141.

[13] Olsson L, Rugbjerg P, Pianale LT, Trivellin C. Robustness: linking strain design to viable bioprocesses. Trends in Biotechnology. 2022 Aug 1;40(8):918-31.

[14] Arora M. Cell culture media: a review. Mater methods. 2013 Sep;3(175):24.

[15] Luna M, Mione F, Kaspersetz L, Neubauer P, Martinez EC, Cruz Bournazou MN. Automated regression of bioreactor models using a Bayesian approach for the operation of robotic platforms. Preprint. SSRN; 2024 Oct 29.

[16] Shukman BV. Random and quasirandom sequences: Numerical estimates of uniformity of distribution. Mathematical and computer modelling. 1993 Oct 1;18(8):39-45.

[17] Neidhardt FC, Curtiss R. Escherichia coli and Salmonella: Cellular and Molecular Biology. ASM Press. 1996.

[18] Bentley WE, Mirjalili N, Andersen DC, Davis RH, Kompala DS. Plasmid-encoded protein: The principal factor in the "metabolic burden" associated with recombinant bacteria. Biotechnology and Bioengineering. 1990;35(7):668–681.

[19] Lee SY, Chang HN. High cell density culture of Escherichia coli. Trends in Biotechnology. 1993;11(11):456–463.

[20] Shuler ML, Kargi F. Bioprocess Engineering: Basic Concepts. Pearson Education. 2017.

[21] Baneyx F. Recombinant protein expression in Escherichia coli. Current Opinion in Biotechnology. 1999;10(5):411–421.

[22] Papoutsakis ET. Equations and calculations for fermentations of butyric acid bacteria. Biotechnology and Bioengineering. 1984;26(2):174–187.

[23] Cooney CL. The fate of oxygen in bioreactors. Trends in Biotechnology. 1981;10(4):101–106.

Optimal Control of PV-Battery Systems in Commercial Buildings Using Tabular Q-Learning

Yingshan Gao and Ziyang Liu
Department of Chemical Engineering, Imperial College London, U.K.

Abstract

The pressing need for decarbonization has significantly expanded the renewable energy sector, with a particular emphasis on achieving Net Zero targets. Among these, solar photovoltaic (PV) energy is projected by the International Energy Agency (IEA) to be a leading renewable source. This necessitates advanced control techniques to balance load demand and supply, thereby reducing grid utilization and accommodating real-time electricity prices. This study employs value-based tabular Q-learning (TQL), adaptable to stochastic systems, to optimize a real-world PV-battery system in a Sainsbury store. The system is trained using a historical dataset over one year with finely-tuned parameters. Performance is evaluated against key performance indicators and compared with existing rule-based control systems. The results indicate a 56% reduction in grid export, albeit with only a marginal decrease in cost and emissions, showcasing improved performance across all metrics during spring and summer, yet limited effectiveness in autumn and winter.

1 Introduction and Background

There is an increasing number of countries across the world committed to reducing greenhouse gas (GHG) emissions to net zero due to the growing concern about climate change, with 70% of countries aligned with this goal[7]. This trend has led to a dramatic expansion in the renewable energy sector worldwide over the past few years, accelerating the energy transition[7]. Among all the key clean energy sources, the 2024 International Energy Agency (IEA) report highlighted that solar photovoltaic (PV) power will become the fastest-growing resource in the next decade, contributing 29% of the overall clean energy deployment in 2035 and playing a leading role in tripling renewable energy capacity[8]. Sainsbury's, as one of the most popular supermarket chains, has committed to the Towards Net Zero pledge[13]. A range of studies have been conducted to reduce reliance on fossil fuels, with ongoing research efforts supporting Sainsbury's supermarket chain in achieving its Net Zero Emission [6].

The practical implementation of PV systems is highly dependent on natural conditions, such as weather, which can result in imbalances between energy demand and supply, particularly when excess PV power remains unused [12]. Integrating PV systems with battery energy storage systems (BESS) is considered a practical and effective solution to ensure stable system operation. This approach has gained considerable popularity in recent years due to its capacity to mitigate fluctuations in energy generation.[12].

Since 2013, there has been a rapid increase in research on reinforcement learning for energy management in various systems, such as electric vehicles and HVAC systems. However, research on energy storage systems remained limited until 2015 [17]. An early study proposed the use of Q-learning with a Softmax action selection strategy to reduce costs while meeting energy demands, demonstrating the initial feasibility of Q-learning for energy management [5]. Later work implemented reinforcement learning frameworks to derive optimal policies for improved decision-making, highlighting Q-learning's ability to address price trend variations [4]. More advanced algorithms, such as hierarchical approaches combined with Q-learning, were developed to adapt to system dynamics and enhance model scalability [9].Recent studies have reviewed the application of traditional Q-learning control frameworks in various energy systems. For instance, Q-learning was applied in a dairy farm to allocate wind power effectively, resulting in a 13.4% reduction in grid imports [1]. In addition, training the traditional Q-learning approach with historical data has proven to be an effective method for cost optimisation.[3]

Most of the studies highlighted the strength of Q-learning in its adaptability to stochastic data due to the model-free and value-based features of the algorithm, providing insights and inspiration to address the difficulties in the Sainsbury's building. While previous research has comprehensively evaluated household grid control, there is a notable lack of studies investigating commercial buildings, which differ significantly in scale and operational schedules from residential systems. Therefore, a real-world supermarket store will be tested using the proposed tabular Q-learning (TQL) method. The controller parameters will be fine-tuned, trained, and evaluated with the provided data. The objective

is to reduce costs and carbon emissions compared to both the original system without a battery energy storage system (BESS) and the rule-based control (RBC) framework previously implemented in the same store of interest [11].

In this paper, the TQL algorithm and the system configuration are briefly introduced. During the application of the algorithm to the system, optimal parameters are designed and refined through sensitivity analysis. The parameter combination with the best performance is proposed, and the results are evaluated and compared with the existing RBC approach using key performance indicators (KPIs).

2 Methodology

A real-world building will be analysed in this paper in Section 3 using the principal model described in this section. The existing rule-based control [11] of this system will be presented which serves as the baseline control, whilst the principles of reinforcement learning will be outlined as the mathematical framework of the proposed control technique in section 3. The controllers in all cases managed the charging/discharging of the battery and thereby import/export of the grid power by a certain power [11]. The primary objective of the controller is to account for the demand for building energy and the excess PV energy to optimise load shifting effectively at each timestep.

2.1 Model description

The general PV-Battery system contains 4 components: the battery, the building, the PV panels, and the grid. Those components would always satisfy the energy balance equation (1).

$$P_{pv}(t) + P^{im}_{grid}(t) + P^{dch}_{bess}(t) = P_{bu}(t) + P^{ex}_{grid}(t) + P^{ch}_{bess}(t) \tag{1}$$

where $P_{pv}(t)$ is the power generated from PV panels and imported into the system, $P_{bu}(t)$ is the power required from the building, $P^{im}_{grid}(t)$ and P^{ex}_{grid} represent the import and export power of the grid respectively, and $P^{ch}_{bess}(t)$ and $P^{dch}_{bess}(t)$ represent the battery charging and discharging rate respectively.

To further assess the system performance, further calculations are needed for the grid energy imported $E^{im}_{grid}(t)$ and exported $E^{ex}_{grid}(t)$ within a timestep of Δt and are described in Equation (2)

$$\begin{aligned} E^{im}_{grid}(t) &= P^{im}_{grid}(t) \times \Delta t \\ E^{ex}_{grid}(t) &= P^{ex}_{grid}(t) \times \Delta t \end{aligned} \tag{2}$$

The energy accumulated in the battery at time t, $E^{acc}_{bess}(t)$, is calculated as follows and is further used to calculate the state of charge (SoC) of the battery in Equation (4), that described the percentage of storage in battery occupied.

$$E^{acc}_{bess}(t) = E^{acc}_{bess}(t-1) + (P^{ch}_{bess}(t) - P^{dch}_{bess}(t)) \times \Delta t \tag{3}$$

$$SoC(t) = \frac{E^{acc}_{bess}(t)}{Battery\ Capacity} \tag{4}$$

2.2 Rule-Based Control

The existing rule-based control of the same system of interest described in [11] is used as the baseline control. There are two control models of rule-based control: the fixed charge and self-consumption modes, and they are assigned to specific time ranges for daily operations. A constant fixed charge power is manually set to the controller, maintained primarily by excess PV and supplemented by grid energy. In the self-consumption mode, charging and discharging of the battery will depend on the excess in PV, it stores energy when there is excess power in PV, and discharges to the building when the PV is insufficient. The SoC limit of the battery was set between 40% and 95%, considering a sufficient backup amount might be required for any emergency. The minimum and maximum power flow of the grid and battery were constrained according to the network constraints from the Distribution Network Operator (DNO) and the battery's physical properties [11].

The allocation of energy in both modes takes the PV energy as the primary source to charge the building and the battery. In contrast, the grid will be used as the last resort to import when both PV and battery have insufficient power, or to export when PV is in excess and the battery is fully charged to effectively achieve load shifting.

The RBC was only designed for 2 periods in a day: 1) From 00:00 AM-06:00 AM, fixed charge mode is in operation, and 2) From 06:00 AM-12:00 AM, the controller is set to be self-consumption mode. The midnight charging prepares the battery to adapt to the PV generation variations during the rest of the day while encouraging charging at a low price period.

2.3 Tabular Q-learning Formulation

The Q-learning formulation is fundamentally based on the Bellman equation[18], which defines the relationship between the Q-value of a state-action pair and the expected cumulative reward. The Q-value is updated iteratively using the following equation:

$$Q(s,a) \leftarrow Q(s,a) + \alpha \left[r + \gamma \max_{a'} Q(s',a') - Q(s,a) \right] \tag{5}$$

where Q(s, a) represents the Q-value for the current state s and action a, reflecting the expected cumulative reward associated with that state-action pair. α is the learning rate, which determines the degree of influence of new information on the existing strategy. r is the immediate reward obtained from

the action. γ is the discount factor, controlling the weight of future rewards relative to immediate rewards. Finally, $\max_{a'} Q(s', a')$ represents the highest Q-value among all possible actions in the next state s'.

As an off-policy algorithm, Q-learning separates the process of learning the optimal policy from the actions carried out during exploration. At each step, the Q-value updates are informed by the maximum Q-value of the next state-action pair in the current Q-table, rather than the action actually taken in the environment. This approach enables the agent to rapidly converge toward the optimal strategy.

3 Case Study

The control method was applied to a real-world commercial building, with a PV-BESS system connected to the grid provided by [11]. The connections of the system components are illustrated in Fig. 1.

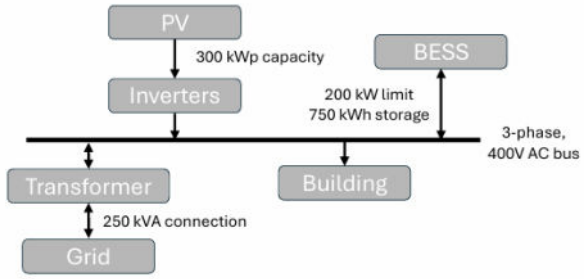


Figure 1: The system configuration of the examined store [11]

Aligning with the energy balance described in Section 2.1, the PV and building load are single direction, where the power flows of BESS and grid can go both ways. The PV panels were installed to the roof of the store, with a peak power capacity of 300kWp and connected to the inverter to supply the 3-phase, 400V AC bus[11]. A Lithium Nickel Manganese Cobalt (NMC) cell was installed, providing a maximum capacity of 750kWh and maximum charge/discharge power of 200kW [11]. A one-year dataset collected from July 2023 to July 2024 was provided by the store, containing real-time data of PV power generation (kW), building load (kW) for every 5 minutes, and electricity price ($£/kWh$) and carbon intensity (gCO_2/kWh) for every hour. This set of data will be used to train the agent of the proposed method.

The effectiveness of control techniques may differ according to regions, due to the differences in weather conditions. Two-day data sample from summer and winter was plotted in Fig. 2, providing the representation of energy gap variation due to seasonal differences in this specific store location.

3.1 Tabular Q-learning (TQL) Implementation

To implement Tabular Q-Learning (TQL) for deriving an optimal battery charging schedule, it is necessary to ensure that both the state and action spaces are discrete.[18] The agent determines actions, specifically the battery charging rate, based on a sequence of states. At the same time, grid import is calculated using the given data and energy balance equations mentioned in Section 2.1, which informs the cost-related reward. Data from the battery system is also employed to evaluate the reward associated with the state of charge (SoC). These calculated rewards are then fed back into the agent, enabling it to iteratively update the Q-table and determine an optimal charging schedule over one year.

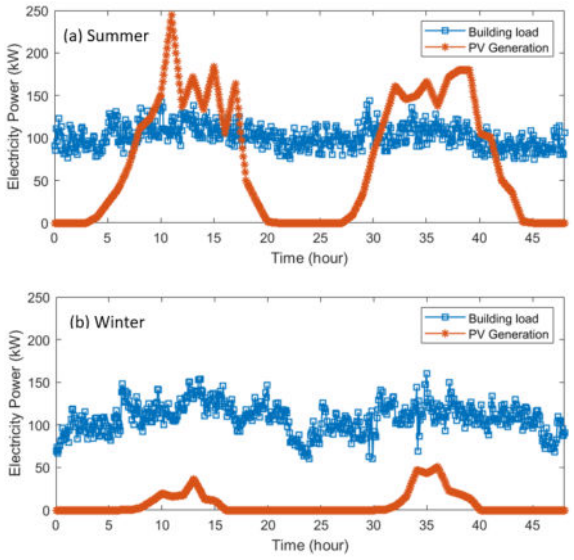


Figure 2: The Building load and PV generation profiles of the sampled two days during (a) summer and (b)winter

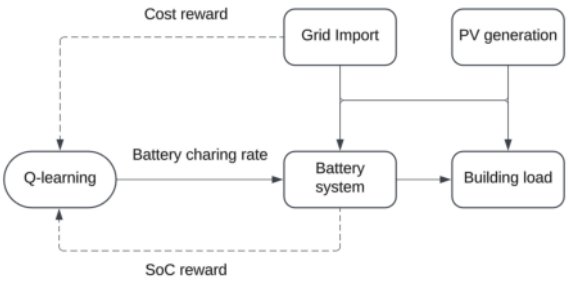


Figure 3: Control logic

3.1.1 State space

This study involves three key state variables: the State of Charge (SoC), electricity price (p_{ele}) and PV excess (P_{excess}). SoC represents the current level of energy stored in the battery, while p_{ele} indicates the cost of electricity from the grid per kWh at a given time. In addition, P_{excess} represents the additional energy produced by the PV system after fulfilling the energy requirements of the supermarket's building load, with the following equation:

$$P_{\text{excess}}(t) = P_{pv}(t) - P_{bu}(t) \tag{6}$$

The state space is defined as:

$$S = \{\text{SoC(t)}, P_{\text{excess}}(t), p_{\text{ele}}(t)\} \tag{7}$$

where t refers to the current time.

The SoC is divided into three discrete ranges: low, medium, and high, represented by the bins [0, 0.2, 0.8, 1]. This segmentation not only helps the model distinguish between different energy levels in the battery but also considers battery health, as it is preferable to maintain the SoC within the range of 0.2 to 0.8, which is treated as a single state.

At the same time, the P_{excess} and p are discretized as well according to the observed data boundaries, allowing for more accurate and effective decision-making within the model. The detailed values for state space are provided in Table 1.

Table 1: *State variables*

State	Values
SoC(t)	[0, 0.2, 0.8, 1]
$p(t)$	[0.1, 0.2, 0.3, 0.4, 0.5]
$P_{\text{excess}}(t)$	[-120, -80, -60, -30, 0, 50, 75, 100, 150, 225]

3.1.2 Action space

The action determined by the Q-learning controller in this study is the battery charging rate. As mentioned in Section, the battery's rated AC charging/discharging capacity reaches up to 200 kW, the action space is defined accordingly as:

$$A = [-200, -150, -100, -50, 0, 50, 100, 150, 200] \tag{8}$$

3.1.3 Reward

The main objective of this study is to minimise costs. Therefore, the first component of the reward function is related to it, which is calculated based on the electricity price and grid import. In practical scenarios, exporting excess energy back to the grid can generate revenue. However, due to the absence of future information in the observation state, the model is unable to accurately distinguish that the export price (£0.06/kWh) is significantly lower than the import price (£0.1-£0.5/kWh). As a result, the agent tends to prioritise selling electricity directly rather than charging the battery for use during peak periods, which ultimately leads to higher overall costs. To address this issue, a cost penalty (c_e) is introduced to limit the grid export.

$$r_{\text{cost}} = \begin{cases} p_{\text{ele}} \times E^{im}_{\text{grid}} \\ c_e \times E^{ex}_{\text{grid}} \end{cases} \tag{9}$$

Another key factor considered in this study is carbon emissions resulting from power usage. However, given the highly variable nature of carbon intensity, which is also correlated with the cost of electricity, the decision was made to exclude a carbon emissions term from the reward function in order to simplify the algorithms.

Moreover, it is essential to include an element in the reward function to ensure that the SoC remains within a physically feasible range. This component is divided into two aspects: the first is a penalty for violating the SoC limits r_{SoC1}, which is triggered when the SoC exceeds the defined range. The second part is a penalty for invalid actions r_{SoC2}, which is applied if the system attempts to charge when the battery is full or discharge when it is empty. The latter is considered more critical, and thus the penalty magnitude is ten times greater than the former. To prevent the SoC from exceeding the range of 0 to 1, a buffer is introduced, setting the penalty threshold within the range of 0.2 to 0.8, which also helps maintain battery health effectively. In addition, a square function is employed in the reward function to penalise SoC violations more severely as the deviation increases.

$$r_{\text{SoC1}}(t) = \begin{cases} (\text{SoC} - 0.8)^2 & \text{if SoC} > 0.8 \\ (0.2 - \text{SoC})^2 & \text{if SoC} < 0.2 \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

$$r_{\text{SoC2}} = \begin{cases} (\text{SoC} - 0.8)^2 \times 10 & \text{if SoC} > 0.8 \text{ and } A > 0 \\ (0.2 - \text{SoC})^2 \times 10 & \text{if SoC} < 0.2 \text{ and } A < 0 \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

Last but not least, weighting coefficients are introduced to balance the various elements of the reward function in order to achieve an optimal scheduling strategy. Details regarding the selection of weighting coefficients and export penalties can be found in Section 4.1.

$$R = -w_1 \times r_{\text{cost}} - w_2 \times (r_{\text{SoC1}} + r_{\text{SoC2}}) \tag{12}$$

3.2 Agent training

The Q-learning agent is trained using the epsilon-greedy algorithm, with an initial exploration rate of 1 and a decay rate of 0.995, to achieve a balance between exploration and exploitation, Initially, the agent focuses on extensive exploration, gradually shifting towards using past experiences as the number of episodes increases.

$$\epsilon = \epsilon \times \beta \tag{13}$$

By analysing the decay of epsilon, it was initially determined that 1000 episodes would be sufficient to meet the exploration and exploitation requirements.Furthermore, the learning rate is defined as 0.5, while the discount rate is set to 0.9. The minimum epsilon value is established at 0.01, with a sensitivity analysis provided in Section 4.3.

The nominal hyperparameters are summarised in Table 2.

3.3 The key performance indicators (KPIs)

To determine the performance of the tabular Q-learning, cost and carbon emission are calculated by the following equations:

Table 2: *Nominal hyperparameters*

Parameter	Symbol	Value
Initial exploration rate	ϵ	1
Learning rate	α	0.5
Discount factor	γ	0.9
Decay rate	β	0.995
Minimum exploration rate	ϵ_{min}	0.01
Number of episode	N	1000

$$\text{Cost} = p_{\text{elec}} \times E^{im}_{\text{grid}} \tag{14}$$

$$E_{\text{carbon}} = I_{\text{carbon}} \times E^{im}_{\text{grid}} \tag{15}$$

where I_{carbon} represents the carbon intensity given in the dataset.

3.4 Sensitivity Analysis

The sensitivity analysis on the reward weighting was conducted to evaluate how parameter changes influence the overall performances and discover possible outcomes on top of the nominal parameters. Experiment design and nominal parameter selection are important parts of the analysis and the orders of magnitude on each reward component is a reasonable starting point. To prevent SoC from violation for safety reason, the order of magnitude of the sum of the SoC violation penalty should be about 2 orders of magnitudes larger than the sum of other terms. After an initial investigation of the data, the cost of each timestep is in the range of approximately [0.1, 15]. Since the SoC is in the magnitude of 10^{-1}, squaring it as designed in Equation (10) and (11) will then result in magnitude of nearly 10^{-2}. With the concepts of the ranges and the expected order of magnitude difference in mind, and the negative sign for a penalty, w_2 is fixed to 1×10^4 while the values of w_1 being examined individually from 10 to 85.

There is also a 'hidden' term in the cost penalty as the grid export penalty specified in Equation (9) to discourage the grid export. The penalty coefficient c_e was selected around the maximum electricity price of about 0.5 £/kWh observed from the dataset, to inform the agent that exporting is even more undesired than importing grid power at price peak. Therefore, the values for examination range from 0.4 to 0.7.

Finally, to get more understanding of learning behaviour, the effects on the minimum epsilon, ϵ_{min} were studied from 0.1 to 0.01. The experimenting values of the parameters of interest are summarised in Table 3

Table 3: *Values of cost penalty coefficients, w_1, export penalty coefficients, c_e and minimum exploration rate, ϵ_{min} to be analysed.*

Parameter	Values of parameter to test
w_1	10, 20, 40, 60, 80, 85
c_e	0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7
ϵ_{min}	0.01, 0.05, 0.1

4 Results and discussions

The training results will be first presented with the tuned parameters and training performances. The annual simulation will be compared with the baseline without battery system, and the rule-based control mentioned in section 2, regarding the KPIs specified in section 3. Seasonal performances of sampled 2 days will be evaluated. Finally, the sensitivity analysis of parameters will be presented to provide further understanding of how parameter changes affect the annual performances.

4.1 Training Results

The tuned parameters are summarised in Table 4. This combination of parameters gives the best annual performance during the sensitivity analysis described in Section 3.5.

Table 4: *Optimal weights of components in the reward function*

Parameter	Symbol	Value
Cost weighting	w_1	80
violation penalty	w_2	1×10^4
Export Cost Coefficient	c_e	0.55

This combination of the reward weighting successfully kept the battery state of charge within its range of [0,1] while achieving the lowest annual cost and carbon emission as shown in Table 5.

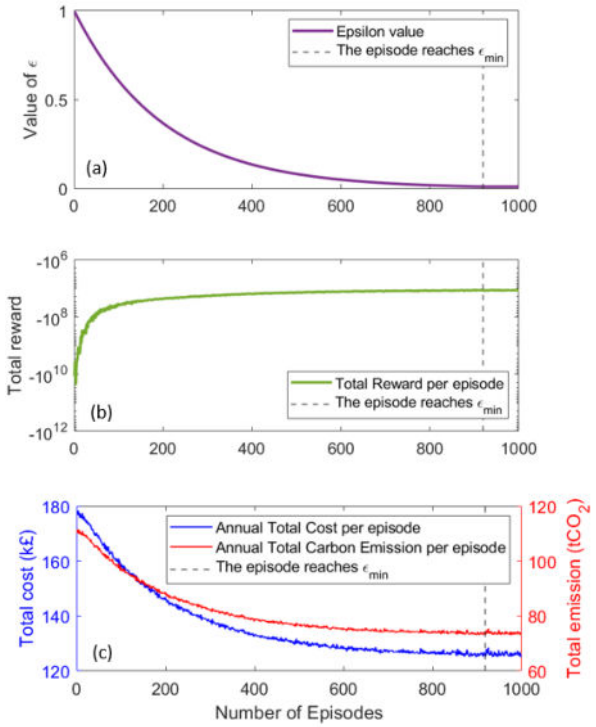


Figure 4: The training curve of the agent down the episode with nominal hyperparameters and a combination of reward weights of the best performance. (a) represents the exploration rate decay ϵ down the episode; (b) shows the total reward convergence; (c) describes the annual electricity cost and carbon emission reduction down the episode

As plotted in Fig 4, the ϵ decayed from the initial exploration rate of 1 to minimum epsilon ϵ_{min} of 0.01 after 919 episodes, which proves the consistent learning of agent throughout the process, and

Table 5: *Annual Performance Summary*

KPIs	Baseline	RBC	TQL	%Reduction from Baseline	%Reduction from RBC
Grid Import (MWh)	559	552	513	8.2	7.1
Grid Export (MWh)	65	43	19	71.2	56.6
Net Import (MWh)	494	509	494	0	2.9
Import cost(k£)	138	130	125	9.2	3.5
Export earnings(k£)	4	3	1	71.2	57.2
Net cost(k£)	134	127	124	7.4	2.4
Carbon emission(tCO_2)	79	75	72	8.8	3.2

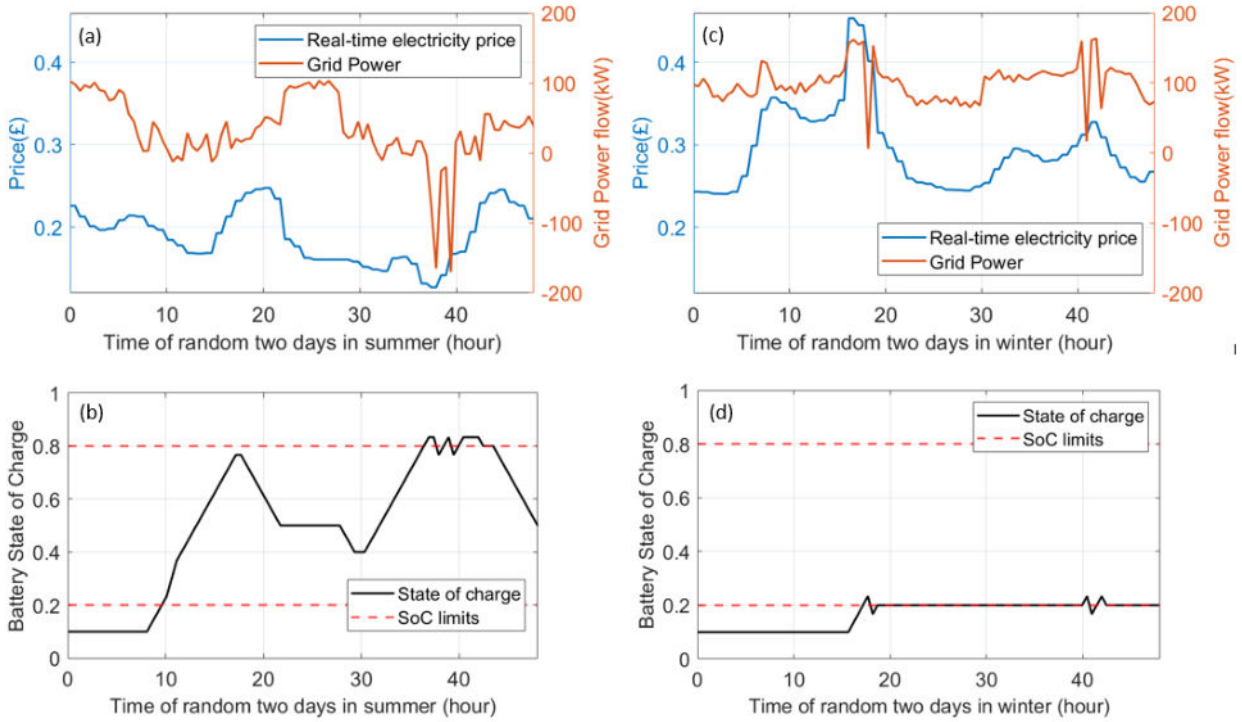


Figure 5: Dynamic change during random two-day examples extracted from the summer and winter periods.

suggested that 1000 episodes could sufficiently converge the exploration rate to its minimum. The total reward received in each episode started from an order of magnitude of -10^{10} and converged to -10^{7} as the training progressed. The total reward from the first 200 episodes was significantly more negative of about 3 orders of magnitudes than the converged total reward, indicating sufficient exploration at the first stage of training. The annual cost of electricity and carbon emission were also effectively reduced as the training progressed, and gradually stabilised towards over time.

4.2 Annual Performance

4.2.1 Overall Comparison

The comparison of annual performance between the system without battery (Baseline), rule-based control (RBC) and tabular q-learning control (TQL) with the KPIs is summarised in Table 5. The rule-based control (RBC) is shown to notably improve the load shifting and reduce cost and carbon emission. Due to the trackable patterns of the PV power generation, building load and electricity price, RBC can significantly improve the performance by carefully scheduling to adapt the patterns, such as avoiding charging at peak price period, or switching to self-consumption mode when PV energy is rich etc. In other words, it benefits from the prevision of overall trend. However, it is less adaptive to address time-to-time changes of the system behaviour and the huge variation, such as seasonal change and unexpected spikes.

In contrast, the tabular Q-learning control (TQL) is more flexible on the dynamics of the system, which significantly reduced unnecessary export by 56% from RBC. However, there is only marginal improvement of cost and carbon emission by TQL of 2.4 % and 3.2 % respectively. The detailed comparisons between two comparisons will be conducted in seasonal analysis, providing response of controllers to weather differences."

4.2.2 Analysis of seasonal dynamics

To further understand the system dynamics and assess how the control adapts to seasonal scenarios, a comprehensive study of specific two-day periods (48 hours) sampled from summer and winter is conducted in this section. These two seasons are a reasonable representation of significant variation in

system states which provide a reasonable basis for evaluation.

The behaviour of the grid power in response to the changes in electricity price and battery state of charge was plotted in Fig. 5. It demonstrated a reasonable response of the grid power, with the electricity price and grid power moving in a counter direction with the price. An obvious load shift can be seen in the 15th - 30th hour in the summer case. During this period, electricity prices are high from hours 15 to 22, resulting in the grid power dropping nearly to zero. As the price began to drop around the 21st hour, the grid load started to increase accordingly. This movement is solid proof that TQL is able to adjust the grid power according to the real-time price.

However, TQL seems to be less adaptive to winter scenarios. Due to the shortage of PV energy and relatively higher prices in winter, grid power is heavily used to supply building loads, causing large penalties for purely charging the building and leading to less availability to charge the battery even when the prices are low. This is due to its lack of consideration of future trends. An example can be seen during hours 10 to 20 in the winter scenario, where the prices before the 15th hour were relatively low but the agent had no realiation of storing charge. This failure of decision-making caused more expensive charging from hours 15 to 20 when the electricity price was at its peak.

At those conflicting points where the SoC is often at the edges of the boundaries, the dynamics of the agent struggling with balancing the electricity price and SoC violation can be observed. In the same example in the winter scenario around the highest price period, importing grid power can gain a high penalty but the battery state of charge is about to drop out of the lower bound, preventing the battery from discharging. Those conflicting scenarios result in sharp fluctuation of the grid power as presented in the graph. Since the agent always receives huge penalties in those relevant steps, it constantly changes to the reversed actions to try to overcome the penalties received from the last step, leading to oscillatory behaviours.

In general, the dynamics illustrated a better control performance in summer but compromised performance in winter.

4.2.3 Seasonal Comparison

Those sampled days are also compared with the rule-based control as summarised in Fig. 6.

The results align with the common sense that: more PV will be generated on sunny summer days, reducing the usage of grid power. At the same time, less PV is in excess in overcast winter days causing more reliance on the grid power. Remarkably, tabular Q-learning significantly reduced the power exported to the grid in spring and summer. This observation is evident in the cases where RBC was forced to export in the early morning more often in spring and summer days. This is particularly where real-time adaptation is more effective in addressing the over-storing. Although the net costs of the two control strategies are closed in spring and summer, the RBC cost was largely discounted by exporting the grid, which is not desired.

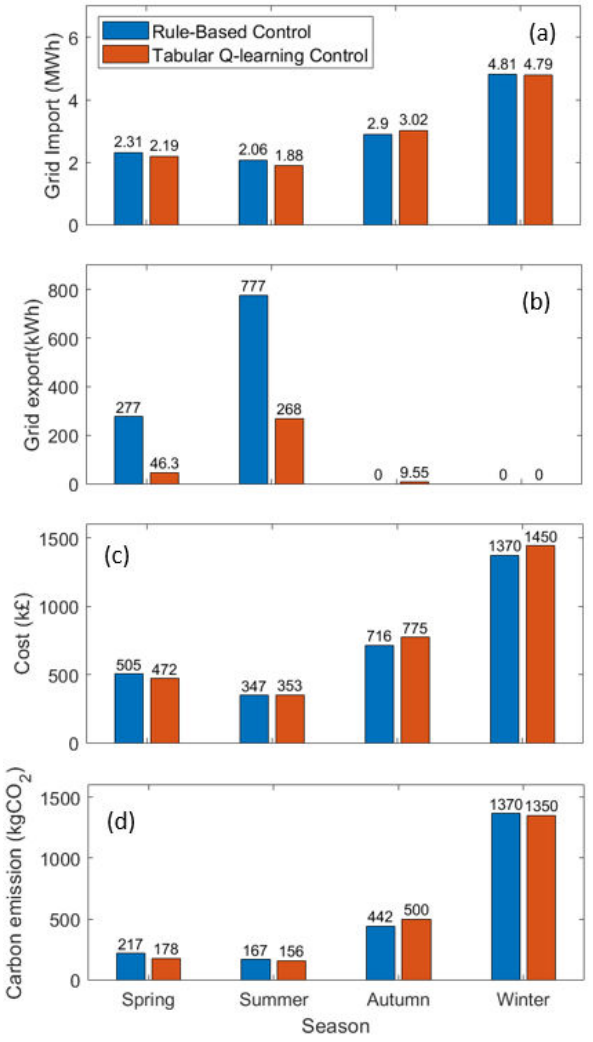


Figure 6: Results comparison between RBC and TQL during random two-day examples extracted from different seasons.

Whereas in autumn and winter days, the performance of tabular Q-learning becomes comparable with RBC. During these seasons, the battery state of charge will often be around the lower violation limit due to the PV generation shortage. Even the agent recognised that importing electricity is costly, it has no choice but to do so to maintain the normal operation of the building and the physical properties of the battery state of charge. In other words, the cost penalty is less effective. RBC in this case takes advantage of anticipating future trends, so that the period of peak price charging can be somehow reduced, resulting in lower costs than TQL with

negligible export earnings in both cases.

In summary, the improvement of TQL from the baseline RBC is remarkable in spring and summer, but more comparable in autumn and winter, resulting in a slight improvement in the overall performance as agreed in Section 4.2.1.

4.3 Sensitivity Analysis

4.3.1 Cost Coefficient, w_1

The designed experiments in section 3.3 were conducted with the listed values, and the results were displayed in Fig. 7 When the cost penalty is low

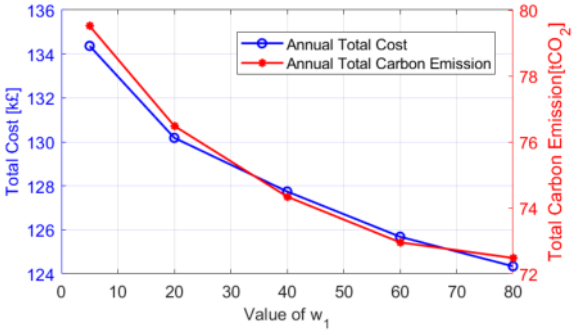


Figure 7: Sensitivity analysis of cost coefficient of reward function w_1, with fixed export penalty c_e of 0.55, trained through 1000 episodes with a minimum epsilon ϵ_{min} of 0.01

(5-40), the weight in the SoC penalty is in control, causing limited utilisation of the battery system and cost and carbon dioxide reduction. As the penalty gets larger, the cost penalty coefficient gets more effective and balanced with the the magnitude of violation penalty coefficient. While the cost penalty achieved 85, the controller failed to cap the SoC limits, indicating over-weighting cost penalty. Therefore, the optimal weight is when $w_1 = 80$, where the penalties were appropriately balanced.

4.3.2 Export Penalty, c_e

The sensitivity analysis on c_e was summarised in Fig.8. At low grid export penalties, the strength is not strong enough to prevent grid export. As the penalty coefficient increases, it reaches an optimal at 0.55, where the export penalty is well-balanced with other components. However, over-penalising the grid export results in an overall increase in annual export, and therefore fails to control the cost and carbon emission.

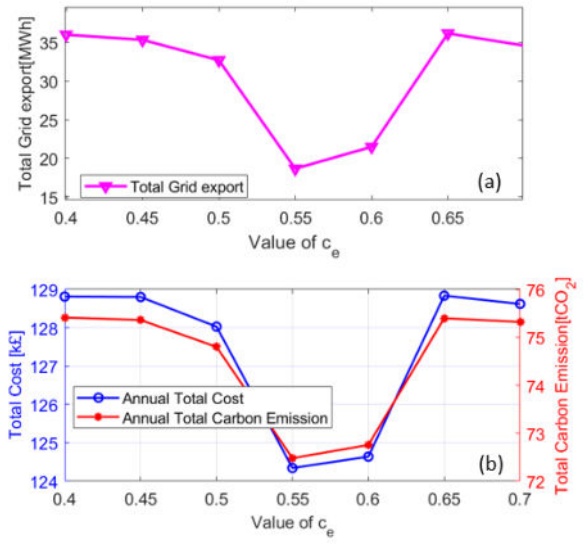


Figure 8: Sensitivity analysis of export penalty c_e, with a fixed cost coefficient w_1 of 80, trained through 1000 episodes with minimum epsilon ϵ_{min} of 0.01

As discussed, the agent only focuses on the current state, which means it pays less attention to the possible drawbacks to the future caused by current actions. While increasing the export penalty reduces the grid export of the immediate step, it does not prevent the grid export at inevitable circumstances, such as SoC violation. In this case, further export will occur to cover up the failure decisions by the previous steps.

4.3.3 Minimum epsilon, ϵ_{min}

The summary of cost and carbon emission responses was plotted in Fig 10, and the dynamic total reward from the 450th episode (when minimum epsilon 0.1 was reached), is displayed in Fig.9.

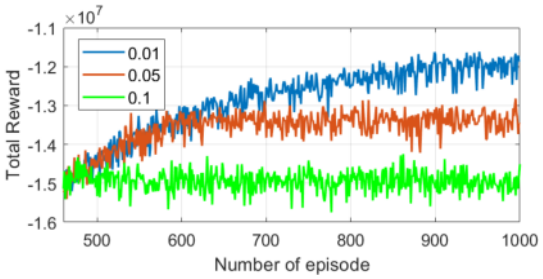


Figure 9: Reward analysis of minimum epsilon ϵ_{min}, with fixed export penalty c_e of 0.55 and fixed cost coefficient w_1 of 80, trained through 1000 episodes

The results demonstrated optimal balance of exploration and exploitation was achieved with the

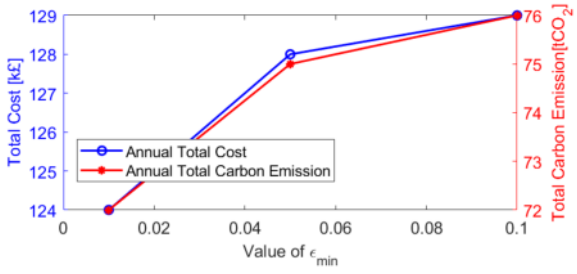


Figure 10: Cost and carbon analysis of minimum epsilon ϵ_{min}, with fixed export penalty c_e of 0.55 and fixed cost coefficient w_1 of 80, trained through 1000 episodes

minimum exploration value of 0.01, observed from sufficient convergence in Fig.4, and optimal performance for carbon and cost in Fig. 10 The initial training with $\epsilon_{min} = 0.1$ failed to capture the SoC violations, indicating a high exploration rate that the agent may still choose the actions that violate the SoC. The performance for an exploration rate of 0.05 is slightly better than 0.1. Although the degree of exploration is still high, the reward is not converged to effectively reduce the total cost and emission, resulting in sub-optimal performance [14]. As such, the minimum exploration rate maintains sufficient exploitation to account for safety, while remaining some exploration even in a hard constraint model. Further analysis in lower minimum exploration rate was not examined as over-constraining the agent to exploitation may reduce its adaptability to other systems that differ geometrically[14]. It emphasises the importance of carefully tuning not only to make it feasible to the current model but also to generalise to global applications.

5 Conclusions

In this study, tabular Q-learning is applied to optimise the control of a PV-battery system in commercial buildings, with electricity prices and carbon intensity of Britain. This study not only analysed the annual and seasonal performance but also thoroughly compared the result with rule-based control in the previous study and baseline. In addition, a sensitivity analysis of the cost coefficient, export penalty, and minimum epsilon was conducted.

Below are the key findings of this study:

1. In terms of annual performance, tabular Q-learning reduced the cost and carbon emissions by 7.4% and 8.8% respectively from the baseline, while decreasing by 2.4% and 3.2% respectively compared to the rule-based control.

2. Upon consideration of seasonal performance, it exhibited better efficiency during the spring and summer, while its effectiveness declined in the autumn and winter. This is attributed to two primary reasons: the shortage of PV and relatively high electricity prices in winter, and the lack of future information. Due to the high electricity price, the agent refuses to import the grid even during off-peak times considering the cost. However, the absence of data on future electricity prices leads to the failure to prepare for peak pricing periods, which results in the agent being forced to charge a high price to avoid the consequence of SoC violation penalties. In contrast, the rule-based control performs better control, benefiting from its fixed charging schedule between 0:00 and 6:00 am. This strategy ensures that the system is charged during periods of lower electricity rates.

3. The introduction of an export penalty has increased the complexity of the agent's task in balancing multiple penalties, leading to oscillations in the outcomes.

6 Outlook

6.1 Training strategy

In the future, incorporating future timestep electricity price information into the state space could significantly improve the agent's performance. This adjustment would eliminate the need for an export penalty, thus reducing the complexity associated with balancing multiple penalties. Without this adjustment, as previously discussed, the agent struggles to balance cost and SoC effectively during the decision-making process.

Although the epsilon-greedy strategy is straightforward to implement and effective for encouraging exploration, it allocates equal probabilities to all actions during random selection. This uniform randomness can lead to repeated choices of suboptimal or even poor actions, thereby reducing exploration efficiency and wasting valuable time. In contrast, the softmax policy improves exploration by assigning probabilQ-values. actions based on their Q-values.[16] Actions with higher Q-values are prioritised, while other actions are still considered with probabilities proportional to their potential value. This improves exploration efficiency and makes the process more focused by prioritising actions with higher potential value.

During the Q-learning update process, the use of the max operator to select the action with the highest Q-value inherently introduces a bias. Due to the significant volatility in electricity prices, the variance in Q-value estimates can be considerable. This variance leads to an accumulation of bias over iterations, causing divergence in the Q-value estimates and oscillatory behaviour in the results. By introducing a bias-corrected term, these challenges can be effectively addressed, leading to reduced estimation bias, faster convergence rates, and the prevention of divergence during the learning process.[10]

6.2 Alternative methods

Deep Q-Networks (DQN) could offer a powerful alternative to traditional Q-learning by utilising convolutional neural networks (CNNs) to approximate the Q-function.[15] This capability allows DQN to effectively handle high-dimensional state spaces, which are challenging for tabular Q-learning methods. Furthermore, DQN employs an experience replay mechanism, which means the agent could store its interactions with the environment in a replay memory. During training, small batches of samples are randomly chosen from this memory to train the network. This random sampling breaks the correlation between consecutive experiences, thereby enhancing training stability and efficiency.

6.3 Further analysis

Adding carbon intensity information to the state space provides a potential solution for achieving a more balanced trade-off between cost and carbon emissions. Alternatively, redefining the primary objective to prioritise the minimisation of carbon emissions would enable the agent to focus on environmental outcomes. For this multi-objective optimisation, a multi-objective distributed Q-learning approach can be employed.[2] In this approach, multiple agents operate concurrently, with each agent dedicated to optimising a specific objective. These agents explore and learn in parallel, which not only reduces the time required for convergence but also enhances the quality of the solutions.

Acknowledgments We would like to express our sincere gratitude to Dr. Max Bird for his generous support and encouragement throughout this project.

References

[1] Nawazish Ali, Abdul Wahid, Rachael Shaw, and Karl Mason. A reinforcement learning approach to dairy farm battery management using q learning. *Journal of Energy Storage*, 93:112031, 2024.

[2] Abbas Ahmadi Hossein Hejazian Ammar Jalalimanesh, Hamidreza Shahabi Haghighi and Madjid Soltani. Multi-objective optimization of radiotherapy: distributed q-learning and agent-based simulation. *Journal of Experimental & Theoretical Artificial Intelligence*, 29(5):1071–1086, 2017.

[3] Krishna Baberwal, Anshul Kumar Yadav, Vikash Kumar Saini, Ravita Lamba, and Rajesh Kumar. Data driven energy management of residential pv-battery system using q-learning. In *2023 IEEE International Conference on Recent Advances in Systems Science and Engineering (RASSE)*, pages 1–6, 2023.

[4] Heider Berlink, Nelson Kagan, and Anna Helena Reali Costa. Intelligent decision-making for smart home energy management. *Journal of Intelligent & Robotic Systems*, 80:331–354, 2015.

[5] Gregor P Henze and Robert H Dodier. Adaptive optimal control of a grid-independent photovoltaic system. *J. Sol. Energy Eng.*, 125(1):34–42, 2003.

[6] Imperial College London. Imperial partnered with Sainsbury's for sustainability research. `https://www.imperial.ac.uk/news/206333/imperial-partnered-with-sainsburys-\ \sustainability-research/`, 2024. Accessed: 13/10/2024.

[7] International Energy Agency (IEA). Net Zero by 2050, 2021. Licence: CC BY 4.0.

[8] International Energy Agency (IEA). World Energy Outlook 2024, 2024. Licence: CC BY 4.0 (report); CC BY NC SA 4.0 (Annex A).

[9] Bingnan Jiang and Yunsi Fei. Smart home in smart microgrid: A cost-effective energy ecosystem with intelligent hierarchical agents. *IEEE Transactions on Smart Grid*, 6(1):3–13, 2014.

[10] D. Lee and W. Powell. An intelligent battery controller using bias-corrected q-learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 26(1):316–322, 2021.

[11] Salvador Acha Nilay Shah Max Bird, Reewa Andraos. Lifetime financial analysis of an mpc retrofit for integrated pv-battery systems in commercial buildings. *Energy and Buildings*, 2024.

[12] Md Masud Rana, Moslem Uddin, Md Rasel Sarkar, GM Shafiullah, Huadong Mo, and Mohamed Atef. A review on hybrid photovoltaic–battery energy storage system: Current status, challenges, and future directions. *Journal of Energy Storage*, 51:104597, 2022.

[13] Sainsbury's. Reduce Carbon Emissions. `https://about.sainsburys.co.uk/sustainability/better-for-the-planet/carbon`, 2024. Accessed: 30/09/2024.

[14] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. Adaptive Computation and Machine Learning Series. The MIT Press, Cambridge, MA, second edition, 2018. Includes bibliographical references and index. Licensed under Creative Commons Attribution-NonCommercial-NoDerivs 2.0 Generic License.

[15] Fuxiao Tan, Pengfei Yan, and Xinping Guan. Deep reinforcement learning: From q-learning to deep q-learning. In Derong Liu, Shengli Xie, Yuanqing Li, Dongbin Zhao, and El-Sayed M. El-Alfy, editors, *Neural Information Processing*, pages 475–483, Cham, 2017. Springer International Publishing.

[16] Arryon D. Tijsma, Madalina M. Drugan, and Marco A. Wiering. Comparing exploration strategies for q-learning in random stochastic mazes. In *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1–8, 2016.

[17] José R Vázquez-Canteli and Zoltán Nagy. Reinforcement learning for demand response: A review of algorithms and modeling techniques. *Applied energy*, 235:1072–1089, 2019.

[18] C.J.C.H. Watkins and P. Dayan. Q-learning. *Machine Learning*, 8(3-4):279–292, 1992.

ASPEN Integrated 1-Dimensional Non-Equilibrium Plasma Model for CO_2 Dissociation

Thomas Charlton and Bide Chen
Department of Chemical Engineering, Imperial College London, U.K.

Abstract A novel, one-dimensional, non-thermal plasma reactor model for CO_2 dissociation, developed in MATLAB and integrated into ASPEN Plus is proposed. The model simulates a dielectric barrier discharge (DBD) reactor, incorporating key dissociation mechanisms, including vibrational and electronic dissociation, as well as a radical chemistry set. Validation against experimental data showed good agreement at low conversion in the variable space: inlet CO_2 mol fraction [0.2% - 1.3%] and specific energy input [1.5 - 6.5 J cm^{-3}]. The model was less effective in resolving the conversion suppression effects at high plasma power input. We optimised a selection of reactor model parameters, semi-empirically aligning our model to experimental results, and contextualised our model against literature reported conversion values at specified experimental conditions. This approach offers a tractable model that captures the complex plasma-kinetic interactions, while enabling efficient ASPEN based process design to scale CO_2 plasma dissociation.

Keywords
Dielectric-Barrier-Discharge, Non-thermal, Plasma, Argon/CO_2 mixture, CO_2 Utilisation, CO_2 dissociation, 1-D model.

1. Introduction

Increasing atmospheric carbon levels have spurred recent advancements in CO_2 utilisation technologies aimed at mitigating climate change.[1-7] Among these, plasma dissociation emerges as a promising solution, offering dual benefits: the production of low-carbon, value-added chemicals and enabling renewable energy storage.[2,8]

The primary dissociation product of CO_2 is carbon monoxide, a core component of syngas. By integrating CO_2 dissociation with renewable hydrogen production, a sustainable pathway for syngas generation is established.[2,3,8] Syngas, a precursor material ubiquitous with chemical and energy industries, saw a global consumption of 31 exajoules in 2023. By supplanting syngas with a sustainable alternative, plasma-driven CO_2 dissociation facilitates a circular, decarbonised approach to meet industrial demands.

The decarbonisation imperative necessitates a shift away from on-demand fossil fuels to inherently volatile renewable energy sources.[9] Electrically driven plasma processes, with their favourable on/off operability, enable conversion of surplus renewable power into energy-dense gas (power-to-gas). [3,6,10] This capability is a vital feature for stabilising a decarbonised energy grid. [4,7]

Advances in CCU infrastructure, such as the Northern Lights project[11,12], underscore the realisation of CO_2 not as a scourge, but as a valuable resource - with a technical capture potential of 4.7-37.5 $GTCO_2yr^{-1}$ by 2050[13].

Plasma dissociation is poised to capitalise on this opportunity. A notable example of the extensive industrial application of plasma dissociation is ozone generation in a dielectric barrier discharge reactor[14]. The same commercial success could potentially be replicated in converting CO_2 into CO. By efficiently breaking the C=O bond non-thermal plasma offers significant energy savings compared to alternate processes.[8] In contrast, thermal decomposition demands temperatures exceeding 1800 °C (often using heat derived from fossil fuels) while photocatalytic processes struggle with low conversion as activating the linear CO_2 molecules requires large overpotentials.[2,15] This work concerns DBD plasma, favouring its simplicity, room temperature and atmospheric pressure operation.

DBD reactors feature a dielectric material separating two electrodes,[16] limiting current density and preventing arcing, which protects the electrodes and maintains efficiency. When a potential is applied across them, numerous (O(6)/s) micro-discharges form which coalesce into spatially localised filamentary channels.[5] DBD reactor efficiency ranges from 2-10%,[2] influenced by reactor geometry, SEI, inlet gas composition and the presence of packing materials or catalysts.[8]

Researchers modelling DBD plasma often use global reaction kinetics for simplicity, treating the system as spatially homogenous.[16,17] While ensuring computational simplicity and tractability, they neglect key plasma-kinetic details.[18] Recognising these limits, some researchers transitioned to spatially resolved models, accounting for the non-uniform distribution of plasma parameters like electron density, ionization degree and species concentrations.[19] Modern models incorporate thermal, electromagnetic and complex electro-hydrodynamic fluid behaviours to better represent the multi-physics of plasma, though computationally intensive and infeasible to solve quickly.[20]

To develop a tractable model suitable for process design, we present a novel approach that integrates a spatially 1-dimensional non-thermal DBD model into ASPEN PLUS. A MATLAB-based ODE solver captures the non-equilibrium plasma kinetics, with results fed into an ASPEN block representing the plasma reactor. This framework enables seamless integration of non-thermal plasma kinetics into broader process design workflows.

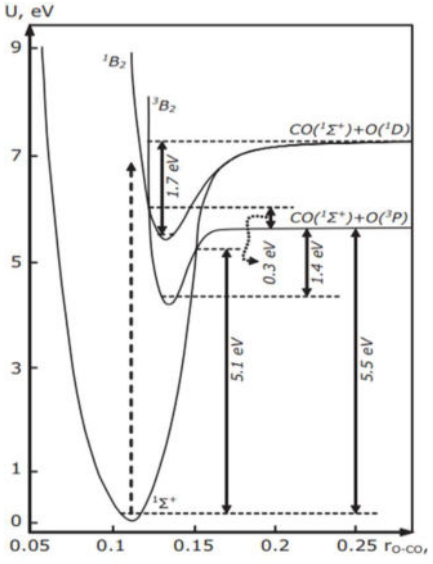


Figure 1. Diagram representing the energy levels associated with the transitions between various CO_2 dissociation states, adapted from Fridman Plasma Chemistry). [21] [$r_{O\text{-}CO}$ nm]

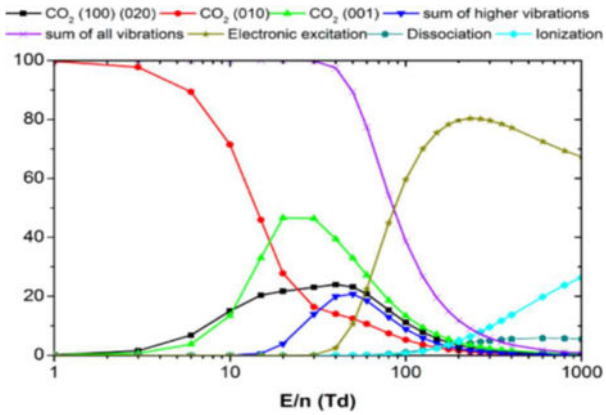


Figure 2. Diagram representing the relative abundance of CO_2 excitation states against reduced electric field strength (Aerts et al.). [14]

2. Plasma-Kinetic Theory

Electrons receive energy from the applied electric field and retain it following collisions, due to their lower mass.[22] This leads to electron temperatures exceeding 10,000K, while molecular translational energies remain near room temperature – the defining characteristic of non-thermal plasma (NTP).[7] Excited electrons asymmetrically distribute their kinetic energy across the system's available degrees of freedom, enabling the activation of CO_2 molecules at ambient temperatures.[10] Vibrational excitation is the most efficient pathway for CO_2 dissociation in NTP ~ 100 Td (Fig. 2).[17] At these conditions, up to 97% of the discharge power is channelled into the vibrational excitation.[23,24] Lower CO_2 vibrational states - populated by electron impact - exchange energy in vibrational-vibrational (VV) collisions that transfer energy to higher levels.[4] Dissociation occurs once the vibrational state energy exceeds the 5.5 eV threshold (Eq. 1).[23]

CO_2 exemplifies this mechanism, with its asymmetric vibrational mode benefiting from a low VT relaxation rate and a much higher VV exchange rate. The presence of CO enhances this process, maximising the population of the efficient $^3B_2$ transition state (Fig. 1).

Electron temperatures of ~1-2 eV are far below the 7-eV threshold required for electronic excitation in CO_2. This explains why the rate of vibrational excitation exceeds the rate of electronic excitation by an order of magnitude. Suppressing wasteful (~8eV) electronic excitation mechanisms (Eq. 2) improves efficiency.[5,17,24]

$$\boldsymbol{CO_2}^*(^1\boldsymbol{\Sigma}^+) \rightarrow \boldsymbol{CO_2}^*(^3\boldsymbol{B_2}) \rightarrow \boldsymbol{CO}(^1\boldsymbol{\Sigma}^+) + \boldsymbol{O}(^3\boldsymbol{P}) \quad -\ (\text{Eq. } 1)$$

$$\boldsymbol{e} + \boldsymbol{CO_2}(^1\boldsymbol{\Sigma}^+) \rightarrow \boldsymbol{CO_2}^*(^1\boldsymbol{B_2}) \rightarrow \boldsymbol{CO}(\boldsymbol{a}^3\boldsymbol{\pi}) + \boldsymbol{O}(^3\boldsymbol{P}) \quad (\text{Eq. } 2)$$

Furthermore, the Fridman alpha model[21] indicates vibrational energy is ideal for endothermic reactions, such as bond breaking. To prevent heat loss and radical recombination, ambient temperatures must be suppressed, as vibrational-translational (VT) relaxation – the reverse process to VV - is gas temperature-dependent. To optimise performance, controlling parameters such as electron temperature, ionisation degree and specific energy input (SEI) is essential.[15]

3. Model and Experimental Methodology

3.1 Plasma Model Assumption

Past studies have used mass and electronic energy balances to develop their plasma-kinetic models, exploring various reactor configurations. Some researchers homogenise the system, simulating the reactor as a CSTR[16,25,26], while others incorporate reaction-diffusion in the plasma space, leading to 2-D coaxial DBD models[27,28]. Based on the experimental plasma generator we use for verification, this study makes key assumptions in deriving the kinetic model. Specifically, the plasma reactor is modelled as an ideal plug flow reactor due to its operation in a turbulent regime (Re ~ 3900) where gas and electron dispersion are negligible because of short residence time in the micro-discharge volume.

3.2 Species and Reactions Considered

The reactant mix consists solely of carbon dioxide and argon, limiting the number of intermediate and product species. A total of 17 species (Table 1) and 44 reactions were identified, based on elementary reactions, such as electron impact, neutral-neutral, ion-neutral and ion-ion reactions, as summarized in the NIST[29] and UMIST databases[30].

Table 1. List of species considered in the model kinetics

Molecules	$\mathbf{CO, CO_2, O_2, Ar}$
Radicals	C, C_2, O
Positive species	$C^+, C_2^+, CO^+ CO_2^+, O^+, O_2^+, Ar^+$
Negative species	O^-, O_2^-, e

3.3 Plasma Material Balance

In the 1-dimensional model, the concentrations of electrons, gas species and reaction rates vary with distance (Fig. 3). The material balance for each species i is expressed as:

$$\frac{dX_i}{dV} = \Sigma_{j=1}^{J} v_{i,j} \times q_j \quad -\ (\text{Eq. } 3)$$

where q_j represents the rate of reaction, defined as:

$$q_j = k_{f,j} \Pi_{i=1}^{I} [X_i]^{v'_{i,j}} - k_{b,j} \Pi_{i=1}^{I} [X_i]^{v''_{i,j}} \quad -\ (\text{Eq. } 4)$$

The forward and backward rate constant are calculated using a modified Arrhenius equation[30] to account for high-temperature deviations. In the DBD plasma reactor, non-equilibrium plasma generates a difference between electron and the gas/ion temperature. For reactions involving electron impact, the Arrhenius rate is evaluated as follows:

$$k_j = A_j \times T_e^{\beta_j} \times \exp\left(-\frac{E_j}{RT_e}\right) \quad -\ (Eq. 5)$$

For reactions that do not involve electron impact:

$$k_j = A_j \times T_g^{\beta_j} \times \exp\left(-\frac{E_j}{RT_g}\right) \quad -\ (Eq. 6)$$

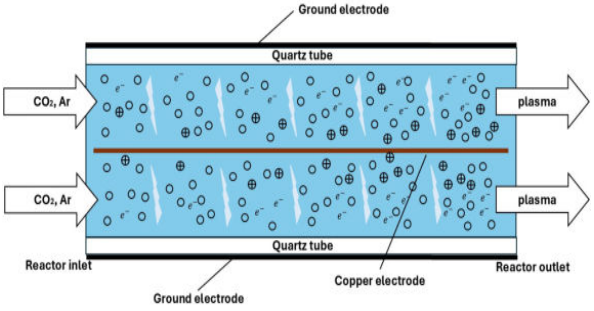


Figure 3. Illustration of 1-D plug-flow model, showing the plasma region, electric field and perpendicular flow direction.

Another significant pathway for CO_2 dissociation is the energy exchange between vibrationally excited carbon

dioxide molecules, which enables them to reach a dissociative state (v_{21}) with energy exceeding the dissociation threshold at 5.5eV. The resonant vibrational excitation rate $k(v_0, v_1)$ is obtained from cross sectional integration performed by the Bolsig+ solver[31] and is a function of electron temperature. The rate of vibrational ladder climbing from lower state (v_A) to higher state (v_B)[21] is semi-empirical, derived by Fridman in Eq. 7.

$$k(v_A, v_B) = k(v_0, v_1) \times \exp\left(-\frac{\alpha(B-A)}{1+\beta \times A}\right) \quad -(\text{Eq. 7})$$

$$\alpha = 0.5, \ \beta = 0 \quad \text{for } CO_2$$

The rate of vibrational CO_2 dissociation is determined by summing all the rates to dissociative level ($v = v_{21}$):

$$q_{vib.} = [CO_2] \sum_{v=0}^{20} k(v_0, v_1) \times \exp\left(-\frac{\alpha(21-v)}{1+\beta \times v}\right) \quad -(\text{Eq. 8})$$

3.4 Plasma Energy Balance

A similar control volume analysis can be applied to the energy balance of both electrons and gas species. Electrons gain kinetic energy by being accelerated in the strong electric field generated between the concentric electrodes. The rate at which energy is supplied to the electrons, denoted as U_{source}, is given by:

$$U_{source} = e \times \epsilon \times n_e \quad -(\text{Eq. 9})$$

Electrons lose energy through collisions. In an elastic collision, electrons transfer momentum to the gas/ion species without reacting[32], meaning kinetic energy is conserved. The frequency of elastic collision is calculated using:

$$v_{e,i} = v_e \times \frac{X_i}{V} \times N_A \times A_i \quad v_e = \sqrt{\frac{3k_B T_e}{M_e}} \quad -(\text{Eq. 10})$$

The loss of electron kinetic energy through elastic collisions is evaluated using the equation provided by Wang et al.[32], assuming that the kinetic energy of the gas atoms in the collision cross-section is uniform.

$$U_{elas} = \sum_{i=i}^{I} \frac{3}{2} \times n_e \times \left(\frac{2M_e}{M_i}\right) \times k_B \times (T_e - T_g) \quad -(\text{Eq. 11})$$

We considered the momentum transfer between electron and neutral gas species (Ar, O_2, CO and CO_2). The effects of electron-ion chemical interactions are accumulated in the inelastic collision term. Enthalpy data was obtained from both the NIST[29] and UMIST databases[30]. The energy loss due to inelastic collisions is given by:

$$U_{inelas,rxn} = \sum_{j=1}^{J} V \times q_j \times \Delta H \quad -(\text{Eq. 12})$$

To account for the vibrational excitation of CO_2 in the energy balance, we included this process to evaluate the energy losses associated with initiating electrons. The rate of excitation ($k_{i,n}$) was generated using the Bolsig+ solver, which employs the Maxwell-Boltzmann distribution of electron energy to estimate the population of vibrationally excited states[31]. The inelastic loss follows:

$$U_{inelas,excite.} = \sum_{i=1}^{I} \sum_{m=1}^{M} k_{i,n} \times E_{excite.} \times X_i \times n_e \quad -(\text{Eq. 13})$$

The overall electron energy balance, with all energy generation and consumption terms follows:

$$\frac{c_e}{N_A} \times \left(n_e \frac{dT_e}{dV} + T_e \frac{dn_e}{dV}\right) = U_{source} - U_{elas} - U_{inelas,rxn} - U_{inelas,excite.} \quad -(\text{Eq. 14})$$

A similar approach was applied to evaluate the temperature of gaseous neutral/ionic species. It was assumed that all heavier gas particles behave as ideal gas and share common T_g, i.e. $T_i = T_g$ for all i.

$$\sum_{i=1}^{I} \left(\frac{dT_i}{dV} \times X_i \times c_i + \frac{dX_i}{dV} \times T_i \times c_i\right) = U_{elas} + U_{inelas,excitation} \quad -(\text{Eq. 15})$$

Note that the excited state gas atoms relaxation (to ground state) leads to increase in gas atom/ion temperature[19].

3.5 Modelling of plasma and afterglow region

The MATLAB model divides the reactor up into plasma and afterglow regions and simulate both as two reactors in series. While plasma region takes in all the proposed balances in section 3.3 and 3.4, afterglow region quenches all electron dependent reactions and energy transfer.

3.6 Sensitivity Analysis – Parameter Fitting

Table 2 List of species considered in the model kinetics

Parameter	Optimal value	Unit
Electron number density	7.743×10^{19}	[/cm^3]
Initial electron temperature	11000	K
Reactor length	10	mm

The initial electron density dominates the modelled conversion. This became apparent in a sensitivity analysis exercise conducted on ASPEN over the range 10^{14} to 10^{22} electrons/cm^3. Beyond a density of 10^{19}/cm^3, a rapid gas temperature rise is predicted – in the explosive VT relaxation region – at this stage, the heightened electron density leads to a higher frequency of elastic collisions, ultimately equilibrating electron temperatures with gas temperatures. The test reactor operates at 315 K[33], leading us to an electron density of 9.73×10^{18}/cm^3 effectively suppressing gas temperatures, while providing the best match to observed CO_2 conversion.

The model predicts typical plug flow conversion behaviour, with respect to the reactor length. The model is entirely insensitive to initial electron temperature, reaching an equilibrium $T_e \sim 11{,}000$K within a few µm of entering the reactor. This is intuitive, high-energy electrons contribute to non-dissociative energy inefficient inelastic collisions at a very high rate, producing high energy products and ionisation events. This is confirmed by analysis of relative reaction rates.



Figure 4. An image of the CP-121 DBD Cold Plasma Demonstrator, CO_2/Ar 1.26% v/v.

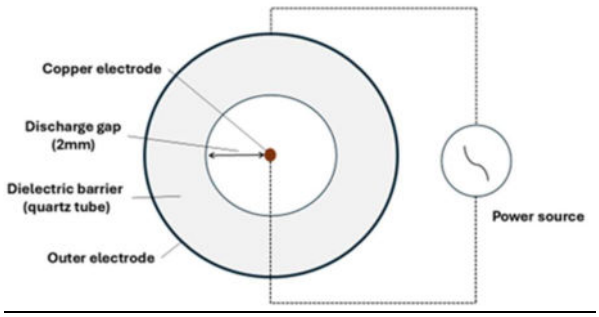


Figure 5 Bird-eye cross-sectional view of the DBD reactor

4. Experimental Verification Set-Up & Methodology

CO_2 conversion, specific energy input (SEI) and efficiency are the metrics commonly cited in literature to assess the CO_2 dissociation plasma reactor performance. Experimental validation based on these two metrics is performed with a DBD cold plasma generator to consolidate the MATLAB model.

$$\text{SEI} = \frac{\text{Power}}{\text{Flowrate}} \left[\text{J cm}^{-3}{}_{\text{plas. gas}}\right] \quad -(\text{Eq. 16})$$

$$\text{Efficiency}, \eta = \frac{\Delta H_{CO2 \rightarrow CO}}{E_{CO}} = \frac{\chi_{CO_2}(\%) * \Delta H_r}{\text{SEI}} \quad -(\text{Eq. 17})$$

Illustration of experiment set-up is shown in figure 6. The reactor is in coaxial configuration, with the two electrodes separated by a dielectric quartz tube. The radial discharge gap between the copper electrode and quartz tube is 2mm (Fig. 5). 9.2 sL/min of flow is required to meet the system threshold to trigger electrical discharge. Argon is used as a supplementary gas, to meet the flow requirement and to serve as inert carrier gas; Argon also sets the basis for the mass spectrometer measurement (Model 200D, ESS), enabling quantification of the product gas composition.

The total gas flowrate is maintained at 9.2 sL/min. CO_2 flowrate, varying from 0.023 to 0.12 sL/min, is mixed with argon before passing through the coaxial DBD plasma. CO_2 flow is controlled by a mass flow controller (MFC) (EL-FLOW Select, Bronkhorst) with accuracy of ± 0.2 mL/min.

A calibration curve was generated which maps the MFC set-point flowrate to the actual flowrate. In each experimental run, CO_2 flowrate is set first. The argon flowrate is manually adjusted to achieve total flowrate of 9.2 sL/min using needle valve controlled by a steel knob on the Bronkhorst mass view meter. The reactor contains four power input setting (0, 15, 40, 60W) which is manually set using an inbuilt rotary switch. The plasma reactor runs at ambient temperature (298K) and 1.7 bar pressure which is monitored using on-line pressure sensor.

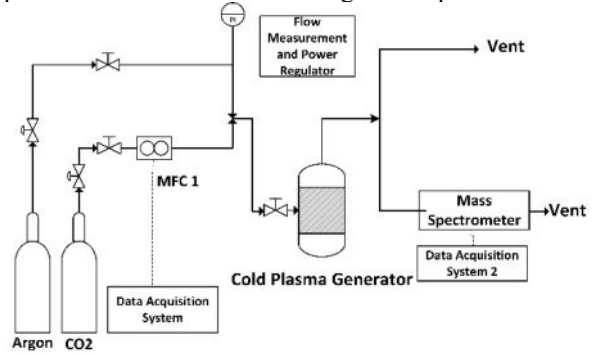


Figure 6 PFD of experimental set-up

5. Results

5.1 Conversion against mol fraction

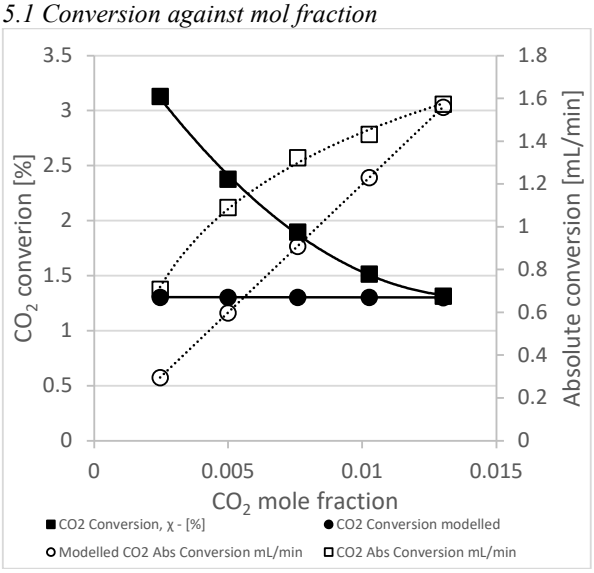


Figure 7. Plots of conversion – percentage and absolute - against inlet CO_2 mol fraction, contrasting model performance (circle markers) against experimental data (square markers). Dashed lines trends the absolute conversion data (white markers) [ml/min]. Solid lines trends the percentage conversion data (black markers) [%].

With rising CO_2 inlet mol fraction, percentage conversion decreases, falling from 3.1% to 1.3% experimentally. The absolute experimental conversion trend is inverse, increasing non-linearly from 0.23ml/min to over 1.6 ml/min. The modelled conversion is constant, at 1.31%, corresponding to a linear increase in absolute conversion with inlet mol fraction. (Fig. 7)

5.2 Conversion and Efficiency against SEI

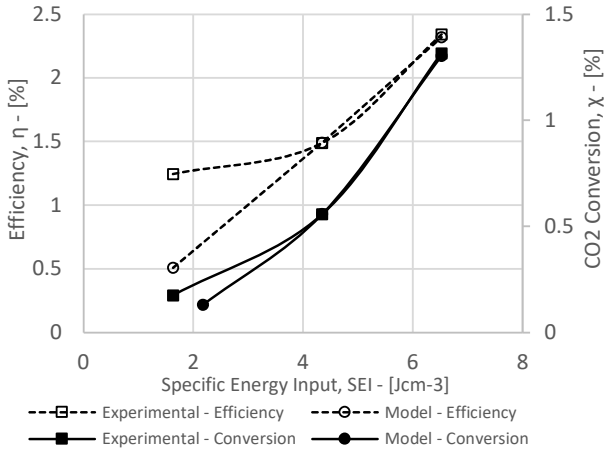


Figure 8. Plots of conversion and efficiency against increasing specific energy input, contrasting model performance (circle markers) against experimental data (square markers). Dashed lines are polynomial fits to the efficiency data (white markers) [%]. Solid lines are polynomial fits to the conversion data (black markers) [%].

Conversion improves significantly, rising from 0.175% to 1.31% - a seven-fold increase – as specific energy input rises from around 1.6 Jcm^{-3} to 6.5 Jcm^{-3}. A similar trend in efficiency is observed, rising from 0.5% (modelled) and 1.3% (experimentally observed) to 2.3%. The modelled outputs present a nearly linear increase trend.

5.3 Modelled Conversion – sensitivity analysis

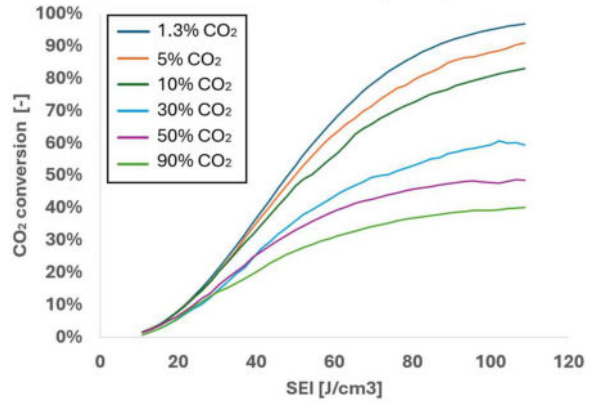


Figure 9. Plots of conversion against increasing specific energy input, determined via model sensitivity analysis, varying the initial CO_2 mol fraction. An S-curve have been traced at all CO_2 fraction, with a high initial increase and levelling at high SEI. Different conversion plateau has been reached, the highest (100% conversion) at 1.3% CO_2 inlet fraction, and the lowest (40% conversion) at 90% CO_2 inlet fraction.

5.4 Modelled Efficiency – sensitivity analysis

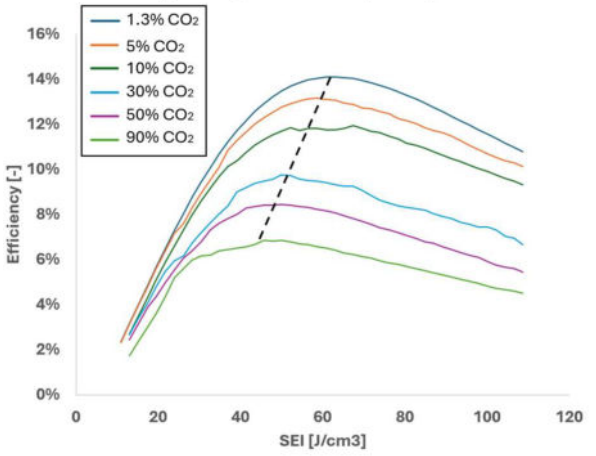


Figure 10. Plots of efficiency against increasing specific energy input, determined via model sensitivity analysis, varying the initial CO_2 mol fraction. The dashed black line crests the locus of stationary points, delineating the maximum efficiencies at each mol fraction. Decrease in critical SEI (obtained at peak efficiency) with increases CO_2 fraction is effected by the increasing collision cross section detailed in section 6.3

6. Discussion

6.1. Limiting reaction kinetics

There is a fundamental difference between modelled and observed species' behaviour. The model assumes plug flow for both electrons and other species, which seems inherently non-physical. Given the homogenous electron density across the reactor length, one might expect electron behaviour to conform to a CSTR configuration. However, as shown in figure 7, the model's use of PFR for electrons is valid, as the CO_2 dissociation process is electron-independent and thus follow pseudo-first-order kinetic equations.

The model originally assumes a second order electron-dependent impact reaction, yet it is reduced to a pseudo-first-order kinetics due to the large excess of electrons. This is reflected in the constant CO_2 conversion in figure 7 irrespective of CO_2 inlet fraction. Therefore, the extent of conversion is solely governed by the residence time, represented by the solution to first order plug flow design equation:

$$\chi = 1 - e^{-k\tau};\ \tau = \frac{V}{v_0} \qquad -(\text{Eq. } 18)$$

Despite the model showing a constant electron density, the experimental results display an overall trend resembling a 0.5th-order kinetics where conversion decreases with rising CO_2 inlet fraction. The trend can be numerically described by the equation

$$\chi = 1 - \left(1 - \frac{k\tau}{2[CO_2]_0^{0.5}}\right)^2 \qquad -(\text{Eq. } 19)$$

The inaccuracy of the rate constant is distinct evidence for the model overlooking some non-trivial reactions, provided by other literature sources. Identification of the rate determining step of the physical reaction and tuning the corresponding rate constant will allow the model to resolve the CO_2 dissociation mechanism, matching reality.

6.2 Effective reactor volume in relation to power input

The non-linear increase of conversion against SEI in both the modelled and experimental results is caused by the expansion of the micro-discharge region in the reactor due to power increase. The number of micro-discharges in the discharge gap is proportional to the power input, ensured by the dielectric barrier distributing charge across its surface. At higher power input, more uniform discharges are established, leading to an increase in the effective reactor volume. This results in longer residence times for molecules in the plasma region, enhancing the CO_2 conversion. This phenomenon can also be quantified with 0.5th-order reaction expression, where conversion is a function of residence time (Eq. 19).

The conversion also directly implies the evolution of efficiency over increasing SEI. If the effective reactor volume is greater, the input energy can be more efficiently delivered to the molecules resulting in the trend displayed in figure 8 of increasing efficiency with conversion.

6.3 Argon contribution to energy efficiency

The model predicts an S-curve for conversion as SEI increases beyond the experimental range (Fig. 9). At high SEI, rapid reactant depletion slows the CO_2-dependent reaction rate, stabilising at a plateau. The final conversion value varies with CO_2 inlet fractions (Fig. 9), indicating the significant role of the Argon carrier gas in CO_2 electron impact reactions. Argon's smaller inelastic collision cross section compared to CO_2, increases the total system inelastic collision frequency as its fraction decreases (e.g., from 1.3% to 90% v/v CO_2) leading to greater power losses and a downward shift in the electron energy distribution. This reduces the population of CO_2 molecules dissociating via electronic and vibrational pathways, lowering both peak energy efficiency and the SEI required to achieve it, as shown by the dashed locus line in figure 10.

6.4 Reaction mechanism efficiency implications

As power input increases, electron gains more energy, enabling population of higher excitation levels that are unattainable at lower electron energies. These accessed levels predominantly react via electronic excitation pathways, much less efficient for CO_2 dissociation compared to lower SEI vibrational dissociation. Therefore, increasing frequency of electronic-excitation-

driven dissociation eventually outweighs the effect of expanding effective reactor volume.

The critical SEI, corresponding to peak efficiency, is a locus of peak efficiency points, shown in figure 10. It can also be obtained at the inflection point of the conversion versus SEI curve shown in Figure 9.

6.5 Literature value comparison with model prediction

To further verify the validity of our model as different operating conditions, a list of Ar-CO_2 DBD conditions data are extracted from different literature and compared against the model-predicted data running under the same condition. The result is shown in the comparison chart (Fig. 11).

CO_2 conversion is selected as the metric for testing the robustness of the model as it can be readily converted into other common metrics, such as energy efficiency.

In the comparison chart, the model and experimental conversions show a variance in alignment. The percentage error varies greatly, with good agreement at low conversion (<5%) but misalignment at high conversion (>5%). It is also shown that most of the model results tend to overestimate the conversion. Grouping the data points by their common controlled variables, a trend has been revealed – the error gap becomes larger at higher operating input power. Across data point 6 to 9, literature reports the conversion levels off at around 10% conversion while model suggests a plateau at near 100% conversion. This indicates plateau suppression mechanisms on increasing CO_2 collision cross-section (as explained in section 5.3) is insufficient in the literature conditions. More electron energy needs to be absorbed in facilitating energy inefficient CO_2 electronic excitation reactions to achieve a lower conversion maximum.

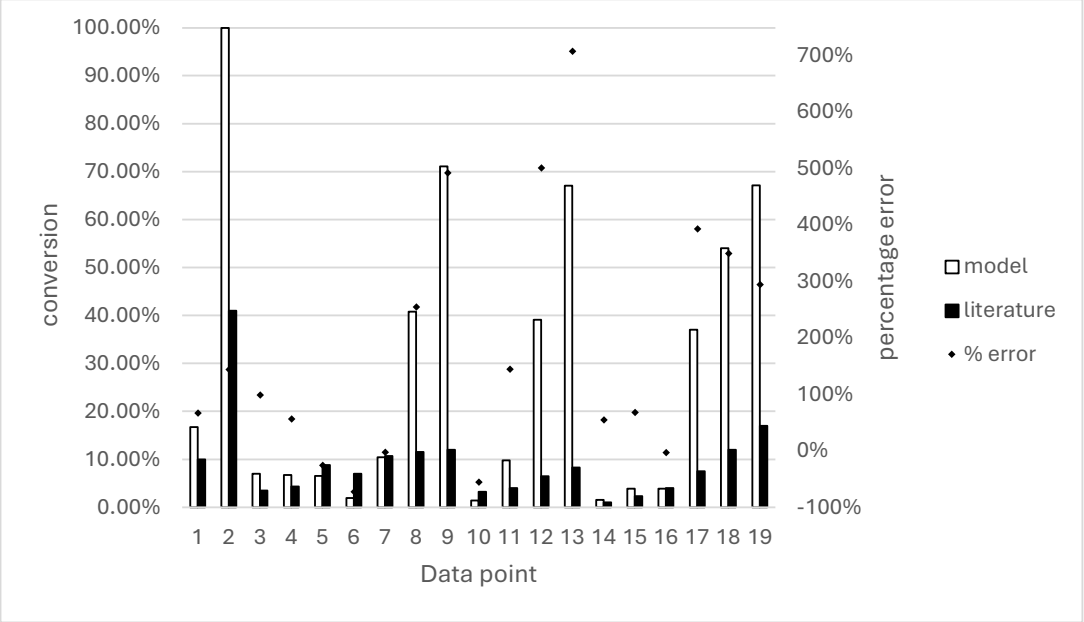


Figure 11. Comparison of reported CO_2 conversion and model simulated values, varied condition such as power, CO_2 fraction, pressure, total gas flowrate

Table 3 Literature reported experimental conditions

Literature value	Input Power (W)	Discharge gap width [mm]	Effective reactor length [mm]	CO_2 fraction [v/v]	Pressure [bar]	Total gas flowrate [L/min]	Reference
1	83	5	50	5%	1	/	[34]
2	80	1.83	200	5%	1	0.3	[35]
3	10	2	200	20%	1	0.4	[36]
4	10	2	200	20%	1.5	0.4	[36]
5	10	2	200	20%	2	0.4	[36]
6	5	2	150	5%	1	0.4	[37]
7	10	2	150	5%	1	0.4	[37]
8	20	2	150	5%	1	0.4	[37]
9	30	2	150	5%	1	0.4	[37]
10	5	3.65	150	10%	1	0.4	[37]
11	10	3.65	150	10%	1	0.4	[37]
12	20	3.65	150	10%	1	0.4	[37]
13	30	3.65	150	10%	1	0.4	[37]
14	50	2	100	100%	1	3	[38]
15	50	2	100	100%	1	2	[38]
16	50	2	100	100%	1	1	[38]
17	50	2	100	100%	1	0.5	[38]
18	50	2	100	100%	1	0.3	[38]
19	50	2	100	100%	1	0.2	[38]

The model also demonstrates an overestimation of the effect of changing flowrate on the conversion. From point 14 to 19, a decrease in total gas flowrate shows an improvement in conversion. While SEI assumes decrease in gas flow has the same effect as increasing power input, Aert stated that given the same SEI change, varying gas flowrate has a more pronounced effect in conversion than varying power input (Aert *et al.*, 2015)[11]. Therefore, even though our model has been optimized with a correlation to respond to SEI change due to power input change, this correlation is unsubstantiated in predicting the change in total gas flowrate. New relation is to be established, via parameters controlling the residence time, to make the model correctly react to total flowrate variation.

7. Conclusions & Outlook

In the development of the 1-D plasma reactor model and subsequent integration into ASPEN, we incorporated elementary plasma-kinetic processes with a radical chemistry set to fully describe the dielectric barrier discharge reactor system. The reactor consists of a plasma and an afterglow region, capturing the shift in dominant reaction mechanisms across these phases.

To validate the model, experiments were conducted using CO_2 inlet fractions between 0.2% and 1.3% [v/v] and specific energy inputs (SEI) ranging from 1.5 to 6.5 Jcm^{-3}. The subsequent model evaluation against experimental results has yielded several key insights.

Ideal plug flow behaviour is an underpinning model assumption, which requires artificial electron density inflation to normalise to experimental results, has the consequence of reaction progression being invariable to electron number. This model behaviour conforms to the uniform electron density of the physical DBD system. Discrepancies in reaction order exhibited by the model suggests that the model neglects certain non-trivial rate equations.

At low SEI, the model aligns well with experimental data, effectively capturing the influence of power input on CO_2 conversion. The U_{source} term of the electronic energy balance equation is the pivotal interface between our model and the physical system. Here, we flatten the complex spatially localised, filamentary discharge maelstrom into an average power input term. While this assumption was effective in describing the energy available to the system, critical information is lost regarding the distribution of this energy. Therefore, we theorise that embracing the neglected phenomenon by developing an expression linking power to micro-discharge formation, and plasma volume could ultimately overcome these limitations.

Despite its shortcomings, our model exposes the sensitivity of plasma dissociation systems to slight variations in reactant composition. Introducing CO_2 significantly increases the system's net inelastic collision cross section. While these collisions drive dissociation, they also serve as major energy sinks, inviting inefficiency at high SEI. Balancing these antagonistic effects is vital to make DBD reactors both efficient and economically viable for industrial application. Our model successfully predicts the resulting decline in efficiency with increasing CO_2 fraction, aligning with literature sources.[11] This finding presents an opportunity for optimization, which we propose must be explored further using the ASPEN platform.

We recommend investigating alternative plasma sources to better understand the shared, underlying dissociation mechanisms. Such an investigation would enable a comparison of DBD systems - integral to this work - with disparate microwave (MW) plasma regimes which operate under low-pressure, low-energy conditions. Comparing the performance of DBD and MW systems under identical conditions, we could infer the relative contribution of each dissociation mechanism and uncover the non-trivial rate behaviour - strengthening our model across a wide range of operating regimes.

By implementing the changes proposed, we anticipate a close alignment of our model output to the experimental results, enhancing its accuracy and applicability. Ultimately, our ASPEN user-model should demonstrate that non-thermal plasma systems can reject the pall of economic uncertainty which hangs over DBD.[24]

Acknowledgements

We would like to extent our gratitude to research associate Lu Ai and PhD candidate Ines Perez Tabarnero.

References

1. Bongaarts, J. (2024). IPCC, 2023: Climate Change 2023: Synthesis Report. IPCC, 184 p., doi: https://doi.org/10.59327/IPCC/AR6-9789291691647. *Population and development review*. doi:https://doi.org/10.1111/padr.12632.
2. Ong, M.Y., Nomanbhay, S., Kusumo, F. and Show, P.L. (2022). Application of microwave plasma technology to convert carbon dioxide (CO2) into high value products: A review. *Journal of Cleaner Production*, 336, p.130447. doi:https://doi.org/10.1016/j.jclepro.2022.130447.
3. Li, S., Ongis, M., Manzolini, G. and Gallucci, F. (2021). Non-thermal plasma-assisted capture and conversion of CO2. *Chemical Engineering Journal*, 410, p.128335. doi:https://doi.org/10.1016/j.cej.2020.128335.
4. Bogaerts, A. and Centi, G. (2020). Plasma Technology for CO2 Conversion: A Personal Perspective on Prospects and Gaps. *Frontiers in Energy Research*, 8. doi:https://doi.org/10.3389/fenrg.2020.00111.
5. Lindon, M.A. and Scime, E.E. (2014). CO_2dissociation using the Versatile atmospheric dielectric barrier discharge experiment (VADER). *Frontiers in Physics*, 2. doi:https://doi.org/10.3389/fphy.2014.00055.
6. Marcantonio, V., De Falco, M. and Bocci, E. (2022). Non-Thermal Plasma Technology for CO2 Conversion—An Overview of the Most Relevant Experimental Results and Kinetic Models. *Energies*, 15(20), p.7790. doi:https://doi.org/10.3390/en15207790.
7. Lebouvier, A., Iwarere, S.A., d'Argenlieu, P., Ramjugernath, D. and Fulcheri, L. (2013). Assessment of Carbon Dioxide Dissociation as a New Route for Syngas Production: A Comparative Review and Potential of Plasma-Based Technologies. *Energy & Fuels*, 27(5), pp.2712–2722. doi:https://doi.org/10.1021/ef301991d.

8. Wang, B., Li, X., Wang, X. and Zhang, B. (2021). Effect of filling materials on CO2 conversion with a dielectric barrier discharge reactor. *Journal of Environmental Chemical Engineering*, 9(6), p.106370. doi:https://doi.org/10.1016/j.jece.2021.106370.
9. United Nations / Framework Convention on Climate Change (2015) Adoption of the Paris Agreement, 21st Conference of the Parties, Paris: United Nations" Aerts, R., Martens, T. and Bogaerts, A. (2012). Influence of Vibrational States on CO 2 Splitting by Dielectric Barrier Discharges. doi:https://doi.org/10.1021/jp307525t.
10. Bongers, W., Bouwmeester, H., Wolf, B., Peeters, F., Welzel, S., van den Bekerom, D., den Harder, N., Goede, A., Graswinckel, M., Groen, P.W., Kopecki, J., Leins, M., van Rooij, G., Schulz, A., Walker, M. and van de Sanden, R. (2016). Plasma-driven dissociation of CO_2 for fuel synthesis. *Plasma Processes and Polymers*, 14(6), p.1600126. doi:https://doi.org/10.1002/ppap.201600126.
11. Gentile, V., Cauchois, G., Ålund, I. and Renzi, N. (2023). Carbon footprint of the Northern Lights JV CO 2 transport and storage value chain.
12. Becattini, V., Luca Riboldi, Burger, J., Nöhl, J., Oeuvray, P., Reyes-Lúa, A., Anantharaman, R., André Bardow, Frattini, L., Fu, C., Mazzotti, M., Roussanaly, S. and Zotică, C. (2024). Rolling-out pioneering carbon dioxide capture and transport chains from inland European industrial facilities: A techno-economic, environmental, and regulatory evaluation. Renewable and Sustainable Energy Reviews, [online] 205, pp.114803–114803. doi:https://doi.org/10.1016/j.rser.2024.114803.
13. Metz, B., Davidson, O., H.C. de Coninck, Loos, M. and Meyer, L. (2005). IPCC special report on carbon dioxide capture and storage. New York: Cambridge University Press.
14. Kogelschatz, U. (2003). Dielectric-barrier Discharges: Their History, Discharge Physics, and Industrial Applications. 23(1).
15. Aerts, R., Somers, W. and Bogaerts, A. (2015). Carbon Dioxide Splitting in a Dielectric Barrier Discharge Plasma: A Combined Experimental and Computational Study. *ChemSusChem*, 8(4), pp.702–716. doi:https://doi.org/10.1002/cssc.201402818.
16. Zheng, H. and Liu, Q. (2014). Kinetic Study of Nonequilibrium Plasma-Assisted Methane Steam Reforming. *Mathematical Problems in Engineering*, 2014, pp.1–10. doi:https://doi.org/10.1155/2014/938618.
17. Aerts, R., Martens, T. and Bogaerts, A. (2012). Influence of Vibrational States on CO 2 Splitting by Dielectric Barrier Discharges. doi:https://doi.org/10.1021/jp307525t.
18. Ponduri, S. (2016). *Fluid modelling of CO 2 dissociation in a dielectric barrier discharge*.
19. Tehrani, D.S., Abdizadeh, G.R. and Noori, S. (2022). Numerical modeling of dielectric barrier discharge actuators based on the properties of low-frequency plasmons. *Scientific Reports*, [online] 12(1). doi:https://doi.org/10.1038/s41598-022-14370-z.
20. Obraztsov, N.V., Subbotin, D.I., Popov, V.E., Frolov, V.Y. and Surov, A.V. (2018). Modelling of heating of plasma-chemical reactor in Comsol Multiphysics. *Journal of Physics Conference Series*, 1038, pp.012137–012137. doi:https://doi.org/10.1088/1742-6596/1038/1/012137.
21. Fridman, A. (2008). *Plasma Chemistry*. Cambridge University Press.
22. Snoeckx, R. and Bogaerts, A. (2017). Plasma technology – a novel solution for CO2 conversion? *Chemical Society Reviews*, [online] 46(19), pp.5805–5863. doi:https://doi.org/10.1039/C6CS00066E.
23. Mayr, M., Hasan, S., Nabi, S. and Jordan (1981). *Soviet Physics Uspekhi The physics of a chemically active plasma with nonequilibrium vibrational excitation of molecules*
24. Fridman, A.A. and Rusanov, V.D. (1994). Theoretical basis of non-equilibrium near atmospheric pressure plasma chemistry. *Pure and Applied Chemistry*, 66(6), pp.1267–1278. doi:https://doi.org/10.1351/pac199466061267.
25. Bogaerts, A., Berthelot, A., Heijkers, S., Kolev, S., Snoeckx, R., Sun, S., ... & Wang, W. (2017). CO2 conversion by plasma technology: insights from modeling the plasma chemistry and plasma reactor design. *Plasma Sources Science and Technology*, *26*(6), 063001.
26. Bogaerts, A., De Bie, C., Snoeckx, R. and Kozák, T. (2017), Plasma based CO_2 and CH_4 conversion: A modeling perspective. Plasma Process. Polym., 14: 1600070. https://doi.org/10.1002/ppap.201600070
27. Thomas G. Beuthe Thomas G. Beuthe and Jen-Shih Chang Jen-Shih Chang 1997 Jpn. J. Appl. Phys. 36 4997
28. A. Barkhordari, S.I. Mirzaei, A. Falahat, A. Rodero,Numerical and experimental study of an Ar/CO2 plasma in a point-to-plane reactor at atmospheric pressure, Spectrochimica Acta Part B: Atomic Spectroscopy,Volume 177, 2021, 106048, ISSN 0584-8547,https://doi.org/10.1016/j.sab.2020.106048.
29. J.A. Manion, R.E. Huie, R.D. Levin, D.R. Burgess Jr., V.L. Orkin, W. Tsang, W.S. McGivern, J.W. Hudgens, V.D. Knyazev, D.B. Atkinson, E. Chai, A.M. Tereza, C.Y. Lin, T.C. Allison, W.G. Mallard, F. Westley, J.T. Herron, R.F. Hampson, D.H. Frizzell, NIST Chemical Kinetics Database, NIST Chemical Kinetics Database, NIST Standard Reference Database 17, Version 7.0 (Web Version), Release 1.6.8, Data Version 2015.09, National Institute of Standards and Technology, Gaithersburg, Maryland, 20899–8320.
30. D. McElroy, C. Walsh, A.J. Markwick, M.A. Cordiner, K. Smith, T.J. Millar, The UMIST database for astrochemistry 2012, A&A. 550 (2013) A36, https://doi.org/ 10.1051/0004-6361/201220465.
31. Hagelaar, G. J. M. (2016). Brief documentation of BOLSIG+ version 03/2016. *Laboratoire Plasma et Conversion dEnergie (LAPLACE), Universit Paul Sabatier*, *118*.
32. Wang, S., Rohani, V., Dupont, P., Pagnon, S., & Fulcheri, L. (2023). A chemical kinetics simulation of plasma-catalytic dry reforming. *Chemical Physics*, *574*, 112017.
33. OMVE (2024). CP-121 Cold Plasma Demonstrator Manual, page 11.
34. Brezmes, Á. O., & Breitkopf, C. (2019). Numerical analysis of atmospheric pressure plasma produced by a dielectric barrier discharge in a mixture of Ar/CO 2. *IEEE Transactions on Radiation and Plasma Medical Sciences*, *4*(4), 498-511.
35. Ramakers, M., Michielsen, I., Aerts, R., Meynen, V., & Bogaerts, A. (2015). Effect of argon or helium on the CO2

conversion in a dielectric barrier discharge. *Plasma Processes and Polymers*, *12*(8), 755-763.

36. Rad, R. H., Brüser, V., Schiorlin, M., Schäfer, J., & Brandenburg, R. (2023). Enhancement of CO2 splitting in a coaxial dielectric barrier discharge by pressure increase, packed bed and catalyst addition. *Chemical Engineering Journal*, *456*, 141072.
37. Jahanbakhsh, M. R., Taghvaei, H., Khalifeh, O., Ghanbari, M., & Rahimpour, M. R. (2020). Low-temperature CO2 splitting in a noncatalytic dielectric-barrier discharge plasma: effect of operational parameters with a new strategy of experimentation. *Energy & Fuels*, *34*(11), 14321-14332.
38. Ozkan, A., Bogaerts, A., & Reniers, F. (2017). Routes to increase the conversion and the energy efficiency in the splitting of CO2 by a dielectric barrier discharge. *Journal of Physics D: Applied Physics*, *50*(8), 084004.

A Preliminary Investigation of Large Language Models (LLMs) as Process Operator Assistants

Abdullah Ali and Jacky Wong
Department of Chemical Engineering, Imperial College London, U.K.

Abstract
Operator mistakes can be catastrophic; Large Language Models (LLMs) address this by enhancing decision-making through contextual reasoning. This preliminary study aims to assess the potential of LLMs as process operator assistants through quantifying their capabilities in recognising P&IDs, understanding plant processes, and predicting and calculating actions. A multi-stage methodology was employed involving first designing a simple process and creating a P&ID in both image format and LLM specific formats (Vector and SFILES2.0) to test component recognition. LLMs' process safety understanding was assessed using cause-and-effect diagrams, where they were prompted to identify causal relationships leading to interrupts and their respective mitigating actions. Finally, the models' ability to predict actions and numerically calculate parameters was evaluated under various defined operating conditions. The study's results identified significant constraints in P&ID recognition, as LLMs were unable to process P&IDs in image form and required specific formats for comprehension. However, once in these formats, LLMs had no difficulty recognising the P&ID and demonstrated strong performance in solving simple, binary cause-and-effect problems and, despite limitations in numerical calculations, were proficient in logically explaining complex issues. In ambiguous, unrestricted scenarios, LLMs exhibited comprehensive troubleshooting capabilities and generated innovative ideas, identifying overlooked causes and challenging unrealistic plant specifications. These findings show LLMs high potential as Q&A advisory tools for inexperienced operators. Further investigation opportunities include P&ID digitisation and integrating LLMs in plant design and re-engineering.

1. Introduction

Operating a chemical or process system is a high stakes environment. Operator mistakes caused by inexperience or oversight can lead to significant fallout. An example being the disaster in Chernobyl occurring through operator oversight of safety protocols during a routine test (Norman, 1986). In less severe cases, mistakes can still often lead to extensive damage to the plant, injury or monetary loss.

Increasing safety in process engineering, essential for the function of global supply chains, is an ongoing goal. A step towards this has been automation, whereby solvers run all the calculations regarding set points, coolants, etc. However, these solvers are very limited in scope and are mainly only for mathematical solving.

The rapid development of AI in recent years has enabled its transition from only being capable of specific tasks to becoming general purpose models. This progress into advanced large language models (LLMs) has allowed AI's use to now be democratised across a wide range of applications. This generalisation also means that less investment is needed to adopt AI for each purpose, since in-house training is so costly, in terms of money and time. The engineering sector is a space that could see great promise from the implementation of such systems.

Having a more flexible, general intelligence operator assistant, with question-and-answer capabilities could be of great help in this space. This would take the form of either a system that would evaluate and reason about contextual knowledge to act automatically or provide practicable recommendations to an operator.

However, the suitability would need to be carefully verified, as this is still a novel development. LLMs have not yet proven themselves to be wholly reliable and are prone to leaps in logic and hallucinations of information (Huang et al., 2024). Hence, this report aims to quantify the extent to which LLMs can act as operator assistants and identify areas for future integration and improvement. To achieve this, this study will quantify different OpenAI models' capabilities in the three steps of problem solving: recognition, understanding and action. In this context, these would take the form of verifying recognition of a P&ID and its components, understanding of the process and its causal relationships and predicting actions and calculations based on simulated operating conditions. A side objective of ascertaining the impact of custom training have on a model's performance would also be carried out. This would be to see if current limitations are due to a lack of reasoning ability or a lack of relevant knowledge.

2. Background

For an LLM to process data, it first breaks down any query into a set of tokens, which can be vowels, words, symbols or numbers. These tokens are then encoded and processed through attention mechanisms to be assigned a location in a highly dimensional space.

This dynamic location is determined by an assigned weighting characterised by importance and recency. (Kudo 2018)

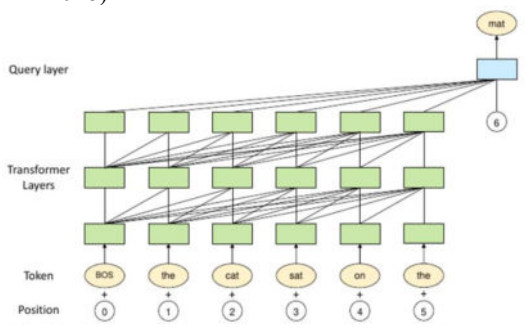


Figure 2.1 LLM Tokenisation Process Taken from (Naveed et al., 2024)

Models are then exposed to vast corpora for generalised learning and then tasked with fine tuning to improve

performance in tasks, as well as zero-shot/few-shot capabilities. All the while, they use self-improving algorithms to iteratively refine themselves.

Access to more data typically improves performance, hence the current drive in training highly optimised algorithms using larger datasets. Even in a logarithmic graph, there has been an exponential trend over the past decades in how much data each subsequent model has been trained on (Roser 2022). Research has followed a similar trend, with a 473-times increase in the number of papers published mentioning LLMs over the past 5 years (Naveed et al., 2024).

It lends itself to reason that the models themselves are developing just as quickly, and are now capable of diverse problem solving, logical reasoning, in-context learning and natural language understanding and generation. OpenAI's ChatGPT has had a monumental impact on the worlds attitude to AI and accelerated the pace of adoption across industries. Recent evaluations indicate that OpenAI's large language models outperform previous systems in reasoning, creativity, and versatility (Bubeck et al., 2023) ,thereby justifying their selection for this study.

Current literature applying LLMs to P&IDs remains limited and emergent. Although P&ID digitisation efforts have advanced (Kim et al., 2022), these have generally been pilot studies focused on converting P&IDs into digital formats rather than evaluating comprehensive recognition and understanding by LLMs. Moreover, recent studies have only examined specialised pre-trained models (Kozeiolek and Kozeiolek, 2024) or provide only conceptual reviews or propositions on potential applications of LLMs in process operation contexts (Nelson et al., 2024), rather than quantifiably evaluating readily available, off-the-shelf solutions like OpenAI's models. A notable study though is the work of Sakhinana, Sannidh and Runkan (2024), which showed the promise of a specially trained LLM for question-answering style analysis on P&IDs, generating accurate and contextually relevant answer.

This study aims to build upon and extend current literature findings by quantifying the capabilities of various ready-to-deploy OpenAI models to provide a more comprehensive evaluation of LLMs performance and potential, specifically as a plant operator's assistant.

3. Methods

A P&ID of a simple CSTR system with two coolant jacket loops was created for assessment purposes. It was isolated so that complex cause and effect relationships did not impact the system or obfuscate results.

Five of OpenAI's most recent models were selected for this investigation: ChatGPT 4, 4o mini, 4o, o1 mini, and o1 Preview. In addition, a separate 'Custom' model was trained using ChatGPT 4o. This involved a knowledge-based learning approach through the provision of an initial training document (Appendix 8.1) with typical P&ID customs and definitions including common process units, alarm systems, etc. The custom model was only used in the recognition and understanding testing sections. The tests used to quantify their respective capabilities have been detailed below.

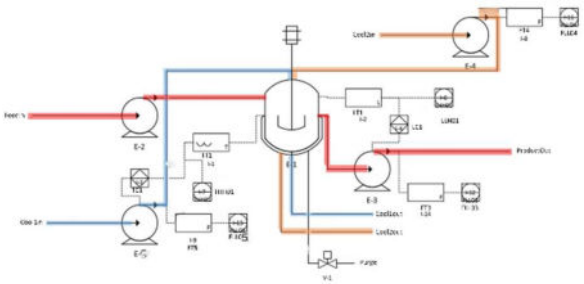


Figure 3.1 Mock P&ID Used in This Study

3.1 Recognition of P&ID

The initial phase in problem-solving is accurately ascertaining the nature of the task. As such, it was first imperative to evaluate whether LLMs could consistently and accurately wholly interpret P&IDs. A direct and streamlined method of familiarising an LLM with a process system involves presenting it with images of P&IDs. However, complex image processing remains a significant challenge for AI systems and so this approach was pre-empted to be insufficient for the goal of reliable interpretation.

Hence, a secondary method was devised to represent P&IDs in a textual format. This was through using a vector format, where the unit connections were described through nodes and edges. Comments accompanied to explain certain non-structural details, such as coolant loop priority. This also allowed for the instrument nodes to be separated from material nodes, preventing potential confusion or overlap in the models' understanding. As an aside, a more compact version of this format has been developed through SFILES 2.0, which is based on the SMILES format (Vogel et al., 2023). This can represent the same vector format but on a single line (aside from the descriptor comments). It wouldn't propose any performance gain over the standard vector format but would be easier to communicate between different interfaces and if storing many P&IDs.

Models were fed both versions, with the same, specific prompt to ensure consistency. Note that neither o1 mini or o1 Preview accept images and so these results are not available. The LLMs were asked to explain their understanding of the P&ID, writing a description of every recognised component, including what they connect to and from, and a stream-by-stream explanation. Each test was carried out three times.

Answers were graded according to an objective mark scheme to again provide consistency. A second component was also graded for. This was the percentage of the marks that were lost were due to not recognising items from the P&ID. In this way, "looking" can be separated from "seeing" (detecting from recognising).

3.2 Understanding of the Process

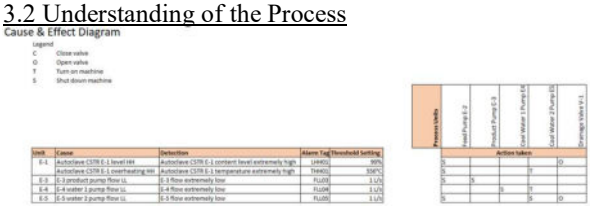


Figure 3.2 Manually Developed Cause and Effect Diagram

To assess the LLMs' understanding of the process, a cause-and-effect diagram was manually developed. This involved identifying particularly hazardous causes capable of causing an interrupt in the process, situations necessitating immediate operator intervention, as well as their associated binary effects for mitigation, as illustrated in the figure above. This diagram would serve as the basis of several levels of tests which were devised to further quantify the models' understanding of the process and its safety operations. Each test was conducted three times per model, and the results were averaged. Notably, at the start of each test, each LLM model was provided with the P&ID in vector format.

3.2.1 Level 1: Restricted Actions and Causes Fed

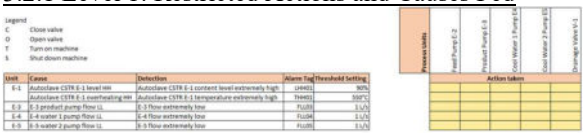


Figure 3.2.1 Level 1 Input

In the first level, the LLMs were presented with the cause-and-effect diagram, which included the causes and potential actions it can perform (i.e., opening/closing valves, starting/shutting down process units) on specific effects but omitted what to do in each cause. Hence, their task was to determine the appropriate effects for each cause, thereby demonstrating their understanding of the relationships between different causes and effects. A marking scheme was employed, whereby models were awarded points for correctly identifying effects, and penalties were imposed for irrelevant or incorrect responses, with weightings reflecting the severity of each error.

3.2.2 Level 2: Restricted Actions and Diagram Headings Fed

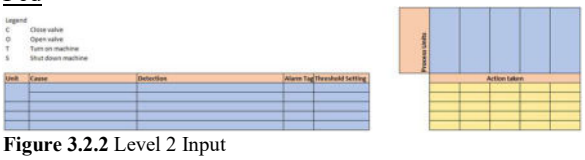


Figure 3.2.2 Level 2 Input

In the second level, the LLMs were provided only with the possible actions they could perform and the headings of the cause-and-effect diagram, without any listed causes or effects. The objective at this stage was to assess whether the LLMs could identify the correct causes likely to trigger a process interruption as well as the system's controllable effects, relying solely on their understanding of the system's safety aspects. A marking scheme was again applied with half of the score being based on the number of correct causes identified relative to the originally created diagram, and the remaining half of the marks was determined by the effects the model identified to address those causes. Additional capped bonus marks were awarded for identifying relevant causes not originally included in the diagram. Conversely, deductions were made for irrelevant or incorrect causes and effects, following the principles established in the first level's marking scheme.

3.2.3 Level 3: Nothing Fed; Unconstrained Problem

In this level of analysis, with the highest level of graduation in independence, the LLMs were tasked with performing a complete cause-and-effect evaluation of the system. Unlike previous levels, no predefined interrupt tables, templates or structured input/output formats were provided. This approach aimed to assess the thoroughness, creativity and practicability of the LLMs thought processes in proposing solutions. Each test was run three times.

After being provided P&IDs, the models were asked to identify potential causes leading to process units exceeding physical limits or failing. Each identified cause should be accompanied with a diagnostic method to verify its likelihood. The models were asked to then analyse the potential broader implications on the system if each cause were to occur. The final step would be to propose practicable solutions to each identified failure, reflecting real-world operational constraints and protocols. The LLMs were encouraged to explore a wide range of potential causes and solutions, to ascertain more closely how thoroughly the system's more nuanced causal relationships were being understood.

Responses were evaluated using a mark scheme adjusted for the more qualitative nature of the task. The five criteria appraised were (1) competency and accuracy, (2) thoroughness and detail, (3) creativity and problem-solving, (4) clarity and structure, and (5) practical relevance.
(1) - Assessed the technical correctness of the response;
(2) - Assessed the range of relevant identified causes, effects and solutions and to which extent they were explored;
(3) - Assessed the ability to propose innovative troubleshooting approaches;
(4) - Assessed the readability and presentation of responses;
(5) - Assessed the alignment of proposed solutions with real-world operational challenges and safety protocol.

3.3 Action Prediction & Calculation

LLMs were then assessed on their ability to apply their understanding to generate predictive actions and calculate parameters given further specification of the process. Again, at the start of each test, each LLM model was provided with the P&ID in vector format and tests were run 3 times with results averaged.

3.3.1 Scenario Analysis

Scen-ario	Description	Invo-lved Units	Action
1	E-1's level at 91%		
2	E-1's level at 55%		
3	E-1's temperature at 551°C		
4	E-1's temperature at 300°C		
5	E-3 flow rate at 1.5 L/s		
6	E-3 flow rate at 0.5 L/s		
7	E-4 flow rate at 2 L/s		
8	E-4 flow rate at 0.25 L/s		
9	E-5 flow rate at 1.75 L/s		
10	E-5 flow rate at 0 L/s		

Table 3.3.1 Scenario Analysis Input

Different scenarios on each of the process's units (Table 3.3.1) based on the causes and effects identified, were designed to assess the LLMs ability to relate each scenario to its respective causes and identify the correct effect actions to take given an input of the manually completed cause-and-effect diagram (Figure 3.2). LLMs were marked on how effectively the proposed actions solved the problem.

3.3.2 "Mastery" questions

LLMs were further tested on a more granular level through process mastery questions. The aim was to assess the LLMs ability to not only generate answers beyond standard binary actions like causes and effects but to also solve and provide logical, clear and in-depth explanations to advanced problems to simulate its role as an operator's advisor. To achieve this, a simple reaction of 2 molecules being reacted to produce 2 products was designed for our process.

$$A + B \rightarrow C + D \qquad \text{(Eqn. 1)}$$

The LLMs were then provided with additional specific information relating to each advanced problem and were marked based on the accuracy of the answer, correctness of method and its explanation to guide an operator through its the answer. Bonus marks were awarded on a discretionary basis for responses with factors that exceeded expectations such as the level of depth provided, or intuitive assumptions made.

3.3.2.1 Physical/Mass questions

Problem Type	Additional Specifications
Stream Table	• 80% Conversion • 0.06 mol/s Feed Rate • 1:1 Stochiometric Pure Feed of A & B
Level High High Alarm Trigger	• 80% Conversion • 50% Tank Level • 0.2 m^3 Reactor Volume • $-k[Ca]$ Rate Equation • $k = 10^4 \times e^{-\frac{54300}{8.314 \times Temperature}}$ (1/s) • 100°C Reactor Temperature (Isothermal)

Table 3.3.2.1 Physical/Mass Problems Additional Specifications

First, the LLMs were tested on their ability to calculate physical properties of the system, namely a stream table for the entire process and whether the LLM can apply this to the previous cause-and-effect relationships regarding the level high high alarm and when it will trigger if the product pump fails under specific conditions. This was done through feeding the LLMs with additional specifications respective to the problem type listed in Table 3.3.2.1 and prompting the LLMs to solve the problem.

3.3.2.2 Energy questions

Problem Type	Additional Specifications
Temperature of Reactor	• 80% Conversion • 50% Tank Level • 0.2 m^3 Reactor Volume • $-k[Ca]$ Rate Equation • $k = 10^4 \times e^{-\frac{54300}{8.314 \times Temperature}}$ (1/s) • Equimolar Pure Feed of A & B • $\Delta Hr^{\ominus}$ = -40 kj/mol • Cp A = 120 j / mol k • Cp B = 150 j / mol k • Cp C = 150 j / mol k • Cp D = 150 j / mol k
Cooling Water Flowrate	• 80% Conversion • 50% Tank Level • 0.2 m^3 Reactor Volume • $-k[Ca]$ Rate Equation • $k = 10^4 \times e^{-\frac{54300}{8.314 \times Temperature}}$ (1/s) • Equimolar Pure Feed of A & B • $\Delta Hr^{\ominus}$ = -40 kj/mol • Cp A = 120 j / mol k • Cp B = 150 j / mol k • Cp C = 150 j / mol k • Cp D = 150 j / mol k • 25°C Inlet Cooling Water Temperature • 70°C Outlet Cooling Water Temperature • 100°C Reactor Temperature (Isothermal)

Table 3.3.2.2 Energy Problems Additional Specification

Secondly, the LLMs were tested on their ability to calculate energy properties of the system, namely temperature of the reactor without cooling and subsequently the cooling water flowrate to keep the system isothermal. This was done through feeding the LLMs additional specifications respective to the problem type listed in Table 3.3.2.2 and prompting the LLMs to solve the respective problems.

4. Results & Discussion

4.1 Recognition of P&ID

A clear progression in recognition capabilities was measured along the evolution of newer models. The metric measuring the percentage of marks lost due to a lack of detection also showed a consistent improvement. This trend underscores the advancements in LLMs' ability to scan complex images for detailed information over time. It is assumed that this trend will extend into the future.

The custom trained model had the best performance in this category, which can be attributed to its specialised training. This prior knowledge enabled the model to focus recognition on features it knew would be relevant, reducing ambiguity and ignoring items.

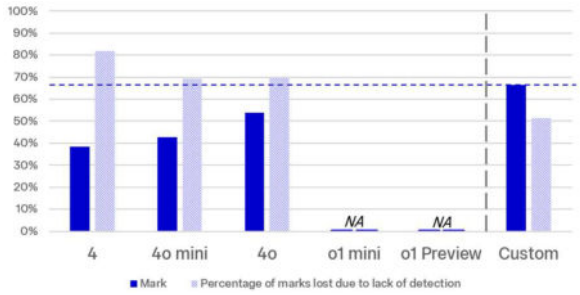


Figure 4.1.1 Image P&ID Recognition Results

One thing to note regarding initially constructing the vector form P&ID is that a method to streamline the typically laboursome process was found. Through providing prose descriptions of each stream to the LLM, it can convert the information into the vector format autonomously.

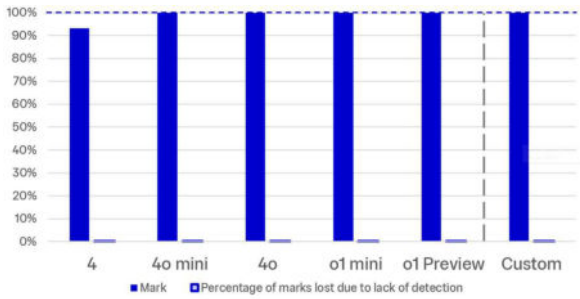


Figure 4.1.2 Vector Format P&ID Recognition Results

In terms of performance, all models except the oldest available (GPT 4, March 2023) achieved a perfect score when interpreting the textual input, showing complete understanding of the P&ID. This result suggests that the vector format offers an avenue for models that is more accessible and interpretable, effectively eliminating the challenges associated with processing the image-based inputs.

A limitation identified was the tendency of models, particularly those without the custom training, to make excess assumptions. For example, pumps were misclassified as heat exchangers, because of assumed naming conventions (e.g. Pump E-3). This is despite a pump symbol clearly being displayed in the P&ID. This highlights LLMs susceptibility to infer meaning where it may not exist, introducing errors.

The vector format was hence chosen for all future tests as it was determined it provided an unambiguous representation of the system. This would act to reduce the impact of interpretive biases, emphasising the importance of structured and precise inputs when using LLMs for process engineering tasks.

4.2 Understanding of the Process

4.2.1 Level 1: Restricted Actions and Causes Fed

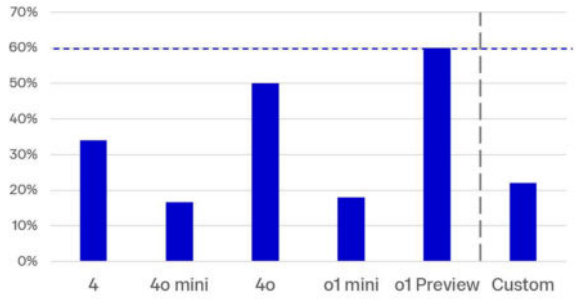


Figure 4.2.1 Average Performance Using Custom Mark Scheme

Overall, the progression from earlier to more advanced models led to notable improvements in performance, from [34%] in Model 4 to [60%] in the O1 model, particularly in their capacity to address low low flow rate causes and to adopt a more systematic and methodical approach to cause-and-effect reasoning. In contrast to the earlier models that attempted to account for all causes and effects simultaneously, later iterations, exemplified by the O1 model [at 60% performance], invested greater time (40–60 seconds) in carefully defining controllable effects before mapping out each cause individually, thereby demonstrating a more strategic and structured reasoning process. This approach yielded a more nuanced understanding of how feed-related factors influenced both level and temperature. These models dealt more effectively with high high temperature and level causes. Notably, in such scenarios, if four out of six models agreed on a specific effect, there was an approximately 80% likelihood of that effect being correct—an achievement that starkly contrasted with the effects identified by the LLMs for the low low flow rate causes, where models rarely achieved consensus and thus struggled to produce correct or consistent effects.

At the opposite end of the spectrum, the mini models exhibited the poorest performance [at 17–18%], displaying marked confusion regarding appropriate actions for specific units, such as confusing valve operations with pump operations—mistakenly shutting down valves instead of merely closing them—and failing to recognise constraints such as the impossibility of shutting down the reactor. These issues, combined with a high variance in responses, reflected a tendency to converge prematurely on simplistic conclusions, considering only one effect per

cause and conflating multiple causes into a single line of thought. The custom model, while marginally better [at 22%] than the mini models—particularly in identifying effects related to overheating and overflowing—nonetheless introduced unnecessary complexity by adding extraneous effects, struggled to correct malfunctioning pumps by turning them on, and remained prone to over-specifying responses to level and temperature difficulties. These persistent errors frequently occurred under conditions of low low flow rate causes and likely stemmed from a training bias introduced by the initial training document, which placed significant emphasis on pump and coolant definitions, prompting the model to propose excessive and extraneous actions.

4.2.2 Level 2: Restricted Actions and Diagram Headings Fed

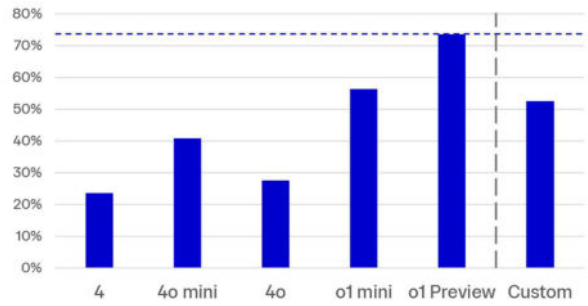


Figure 4.2.2 Average Performance Using Custom Mark Scheme

Across the models examined, their capacity to accurately identify underlying causes generally surpassed their ability to recommend their respective effects, with performance ranging from [24%] in Model 4 to [74%] in the O1 model. While individual models frequently pinpointed numerous core causes, the actions they suggested were often erroneous, incomplete, or tangential, a pattern especially evident in the earlier iterations [with Model 4 at 24% and 4o at 28%]. Notably, it appeared that even when cause identification was successful, the likelihood of selecting the correct respective mitigating effects often remained lower than compared to the level 1 understanding of the process, suggesting a limitation in the models' computational capacity to seamlessly translate accurate cause recognition into appropriate effects. In contrast, the more advanced model archetypes [such as O1 at 74%] not only reduced the incidence of irrelevant or incorrect cause identification but also recognised additional significant factors, such as low-low reactor levels that could lead to dry-running pumps, that had been overlooked in the manually created cause-and-effect diagram. These advanced models proposed more appropriate and contextually relevant corrective effects, echoing improvements observed in earlier experiments. The mini models performed considerably better in this assessment [with 4o mini at 41% and o1 mini at 56%] compared to previous tests, largely due to their enhanced cause identification capabilities, although they continued to struggle with implementing the correct mitigating effects. Persistent difficulties were observed in relation to flowrate causes, operation of the drainage valves, and cooling issues, with many models initially misidentifying extremely high flow rates as a cause and misunderstanding the relationship between increased cooling flow and reactor temperature reduction. Later models, exemplified by the O1 configuration [achieving 74%] and the Custom model [at 53%], did not replicate this misconception, indicating an evolution towards more accurate causal reasoning and action determination.

4.2.3 Level 3: Nothing Fed; Unconstrained Problem

The qualitative, open-ended nature of the test revealed significant variations in the models' reasoning, diagnostic, and problem-solving abilities.

Both o1 models demonstrated high levels of technical accuracy and comprehensiveness. They effectively identified even the more nuanced instrumentation and unit relationships, including feedback loops and notably pump failures. Pump failures had so far been the most difficult challenge to address for most models in all tests. However, the o1 Preview model provided a very comprehensive troubleshooting process, including recommending line walks, instrumentation checks, etc. Providing detailed and practical solutions, the o1 models have demonstrated a strong understanding of complex process interactions and a detail-oriented thought process. However, they did take a significant amount of time to give answers, with the longest being 53 seconds.

The mini models (o1 mini and 4o mini) displayed information in a clear and easy to follow way but did focus on fewer causes and didn't explore all failure points, especially in the case of 4o mini. Continuing with 4o mini, the time saving nature compared to the main model resulted in concentrating on the most direct causes. This often led to neglect of secondary failure points and downstream effects. Hence, despite acting the most quickly, a trait desired for a real time operator assistant, this lack of cutting of corners may affect their suitability for the role. They may however still be suited for operator training or educational settings, in contexts where straightforward and easily interpretable information is prioritised. Here, the mini models' ability to communicate findings simply and clearly would act in their favour.

The 4o models were found to have notable deficiencies in depth and creative problem solving. They tended to oversimplify diagnoses, neglecting critical relationships and dependencies, such as the necessity of activating the second cooling loop once the first had reached maximum capacity. Additionally, the models focused on simple fixes without due care to consider backup systems or alternative method resolution. These shortcomings indicate that earlier models may lack depth for addressing complex failure scenarios effectively.

The Custom version of the 4o model however performed significantly better, proving the importance of targeted training. It followed a more informed chain of logic,

resulting in recommendations that were more actionable and less vague. This was expectedly because of the heightened level of contextual knowledge.

Alarm	Possible Causes	How to Distinguish Causes	Likely Effects
LHH01	High level in reactor (E-1), valve malfunction (V-1), pump (E-3) failure	Check if level transmitter (I-2) is working. Verify pump (E-3) status. Inspect valve (V-1) behavior.	Reactor overflow, flooding, process shutdown.

Cause and Effect Table for Alarm LHH01

Alarm	Possible Cause	How to Distinguish Causes	Likely Effects	How to Solve
LHH01	Excessive feed rate into reactor (E-2)	Check feed flow rate from E-2	Reactor overfills, potential spill, safety hazard	Reduce feed rate from E-2, adjust feed pump settings
	Insufficient output flow from reactor (E-3 failure)	Check flow rate via I-14 (E-3 flow transmitter)	Reactor level increases, decreased product output	Repair or restart E-3 pump
	Level controller failure (I-4 malfunction)	Inspect level controller I-4 operation	Level not maintained, possible overfilling	Repair or recalibrate level controller I-4
	Blockage or closed valve in output line	Inspect output lines and valves	Level increases, possible pressure build-up	Clear blockage, open any closed valves
	Pump E-3 failure	Check operation of E-3 pump	No product output, reactor level increases	Repair or replace E-3 pump
	Instrument failure (I-2 faulty measurement)	Cross-check level with manual measurement	False alarm or missed actual high level	Calibrate or replace level transmitter I-2

Figure 4.2.3 Comparison between a section of Sample C&E Level 2 Answers between GP 4o mini (top) and GPT o1 Preview (bottom) for a high-high level warning

The evaluation showcases the significance of advancing models to address open-ended engineering problems. The newer, more advanced models, particularly o1, exhibited greater depth, creativity and practical relevance. This is compared to earlier models that were limited by a lack of exploratory reasoning, limiting their effectiveness in comprehensive problem solving.

The holistic and comprehensive reasoning displayed by o1 suggests that leveraging state-of-the-art LLMs could also find use outside of acting as operator assistants. This would be in contributing to wider, non-time sensitive tasks such as strategic safety planning or protocol development, which was originally beyond the scope of the investigation.

4.3 Action: Restricted Actions and Diagram Headings Fed

4.3.1 Scenario Analysis

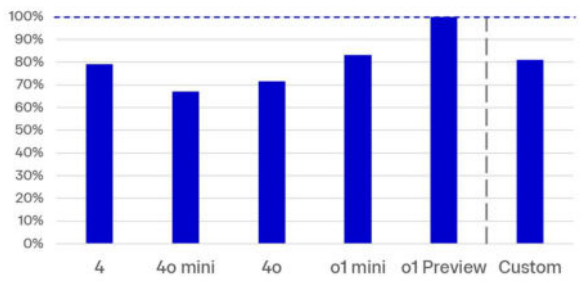


Figure 4.3.1 Average Performance Using Custom Mark Scheme

When presented with a set of predefined scenarios, the LLMs' abilities to recommend appropriate action were evaluated. This assessment aimed to determine how well models could apply their understanding to effectively solve specific operational issues, a task typical of a process operator.

The o1 family again consistently outperformed others in reasoning through the scenarios and proposing the technically accurate solutions. Despite again demonstrating superior problem-solving capabilities, the increased reasoning depth came at the expense of response times, taking significantly longer to generate answers than the other, older models. It seems like accuracy comes at the cost of speed.

Other off-the-shelf models faced consistent challenges with nuanced cause-and-effect relationships, particularly concerning pump flowrate issues. In these instances, answers often defaulted to simplistic solutions, such as suggesting turning the pump back on. The underlying causes or alternative action weren't being addressed.

The trained model again exhibited a better ability to align its recommendations with those in the cause-and-effect table than its untrained counterpart (GPT 4o). Detailed, actionable recommendations were provided. This highlights that targeted training can positively impact performance in scenarios requiring contextual understanding.

The slower response times of the o1 variants, while manageable in training or simulation, could pose challenges in real-time operational support. This trade-off of response time against depth will be a critical consideration for the deployment of LLMs in time-sensitive engineering environments. It also highlights the need for future research to optimise for processing speed without a compromise in reasoning depth.

4.3.2 "Mastery" questions

4.3.2.1 Physical/Mass: Stream Table

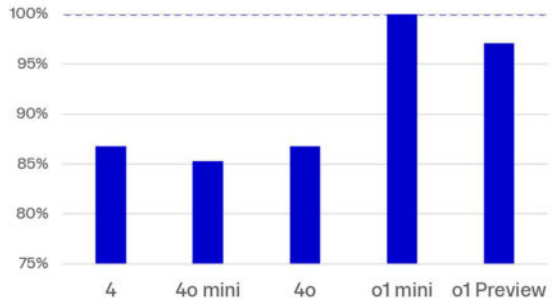


Figure 4.3.2.1.1 Average Performance Using Custom Mark Scheme

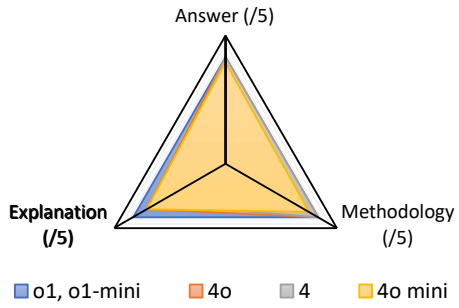


Figure 4.3.2.1.2 Performance Based on Evaluation Metrics

The performance data shows impressive results across all models, with scores in the high range of 85-100%. The o1 variants achieved nearly perfect scores and consistently outperformed other models across all metrics. They were the only models that achieved a 5/5 in the explanation metric by providing the most comprehensive stream table, which is a crucial aspect of presenting chemical engineering data. The base model, 4 and its variants (4o, 4o mini) on the other hand, performed similarly well in the 85-87% range, but scored less in the explanation metric and provided less detailed stream tables. These data showed how the o1 family's capacity to provide more comprehensive and detailed stream tables is a major benefit in making technical solutions more thorough and user-friendly.

4.3.2.2 Physical/Mass: High-High Level Alarm

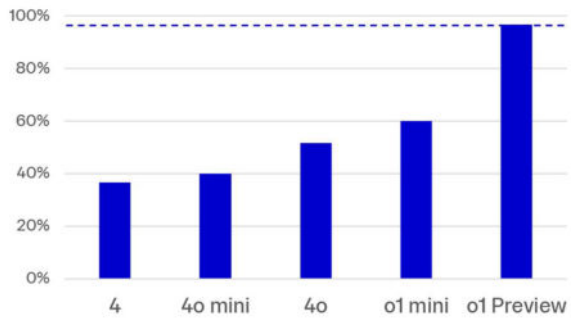


Figure 4.3.2.2.1 Average Performance Using Custom Mark Scheme

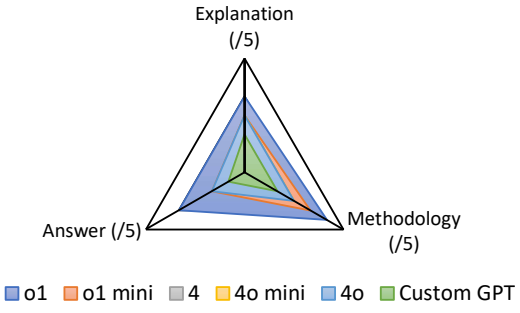


Figure 4.3.2.2.2 Performance Based on Evaluation Metrics

The data shows significant performance variations across models, with o1 being the clear leader at 97% average (93-100% across trials), distinguished by its correct equations, proper substitutions, and unique ability to identify and correct unrealistic activation energy values. On the other end of the spectrum, lower-performing models such as model 4 (37% average) struggled with incorrect mass balances, improper assumptions, and equation errors. Mid-range performers showed interesting patterns: 4o scored (52% average) despite some calculation errors, while o1 mini maintained steady performance (60%) showing good methodology but consistent issues with calculations. Common problems among lower-performing models included arbitrary assumptions, calculations that are off by 3 or more orders of magnitude, and minimal explanations lacking justification steps. The key differentiator across all models was their ability to implement the correct methodology rather than unnecessary theoretical understanding, with higher-performing mode, o1, providing systematic, step-by-step explanations with clear assumptions, while lower-performing ones (4o mini, model 4) failed to justify their calculations and assumptions. This pattern was particularly evident in methodology scores, which showed the highest variability compared to more consistent but generally low answer accuracy scores (0-2/5 for most models except o1). This provided further evidence that o1 is methodologically sound and accurate in comparison to the other models.

4.3.2.3 Energy: Reactor Temperature

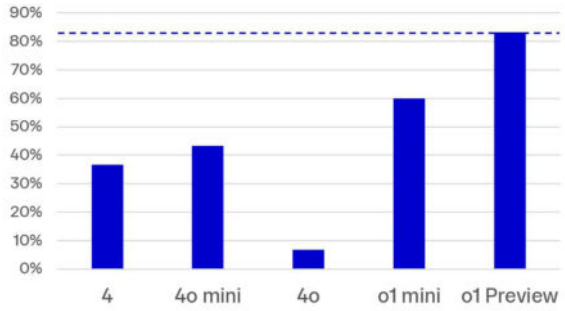


Figure 4.3.2.3.1 Average Performance Using Custom Mark Scheme

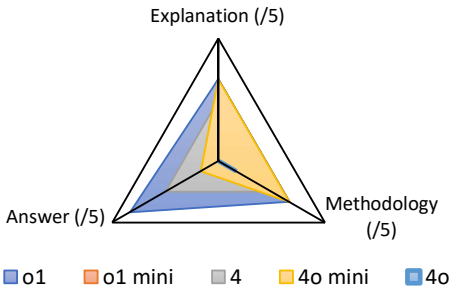


Figure 4.3.2.3.2 Performance Based on Evaluation Metrics

The o1 model family (o1 and o1 mini) is the clear leader across the LLMs. They scored the highest average performance and achieved a 5/5 in the answer's metric, signifying that the models excelled in calculating by making accurate assumptions based on practical chemical engineering situations. In contrast, the 4-series models revealed telling performance gaps - the base 4 model's terrible performance at 37%, while 4o performed critically poorly at 7%. These low scores reflect their fundamental methodology weaknesses (2/5 or 3/5). These models made concerning assumptions such as incorrect heat capacity mixtures, fundamentally wrong energy balances and incorrect constant heat of

reaction. In contrast, the 4o mini scored similarly to the o1 family in methodology, suggesting that it has a strong theoretical framework but lack consistency in its calculations to achieve the correct answer.

4.3.2.4 Energy: Cooling Water

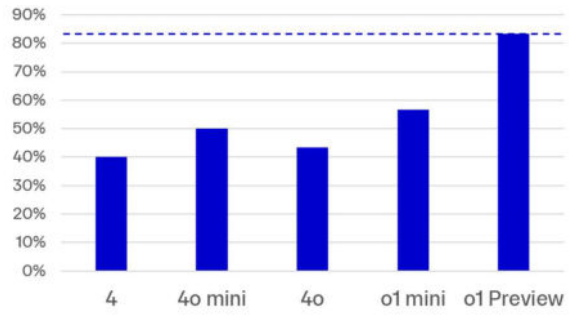


Figure 4.3.2.4.1 Average Performance Using Custom Mark Scheme

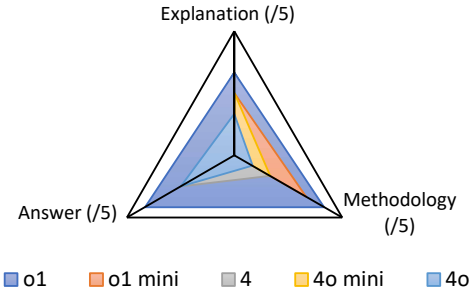


Figure 4.3.2.4.2 Performance Based on Evaluation Metrics

The base 4 model's scoring distribution (3/5 Answer, 2/5 Method, 1/5 Explanation) highlights a consistent weakness in its ability to combine multiple concepts - for instance, while it could sometimes handle basic heat calculations, it incorrectly assumes an unnecessary flowrate to solve the problem. The o1 model on the other hand, were able to score almost perfectly across all metrics. This is because o1 could consistently provide an accurate answer while making the correct assumptions and presenting its data thoroughly to the user. The mini variants provided an interesting case for this specific problem - o1 mini's (4/5) in methodology but (0/5) in answers, 4o mini also scored (0/5) in the answers. This is because while they provided a strong foundation of mass and energy balance equations – they do not attempt to solve it – which then makes this model less user friendly for this specific problem. Through holistically observing the data, one trend becomes clear - o1 consistently produced cooling water flow rates within reasonable tolerance, while other models deviated by orders of magnitude or failed to produce numerical answers entirely.

5. Conclusions

This study's findings highlight both the potential and challenges of applying LLMs as an operator's assistant within process plants. Across all evaluated tasks, the more advanced LLMs exhibited improved reasoning, clarity, and depth, indicating that ongoing developments could further enhance their utility as industrial support tools.

In the recognition of P&IDs, LLMs were restricted by image-based inputs, requiring extensive manual reformatting into a processable format for LLMs. This highlights the need for more robust P&ID importation methods—such automated conversion into standardised vector formats—to reduce manual effort, improve efficiency, and facilitate the integration of LLMs into more sophisticated, real-world plants.

Regarding process understanding, the models generally performed well in identifying primary cause-and-effect relationships, particularly for common major issues such as overheating and overfilling. Confidence in predictions increased significantly when multiple models reached consensus on an outcome for these causes. More advanced LLMs also offered nuanced insights, recognising less obvious causes and suggesting a comprehensive range of troubleshooting measures behind problems, thereby with a thought process akin to an experienced operator. Conversely, earlier and more compact models often oversimplified assumptions or misinterpreted issues, underscoring the importance of tailored prompting and fine-tuning.

LLMs demonstrated strong performance in simple, binary and conceptual decision-making, providing comprehensive and logical explanations, but struggled with precise numerical calculations and complex non-linear relationships, where traditional PID controllers are likely faster, more reliable and accurate. Nonetheless, their value as educational and advisory tools is noteworthy. They can guide operators through complex troubleshooting processes in an easily accessible question-and-answer format, expanding opportunities for operator training and decision support.

In almost all tasks, the custom-trained LLM outperformed its base GPT 4o. This showcases the potential of domain-specific fine-tuning to bridge knowledge gaps and enhance the applicability of LLMs for specialised engineering tasks. The findings suggest that incorporating targeted training modules can significantly elevate the operational utility of LLMs.

Overall, while LLMs are not set to replace human operators or established control systems, they can serve as valuable complements — enhancing safety and efficiency through supporting problem-solving and decision-making to assist less experienced personnel.

6. Future

The immediate next steps would be to validate LLMs against more complex, non-linear causal relationships within more sophisticated processes (i.e. those with more stages, separations or side reactions). This also includes exploring models beyond those developed by OpenAI, as well as alternatives including artificial general intelligence (AGI) models that are expected to better handle ambiguity, once they become available.

Looking ahead, a key area lies in advancing P&ID digitisation. Most P&IDs currently exist in physical or image-based formats, so their recognition is the key limitation of implementing LLMs to processes. Emerging tools, such as ChatP&ID developed by a group led by Artur Schweidtmann, which converts P&IDs into Neo4j graphs (akin to vector formats) (Digitisation Companion. 2024). An innovative end-to-end P&ID digitisation method created from another leading research group (Kim et al., 2022) demonstrates the potential of using pretrained AI to streamline this crucial recognition stage and enable widespread, context-specific testing.

A further opportunity is the integration of LLMs in plant re-engineering and design workflows. Given the creative, idea generation and comprehensive thinking capacity of advanced models, they could assist engineers by suggesting safety measures, advising on design feasibility, and identifying overlooked cause-and-effect relationships. This would be an opportune and direct implementation of this report's findings given this advisory role would be in a non-time-sensitive context, enhancing the quality and safety of process plant designs. There is growing interest in this application, with engineering firms (e.g., ABB) expressing interest, and emerging research (e.g., Balhorn et al., 2024) demonstrating LLM-based methods for autonomously correcting P&IDs by adding or refining process components and units. Few studies currently exist so this remains a nascent area with significant scope for future development.

Finally, integrating LLMs with dedicated mathematical solvers could address their current limitations in numerical precision. While LLMs excel at logic and conceptual reasoning, they struggle with precise calculations. By directing specific problem parameters to specialised solver modules, LLMs can provide overarching guidance while ensuring accurate analytical outcomes. This integration would yield a more comprehensive advisory tool that supports operators with both high-level strategic advice including troubleshooting and accurate quantitative predictions.

7. References

[1] Norman, Colin. (1986) *Chernobyl: Errors and Design Flaws.* Science 233, no. 4768: 1029–31. [online] Available at: http://www.jstor.org/stable/1697520.

[2] Huang, X., Ruan, W., Huang, W., Jin, G., Dong, Y., Wu, C., Saddek Bensalem, Mu, R., Qi, Y., Zhao, X., Cai, K., Zhang, Y., Wu, S., Xu, P., Wu, D., Freitas, A. and Mustafa, M.A. (2024). A survey of safety and trustworthiness of large language models through the lens of verification and validation. Artificial Intelligence Review, 57(7). doi: https://doi.org/10.1007/s10462-024-10824-0.

[3] Kudo, T. (2018). Subword Regularization: Improving Neural Network Translation Models with Multiple Subword Candidates. Meeting of the Association for Computational Linguistics. doi:https://doi.org/10.18653/v1/p18-1007.

[4] Naveed, H., Ullah Khan, A., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N. and Mian, A. (2024). *A Comprehensive Overview of Large Language Models*. [online] Available at: https://arxiv.org/pdf/2307.06435.

[5] Roser, M. (2022). The brief history of artificial intelligence: The world has changed fast – what might be next? [online] Our World in Data. Available at: https://ourworldindata.org/brief-history-of-ai.

[6] Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. and Zhang, Y. (2023). *Sparks of Artificial General Intelligence: Early experiments with GPT-4*. [online] Available at: https://arxiv.org/pdf/2303.12712.

[7] Byung Chul Kim, Hyungki Kim, Yoochan Moon, Gwang Lee, Duhwan Mun, (2022). *End-to-end digitization of image format piping and instrumentation diagrams at an industrially applicable level*. doi: https://doi.org/10.1093/jcde/qwac056

[8] Koziolek, H. and Koziolek, A. (2024). LLM-based Control Code Generation using Image Recognition. Proceedings of the 1st International Workshop on Large Language Models for Code, pp.38–45. doi: https://doi.org/10.1145/3643795.3648385.

[9] Decardi-Nelson, B., Alshehri, A.S., Ajagekar, A. and You, F. (2024). Generative AI and process systems engineering: The next frontier. *Computers & Chemical Engineering*, [online] 187, p.108723. doi: https://doi.org/10.1016/j.compchemeng.2024.108723

[10] Srinivas, S.S., Sannidhi, G. and Runkana, V. (2024). Towards Human-Level Understanding of Complex Process Engineering Schematics: A Pedagogical, Introspective Multi-Agent Framework for Open-Domain Question Answering. arXiv. doi: https://doi.org/10.48550/arxiv.2409.00082.

[11] Vogel, G., Hirtreiter, E., Balhorn, L.S. and Schweidtmann, A.M. (2023). SFILES 2.0: an extended text-based flowsheet representation. Optimization and Engineering, [online] 24(4), pp.2911–2933. doi:https://doi.org/10.1007/s11081-023-09798-9.

[12] Digitization Companion. (2024). [online] Available at: https://www.digitization-companion.com/.

[113] Balhorn, L. S., Caballero, M., & Schweidtmann, A. M. (2024). Toward autocorrection of chemical process flowsheets using large language models. Computer Aided Chemical Engineering, 3109–3114. doi: https://doi.org/10.1016/b978-0-443-28824-1.50519-6

[14] Allen-Zhu, Z., Yuanzhi, L. (2023). Physics of Language models: Part 1, Learning Hierarchical Language Structures. [online] Available at: https://arxiv.org/pdf/2305.13673v3

Multi-Objective Real-Time Optimisation of Load Allocation for Gas-Oil Separation Facilities

Aromal Jawaharlal[1] and Chris Z W Lau[1]
[1]*Department of Chemical Engineering, Imperial College London, U.K.*

Abstract
Gas-oil separation facilities require real-time optimisation (RTO) of operational expenditure (OPEX) and emissions to ensure economic viability and compliance with environmental regulations. This research evaluates the effectiveness of Bayesian optimisation (BO) and Extremum seeking control (ESC) in minimising OPEX and emissions in an gas-oil separation facility, modelled using Aspen HYSYS. This is critical for maintaining optimal feed composition in real industrial facilities, thereby maximising operational performance. A sensitivity analysis was conducted on benchmarked industry standards for each RTO method to determine the optimal hyperparameters, which were meticulously tuned for this specific process. Key hyperparameters for BO included the sampling method, kernel, acquisition function, and noise variance, while for ESC, they included the initial value, learning rate, and perturbation amplitude. Disturbances were simulated to assess the response of both methods. Results indicated that the tuning of hyperparameters was effective, with both BO and ESC outperforming their respective benchmarks. BO proved superior as a global optimiser, excelling at handling significant deviations from the set point, whereas ESC was more effective as a local optimiser, adept at managing minor steady-state deviations.

1 Introduction

Gas-oil separation facilities are critical in the upstream production of oil and gas, where crude oil, natural gas, and water are extracted from underground reservoirs. These facilities aim to efficiently separate each component, ensuring that the resulting products meet quality standards while minimising both operational expenditure (OPEX) and emissions.

Consequently, given the scale of these process, focus has shifted on optimising a single facility. A key aspect that could be improved upon is the feed allocation into the separation facility; balancing the inputs (crude oil, natural gas and water), which can reduce the need for flaring (Elhagar et al., 2020) and ensuring that the process is adequately equipped to handle disturbances effectively. Advancements in flowsheet simulators (Caballero and Grossmann, 2006) and improved algorithms for process optimisation have equipped engineers with the ability to efficiently model and simulate complex processes, including those found in gas-oil separation facilities. These advancements not only revolutionised how processes are initially modelled and designed but also enabled the use of advanced optimisation techniques such as Bayesian optimisation (BO) (Jones et al., 1998) and Extreme seeking control (ESC) (Krstić and Wang, 2000). The introduction of high-accuracy simulations (Rubio, Chamoin and Louf, 2021), advanced computational systems and real-time data acquisition have allowed for the effective application of these methods into the optimisation and control of complex engineering processes. These techniques emerged as a response to specific challenges in optimisation and control, BO was developed in the mid 20th century (Mockus, 1989) as a method for efficient search in expensive and highly dimensional functions. While traditional optimisation methods often struggle with computational costs (Liang et al., 2024) and their inability to handle complex problems effectively (Samanta, 2016), BO is a global search method which identifies solutions by constructing a probabilistic model using a surrogate function to approximate the objective function. This surrogate function balances the exploration of new areas and exploitation of promising regions to converge to an optimal solution.

Similarly, the concept of ESC can be traced back to the 1920s (Nesic, Tan, and Mareels, 2010), when it was initially developed to address the challenge of optimising dynamics systems in real time (Tan et al., 2010) without relying on mathematical models. ESC works by introducing small, controlled variations to the system inputs and analysing the estimated gradient to iteratively guide the model towards an optimal point.

This approach proved particularly effective for managing systems with uncertain dynamics (Pokhrel and Eisa, 2023), where traditional control strategies often fail (Ayten et al., 2024). Building upon these models, this paper focuses on evaluating how BO and ESC can be employed to enhance real-time operational decisions in a gas-oil separation facility. In particular, this research investigates the effectiveness of these methods in handling disturbances, guiding system parameters towards optimal points, and improving overall operational reliability. By comparing the performance of BO and ESC under varying operational conditions, this work aims to identify their respective strengths, limitations and potential for real-world application, ultimately aiding in the development of more robust and efficient optimisation strategies for industrial scale gas-oil separation systems.

2 Background

The global oil and gas industry faces growing pressure to align production activities with increasingly stringent environmental standards and profitability

metrics (Fernández-González, Puime-Guillén, and Vila-Biglieri, 2022). In response, the upstream sector; responsible for the exploration and production of oil and gas, has intensified efforts to improve process efficiency while simultaneously reducing environmental impacts (IEA, 2023). A key consideration is the day-to-day operations of gas-oil separation facility (GOSP), shown in Figure 1. Such facilities are integral in the separation of crude oil, natural gas and water, and so enhancing the performance and sustainability of a GOSP is essential in achieving both operational efficiency and environmental compliance. Historically, GOSPs have been designed and operated based on steady state assumptions and heuristics (Al-Dossary et al., 2020) as opposed to rigorous mathematical or data driven

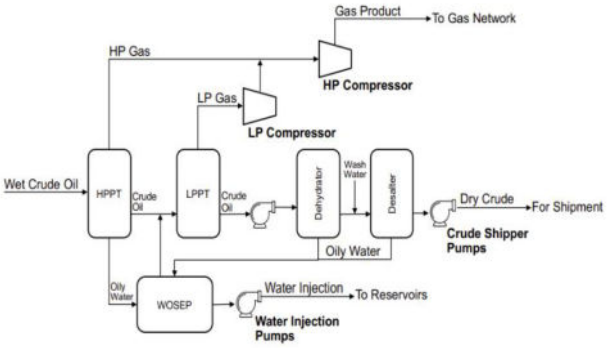


Figure 1: Block flow diagram of a gas-oil separation facility (GOSP) showing how crude oil, natural gas and water are separated (Bahamdan et al., 2024)

models. The expansion of enterprise-wide optimisation (EWO) has shifted the perspective to optimising entire production networks rather than individual units in isolation (Grossmann, 2012). Moreover, the complexity of modelling multiphase flow, phase separation inefficiencies, and operational dynamics between different pieces of equipment has historically been challenging (Pourahmadi Laleh, Svrcek, and Monnery, 2012). Traditional optimisation methods which rely on simplified analytical methods or linear approximations struggle to capture the nonlinear characteristics of the GOSP network (Ghazal, He, and White, 2024). In recent years, advancements in sophisticated modelling and optimisation techniques (Andreasen, 2020) have allowed engineers to accurately simulate complex operations with increasing precision. Consequently, advanced techniques like Bayesian optimisation and Extremum seeking control have gained notoriety due to their ability to navigate complex search spaces (Pokhrel and Eisa, 2023); (Chang and Gao, 2024). Bayesian optimisation, in particular, uses a probabilistic surrogate to model uncertainty in the objective function, balancing exploration of new, unexplored, regions with the exploitation of high-performance areas. It has shown promise in various domains, including chemical process optimisation (Shields et al., 2021) and reservoir management (Mujumdar and Nirmala, 2007), allowing for a systematic solution to high dimensional and expensive to evaluate functions. Extreme seeking control is a model free real-time optimisation technique that introduces controlled perturbations (often called dithers) to inputs, infers gradient information from output responses, and iteratively guides the system toward optimal points, making it an essential tool for real time control applications. Furthermore, recent progress in multi-objective optimisation frameworks have emerged that simultaneously address conflicting goals (Sharma and Kumar, 2022), such as minimising OPEX while reducing emissions, within large scale GOSPs. This allows engineers to consider trade-offs, aligning operational strategies with environmental targets and financial objectives all together. The subsequent work aims to further investigate and compare the use of Bayesian optimisation with Extreme seeking control in guiding operational considerations within a gas-oil separation facility. By utilising these techniques, this research seeks to enhance the robustness, reliability and sustainability of upstream operations.

3 Methodology

3.1 Multi-Objective Function

3.1.1 Annual OPEX

The annual operational expenditure (OPEX) is the primary focus in the optimisation. (Bahamdan, Shah, and del Rio-Chanona, 2024) states that the operation of pumps and compressors contribute to 80% of the GOSP's annual OPEX ($£\ million$) signifying the importance of its optimisation.

$$\begin{aligned} Annual\ OPEX\ &= Total\ Power \\ &\times Cost\ of\ Electrcity \\ &\times Operation\ Time \\ &\times 10^{-6} \end{aligned} \quad (1)$$

Where the total power (in kW) was obtained from the pumps and compressors, the cost of electricity is $£0.245/kWh$ (Ofgem, 2024), and the operation time is $8150\ h$ per year (Douglas, 1988).

3.1.2 Annual Emission

The annual emissions emitted from the plant are attributed to the flaring of excess natural gas due to overallocation into the GOSP. Emissions associated with power generation have been accounted for within the OPEX. Complete combustion of natural gas under standard conditions within the plant is assumed, hence the emission factor.

$$\begin{aligned} Annual\ Emission = &\ Excess\ Gas\ Product \\ &\times Gas\ Density \\ &\times Combustion\ Emission\ Factor \\ &\times 365 \times 10^{-3} \end{aligned} \quad (2)$$

The total annual emissions were estimated in t_{CO_2}. The excess gas product ($m^3 day^{-1}$) represents the total amount of gas product exceeding the max gas capacity of the facility ($m^3 day^{-1}$). The gas density ($kg\ m^{-3}$) was given in Aspen Hysys and a combustion factor of 2.75 ($kg\ CO_2\ (kg\ gas)^{-1}$).

3.1.3 Solving Multi-Objective Function

To optimise the multi-objective function, the epsilon constraint method was used. Annual OPEX was chosen as the primary objective function due to its critical role in enhancing profitability. Epsilon values were determined based on emission trends in the oil and gas industry (World Economic Forum, 2024). By iteratively running the optimisation with different epsilon values, a Pareto front was established. The optimal Pareto solution was identified by finding the point with the smallest Euclidean distance from the utopian value, which is the best possible values for both objective functions.

3.2 Problem Initialisation

$$\begin{gathered}\min AnnualOpex(\boldsymbol{x}) \\ st.\,Annual\ Emission(\boldsymbol{x}) \leq \varepsilon \\ where\ \epsilon\ \in \{0, 50, 100, 150, 200\}\end{gathered} \tag{3}$$

3.3 Constraints

When modelling the process dynamics of the plant, it was considered to be an expensive Blackbox function; hence, we limited the total iterations to 40 to reflect the computational complexity and resource-intensive nature of the plant. Alongside this, bound constraints were set on the inlet flowrates of each component entering the GOSP to reflect the physical constraints of the system such as equipment capacity and operational safety limits.

3.4 Simplification for Aspen Model

In practise, multiple plants are distributed across an oil field to handle production from a cluster of wells. For the purposes of this research, the model assumes the use of a single GOSP, which was thoroughly simulated using Aspen HYSYS. The simulation incorporates an idealised system with perfect separation integrated into it.

3.5 Gaussian Process

A Gaussian Process (GP) is a non-parametric, probabilistic method that utilises available data to approximate an unknown function. It is defined as a collection of random variables, any finite subset of which jointly follows a multivariate Gaussian distribution. A GP is fully defined by a prior mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{K}_y$, the covariance matrix accounts for assumptions of smoothness, complexity and similarity between inputs.
For any finite set of inputs $\boldsymbol{X} = \{x_1, x_2, \ldots, x_n\}^T$ The corresponding outputs $\boldsymbol{f} = [f(x_1), f(x_2), \ldots, f(x_n)]^T$ are jointly Gaussian distributed

$$\boldsymbol{f} \sim \boldsymbol{N}(\boldsymbol{\mu}, \boldsymbol{K}_y) \tag{4}$$

where

$$\begin{gathered}\boldsymbol{\mu} = [\mu(x_1), \mu(x_2), \ldots, \mu(x_n)]^T, \quad \boldsymbol{\mu} \in R^n \\ \boldsymbol{K}_{ij} = k(x_i, x_j) + \sigma_n^2\, \delta_{ij} \quad \boldsymbol{K}_y \in R^{n \times n} \\ for\ i, j\ =\ 1, 2, \ldots, n\end{gathered}$$

$k(x_i, x_j)$: The kernel function
σ_n^2: The variance of the observed noise
δ_{ij}: The Kronecker delta, is 1 if $i = j$ and 0 otherwise

The characteristics of the kernel impact the shape and performance of the model. A widely favoured choice is the use of the Radial Basis Function (RBF) kernel due to its well-established properties and mathematical resemblance to the Gaussian distribution, which allows for the modelling of smooth and continuous functions effectively. The RBF kernel is mathematically defined as:

$$k_{\text{RBF}}(x_i, x_j) = \exp\left(-\frac{\| x_i - x_j \|^2}{2l^2}\right) \tag{5}$$

Other common choices include Matérn 5/2, Linear Kernel and the White Kernel shown in equations (6), (7), (8)

$$k_{\text{Matérn5/2}}(x_i, x_j) = \sigma^2\left(1 + \frac{\sqrt{5}r}{l} + \frac{5r^2}{3l^2}\right)\exp\left(-\frac{\sqrt{5}r}{l}\right) \tag{6}$$

$$k_{\text{Linear}}(x_i, x_j) = \sigma^2(x_i \cdot x_j) \tag{7}$$

$$k_{\text{White}}(x_i, x_j) = \begin{cases}\sigma^2 & if\ x_i = x_j \\ 0 & if\ x_i \neq x_j\end{cases} \tag{8}$$

where l denotes the length scale hyperparameter. The length scale controls the rate at which the similarity between two points x_i and x_j, decreases as the distance between them increases. r represents the Euclidean distance between two points.

The kernel parameters, such as length-scale and signal variance, were determined through Maximum Likelihood Estimation (MLE). A process in which l is optimised to maximise the log marginal likelihood of the observed data. The log marginal likelihood is given by:

$$\begin{aligned}\log\ p(\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{\theta}) = &-\frac{1}{2}\boldsymbol{y}^T\boldsymbol{K}_y^{-1}\boldsymbol{y} - \frac{1}{2}\log\ |\boldsymbol{K}_y| \\ &-\frac{n}{2}log2\pi\end{aligned} \tag{9}$$

The GP's prediction at $\boldsymbol{x}$ corresponds to the posterior predictive distribution at $\boldsymbol{x}$, conditioned on the observed data (**X**,**y**). Due to the closed nature of Gaussian Processes under conditioning, this distribution means multivariate Gaussian.

Consequently, the predictive mean and uncertainty are as follows:

$$\mu(\boldsymbol{x}) = K(\boldsymbol{x}, X)\boldsymbol{K}_y^{-1}\boldsymbol{y} \quad (10)$$

$$\sigma^2(\boldsymbol{x}) = K(x, x) - K(\boldsymbol{x}, X)^T\boldsymbol{K}_y^{-1}K(\boldsymbol{x}, X) \quad (11)$$

3.6 Bayesian Optimisation with GP

Bayesian Optimisation (BO) is an iterative optimisation strategy suited for problems where the objective function $f(x)$ is expensive to evaluate, non-convex or black box in nature. BO involves constructing a probabilistic surrogate model of the objective function, in this context the surrogate model is represented by a Gaussian Process. The differentiability of the GP's predictive distribution makes it well suited for use within BO methods as it enables the efficient optimisation of acquisition functions using inexpensive gradient-base techniques, in contrast to traditional methods such as grid search or random search, which are commonly employed for non-differentiable models. At each iteration, an acquisition function is evaluated based on the GP model which determines the most promising location ($\boldsymbol{x}^*$) to sample from the true function, balancing both exploration and exploitation. The black-box function is evaluated at the selected point ($\boldsymbol{x}^*$), and the new output $\boldsymbol{y}^* = f(\boldsymbol{x}^*)$ is introduced to the training set. The surrogate Gaussian Process is then updated by incorporating the new data point ($\boldsymbol{y}^*, \boldsymbol{x}^*$), which refines the model's predictions by recalculating the posterior distribution. This process improves the GP's accuracy and confidence for subsequent iterations, enabling for better informed decision making in the optimisation process.

3.7 Acquisition functions

An acquisition function defines the search strategy in Bayesian Optimisation, determining where to evaluate next to effectively optimise the objection function. To explore how acquisition functions guide the optimisation process, we consider the three most widely used strategies in BO: Expected Improvement (EI), Probability of Improvement (PI), and the Lower Confidence Bound (LCB). Each of these functions provide a unique balance between exploration and exploitation, leveraging the predictive distribution from the Gaussian Process to achieve this. Let $\boldsymbol{f}^*$denote the optimum value of the objective function so far and define $\boldsymbol{z} = \frac{\mu(\boldsymbol{x}) - \boldsymbol{f}^*}{\sigma(\boldsymbol{x})}$ as the standardised measure of improvement. With Φ (z) and ϕ(z) representing the cumulative distribution function (CDF) and probability density function (PDF) of the standard normal distribution, respectively. The performance of these three acquisition functions was evaluated, and the impact of the number of iterations was also tested. A detailed comparison is provided in the results section, where their effectiveness in balancing exploration and exploitation is discussed.

Closed form expression for Expected improvement (EI), Maximum Probability of Improvement (MPI) and Lower Confidence Bound under GPs:

$$EI(x) = \begin{cases} \sigma(x)[z\Phi(z) + \phi(z)], & \sigma(x) > 0 \\ 0, & \sigma(x) = 0 \end{cases} \quad (12)$$

$$MPI(x) = P(f(x) > f^*) = \Phi(\mathrm{z}) \quad (13)$$

$$LCB(x) = \mu(x) + \kappa\sigma(x) \quad (14)$$

where $\kappa > 0$ is a parameter to manage exploitation-exploration.

3.8 ESC

Extreme Seeking Control (ESC) is a model free, real time optimisation technique designed to guide a system towards the optimal operating point that maximises or minimises performance. A key advantage of ESC over model-based control methods is that it does not rely on explicit mathematical models of a given process, this makes ESC particularly applicable in scenarios where system dynamics are complex, nonlinear or difficult to model accurately. The approach involves iteratively probing the system's response to small perturbations (dithers) in the input and uses these measurements to estimate local gradients, steering the system towards an optimal point.

Consider a single-input, single-output (SISO) system described by an unknown function $y = f(x)$ where $x \in R$ is the control input, and $y \in R$ is the measurable output. The objective is to find the optimal input x^*, given by $x^* = arg\min_x f(x)$.

ESC achieves this by introducing small, time dependent perturbations, sinusoidal in this case, around a slowly varying control input $u_0(t)$ which is subsequently applied to the system:

$$\mathrm{u(t)} = \mathrm{u_0(t)} + \mathrm{asin(\omega t)} \quad (15)$$

$$y(t) = f\big(\mathrm{u_0(t)} + \mathrm{asin(\omega t)}\big) \quad (16)$$

The key insight of ESC is to extract gradient information $f'(\mathrm{u_0(t)})$ from the output signal. By multiplying $y(t)$ by the sinusoidal function and applying a low pass filter (denoted by $L\{\cdot\}$) an approximation of the local gradient can be found:

$$f'\big(u_0(t)\big) \approx \frac{2}{a} L\{y(t)sin(\omega t)\} \quad (17)$$

Given an estimate of the gradient, the input $\mathrm{u_0(t)}$ is updated in a direction that improves performance. For minimisation problems, the adaptation law dictates that if $f'\big(u_0(t)\big) > 0$ this implies that increasing $u_0(t)$ would raise the objective function $f(x)$, therefore, to achieve minimisation $u_0(t)$ must move in the opposite direction. This principle ensures that at every step, the

model moves $u_0(t)$ closer to the optimum x^*. A gradient ascent rate can be employed to manage the improvement of $u_0(t)$, given by $\frac{du_0(t)}{dt} = \gamma f'(u_0(t))$.

3.9 Benchmarking

Bayesian Optimisation (BO) was benchmarked using a general optimisation algorithm for material systems (Liang et al., 2020). Extremum Seeking Control (ESC) was benchmarked using the ESC design principles outlined in for Simulink Control Design (MathWorks, 2024). Key metrics such as convergence speed, accuracy, and computational efficiency were compared to variations of each model.

Table 1: *Comparison of final Bayesian Optimisation (BO) and Extremum Seeking Control (ESC) models to their respective benchmarks. This study investigated: annual operational expenditure (OPEX), annual emissions, computation time.*

RTO Model	Final Annual OPEX ($£\ Million$)	Final Annual Emission (t_{CO_2})	Computation Time (s)	Difference in Annual OPEX from Benchmark ($£\ Million$)	Difference in Annual Emission from Benchmark (t_{CO_2})	Difference in Computation time from Benchmark (s)	Final Epsilon Value
BO	9.85	0	797	-1.63	0	-46	200
ESC	4.26	0	1991	-16.8	0	+259	0

4 Results

4.1 Forming the Ideal Hyperparameters

For the final Bayesian optimisation model, Latin Hypercube sampling ensured a better representation of the domain for the initial Gaussian Process (GP) model. Expected Improvement function was chosen due to faster convergence rates and improved optimal solutions. The Matern52 kernel was employed because of its cohesiveness and flexibility with the other features of the final model. Although limited noise variance was applied to the GP model, 5 restarts were applied for a thorough exploration of the domain.

For the final ESC model, the initial value was (190, 1, 1), obtained through adaptive stratified sampling of the domain space. The perturbation amplitude selected was 10% of the smallest component flowrate in the initial value. The perturbation frequency was maintained at the same level as the benchmark to ensure the stability and efficiency of the model. The learning rate was set to 0.6 as to ensure sufficient progress is made with each iteration and prevent convergence errors.

4.2 Comparison of Bayesian Optimisation and Extreme Seeking Control

As shown in Table 1, BO converged at iteration 34 with no further improvements, while Extremum Seeking Control (ESC) continually improved with each iteration. Table 1 shows that BO converged at $£9.85\ million$ compared to ESC's $£4.26\ million$, indicating ESC is superior at reaching the global minima. However, computation times for BO were $1194s$ faster than ESC, making it more effective for a quicker response to disturbances in the system.

Both BO and ESC models were improved from the respective benchmarking models, however for ESC, the computation time increased for the optimal convergence, BO had an epsilon value of $200\ t_{CO_2}$, which was considerably higher than that of ESC suggesting BO can be used for global exploration and handle higher tolerance levels, whereas ESC requires tighter control for stable results.

4.3 Mitigation of Model Errors

Aspen HYSYS flowsheets use iterative numerical methods to solve complex thermodynamic models, which can affect convergence precision. This was measured by conducting 3 repeats of the same Aspen input parameter. The uncertainty for the Aspen output can be inferred by the variance of the Annual OPEX, which is $\pm$ £ 23, which is very insignificant. As a result, this was neglected to improve computational efficiency.

Convergence errors were common with ESC especially when iterative values neared the component feed flowrate bounds, even with strict restrictions. To mitigate this, backoffs were applied to the boundary constraints as a function of the perturbation amplitude, which was the primary contributor to the iterative step size.

ESC was also prone to getting stuck at specific points and showed no improvement in subsequent iterations. This was predominantly dictated by the initial value selected for the optimisation, exemplifying the importance of this feature.

5 Discussion

5.1 Forming the Bayesian Optimisation and Extreme Seeking Control Models

To form the Bayesian optimisation (BO) and Extremum seeking control (ESC) models, the ideal hyperparameters must be selected. To understand the effect of different hyperparameters, sensitivity analysis was conducted on both models, using their respective benchmarking models for comparison, to improve performance. Noise was simulated in both models indirectly to test for robustness.

5.1.1 Bayesian Optimisation Sensitivity Analysis

5.1.1.1 Sampling Methods

Latin Hypercube sampling (LHS) enhanced BO by providing uniform coverage of the parameter space, which led to more informative samples and a better initial GP model with reduced variance in estimates. In contrast, random sampling had redundant coverage and higher variability resulting in slower convergence and reduced effectiveness compared to LHS. This led to faster convergence with LHS converging at 20 iterations compared to random sampling which converged at 26 iterations, shown in Figure 2. With LHS, the facility can save more than £1.52 $million$.

Therefore, LHS is the systematic choice for BO in complex systems like gas-oil separation process.

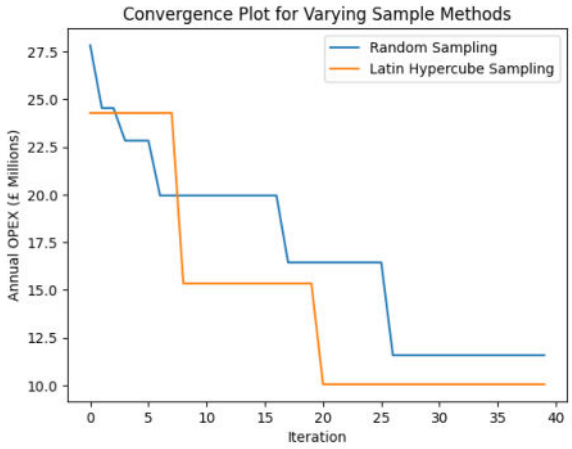


Figure 2: Annual Operational Expenditure (OPEX) convergence plot for Bayesian optimisation given different sampling methods: Random and Latin Hypercube Sampling.

5.1.1.2 Kernel

As shown in Figure 3, Linear and White kernels converged quicker at iteration 12 and 18 respectively compared to Radial Basis Function (RBF) and Matern52 which converged at iteration 26 and 36 respectively. Linear kernel simplifies the model by assuming a linear relationship so only has scale (influence of each feature) to tune, increasing the convergence speed and being less computationally intensive. White kernel also simplifies the model by incorporating the noise within the kernel so the noise can be separated from the true objective function for the GP model. RBF and Matern52 kernels take longer to converge due to their complexity, as well as having more hyperparameters to tune when modelling the objective function than Linear and White. Both have length-scale and signal variance, but Matern52 allows for flexibility in modelling the GP model unlike RBF which is can only model extremely smooth.

As a result, Matern52 was selected as the best kernel for BO for modelling and convergence in the GOSP.

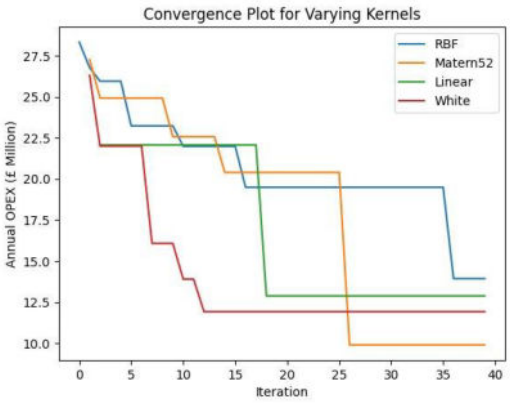


Figure 3: Annual Operational Expenditure (OPEX) convergence plot for Bayesian optimisation given different kernels: Radial Basis Function (RBF), Matern52, Linear and White

5.1.1.3 Acquisition Function

As shown in Figure 4, Expected Improvement (EI) function outperformed others in convergence value and speed by converging at iteration 7 with an Annual OPEX of £13.0 $million$ due to its effective balance of exploration and exploitation, since it focuses on areas of high uncertainty and potential improvement. Consequently, it is more effective at finding the global optima. The Lower Confidence Bound (LCB) function performed well but was less effective than EI, emphasising exploration over exploitation, which could be useful to not get stuck at local optimums. The Maximum Probability of Improvement (MPI) function had the slowest convergence and least optimal result, because it is highly exploitative focusing on regions with the highest probability of improvement, suggesting it is likely to get stuck at local optimums quite often.

These findings show that EI is the most effective for the load allocation optimisation.

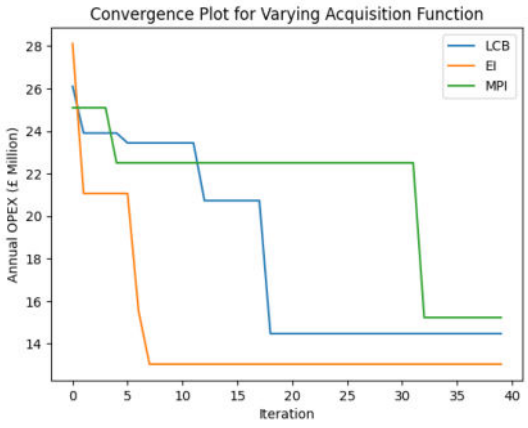


Figure 4: Annual Operational Expenditure (OPEX) convergence plot for Bayesian optimisation given different acquisition functions: Lower Confidence Bound (LCB), Expected Improvement (EI) and Maximum Probability of Improvement (MPI).

5.1.1.4 Iterations

Figure 5 shows that fewer iterations yielded a better convergence. This can suggest with more iterations, the optimisation process overfits to the noise present in the data, resulting in poorer modelling and thus worse convergence. As such, fewer iterations may overemphasise exploitation over exploration, which can explain the sharp decline in Annual OPEX values. However, this is not good enough since there is a large domain space and thorough exploration is required to find the global optimum. More iterations increase the reliability of the model, reducing the impact of individual noisy data.

Therefore 35 iterations were made for BO, constrained by the expensive backbox assumption.

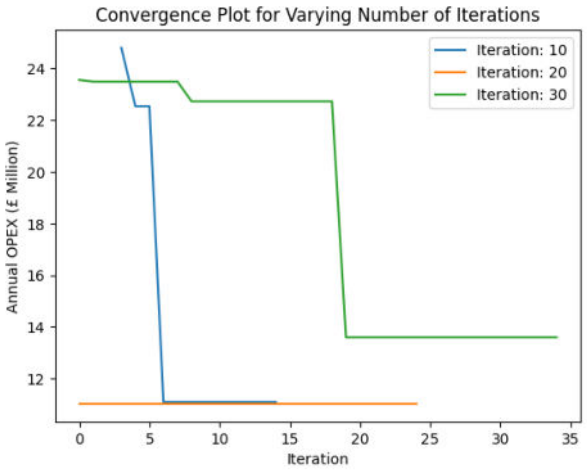


Figure 5: Annual Operational Expenditure (OPEX) convergence plot for Bayesian optimisation given different number of iterations: 10, 20, 30 (excluding initial sampling)

5.1.1.5 Noise Variance

BO demonstrated robust performance in the presence of varying noise variances. Convergence speed was quicker when some noise variance was added (0.1 and 1.0) given that it converged at iteration 6 and 7 respectively, shown in Figure 6. This can be because of BO's integration of noise variance into its surrogate models which allows it to balance exploration and exploitation effectively. This mitigates fluctuations and focuses on the true optimisation function.

The ability to handle high noise variance provides insight into BO's resilience against major disturbances within the process system. Notably, BO maintained convergence values akin to models with no noise variance when operating with noise variance of 10.0, underscoring its reliability in its application to gas-oil separation facilities.

As such, a small noise variance of 0.1 was integrated into BO to avoid overfitting and build a more realistic model with uncertainty.

5.1.1.6 Restarts

Increasing the number of restarts in BO significantly enhanced performance by building the best initial GP model, which led to better convergence. As illustrated in Figure 7, the final convergence value improved from £16.6 *million* to £10.0 *million*. However, increasing the number of restarts raised the computational cost, creating a trade-off between solution quality and resource utilisation. To enhance the likelihood of reaching the global optimum, 5 restarts were employed for the BO framework.

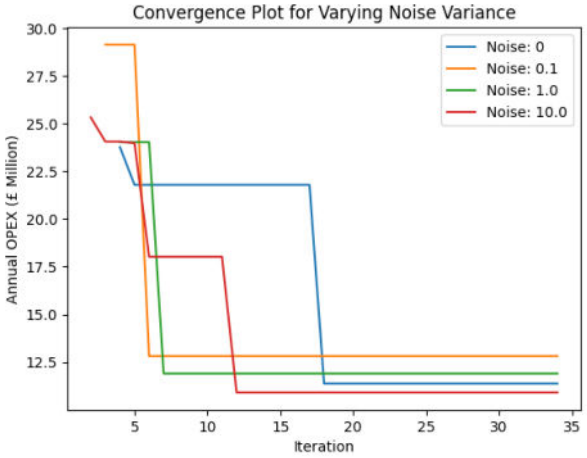


Figure 6: Annual Operational Expenditure (OPEX) convergence plot for Bayesian optimisation given different noise variance: 0, 0.1, 1.0, 10

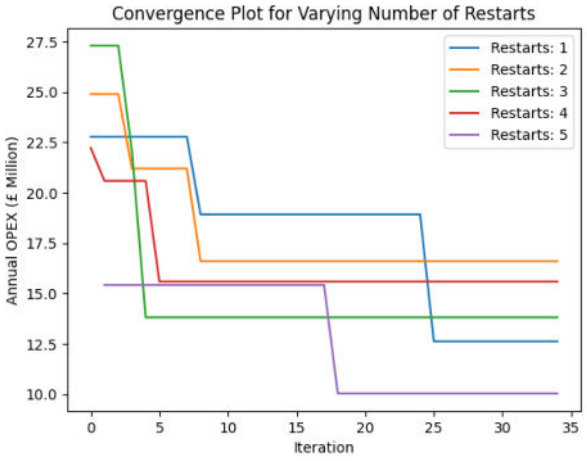


Figure 7: Annual Operational Expenditure (OPEX) convergence plot for Bayesian optimisation given different number of restarts: 1,2,3,4,5.

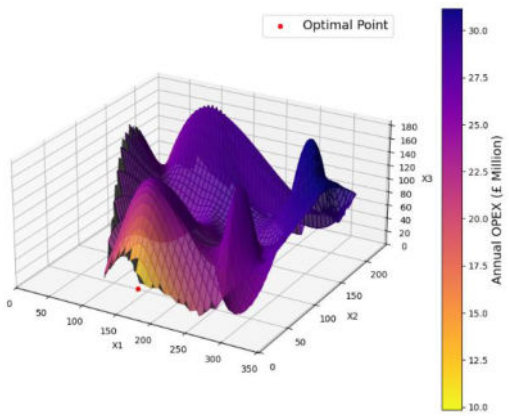


Figure 8: Uses observed data from Bayesian Optimisation (BO) to show how Annual Operational Expenditure (OPEX) changes with (x_1, x_2, x_3) which are crude oil, sour gas and water flowrates respectively. The optimal value for final BO model is $(180, 1.67, 42.8)$ labelled with a red point, given at epsilon = 200 t_{CO_2}.

5.1.2 Objective Function

Using the observed data from the BO method, the constrained multi-objective function can be seen in Figure 8, which is shown to be highly non-linear, shown by the concavity and convexity displayed within the predefined domain. It is also evident that for most feed flowrate combinations, the OPEX is above £25 million. This demonstrates the importance of real-time optimisation since sudden deviation from the optimal OPEX is likely for small perturbations in the feed. Another feature that can be observed is OPEX is lower when the crude oil feed flowrate is at least 50 times more than the natural gas and water. This can be explained because there are 4 less pumps and compressors being operated so less power is required for the process. The non-linearity of the crude oil pump curve also minimised the total power required in this scenario.

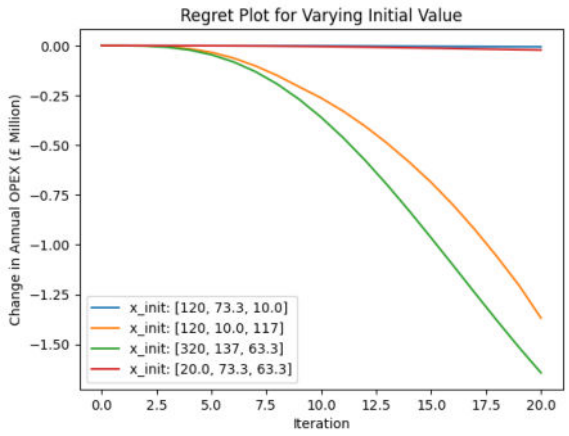


Figure 9: Annual Operational Expenditure (OPEX) convergence plot for Extremum Seeking Control given different initial values

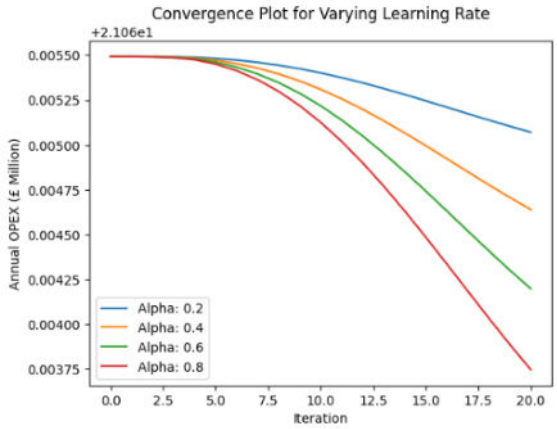


Figure 10: Annual Operational Expenditure (OPEX) convergence plot for Extremum Seeking Control given different learning rates

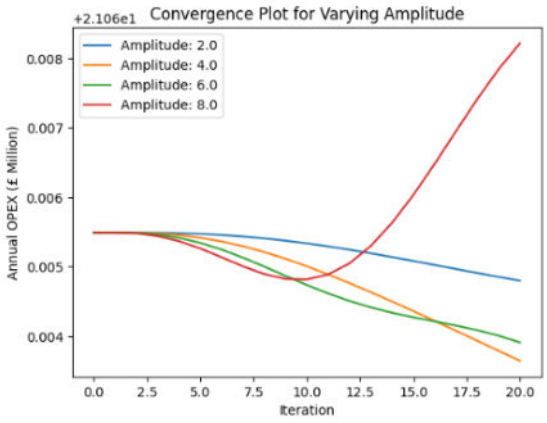


Figure 11: Annual Operational Expenditure (OPEX) convergence plot for Extremum Seeking Control given different initial values

5.1.3 ESC Sensitivity Analysis

5.1.3.1 Initial Value

Using stratified sampling to obtain different points within the feed flowrate domain, the change in annual OPEX in the form of a regret plot was observed. The plot clearly showed that the initial point can affect how well the ESC converges, shown by Figure 9. At initial points (120, 73.3, 10.0) and (20, 73.3, 63.3) there were minimal change showing ESC had extremely slow convergence suggesting it is highly non-linear around these values. The other points showed great improvement because they were in proximity of the optimum, so convergence was quicker and highly efficient. This is described by a maximum improvement of £1.7 million at initial value (320, 137, 63.3). Therefore, it would be important to test different initial values on the ESC algorithm to identify which would maximise performance. As such, adaptive stratified sampling was employed to locate a suitable optimal starting value for ESC, which was (190, 1, 1). However, this process is highly resource intensive and has a high likelihood of converging to a local optimum.

5.1.3.2 Learning Rate

From Figure 10, It is evident that increasing learning rate from 0.2 to 0.8 improved the annual OPEX by £1480. This trend suggests learning rate is directly proportional to convergence rate. However, when a high learning rate of 0.8 was used experimentally, there was an increase in the rate of convergence errors within Aspen Hysys model, so it is highly unstable. This is because the model can lead ESC to overshoot and oscillate around the minimum value, increasing the likelihood of errors occurring in Aspen Hysys, especially when minute changes are made to the system.

As a result, a learning rate of 0.6 was selected to maximise improvement whilst limiting the errors received.

5.1.3.3 Perturbation Amplitude

Increasing the perturbation amplitude led to better convergence initially from 2.0 to 6.0. Since ESC relies on gradient estimation for improvement, having larger amplitudes also increases the estimated gradient, which allows for better navigation of the search space. Above an amplitude of 8, the system displays more erratic behaviour. ESC is shown to diverge rather than converge which is because of the non-linearity of the objective function, pushing the system to different operating regions, destabilising the system. This is further explored when amplitude is increased to 14.0 where the operating region the system is operating in has improved, leading to better convergence. However, this scenario is not always likely as the opposite is just as likely to occur.

By introducing a perturbation amplitude of 14.0, disturbance can be simulated for ESC. As you can see from Figures 11, larger disturbances resulted in instability for ESC. This means that the perturbation amplitude should aim to prioritise exploitation rather than exploration and as such a lower amplitude should be used. This is so that there is more precision in the iterative steps taken.

As a result of this, a perturbation amplitude dependent on the minimum flowrate in the initial value is used, to ensure the convergence is around the initial value.

5.1.3.4 Perturbation Frequency

Higher perturbation frequency led to greater convergence because the ESC algorithm was better at adapting quicker to changes in the system. This is due to enhanced gradient estimation leading to more precise adjustments from the algorithm. As you can see, the range of frequencies tested were small because higher frequencies can lead to increased sensitivity to noise. The spike at perturbation frequency 0.4 is a mistake due to phase lag leading to an incorrect gradient estimation and subsequently divergence. A perturbation frequency of 1.0 had the greatest convergence so this was used for the final ESC model.

5.2 Final Model Analysis

5.2.1 Pareto Fronts

Table 2: Pareto Set of Solutions for Final BO model, given x_1, x_2, x_3 are crude oil, natural gas and water flowrates respectively

Epsilon	x_1	x_2	x_3	Annual OPEX ($£\ Million$)	Annual Emission (t_{CO_2})
0	99.4	13.0	162	19.1	0
50	300	4.96	121	16.4	0
100	287	40.5	0.787	17.8	0
150	140	11.3	50.5	16.5	0
200	181	1.66	42.8	9.85	0

Table 2 shows that while BO effectively reduced Annual OPEX, it exhibits objective dominance, overshadowing the constraint on Annual Emissions. Despite increasing the epsilon constraint, emissions remained constant at 0 t_{CO_2} despite Annual OPEX decreasing by 48.4%, highlighting that the constraint has not influenced the optimisation meaningfully. Increasing epsilon improved the solution despite OPEX dominance, as it provided additional flexibility in the optimisation process. This situation highlights the challenge of balancing both objectives, indicating a need to adopt a multi-objective optimisation approach which accounts for the emission's impact.

For ESC, there was only one pareto solution, showing a high dependency on the initial value. This indicates that ESC is heavily influenced by starting conditions, limiting its ability to explore the search space and find diverse optimal solutions. Thus, careful selection and tuning of initial conditions are crucial for achieving reliable and effective optimisation outcomes. This finding highlights the importance of initial value sensitivity in ESC performance and reliability. This can be seen in the final solution for ESC, where the feed flowrates deviated only ± 0.001 from the initial value displaying very little optimisation effectiveness. Therefore, ESC is more suitable for optimisation around the set point since it can keep it constant since it cannot handle large disturbances.

5.2.2 Epsilon

BO had a higher epsilon value due to its probabilistic approach, emphasising global exploration and handling uncertainties. In contrast, ESC had lower epsilon values because of its real-time, feedback-based optimisation with precise convergence criteria, necessitating tighter control for stable and accurate results.

6 Conclusion

6.1 Key Findings

Both Bayesian Optimisation (BO) and Extremum Seeking Control (ESC) demonstrated significant improvements over the industrial optimisation benchmarks for each model, underscoring the effectiveness of the hyperparameters chosen through sensitivity analysis, particularly for gas-oil separation facilities.

ESC significantly outperformed BO in converging towards the global optimum for annual OPEX. While BO achieved a considerably lower annual OPEX, ESC's major drawback is its high dependency on the initial value for optimisation.

BO excelled in managing significant deviations from the set point. This was especially evident when a large noise variance was introduced to annual OPEX to simulate disturbance. In contrast, ESC proved more adept at optimising minor, steady-state deviations since

applying a high perturbation amplitude to simulate disturbance led to instability in the system.

6.2 Key Limitations

The primary limitation of this research was considering only a single facility, despite the gas-oil separation process typically involving multiple facilities operating concurrently in clusters. Consequently, the immediate application of these RTO methods is not feasible, as additional factors such as inter-facility mixing and facility variations must be considered.

Additionally, the computational intensity of operating the RTO methods restricted the Blackbox samples to 40. This limitation impacted the BO model, as its performance is highly dependent on the number of iterations.

6.3 Improvements to Models

To enhance the BO model, more advanced surrogate models, such as neural networks and decision trees, can be employed to create a more accurate representation of the multi-objective function. Additionally, exploring and testing improved acquisition functions like Expected Hypervolume Improvement (EHVI) could significantly benefit the model. Unlike the epsilon constraint method, which prioritises one objective over others, EHVI simultaneously considers both objective functions (Yang et al., 2019). This approach dynamically balances trade-offs, thereby effectively forming the Pareto front, negating the need for the epsilon-constraint method.

ESC could integrate a more refined search method, such as Genetic Algorithms (GA), to navigate the search space for potential initial values. GA starts with a population of candidate solutions, evaluates their fitness before iteratively refining them through selection, crossover and mutation (Mathworks, 2024). The wide exploration used prevents premature convergence to a local optimum, increasing the likelihood of global convergence.

Combining BO with ESC should also be considered. BO is shown to be more effective at global optimisation so inputting its convergence value into the initial value for ESC can improve the performance of the optimisation.

6.4 Further Research

Exploration into more multi-objective optimisation methods such as Multi-Objective Particle Swarm Optimisation (Fitas, 2024) and Multi-Objective Evolutionary Algorithms (Pătrăușanu et al., 2024) should be researched for their global search capabilities and versatility.

Investigation into the scalability of the RTO models considering multiple facilities must be explored for practical application.

7 Acknowledgements

We thank Abdullah Bahamdan for his continuous support and guidance in the completion of this research project.

8 Supplementary Information

Access the code and Aspen Hysys file for this paper through the following link: https://github.com/AromalJawaharlal/Research-Project.git to see more sensitivity analysis plots.

9 References

Al-Dossary, B., AL-Naser, M., Al-Dogail, A., Gajbhiye, R., Al-Qathmi, A. and Mahmoud, M., 2020, January. Decision support system for optimizing GOSP operation. In *International Petroleum Technology Conference* (p. D033S232R001).

Alhuraifi, M., Ghazal, A., He, Y. and White, R., 2024, February. Improving Energy Efficiency and Reducing Carbon Footprint of Oilfields Using a Rigorous Multi-GOSP Optimization Platform. In *International Petroleum Technology Conference* (p. D031S100R003).

Andreasen, A., 2020. Applied Process Simulation-Driven Oil and Gas Separation Plant Optimization Using Surrogate Modeling and Evolutionary Algorithms. *ChemEngineering*, 4(1), p.11. doi:10.3390/chemengineering4010011.

Bahamdan, A., Shah, N., and del Rio-Chanona, E.A., 2024. Surrogate-based mixed-integer linear programming model for decarbonization of an integrated gas-oil separation network. *Proceedings of the 34th European Symposium on Computer Aided Process Engineering (ESCAPE34)*, Florence, Italy, 2–6 June 2024. Elsevier.

Caballero, J.A. and Grossmann, I.E., 2006. Integration of modular process simulators under the Generalized Disjunctive Programming framework for structural flowsheet optimization. *Computers & Chemical Engineering*, 30(3), pp.409-422. doi:10.1016/j.compchemeng.2005.10.007.

Chang, Y.C. and Gao, S., 2024. Extremum-Seeking Action Selection for Accelerating Policy Optimization. *arXiv preprint arXiv:2404.01598*. Available at: https://arxiv.org/abs/2404.01598.

Elhagar, F., Elhagary, M., Ahmed, H. & Al-Zahrani, A., 2020. Enhancing heavy hydrocarbon recovery in Gas-Oil Separation Plants to reduce flaring. *Springer Journal of Petroleum Exploration and Production Technology*, 10(4), pp. 1503–1514. doi:10.1007/s40095-020-00363-z.

Fernández-González, R., Puime-Guillén, F., and Vila-Biglieri, J.E., 2022. Environmental strategy and the petroleum industry: a sustainability balanced scorecard approach. *Journal of Petroleum Exploration and Production Technology*, 13, pp.763–774. doi:10.1007/s13202-022-01543-9.

Grossmann, I.E., 2012. Advances in mathematical programming models for enterprise-wide optimization. *Computers & Chemical Engineering*, 47, pp.2-18. doi:10.1016/j.compchemeng.2012.06.037.

International Energy Agency (IEA), 2023. Emissions from Oil and Gas Operations in Net Zero Transitions. Available at: https://www.iea.org/reports/emissions-from-oil-and-gas-operations-in-net-zero-transitions [Accessed 6 December 2024].

Jones, D.R., Schonlau, M. and Welch, W.J., 1998. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13(4), pp.455-492. doi:10.1023/A:1008306431147.

Krstić, M. and Wang, H.H., 2000. Stability of extremum seeking feedback for general nonlinear dynamic systems. *Automatica*, 36(4), pp.595-601. doi:10.1016/S0005-1098(99)00222-1.

Liang, J., Lou, Y., Yu, M., Bi, Y. and Yu, K., 2024. A survey of surrogate-assisted evolutionary algorithms for expensive optimization. *Journal of Membrane Computing*, 4(3), pp.197-213. doi:10.1007/s41965-024-00165-w.

Liang, Q., Gongora, A.E., Ren, Z., Tiihonen, A., Liu, Z., Sun, S., Deneault, J.R., Bash, D., Mekki-Berrada, F., Khan, S.A. and Hippalgaonkar, K., 2021. Benchmarking the performance of Bayesian optimization across multiple experimental materials science domains. *npj Computational Materials*, 7(1), p.188.

Mockus, J., 1989. Bayesian Approach to Global Optimization. Dordrecht: Kluwer Academic.

Mujumdar, P.P. and Nirmala, B., 2007. A Bayesian stochastic optimization model for a multi-reservoir hydropower system. *Water Resources Management*, 21(9), pp.1465-1485. doi:10.1007/s11269-006-9094-3.

Nesic, D., Tan, Y., and Mareels, I., 2010. Extremum Seeking from 1922 to 2010. *Proceedings of the 29th Chinese Control Conference*, pp.14-26. doi:10.1109/ChiCC.2010.5572972.

Pokhrel, S. and Eisa, S.A., 2023. Gradient and Lie Bracket Estimation of Extremum Seeking Systems: A Novel Geometric-based Kalman Filter and Relaxed Time-dependent Stability Condition. *International Journal of Control, Automation and Systems*, 21(12), pp.3839–3849. doi:10.1007/s12555-023-0224-y.

Pourahmadi Laleh, A., Svrcek, W.Y., and Monnery, W.D., 2012. Computational Fluid Dynamics-Based Study of an Oilfield Separator—Part I: A Realistic Simulation. *Oil and Gas Facilities*, 1(6), pp.58-67. doi:10.2118/161212-PA.

Rubio, P.B., Chamoin, L. and Louf, F., 2021. Real-time data assimilation and control on mechanical systems under uncertainties. *Advanced Modeling and Simulation in Engineering Sciences*, 8(4). doi:10.1186/s40323-021-00188-3.

Samanta, D., 2016. Solving Optimization Problems. Soft Computing Applications (IT60108), Indian Institute of Technology Kharagpur.

Sharma, S. and Kumar, V., 2022. A Comprehensive Review on Multi-objective Optimization Techniques: Past, Present and Future. *Archives of Computational Methods in Engineering*, 29(5), pp.5605-5633. doi:10.1007/s11831-022-09778-9.

Shields, B.J., Stevens, J., Li, J., Parasram, M., Damani, F., Alvarado, J.I.M., Janey, J.M., Adams, R.P. and Doyle, A.G., 2021. Bayesian reaction optimization as a tool for chemical synthesis. *Nature*, 590(7844), pp.89-96. doi:10.1038/s41586-021-03213-y.

Tan, Y., Moase, W.H., Manzie, C., Nešić, D., and Mareels, I.M.Y., 2010. Extremum Seeking from 1922 to 2010. *Proceedings of the 29th Chinese Control Conference*, pp.14-26.

The MathWorks, Inc., 2024. Extremum Seeking Control. Available at: https://uk.mathworks.com/help/slcontrol/ug/extremum-seeking-control.html [Accessed 12 November 2024].

The MathWorks, Inc., 2024. What is the Genetic Algorithm? Available at: https://uk.mathworks.com/help/gads/what-is-the-genetic-algorithm.html [Accessed 12 November 2024].

Turton, R., Bailie, R.C., Whiting, W.B., Shaeiwitz, J.A., and Bhattacharyya, D., 2012. Analysis, Synthesis, and Design of Chemical Processes. 4th ed. Upper Saddle River, NJ: Prentice Hall. Available at: http://repository.um-palembang.ac.id/id/eprint/9020/1/Conceptual%20Design%20of%20Chemical%20Processes%20%28%20PDFDrive.com%20%29.pdf

Optimisation of Translational Efficiency in a Pichia Pastoris Cell Free System

Gavin Murugan and Thagshanen Thillairajah
Department of Chemical Engineering, Imperial College London, U.K.

Abstract
Cell-free protein synthesis is an in-vitro process to express proteins, which removes the constraints of membranes, cell walls and the metabolic functions involving them. The roles of Kozaks and additives and their respective effects on translation have been studied and proven to be significant, warranting further exploration of their synergy and optimisation of translational efficiency. E. Coli, containing AAAAAA and GCCACC Kozak sequences was cultured and centrifuged for DNA extraction and purification via a Maxiprep kit. These Kozaks were tested with a control sequence alongside testing of additive combinations with Pichia pastoris as a lysate. Results indicated that the AAAAAA Kozak and the Spermidine and PEG6000 additive combinations achieved the greatest translational rates and yields. Similar results were obtained when testing the combination together with the AAAAAA Kozak. The behaviour can be attributed to eukaryotic yeast strains preferring adenine rich upstream tracts from the start codon in previous investigations and literature. Spermidine's nature as a polyamine and PEG6000's macromolecular crowding effect provide optimal conditions for mRNA translation and efficient folding of resultant proteins, without interfering with one another's activity. Genetic modification of Kozak sequences and use of complementary additives were deemed beneficial for organism-specific enhancement of translational efficiency.

---

1 Introduction

1.1 Cell Free Protein Synthesis

From 2012 to 2017, the market for biological products, which includes therapeutic proteins, increased by 70% [1], illustrating the increase in demand for processes such as protein synthesis industrially. With the pharmaceutical industry investing more with projections to keep increasing [2], innovation to produce complex and mammalian proteins, such as hormones, growth and blood factors, enzymes, etc., is continuously being researched and advanced.[3] With the advantage of biological systems being able to express larger and more complex proteins than chemical synthesis [4], modifications to folding techniques and the recombinant genetic engineering of different hosts, including eukaryotes, are being favoured [5]. Cell-free protein synthesis (CFPS) is one such approach, which provides the benefits of biological expression, whilst reaching full volume fraction of a reactor, a feature which only chemical based systems had before and therefore achieved 80-95% more effective use of reactor volumes, in comparison to cell-based synthesis. [6]

Cell free protein synthesis (CFPS) involves protein synthesis without the use of living cells, first discovered almost 70 years ago. Since then, CFPS has resulted in a number of applications, from small scale as functional genomics to a larger scale such as antibody production. [7]. A key part of the process of CFPS is the elimination of the cell walls and membranes of cells via lysis and eliminating gene regulation by separation techniques, such as centrifugation. The following lysate consists of having ribosomal and nucleic acid content, which can then be enhanced for protein synthesis with additional salts, polyamines, creatine compounds systems, energy systems to synthesize proteins outside living cells. Specifically engineering gene can be supplemented to control the synthesis within the entire solution. The benefits within CFPS are that the protein synthesis time is very short compared to traditional cell protein synthesis; CFPS processing time has a duration of a few hours in comparison to traditional cell protein synthesis, taking days. [8]

1.2 Research Objectives

The main goal of this project is to investigate the optimum conditions to ensure protein yield is at its maximum. These are proceeded by altering additives added within the master mix and modifying Kozak sequences which allows a more enhanced protein synthesis. For this objective to be achieved, the Kozak sequences will be investigated, which will then be followed by combinations of additives. The results from both tasks will then be used to ultimately create a combination of Kozak and additives, which will be tested.

2 Background

Cell-free protein synthesis (CFPS) is transforming protein production by moving away from the constraints of traditional, cell-based methods. For example, this approach is particularly useful for making proteins that are complicated or even toxic to express in living cells, which is a hurdle in conventional systems [9]. Since it bypasses the metabolic processes involving membranes and cell walls, CFPS offers a flexible and rapid platform for protein synthesis. This system can be easily tailored by adding specific components or adjusting the reaction environment, enabling optimisation of conditions to increase transcriptional and translational efficiencies.
The yeast strain, Pichia pastoris (P. pastoris), is well-established for recombinant protein production due to its fast replication [10], ability to accommodate post-translational modifications and generate high yields of protein [11]. Integrating P. pastoris into CFPS systems combines these advantages with the flexibility of a cell-free environment. However, to maximize protein yields, strategic optimization is required, such as the manipulation of translation initiation sequences and the inclusion of specific additives. Additionally, proteins, which may not necessarily be synthesised efficiently in bacterial and viral strains or other yeast strains, have been proven to be expressed in P. pastoris. [12]. It was

therefore decided that P. pastoris would be ideal for an experimental lysate in this investigation.

2.1 Kozak Sequences

A critical factor in translation initiation is the Kozak sequence, a specific nucleotide sequence surrounding the start codon, which enhances ribosome binding and translation efficiency by guiding ribosomes to the start codon. [13] The nucleotide composition around the start codon can significantly impact translational efficiency [14]. This research focuses on two Kozak sequences, AAAAAA and GCCACC, to investigate their effects on translation rate and protein yield in P. pastoris-based CFPS. AAAAAA was selected as one of the Kozaks due to its rich adenine content, which is preferred in the 5' UTR region by eukaryotic strains [15][16]. GCCACC is deemed a general 'consensus' Kozak, due to its bases being prevalent between the -11 position and the -1 region upstream of the start codon, as depicted in Fig 1, and was therefore elected as the second Kozak to test. A positive control Kozak, GAAACC, was tested prior to this study and yielded successful protein expression, with and without additives, therefore being optimal as a basis Kozak for comparison and statistical analysis.

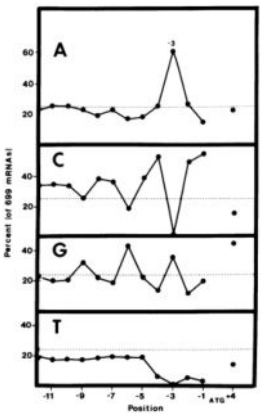


Figure 1.1 Frequency of A, C, G and T around the ATG initiator codon in 699 vertebrate mRNAs [14]

While CFPS systems using P. pastoris lysate present many advantages, they also pose challenges such as lack of stability control of the process as well as inefficient translation initiation. It has been reported that CFPS using organisms, such as E. Coli, is limited as translation are approximately "20 folds more dilute" in the cell free system, with distance between ribosomal sequences being much greater than in cellular translation. [17] However, with the presence of additives, intracellular conditions can be mimicked, supporting enzyme activity and the stability of translational machinery.

2.2 Additives

Additives can play a very important part when it is contributed within the CFPS . By adding specific additives, namely PEG 6000, Spermidine, Putrescine, and Choline Chloride, a significant difference can be made in the system's performance. They create a stable environment for translation, help proteins fold the right way, and keep the entire synthesis process running smoothly. For example, PEG 6000, a type of polyethylene glycol, acts like a molecular crowding agent, congesting and helping proteins to fold efficiently within constrained regions, replicating the natural, crowded conditions found inside cells [18]. It additionally forms hydrogen bonds with water molecules, reducing the free water available to dissolve expressed proteins, thus precipitating and crystallising the protein product. [19] On a similar note, Spermidine and Putrescine, which are polyamines, are vital for stabilizing nucleic acids and ribosomal structures, by electrostatically neutralising the phosphate groups on the acids with their protonated amine groups, ensuring that translation can proceed reliably in an environment, which is not overly negatively charged. [20] Choline Chloride, acts similar to the polyamines as it dissociates into ions which then interact with phosphate groups and water ions, affecting overall charge and reducing the solubility and increasing the precipitation of proteins, similar to PEG 6000. [21] By combining these additives, we can create a more supportive environment in the CFPS that can potentially maximise efficiency and protein yield.

Optimizing CFPS for P. pastoris involves systematic experimentation with Kozak sequence variations and additive concentrations. Testing different nucleotide sequences around the start codon can identify the most effective combinations for maximizing translation initiation [22] Additionally, conducting dose-response studies on various additives can reveal optimal concentrations for maintaining reaction stability and protein yield [23].

3 Methodology

3.1 LB Broth Preparation

For the desired bacteria to grow, a Lysogeny Broth (LB Broth) was required as the LB Broth is known as a fast growth medium with rich nutritional value for microbes, including E. Coli [24]. The recommended concentration of LB Broth is 25g/L, as noted by the manufacturers' manual [25], hence 12.5 g of LB Broth was weighed via a measuring scale and 500ml of deionized milli-Q water was placed in 2 large conical flasks for each Kozak. Both the LB Broth and deionized water were then mixed and a sterile foam plug was inserted on top of the flasks, allowing for steam to enter and decontaminate, when sent to autoclave for 12 hours. Autoclave tape was placed onto the flasks to clarify the sterile condition, which was denoted by dark strips forming on the tape strips.

3.2 Cell Culture Growth

Stored E. Coli cultures containing each Kozak were taken from a -80ºC freezer and were placed in a frozen ice box, to maintain its cold temperature. The work surface was wiped with 70% ethanol solution and a Bunsen burner was lit near the surface, to ensure minimal contamination. $5\mu l$ of kanamycin antibiotic solution was put into 2 tubes, using a micropipette. Given that E-Coli is resistant to this antibiotic, the antibiotic eliminated all other bacteria, reducing uncertainties. 50ml of LB Broth was poured into each tube using a pipette controller and following this,

bacteria from the two cultures was transferred from the culture, using an inoculation loop. The lid was then loosely closed, ensuring airflow for optimal conditions for growth, and the covered with the autoclave tape. The tube is then placed in a thermomixer and was mixed for 8 hours at 37°C. Both the AAAAAA Kozak culture (hosting Plasmid 54) and GCCACC Kozak culture (hosting Plasmid 55) were then transferred from the tube into the larger conical flasks for further growth during incubation in the thermomixer for 8 hours again.

3.3 Maxiprep

To ensure successful extraction and preparation of plasmids 54 and 55 for CFPS, the 500 mL mixtures of LB broth containing each plasmid in the conical flasks were initially distributed into 10 separate 50 mL conical tubes each.. Due to evaporation during autoclaving and incubation, the actual volume of the LB broth solutions containing the cultures in each flask after cell growth was slightly less than the initial 500 mL. This resulted in the last 50 mL tube having a lower volume than the others.

To maintain balance within the centrifuge, it was critical to match the mass of this reduced-volume tube with a tube filled with water of equivalent mass. This was achieved by measuring the mass of each tube using a laboratory scale and adjusting accordingly. Each 50 mL tube containing the LB broth and culture mixture with either plasmid 54 or 55 was placed in the centrifuge. A tube containing water was used to balance the centrifuge load, ensuring that the mass matched the tube with the reduced-volume LB broth.

Centrifugation was performed at 6,000 rpm and 4°C for 15 minutes. After centrifugation, the supernatants were carefully drained from each tube, leaving behind two pellets containing the DNA with the two Kozaks. To each pellet, 10 mL of Buffer Solution A was added, and each pellet was mixed using a Vortex Genie-2 mixer to ensure thorough dissolution of the solid debris. This step was repeated for each remaining 50 mL batch corresponding to plasmids 54 and 55. [26]

Following the dissolution of debris, 10 mL of Buffer Solution B was added to each tube. The mixtures were inverted gently to mix and then left to rest at ambient room temperature for 5 minutes. Lastly, 10 mL of Buffer Solution C was added to the tubes, and the mixtures was left undisturbed for 30 minutes to allow for adequate processing.

The maxiprep procedure aims to ensure the effective cell breakdown and preparation of plasmid DNA, isolating the essential components necessary for cell-free protein synthesis (CFPS). The buffers and processing steps facilitate the release and purification of plasmid DNA, preparing it for downstream applications in CFPS, for which optimal yield and quality are essential.

3.4 Master Mix Preparation

Prior to initiating the experiments, an ice box was prepared to store the following common additive solutions: HEPES-KOH, Magnesium Glutamate, NTPs Mix (ATP, CTP, GTP, UTP), Creatine Phosphokinase, RNA-free water and Lysate. An electric heater was set at 70°C to pre-warm the RTS Amino Acid mix which was left for 5 minutes before being inserted into the master mix.

The final concentrations considered for all the components were obtained from a literature review. [27] For all the components apart from the tested additives, Table 3.4.1 determines which concentrations are optimal for testing during the experiment. All solutions were prepared for 15μL to match the volume capacity of each well on the well plate. Equation 3.4.1 and Table 3.4.2 were used to calculate the necessary volumes for the dilution of stock solutions to reduce them to the final concentrations.

Table 3.4.1 Improved concentrations of reaction components in the coupled Pichia Pastoris CFPS System [27]

Component	Improved Concentration
HEPES KOH pH 7.4	60 mM
K-glutamate	0 mM
Mg-glutamate	6 mM
DTT	2 mM
Amino acids mixture	0.75 mM
ATP	1 mM
NTP mixture (CTP/GTP/UTP)	1 mM
Creatine phosphate	45 mM
Creatine phosphokinase	0.27 mg/ml
RNase Inhibitor	200 U/ml
T7 RNA polymerase	100 U

$$V\ [\mu L] = \frac{C_f\ [mM]}{C_s\ [mM]} \times V_T[\mu L] \qquad (3.4.1)$$

Where: V is the volume required, C_f is the final concentration, C_s is the stock concentration and V_T is the total volume of the wells.

Table 3.4.2 The stock concentrations and its desired final concentration of the components within the master mix

Salts and Polyamines		
Component	Stock Concentration	Final Concentration
HEPES-KOH, Ph 7.6	1000.00 mM	60.00 mM
Magnesium Glutamate	1000.00 mM	6.00 mM
DTT	100.00 mM	2.00 mM
RTS Amino Acid Mix	8.40 mM	1.26 mM
NTPs Mix	25.00 mM	1.00 mM
ATP	100.00 mM	1.00 mM
CTP	100.00 mM	1.00 mM
GTP	100.00 mM	1.00 mM
UTP	100.00 mM	1.00 mM
CrP/CrK System		
Component	Stock Concentration	Final Concentration
Creatine Phosphate	500.00 mM	45.00 mM
Creatine Phosphokinase	4.00mM	0.27 mM
Transcription/ Translation		
Component	Stock Concentration	Final Concentration
RNase Inhibitor	40.00 U/μL	0.20 U/μL
T7 RNA polymerase	200.00 U/μL	2.00 U/μL
Template DNA	69.91 nM	10.00 nM

The units for Table 3.4.2 define as follows: milli-Molar [mM], % of weight over volume [%(w/v)], milligram per milliliter [mg/mL], units per microliter [U/µL]

$$V_{Master} = C(N+1)\sum V_{i,j} \qquad (3.4.2)$$

Where: V_{Master} is the master mix volume, C is the number of variables/combinations, N is the number of repeat readings, V_i is the volume of each component j, assumed to for a single testing.

To prepare the master mix, the calculations from Equation 3.4.2, based on the data in Table 3.4.2, was used. The components were added sequentially into a 2 mL tube in the following order: HEPES solution, magnesium glutamate, DTT, ATP, CTP, GTP, UTP, creatine phosphate, creatine phosphokinase, RNase inhibitor, T7 RNA polymerase, and template DNA. Since the master mix is prepared for distribution across all PCR tubes, its total volume will be scaled up significantly, following the principles of Equation 3.4.2.

3.5 Preparation for Kozak Trials

Negative and positive controls were introduced for Kozak testing. The negative control was a sample, which contained only the master mix and lysate, whereas the positive control contained the GAAACG Kozak (Plasmid MGA). The Kozaks being tested were GCCACC (Plasmid 54) and AAAAAA (Plasmid 55).

Four PCR tubes were aligned and labelled for all four variables: Negative, Positive, Plasmid 54, and Plasmid 55. The following solutions were added in chronological order: DNA from its corresponding origin, RNA-free water, master mix, and finally lysate. It was crucial that lysate was added last to prevent early initiation of transcription and translation.

Since three repeated readings were be conducted for each variable, it was ensured that 45 mL of each mixture was prepared for each PCR tube. The master mix was to have a total volume equal to 16 times the standard volume requirement for the mix. During this experiment, three repeated samples were taken (N=3). Note that the reason an additional volume was included, when preparing the repeat samples, was to ensure there was sufficient mixture for all PCR tubes. This avoided the scenario of insufficient mixture.

$$V_{PCR} = \left(\frac{N}{C \times (N+1)}\right) V_{Master} + V_{Lysate} = 45\ \mu L \quad (3.5.1)$$

Where: V_{PCR} is the volume of the PCR tubes, N is the number of repeat readings, C is the number of combinations, V_{Master} is the master mix volume and V_{Lysate} is the volume of the lysate

3.6 RNA Testing

To ensure the validation of the Kozak sequences' effect on translation, an RNA test was conducted. This test consisted of utilising the in vitro transcription kit, through which RNA was extracted from the DNA via buffer cleansing and centrifugation. A concentration check was then performed to validate the presence of RNA, which was subsequently mixed with the master mix, followed by the lysate, and sent to the ClarioStar Plate Reader for testing

3.7 Preparation for Additive Combinatorial Trials

During this experiment, 4 different additives were combined in different permutations, each being a trial for measuring its effect on translational efficiency. Using Figure 3.7.1, the optimal concentrations for each additive were determined and summarized in Table 3.7.1.The concentration was determined by the one which produced the highest fluorescent at its peak.

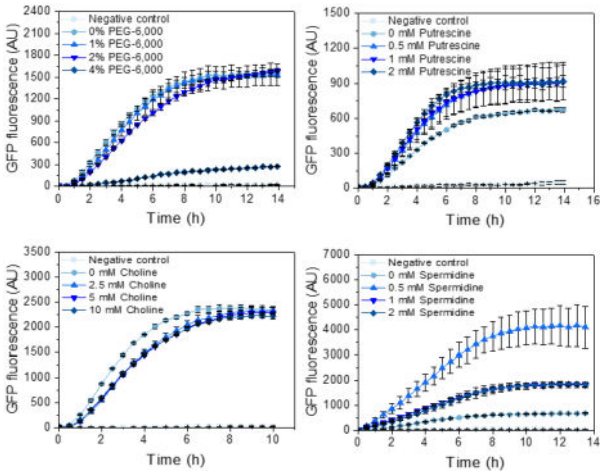


Figure 3.7.1 GFP readings of the of the additives at different concentrations [28]

Table 3.7.1 Summary of the concentrations of additives which will be investigated

Additive	Concentration
PEG 6000	1%w/v
Choline Chloride	2.5mM
Spermidine	0.5mM
Putrescine	2mM

Using these four variables, a total of 17 combinations were produced in order to be tested, including the positive and negative controls. These combinations were prepared in separate PCR tubes, to which the master mix was added, following the same addition procedure as the Kozak testing. It should be noted that the combinations were pipetted first into the PCR tubes, adding the concentrations of each additive associated with the combination. The conversion of stock to desired concentrations followed the same principles as Equation 3.4.1 and was calculated using Table 3.7.1.

3.8 Testing for both combination of additives and Kozaks

Once the best Kozak was determined via the Kozak testing and all the additive combinations were tested, the three best combinations of additives were mixed with the best performing Kozak, to optimise the translational efficiency for cell-free protein synthesis. This final procedure began with each additive combination being pipetted first into the PCR tubes followed by the DNA containing the best Kozak, after which the mandatory components for the testing process were added, namely the master mix and lysate.

3.9 Method of testing

During all trials, the wells were covered with a layer of transparent optical sealing membrane, to prevent the mixtures from contaminating other wells. This was then inserted into the ClarioStar Plate Reader ready to be tested, which measure the green fluorescent protein (GFP) fluorescence caused by the expression of the gfp gene in each DNA. Shown in Figure 3.9.1, is a picture of gfp production within a PCR tube, in which the plate reader will detect the fluorescence. Readings were conducted in intervals of 5 min across a 14-hour period, producing a spreadsheet of results.

4 Results and Discussion

4.1 Method of Analysis

During the analysis of the results of each experiment, all measurements were collated but for each displayed figure, the findings are presented in intervals of 20min for clear visual representation of the behaviours and their error bars. The errors were calculated as the standard deviation between each three repeats for all data points. GFP fluorescence was measured in arbitrary units but is indicative of the general pattern of protein expression as the folding of GFP results in its fluorescence.

4.2 Evaluation of Kozak Sequence Efficiency in In Vitro Transcription

Figure 4.2.1 displays the rates at which GFP fluorescence increased during the trials for different Kozak Sequences. Each protein expression by the different Kozaks reached a steady state within approximately 10 hours from the plate reader's initial readings, suggesting a longer time period than that conveyed in literature. [29] However, it should considered that the temperature at which CFPS was occurring was maintained at ambient temperature of approximately 25°C, which is lower than typical maintained CFPS systems at 30°C or higher. [29]. The lower temperature would have played a role in reducing the kinetic rate and therefore slowing the translational efficiency, achieving optimal fluorescence after a longer duration.

AAAAAA presented the greatest rate at which fluorescence is increased, which may be observed as the greatest protein translation rate. As communicated in existing research, yeast strains appear to have a preference for adenine-rich patterns in the 5' UTR sequence, especially in the -6 to -1 positions from the start codon in mRNA.[15] [16] Figure 4.2.1 exemplifies this behaviour with statistically significant results to denote AAAAAA as an optimal Kozak, in the case of P. pastoris and potentially other eukaryotic systems.

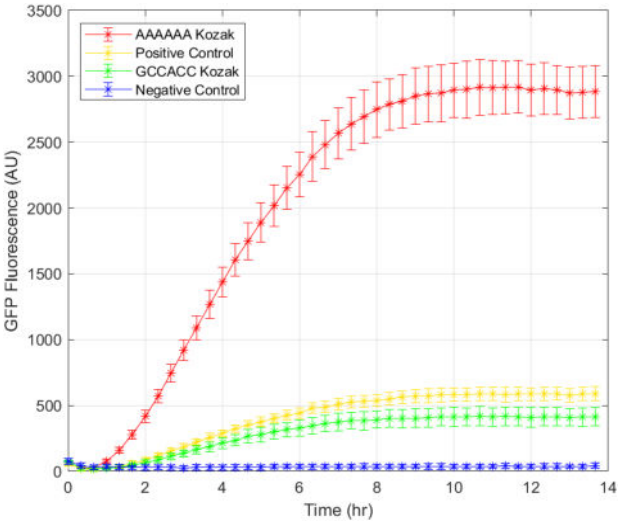


Figure 4.2.1 GFP Fluorescence Over Time for Different Kozak Sequences in a Cell-Free Expression System

Whilst the positive control (GAAACC) performed slightly better than the GCCACC Kozak, repeated trials depicted in Figure 4.2.2 illustrate the opposite, indicating that the rates were extremely similar despite the positive control being more adenine rich. Although it is suggested that eukaryotes prefer adenine rich upstream sequences for protein translation, Kozak's studies express the prevelence of guanines and cytosines in Kozaks across various organisms, with her research encompassing more organisms' DNA. Whilst it is difficult to compare which is definitively superior for

translational efficiency in a cell-free P. pastoris system, their similar activity may be attributed to the significant similarity in bases with only a difference in the second and third bases (-4 and -5 positions).

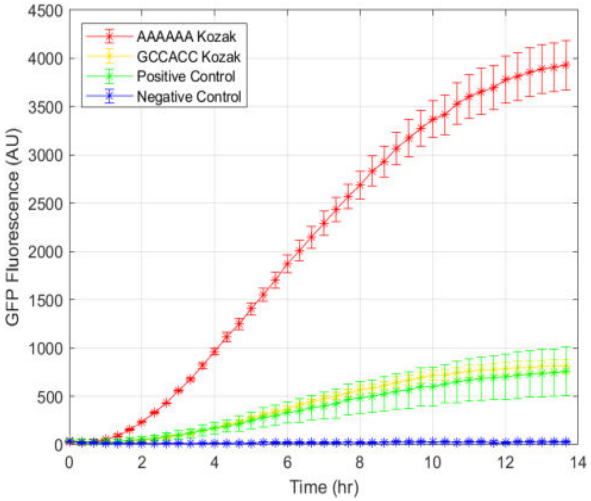


Figure 4.2.2 Analysis of Kozak sequences in which the positive control performed better than GCCACC Kozak

GCCACC is traditionally considered more efficient compared to AAAAAA in Kozak's research but underperformed. AAAAAA may facilitate a more stable pre-initiation complex (PIC) for translation initiation when guiding ribosomes to the start codon and may be tailored to the specific eukaryotic ribosomes' rRNA and ribsomal proteins, which aid in the translation of mRNA. Nevertheless, it should be highlighted that optimsation for organism-specific Kozak sequences would be extremely beneficial, especially in a cell free system.

4.3 Validation of Translational Impact

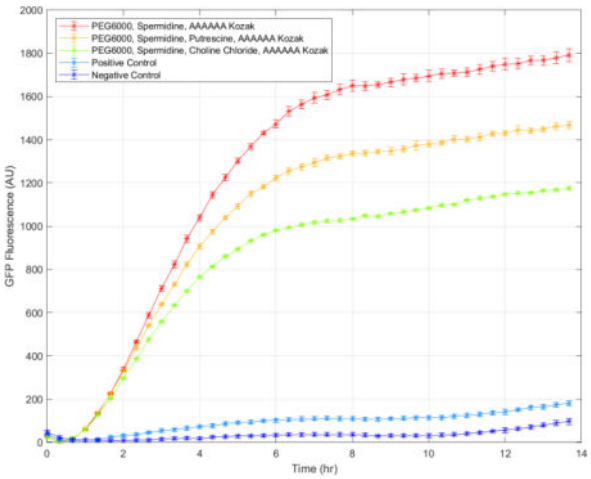


Figure 4.3.1 Fluorescence Over Time for RNA-Driven Translation with Different Kozak Sequences in a Cell-Free System

By testing RNA with the Kozak sequences, a system more akin to mRNA translation is emulated, allowing the investigation of the Kozak sequences directly on translation. Figure 4.3.1 conveys an increase in fluorescence for both the AAAAAA Kozak and GCCACC Kozak with statistical significance, thus validating Kozak's theory and informing the influence of Kozak sequences on translational efficiency. Although there is a sudden increase for all trials at $t = 10$ h, given that the increase is systematically similar for each trial, this could be assessed as a systematic detection error within the plate reader and that the fluorescence should have stagnated to a point of no significant increase with GFP translation reaching steady state

4.4 Evaluation of Additive Combinations for Optimising Cell-Free Protein Synthesis

17 different combinations including the positive and negative controls, which both had no additives, were tested to assess the optimal combinations for translational efficiency. It should be acknowledged that the concentrations for each additive were fixed, determined previously from prior research [28]. Different concentrations for each additive could have potentially had increased synergy with the concentrations of other additives. The tested combinations, therefore, could only be compared to each other and may categorically be suggested to be optimal but cannot be conclusively claimed be the optimal combinations of additive concentrations.

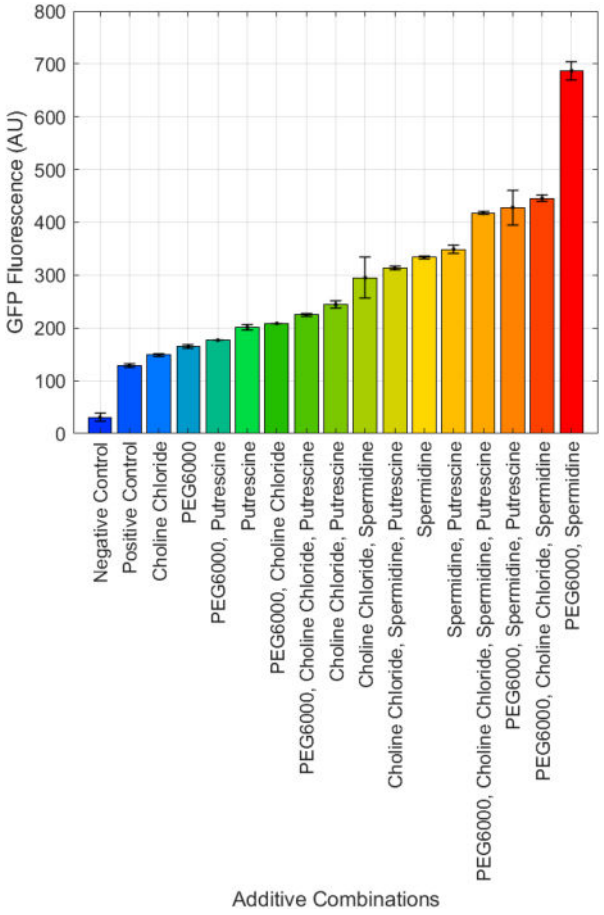


Figure 4.4.1 GFP reading of peak fluorescence within each additive

Figure 4.4.1 portrays the fluorescence achieved at the average time taken for each additive combination to yield its greatest fluorescence optimum, which was at 13h. As depicted, PEG6000 and Spermidine yielded the highest fluorescence and initial rate, illustrated in Figure A1 (Appendix), outperforming combinations, which included Putrescine and Choline Chloride. This can be explained by PEG6000's neutral nature, which ensures it does not interfere with Spermidine's amine groups stabilizing interactions with nucleic acids. PEG6000, therefore, is able to perform molecular crowding without affecting charges, complementing Spermidine's effects whilst improving the translational efficiency.

Spermidine consistently performed well as an additive, featuring in the top 5 combinations, which can attributed to its three positively charged amine groups neutralising negatively charged phosphate groups on nucleic acids. This provides optimally charged conditions for ribosomes to bind to mRNA and translate proteins at higher rates. PEG6000 also featured in the top 4 combinations due to its ability as a crowding agent. However, although PEG6000's hydrophilic nature and crowding effect promotes protein precipitation and folding in confined spaces, the crowding effect may create local pockets of high concentration GFP, which emit high intensity fluorescence in a non-uniform distribution. The plate reader may capture these hotspots rather than acknowledging the homogeneous concentration of protein throughout the solutions in each well. Despite this, it is reasonable to assume that the folded protein in these local pockets contribute significantly to the observed fluorescence and are, therefore, representative of the translational performance.

Although Putrescine and Choline Chloride do not complement Spermidine as well as PEG6000, they are part of the next best combinations. Their lower synergy with Spermidine and PEG6000 may be due to the additional amine groups and ions respectively over neutralising the nucleic acids' charges. Ribosomes prefer some negative charge for effective binding and reading of the mRNA for translation, especially when negatively charged rRNA is required to balance with positively charged ribosomal proteins to maintain the ribosomal activity [30]. Despite the lower synergy, the PEG6000, Choline Chloride and Spermidine and PEG6000, Spermidine and Putrescine combinations were deemed statistically significant both compared to the controls and to Spermidine as a singular additive. They were therefore chosen to be included in the best three combinations to test with the optimal Kozak for the following trials.

The combination with all additives was not chosen, despite being the 4th categorically best combination and having a greater initial rate compared to the combinations with Putrescine or Choline Chloride alongside PEG6000 and Spermidine. Although all additives pose sufficient electrostatic, crowding and precipitation effects for increased translational efficiency, Figure A1 (Appendix) illustrates the rapid reduction in its GFP fluorescence rate, which could be characteristic of exacerbated protein dehydration by the confluence of PEG6000 and Choline Chloride, which would lead to post-translational degradation and instability. Furthermore, the significant amount of amine groups and positive ions from three of the additives may have led to an overly positively charged environment for ribosomes to stay intact and translate the mRNA into proteins, which also would have struggled to fold in a charged environment, thus causing a reduction in translational efficiency. This behaviour in the results informed the decision to exclude the combination of all additives from the optimised testing.

4.5 Evaluation of the optimal Kozak and Additive Combinations

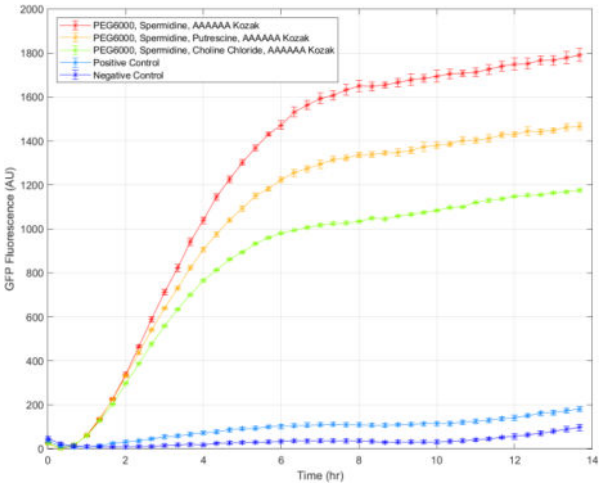


Figure 4.5.1 GFP Fluorescence Over Time for Different Additive Combinations in a Cell-Free System with the AAAAAA Kozak Sequence

Merging the best three additive combinations with the best Kozak, AAAAAA, resulted in a similar pattern to that of the additives' behaviour. However, by comparing the ratio between the results with their respective controls, it was determined that the ratio between the PEG6000 and Spermidine combination with AAAAAA Kozak and the positive control (GAAACC Kozak with no additives) was approximately 1.5 times greater than that between the AAAAAA Kozak with no additives and the positive control. Additionally, the ratio for the optimised testing was approximately 1.7 times greater than the ratio between the PEG6000 and Spermidine combination with the control Kozak to the positive control. The additive combinations, consequently are evaluated to be complementarily enhancing for the optimal Kozak sequence.

It is, however, difficult to make an justifiably sound comparison between these ratios as the protein synthesis is subject to the conditions of each specific trial and should be treated as isolated cases. Whilst conditions remained almost identical for each experiment, errors could have accumulated in preparation, pipetting and precise temperature control. The plate reader, which measures in arbitrary units, may additionally cause erroneous influence on the results as the detection mechanism displays a correction in the initial hour of

measurement in each experiment, displayed in each figure. Optical errors could have also arisen from refraction of light by the additives and seal membrane. Despite this, statistically, the results are significant in each isolated experiment, conveying that the combination of PEG6000 and Spermidine with the AAAAAA Kozak optimises translational efficiency given its high initial rate and achieved fluorescence.

4.6 Statistical Analysis

T-testing was performed for each resultant dataset to observe whether the results differ significantly from the controls. A two-sample test was therefore performed with the acknowledgment that the datasets have unequal variances , meaning the test followed Welch's t-test, which assesses the hypothesis that the datasets have the same mean as the controls.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad (4.6.1)$$

Where:
t *is the test statistic*
$\bar{x}_1, \bar{x}_2$ *are the means of the tested result dataset and the control dataset*

s_p^2 is the pooled variance and is calculated using equation 4.6.2*:*

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \qquad (4.6.2)$$

Where:
s_1^2, s_2^2 *are the variances of the tested result dataset and the control dataset*

The p-values were then determined by applying the test statistics, calculated using Equation 4.6.1,to a t-distribution, with viability to reject the hypothesis being dependent on the p-value being < 0.05.

All t-tests were conducted using MATLAB and compiled in the Appendix/Supplementary File from Table A1 until Table A10.

5 Conclusion

In this study, the AAAAAA Kozak sequence outperformed the GCCACC sequence in terms of translational efficiency. This could also indicate that within eukaryotic cell- free systems, there is a preference for adenine upstream of the start codon, support by literature [31], which suggest adenine at the -3 position provides effective translation. This indication contrasts with Kozak's research, which conveys guanine and cytosine's prevalence upstream of the start codon in natural mRNA. Given that AAAAAA performs better than the natural Kozaks suggested in her research, the results show potential for organism-specific preferences to be implemented in genetic modification of Kozaks for greater translational efficiency than natural cellular Kozaks.

Spermidine and PEG6000 proved to be significantly better in increasing translational efficiency in comparison to other combinations and was shown to complement the AAAAAA Kozak well. The ratio between this combination and the control data denote the potential for additives to be widely adopted in cell free systems. Results also indicated that careful balancing of charges should be considered when applying additives to the system, as their activity, whilst being complementary for the best combination, can interfere with one another.

6 Outlook

Whilst the results obtained display promise for future research and endeavours, it is critical to assess the improvements and alternative pathways to take with optimisation of translational efficiency. For future trials with the same process in mind, a standard curve for the relationship between GFP fluorescence and GFP concentrations would be beneficial in understanding translational reaction kinetics. This would be completed by using known concentrations of GFP and plotting the concentration against its respective fluorescence. With regards to using the same processes, it would clarifying to test additive combinations with GCCACC as the effects of best combinations with GCCACC may differ to that with AAAAAA. Furthermore, additional additives could be introduced to the combinations such as Ficoll 70 and 400, which display promising enhancement to translational efficiency an may further optimise the cell free system. [32] Manipulation of the Kozak sequence would equally be significant to explore, given the extensive research on nucleotides upstream from the start codon. This can be investigated by lengthening the sequence as it is suggested that guanine and cytosine may be optimal for longer Kozaks (Kozak reference). However, given the yeast preference for adenine rich upstream tracts, observing the change of a poly-A Kozak one nucleotide at a time could help optimise for a P. pastoris – specific high translational Kozak sequence.[33]

Optimisation can be applied computationally too with design of experiments. By taking the datasets and applying polynomial curves of best fit, arbitrary multivariable equations relating Kozaks and additive permutations to fluorescence over time can be determined and applied to a machine learning model. By optimisation such as Bayesian, optimal Kozak and additive selection can be achieved as well as evaluating the specific times of fluorescence optima.

To consider industrial application, reactor scale cell free protein synthesis would be beneficial to consider in future studies, as higher volumes may significantly impact interactions between ribosomal components, mRNA and DNA. Molecular crowding and resource availability would also differ, potentially influencing the effectiveness of Kozak sequences. Investigation of the variables at scale would provide insight optimisation of translational efficiency in an industrial setting.

7 Acknowledgement
We would like to thank our supervisors, Prof Karen Polizzi and Rui Wu for their support, guidance and preparation of genetic material, which was used in this investigation.

References

[1] D. Kent, S. Rickwood, and S. Di Biase, *Disruption and Maturity: The Next Phase of Biology*, IqviaCom, p. 24, 2017. [Online]. Available: www.quintilesims.com.

[2] *Evaluate*, "Total global spending on pharmaceutical research and development from 2014 to 2030 (in billion U.S. dollars)," Statista, Statista Inc., 2024. [Online].Available:https://www.statista.com/statistics/309466/global-r-and-d-expenditure-for-pharmaceuticals/.

[3] K. T. Karunakaran, "Recent Progress in Biopharmaceutical Drugs Research and Development," *International Journal of Microbiology and Biotechnology*,vol.2,no.2,p.75,Jan.2017,doi: https://doi.org/10.11648/j.ijmb.20170202.15.

[4] C. H. Chiba, M. C. Knirsch, A. R. Azzoni, A. R. Moreira, and M. A. Stephano, "Cell-free protein synthesis: advances on production process for biopharmaceuticals and immunobiological products," *BioTechniques*, vol. 70, no. 2, pp. 126–133, Feb. 2021, doi: https://doi.org/10.2144/btn-2020-0155

[5] R. Assenberg, P. T. Wan, S. Geisse, and L. M. Mayr, "Advances in recombinant protein expression for use in pharmaceutical research," *Current Opinion in Structural Biology*, vol. 23, no. 3, pp. 393-402,Jun.2013,doi:https://doi.org/10.1016/j.sbi.2013.03.008

[6] J. R. Swartz, "Transforming biochemical engineering with cell-free biology," *AIChE Journal*, vol. 58, no. 1, pp. 5–13, Dec. 2011, doi: https://doi.org/10.1002/aic.13701.

[7] Y. Cui, X. Chen, Z. Wang, and Y. Lu, "Cell-Free PURE System: Evolution and Achievements," *BioDesign Research*, vol. 2022, Jan. 2022, doi: https://doi.org/10.34133/2022/9847014.

[8]X. Jin and S. H. Hong, "Cell-free protein synthesis for producing 'difficult-to-express' proteins," *Biochemical Engineering Journal*, vol. 138, pp. 156–164, Oct. 2018, doi: https://doi.org/10.1016/j.bej.2018.07.013.

[9] J. F. Zawada *et al.*, "Microscale to manufacturing scale-up of cell-free cytokine production-a new approach for shortening protein production development timelines," *Biotechnology and Bioengineering*, vol. 108, no. 7, pp. 1570–1578, Mar. 2011, doi: https://doi.org/10.1002/bit.23103.

[10] M. Karbalaei, S. A. Rezaee, and H. Farsiani, "Pichia pastoris: A highly successful expression system for optimal synthesis of heterologous proteins," *Journal of Cellular Physiology*, vol. 235, no. 9, pp. 5867–5881, Feb. 2020, doi: https://doi.org/10.1002/jcp.29583

[11] J. M. Cregg, T. S. Vedvick, and W. C. Raschke, "Recent Advances in the Expression of Foreign Genes in Pichia pastoris," *Nature Biotechnology*, vol. 11, no. 8, pp. 905–910, Aug. 1993, doi: https://doi.org/10.1038/nbt0893-905.

[12] . P. L. Cereghino, J. L. Cereghino, C. Ilgen, and J. M. Cregg, "Production of recombinant proteins in fermenter cultures of the yeast Pichia pastoris," *Current Opinion in Biotechnology*, vol. 13, no. 4, pp. 329–332, Aug. 2002, doi: https://doi.org/10.1016/s0958-1669(02)00330-0.

[13]C. Ambrosini, F. Garilli, and A. Quattrone, "Reprogramming translation for gene therapy," *Progress in Molecular Biology and Translational Science*, 2021, doi: https://doi.org/10.1016/bs.pmbts.2021.01.028.

[14] M. Kozak, "An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs," *Nucleic Acids Research*, vol. 15, no. 20, pp. 8125–8148, 1987, doi: https://doi.org/10.1093/nar/15.20.8125.

[15] J. Li, Q. Liang, W. Song, and M. A. Marchisio, "Nucleotides upstream of the Kozak sequence strongly influence gene expression in the yeast S. cerevisiae," *Journal of Biological Engineering*, vol. 11, no. 1, Aug. 2017, doi: https://doi.org/10.1186/s13036-017-0068-1.

[16] J. Zeng *et al.*, "Characterization and optimization of 5′ untranslated region containing poly-adenine tracts in Kluyveromyces marxianus using machine-learning model," *Microbial Cell Factories*, vol. 23, no. 1, Jan. 2024, doi: https://doi.org/10.1186/s12934-023-02271-3.

[17] G. Rosenblum and B. S. Cooperman, "Engine out of the Chassis: Cell-Free Protein Synthesis and its Uses," *FEBS letters*, vol. 588, no. 2, pp. 261–268, Jan. 2014, doi: https://doi.org/10.1016/j.febslet.2013.10.016.

[18] I. Kuznetsova, K. Turoverov, and V. Uversky, "What Macromolecular Crowding Can Do to a Protein," *International Journal of Molecular Sciences*, vol. 15, no. 12, pp. 23090–23140, Dec. 2014, doi: https://doi.org/10.3390/ijms151223090.

[19] PubChem, "Concept 8753," *pubchem.ncbi.nlm.nih.gov*. https://pubchem.ncbi.nlm.nih.gov/compound/Polyethylene-Glycol

[20] L. van Dam, “Polyamine-nucleic acid interactions and the effects on structure in oriented DNA fibers,” *Nucleic Acids Research*, vol. 30, no. 2, pp. 419–428, Jan. 2002, doi: https://doi.org/10.1093/nar/30.2.419.

[21] “Registration Dossier - ECHA,” *Europa.eu*, 2023. https://echa.europa.eu/registration-dossier/-/registered-dossier/5360/4/22 (accessed Dec. 12, 2024).

[22] Grégory Boël *et al.*, “Codon influence on protein expression in E. coli correlates with mRNA levels,” *Nature*, vol. 529, no. 7586, pp. 358–363, Jan. 2016, doi: https://doi.org/10.1038/nature16509

[23] J. Zeng *et al.*, “Characterization and optimization of 5′ untranslated region containing poly-adenine tracts in Kluyveromyces marxianus using machine-learning model,” *Microbial Cell Factories*, vol. 23, no. 1, Jan. 2024, doi: https://doi.org/10.1186/s12934-023-02271-3

[24] Broth,” *Journal of Bacteriology*, vol. 189, no. 23, pp.8746–8749, Sep.2007,doi:https://doi.org/10.1128/jb.01368-07.

[25]“VWR®, LB media, pre-mixed,” *VWR*, 2024. https://uk.vwr.com/store/product/8562202/vwr-lb-media-pre-mixed

[26] Qiagen, P. Mini, M. Midi, and Kits, “Quick-StartProtocol Sample & Assay Technologies,” 2011. Available:https://www.qiagen.com/us/resources/donnload.aspx?id=c164c4ce-3d6a-4d18-91c4-f5763b6d4283&lang=en

[27] A. J. Spice, R. Aw, D. G. Bracewell, and K. M. Polizzi, “Improving the reaction mix of a Pichia pastoris cell-free system using a design of experiments approach to minimise experimental effort,” *Synthetic and Systems Biotechnology*, vol. 5, no. 3, pp. 137–144,Sep.2020,doi:https://doi.org/10.1016/j.synbio.2020.06.00

[28] R. Wu, *“Investigating the concentration of additives in cell free protein synthesis” unpublished doctoral research*, 1st ed., vol. 1. Imperial College London, 2024.

[29] Y. 29, T. Wang, X. Chen, and Y. Lu, “IRES-mediated Pichia pastoris cell-free protein synthesis,” *Bioresources and bioprocessing*, vol. 10, no. 1, Jun. 2023, doi: https://doi.org/10.1186/s40643-023-00653-4.

[30] D. V. Fedyukina, T. S. Jennaro, and S. Cavagnero, “Charge Segregation and Low Hydrophobicity Are Key Features of Ribosomal Proteins from Different Organisms,” *Journal of Biological Chemistry*, vol. 289, no. 10, pp. 6740–6750, Mar. 2014, doi: https://doi.org/10.1074/jbc.m113.507707.

[31] “Kozak Consensus Sequence - an overview | ScienceDirect Topics,” *www.sciencedirect.com*. https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/kozak-consensus-sequence

[32] B. R. Fritz, O. K. Jamil, and M. C. Jewett, "Implications of macromolecular crowding and reducing conditions for in vitro ribosome construction," *Nucleic Acids Research*, vol. 43, no. 9, pp. 4774–4784, May 2015, doi: 10.1093/nar/gkv329.

[33] S. D. Petersen, J. Zhang, J. S. Lee, T. Jakociunas, L. M. Grav, H. F. Kildegaard, J. D. Keasling, and M. K. Jensen, "Modular 5'-UTR hexamers for context-independent tuning of protein expression in eukaryotes," *Nucleic Acids Research*, vol. 46, no. 21, p. e127, Nov. 2018, doi: 10.1093/nar/gky734.

Exploration of viscosity trends in ZIF-8–ionic liquid dispersions at varied mass loading for type III porous liquid development

Olusola Deinde, Chelsea Okezie-Akwiwu
Department of Chemical Engineering, Imperial College London, U.K.

Abstract

Type III porous liquids (PLs), consisting of microporous solid framework particles dispersed in pore-excluded liquids, combine the benefits of permanent microporosity with liquid flow properties and kinetics. They show promise for gas separation applications and for industrial pumped processes favouring low viscosity at high mass loadings. However, the relationship between mass loading and viscosity of type III PLs is an area which has not yet been explored. This study aims to address the gap by testing the viscosity of ZIF-8 and silica nanoparticle dispersions, over a 0% to 10% mass loading range, in eight ionic liquids (ILs). Silica-IL dispersions were included as a non-porous control to isolate the effects of porosity and particle-liquid interactions. Dispersion stability testing was performed over 14 days to evaluate the integrity of the dispersions. Viscosity measurements ranged from 50-280 cP and insights into viscosity trends were hypothesised from solid-fluid interactions. Results were dependant on cation functional group with six dispersions exhibiting predicted viscosity increases at low mass loadings with deviations from theory at higher loadings. The remaining two dispersions displayed unconventional viscosity behaviours, with viscosity decreases upon addition of the solid and dual viscosity trends. Based on this, this study details key findings regarding microporous dispersion viscosity trends and type III PLs.

1. Introduction

Research has increasingly been focused towards discovering novel forms of porous media, given the rise in sorbent applications for their industrial versatility, efficiency and sustainability. Conventional liquid sorbents, such as amine solvents used for CO_2 capture, face degradation challenges and have high regeneration costs, while solid sorbents can have poor handling efficiency and slow transfer kinetics. In 2007, O'Reilly et al [1] presented the concept of porous liquids (PLs) which possess permanent microporosity. PLs have the potential to overcome the current sorbent limitations by combining the advantages of porous solids, like selective sorption, and conventional liquids, such as fluidity and enhanced mass transfer and kinetics.

Thus far, four distinct types of porous liquids have been classified. Type I porous liquids are engineered as single-phase liquids systems with inherent cavities within its structure [2]. Type II PL's maintain porosity by using cavity- excluded solvents to dissolve porous organic cages. Type III, in contrast, are homogeneous dispersions of microporous solid frameworks suspended in cavity-excluded liquids. Neat, melted microporous frameworks [3] represent type IV PL's. Literature details the use of solid porous materials such as ZSM-5, ZIF-67 and aluminium fumarate with bulky liquids such as [DBU-PEG][NTf_2] and PDMS for synthesis of PLs [4][5][6].

PL fluidity offers significant advantages for a wide range of industrial applications by overcoming the handling limitations of their solid counterparts. At the forefront of the field, research efforts have focused on the integration of PLs into gas capture, and separation processes. Molecular simulations have been conducted to analyse PL gas adsorption capacities, alongside the associated gas transfer thermodynamics [7]. Further experimental work has demonstrated that PLs can capture not only gaseous but non-gaseous species [8], widening their potential for applications such as gas scrubbing. Thin PL films have been experimentally coated on pipe linings, resulting in faster diffusion kinetics for gaseous processes [9]. Additionally, synthesis of mixed matrix membranes using PLs showed potential for water purification applications, after showing high separation performance [10].

Pumping of PLs has been considered for seamless integration with existing fluid-based processes [11]. A limitation, however, is that high liquid viscosities raise flow resistance, resulting in an increased pressure drop along the pipeline. This elevates energy demands in the system as greater pump power is required to sustain fluid flow. Typical PL viscosities range from 20-140 cP [12], so lower viscosity PLs [13] are desired to promote scale up for continuous flow processes and avoid inflated production costs. While literature supports that increased porous solid loading increases PL gas adsorption capacity [2] [4], little to no information is known about the impact of increasing mass loading on PL viscosity.

Currently, application of type I and II PL's are limited due to their unfavourable characteristics at room temperature and complex preparation methods, respectively [2]. Type IV PL's typically require temperatures of 400°C [13] to reach melting point, hence type III PL's are currently closest to implementation [14] given their ease of synthesis and use at room temperature. The challenge associated with type III porous liquid development lies in establishing a stable dispersion with low viscosity across a

range of mass loadings, thereby increasing the porous capacity available for applications.

The aim of this project is to provide critical insights into the relationship between mass loading, viscosity and dispersion stability in type III PLs for future research, due to the gap in literature. In this study, ZIF-8 was synthesised, characterised and dispersed in 8 ionic liquids: [EMIM][NTf_2],[BMIM][NTf_2],[HMIM][NTf_2], [Epy][NTf_2], [Bpy][NTf_2],[Hpy][NTf_2], [$P_{6,6,6,14}$][NTf_2] and [$P_{1,4,4,4}$][NTf_2]. These dispersions were prepared at mass loadings of 2.5%, 5.0%, 7.5% and 10.0% and used to test the effect of mass loading on PL viscosity.To compare the impact of functional group changes across the ZIF-8 and silica dispersions, the ionic liquid anion [NTf_2]$^-$ was held constant. Analogous ionic dispersions were synthesised using non-porous silica nanoparticles. The silica was selected for comparison due to its lack of both porosity and metal ion centres. This approach aimed to isolate and evaluate the contributions of IL-metal cation interactions and porosity to the observed viscosity trends in ZIF-8-based porous liquids. In removing these variables, the silica IL dispersions provided a baseline for determining whether the viscosity trends associated with ZIF-8 arose from its permanent porosity and metal-centre interactions or from general mass loading effects. Dispersion stability tests were conducted on all PLs and silica-IL dispersions over 14-days to assess their integrity and homogeneity over time. High throughput automation techniques were implemented to accelerate the PL synthesis and viscosity testing processes. Adapting the work of Soh et al. [15], pipetted viscosity tests were run to correlate PL dispense mass data to viscosity.

2. Methodology

2.1. Materials

2.1.1 Materials list

All chemical reagents acquired from commercial suppliers were used as received, without further purification.

2-Methylimidazole (99%) and zinc nitrate hexahydrate (98%) were procured from Thermo Fisher. Methanol (100%) was obtained from VWR.

3-Ethyl-1-methyl-1*H*-imidazolium bis(trifluoromethylsulfonyl)amide ([EMIM][NTf_2]), 1-Butylpyridinium bis(trifluoromethylsulfonyl)imide ([Bpy][NTf_2]), and Tributylmethylphosphonium bis(trifluoromethanesulfonyl)imide ([$P_{1,4,4,4}$][NTf_2]) were purchased from Fluorochem. 1-Butyl-3-methylimidazolium bis(trifluoromethanesulfonyl)imide ([BMIM][NTf_2]) (>98%), 1-Hexyl-3-methylimidazolium bis(trifluoromethanesulfonyl)imide ([HMIM][NTf_2]) (>98%), and Trihexyl(tetradecyl)phosphonium bis(trifluoromethylsulfonyl)amide ([$P_{6,6,6,14}$][NTf_2]) were sourced from Tokyo Chemical Industry. Additionally, 1-Ethylpyridinium bis(trifluoromethylsulfonyl)imide ([EPy][NTf_2]) and 1-Hexylpyridinium bis(trifluoromethylsulfonyl)imide ([HPy][NTf_2]) were obtained from IoLiTec.

Monodispersed, non-porous silica nanoparticles (150 nm) were purchased from Sigma-Aldrich.

2.1.2 Material selection rationale

Zeolitic Imidazole Framework-8 (ZIF-8), a prototypical metal organic framework (MOF) composed of tetrahedrally coordinated zinc atoms linked by 2-methylimidazole ligands [16], was selected for this study due to its permanent porosity and its ability to form stable type III porous liquids at room temperature [4]. Additionally, it has demonstrated compatibility in forming stable dispersions with various ionic liquids [14][17][18].

Nonporous silica nanoparticles (SiO_2) were chosen as a baseline for comparison due to their lack of porosity and metal ion centres. This selection aimed to determine whether the observed trends in ZIF-8 arise from its permanent porosity and metal-centre interactions or from general mass loading effects. Given that the expected particle size of ZIF-8 was approximately 60 nm, a similar size range was considered ideal for comparison. Although not an exact size match, 150 nm nonporous silica was selected as the closest available alternative to the 60 nm ZIF-8 particles. As both sizes fall within the same order of magnitude, the 150 nm particles were considered sufficiently comparable for this study.

ILs were chosen as the continuous phase in the dispersion due to their negligible vapour pressure and pore exclusion capabilities [3]. In this study, the anion of the ILs was kept constant while the cation was varied. Varying the anion would alter the thermal properties of the IL, which falls outside the scope of this work [19]. Conversely, cation size is known to influence viscosity, and variations in cation functional group are likely to result in notable observed trends [20]. The ILs [BMIM][NTf_2], [Bpy][NTf_2] and [$P_{6,6,6,14}$][NTf_2] have previously been previously studied in the literature [4] [18]. This study aims to investigate the relationships between these specific functional groups to better understand how changes in chain length impact porous liquid viscosity.

2.2. Material synthesis and formulation

2.2.1 Synthesis of ZIF-8

ZIF-8 nanocrystals were synthesised based on the methodology provided by Wiebcke et al. [21], and was scaled up to produce 5 g of ZIF-8. A solution of zinc nitrate hexahydrate (16.34 g, 54.9 mmol) in methanol (1.1 L) was rapidly added to a solution of 2-methylimidazole (36.08 g, 439 mmol) in methanol (1.1 L) at room temperature. The mixture was stirred at 250 rpm for one hour at room temper-

ature. To separate the ZIF-8 crystal precipitate from the colloidal dispersion, the mixture was centrifuged for 10 mins at 5100rpm, then washed and centrifuged with fresh methanol once more. The product was left open to dry in a fume cupboard for 48 hours, until the methanol had evaporated. The crystals were ground with a pestle and mortar, and the final ZIF-8 powder was stored in a sealed vial.

2.2.2 Dispersion Formulation

IL-methanol mixtures were formulated in 60 ml tared syringes at a concentration of 0.22 g(IL)/ml(IL+MeOH) for greater compatibility with Chemspeed Flex ISYNTH robot. A 15% excess margin was included to account for pipette aspiration depth.

To calculate the amount of ZIF-8 needed in each sample, enough ZIF-8 was needed for three readings of each mass loading ($wt\%$), which equates to about 1.25 ml for each sample. The required mass of ZIF-8 was calculated using equation (1)

$$m_{\text{ZIF-8}} = 1.15 \cdot \frac{(wt\% \cdot m_{\text{IL}})}{1 - wt\%} \tag{1}$$

Where the mass of IL is shown in equation (2). Densities of ILs (ρ_{IL}) where provided by the commercial suppliers.

$$m_{\text{IL}} = \rho_{\text{IL}} \cdot 1.25\,\text{ml} \tag{2}$$

The required volume of methanol was based on the target concentration (c_{IL}) of 0.22 g(IL)/ml(IL+MeOH)

Volume of the IL-Methanol solution in each stock vial is given by equation (3)

$$V_{\text{stockvial}} = \frac{m_{\text{IL}}}{c_{\text{IL}}} \tag{3}$$

Appendix A details the ZIF-8 masses, IL masses and methanol volumes used in each sample.

To outline how this data was used, [EMIM] [NTf_2] is demonstrated as an example. 8.80 g of [EMIM] [NTf_2] was weighed in the syringe, and the volume aspirated was recorded. Methanol was then aspirated into the syringe to match the target mixture volume of 39.5 ml. The resultant IL-methanol mixture was dispensed into a 40 ml vial and sealed with parafilm to prevent methanol evaporation. This procedure was conducted for each of the eight selected ILs.

The stock vials were labeled A-H to represent each ionic liquid, and 1-4 to denote the dispersion formulation mass loadings of 2.5%, 5.0%, 7.5%, and 10%. Testing was conducted in two batches: A1-D4 and E1-H4. All 40 ml vials were transferred to a 3D printed 11-well plate, and the sixteen 8 ml product vials were arranged in a 4x4 formation on a 24-well plate on the Chemspeed platform. The set up is illustrated in Appendix A. The solid to be dispersed was loaded into the Chemspeed dispense container, and the solvent lines were primed with methanol for rinsing between transfer steps. Dispense mass and volume values for 1.25 ml of product are listed in Appendix A. The GDU-Pfd device gravimetrically transferred the solid mass to the 8 ml product vials, with a dispense tolerance of ± 1 mg. Subsequently, the needle-head volumetrically transferred the IL-methanol mixtures to the same vials. For product vials requiring more than 8 ml, the initial volume dispense was limited to 8 ml, with the remainder added after a small amount of methanol was evaporated off. All vials were vigorously shaken to obtain a homogeneous suspension, and the methanol was evaporated off overnight using an 24-well Equivap low-down evaporator.

2.3. Characterisation and Testing

2.3.1 Solid Characterisation

Thermogravimetric analysis (TGA) was performed on a TA TGA 550 instrument with a gas blending module. It was run under compressed air, at a ramp rate of 10°/min, and a temperature range of 21 to ∼900°C. To verify the crystal structure, X-ray diffraction (XRD) patterns were obtained using an Aeris Benchtop XRD system from Malvern Panalytical. The 2 theta values ranges from 5 to 70°, with a step size of ∼0.01° and ∼0.1 s per step. Dynamic light scattering (DLS) was performed on a Wyatt Mobius DLS. The sample was dispersed in methanol, and the acquisition time was set to 5s, with 5 acquisitions. The particle size was further visualised using a a Zeiss sigma 300 scanning electron microscope (SEM). The electron beam was set to a voltage of 2 kV at a working distance of 2.4 mm, taking images at 100k magnification. This also provided morphological characteristics.

2.3.2 Viscosity Testing

The viscosity testing of dispersions was performed using the Opentrons OT-2 robot, using the P1000 single-channel GEN2 pipette head and 1000 µl tips. The experimental setup is illustrated in Figure Appendix B. After 10 min of sonication, 8 ml dispersion vials, and pre-weighed 2 ml integrated micro-insert vials were loaded onto the platform in a 24-well plate and 48-well plate, respectively. Python scripts, detailed in Appendix B, directed the pipetting process.

The Opentrons aspirated 450 µl of dispersion from each 8 ml source vial at a rate of 25 µl/s and dispensed 300 µl into the corresponding 2 ml vial at 50 µl/s. Excess dispersion was returned to source vial, and the tip was discarded. The mass of each dispensed dispersion was recorded, and the procedure was repeated three times per batch.

A viscosity calibration curve was generated, using the same method, with liquids of known viscosity, including 500 cSt silicone oil, 1000 cSt silicone oil and ionic liquids [EMIM][NTf_2],[BMIM][NTf_2],[HMIM][NTf_2],

[Epy][NTf_2], [Bpy][NTf_2], [Hpy][NTf_2], [P6,6,6,14][NTf_2] and [P1,4,4,4][NTf_2]. The calibration curve is shown in Figure (2.3.1). The resulting exponential fit equation was used to calculate the viscosity values of the ionic liquid dispersions.

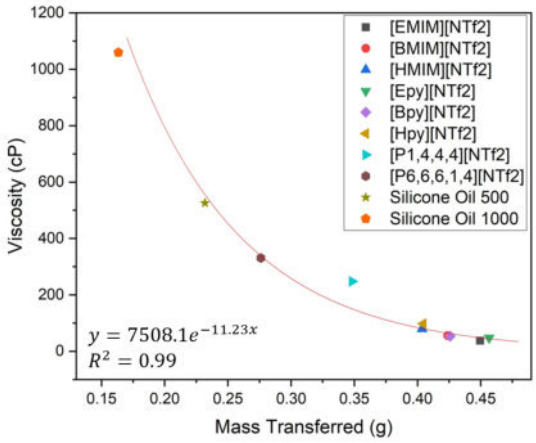


Figure 2.3.1: Calibration curve of neat ionic liquids with exponential fit curve equation

2.3.3 Dispersion Stability Monitoring

Following the viscosity testing in the Opentrons, the dispense vials were placed in custom 3D-printed monitoring racks and monitored for 14 days using a Raspberrry Pi Zero equipped with a with Pi Camera v2. This set up is shown in Appendix C.

2.4. Theoretical Analysis

2.4.1 Batchelor Equation

Theoretical trends for viscosity with mass loading were predicted using the Batchelor equation [22], shown in equation (4), where η is the dynamic viscosity of the dispersion, η_s is the dynamic viscosity of the continuous phase, and ϕ is the volume fraction. The Batchelor equation assumes particles are Brownian spheres in medium-concentration suspensions($\phi < 0.15$) and is suited to monodisperse systems.

$$\frac{\eta}{\eta_s} = 1 + 2.5\phi + 6.2\phi^2 \tag{4}$$

This equation predicts an increase in dynamic viscosity of a dispersion with an increase in the volume fraction of solid. It predicts a non-linear, quadratic increase in viscosity with volume fraction. The volume fraction was calculated using equation (5).The volume fraction and mass loadings are percentages, and density has units of kg/m^3.

$$\phi = \frac{\text{wt\%}}{\text{wt\%} + (100 - \text{wt\%})\left(\frac{P_{\text{solid}}}{P_{\text{liquid}}}\right)} \tag{5}$$

The volume fraction and mass loadings ($wt\%$) are percentages, and density has units of kg/m^3.

3. Results and Discussion

3.1. Material Characterisation

Following ZIF-8 synthesis, the material was characterised using X-ray Diffraction (XRD), Thermogravimetric Analysis (TGA), Dynamic Light Scattering (DLS) and Scanning Electron Microscopy (SEM). The results of the ZIF-8 characterisation are in Figures 3.1.1 (a)-(d)

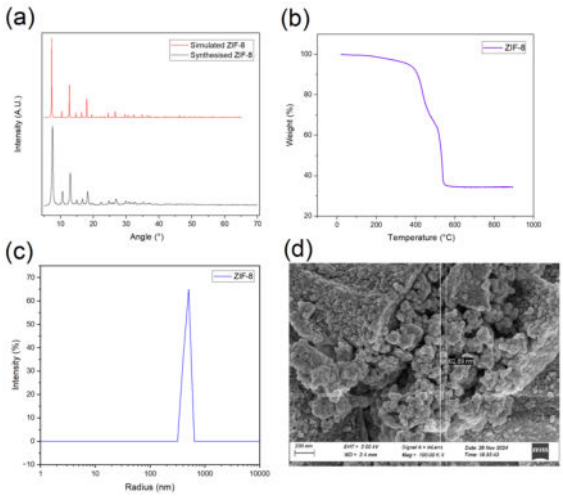


Figure 3.1.1: (a) X-ray Diffraction pattern of synthesised ZIF-8 compared to simulated data of known structure (b) Thermogravimetric Analysis of synthesised ZIF-8 sample (c) Dynamic Light Scattering reading of ZIF-8 dispersed in methanol (d) Scanning Electron Microscopy image of synthesised ZIF-8 at 100 magnification

3.1.1 X-ray Diffraction (XRD) of ZIF-8

The XRD pattern of the synthesised ZIF-8 (Figure 3.1.1(a)) showed strong diffraction peaks at 7°, 10°, 12° and 18°, which matches very closely to the simulated data for the known ZIF-8 structure. This suggests that a high-purity crystal structure of the correct compound has been made.

3.1.2 Thermogravimetric Analysis (TGA)

From (Figure 3.1.1(b)) , the TGA closely aligns with literature [23]. Both literature and experimental data show a 'hump' on the ZIF-8 graph, which may be attributed to the thermal decomposition competing with oxidation and forming intermediates [24]. Both literature and experimental data plateau at 36% which is attributed to residual zinc oxide after thermal decomposition. The pattern shows a drop in weight from ∼450°C, indicating strong thermal stability.

3.1.3 Particle Size Analysis

From (Figure 3.1.1(c)), the Dynamic Light Scattering (DLS) shows a modal particle size of 500nm. This

is unexpected, as the anticipated particle size was approximately 60nm. After filtration of the sample using a 250 nm filter, the absence of particles below this size range supported the measurement.

The SEM in (Figure 3.1.1(d)) shows individual particles sized ∼60nm, but the lack of definition between particles indicated aggregation had occurred. It is likely that these aggregated particles were detected by the DLS, leading to the large particle size reading. Literature explains that DLS provides a distribution based on the scattering intensity of particles present in the entire system, regardless of agglomerates [25]. This means its readings can be unreliable for smaller nanoparticles, because DLS favours larger particle sizes[26] . One that conclude that without proper particle definition, small nanoparticles will not be detected properly by DLS.

3.2. ZIF-8-IL viscosity trends

3.2.1 ZIF-8-Imidazolium IL viscosity trends

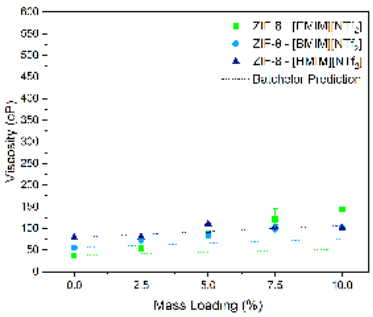


Figure 3.2.1: Viscosity (cP) against mass loading(%) trend for ZIF-8-[EMIM][NTf_2], ZIF-8-[BMIM][NTf_2] and ZIF-8-[HMIM][NTf_2] dispersions with comparison to theoretical predictions derived from the Batchelor equation

Figure (3.2.1) examines the relationship between mass loading of ZIF-8 and viscosity for ZIF-8-[EMIM][NTf_2] dispersions, ZIF-8-[BMIM][NTf_2] dispersions, and ZIF-8-[HMIM][NTf_2] dispersions. Generally, viscosity increases with mass loading, which isconsistent with prediction by the Batchelor equation. While ZIF-8-[EMIM][NTf_2] dispersions and ZIF-8-[BMIM][NTf_2] dispersions always increase over the range, ZIF-8-[HMIM][NTf_2] dispersions shows an unexpected 10% decrease from 111 cP to 101 cP between 5% and 10% loading.The decline in the viscosity may be due to zinc ion aggregation on the ZIF-8 surface interrupting the solid-liquid intermolecular interactions [27]. The ZIF-8-[EMIM][NTf_2] and ZIF-8-[BMIM][NTf_2] dispersion deviations from their Batchelor predictions could suggest the presence of attractive Coulomb forces that enhance the viscosity of the system [28].

At lower mass loadings (0-2.5%), viscosity consistently increases with IL cation chain length ($[EMIM]^+ < [BMIM]^+ < [HMIM]^+$), consistent with theoretical expectations. However, beyond 2.5%, the trend reverses, with ZIF-8-[EMIM][NTf_2] showing the steepest viscosity increase, surpassing both ZIF-8-[BMIM][NTf_2] and ZIF-8-[HMIM][NTf_2]. The reversal may be a result of chain length altering the dispersion stability at higher loadings [29]. However, literature indicates that longer alkyl chains on the cation increase viscosity on imidazolium-based ionic liquids [29] [30]. It is worth investigating further to verify the relationship observed in ZIF-8-[HMIM][NTf_2].

3.2.2 ZIF-8-Pyridinium IL viscosity trends

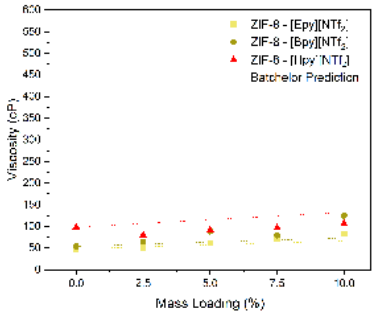


Figure 3.2.2: Viscosity (cP) against mass loading(%) trend for ZIF-8-[Epy][NTf_2], ZIF-8-[Bpy][NTf_2] and ZIF-8-[Hpy][NTf_2] dispersions with comparison to theoretical predictions derived from the Batchelor equation

In Figure (3.2.2), while viscosity generally increases with mass loading, as seen with the imidazolium-based IL dispersions, exceptions are observed. In the ZIF-8-[Hpy][NTf_2] system, viscosity decreases from 98 cP at 0% to 80 cP at 2.5%, and a decrease is also observed in ZIF-8-[Bpy][NTf_2]from 88 cP at 5% to 79 cP at 7.5%. The consistent increase for ZIF-8-[Epy][NTf_2] aligns with theoretical expectations in the Batchelor equation, but the non-quadratic increase once again indicates the presence of intermolecular interactions. The declines for [Bpy][NTf_2] and [Hpy][NTf_2] suggest disruption in ZIF-8-IL interactions, which are likely attributed to aggregation. These findings suggest ZIF-8-[Epy][NTf_2] is the most reliable candidate for applications requiring steady viscosity control, but from laboratory experience, handling neat [Epy][NTf_2] is temperamental, as it solidifies very easily when in contact with moisture. However, this is not seen when it is in dispersion. If [Epy][NTf_2] were to be used in industry, robust stabilization treatments would be required.

Across the pyridinium-based ILs, as chain length increase, viscosity also increases generally. Across fixed mass loadings, the ($[EMIM]^+ < [BMIM]^+ < [HMIM]^+$) trend is maintained up until 10% mass loading. *Brennecke et al.* [31] observed an

increase in viscosity with nitrogen atom alkyl chain length of the cation results in an increase in viscosity. This aligns with the alkyl chain length trend in Imidazolium based ionic liquids. This is expected as Pyridiniums and Imidazoliums are both heterocyclic compounds with Nitrogen atom(s) present in rings. This similarity in structure, and the similar viscosity of the neat ILs means there is an expectation for their trends and viscosity values to be similar, which was seen, despite the irregularities.

3.2.3 ZIF-8-Phosphonium IL viscosity trends

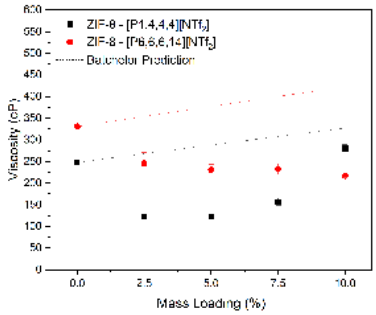


Figure 3.2.3: Viscosity (cP) against mass loading(%) trend for ZIF-8-$[P_{1,4,4,4}][NTf_2]$, ZIF-8-$[P_{6,6,6,14}][NTf_2]$ dispersions with comparison to theoretical predictions derived from the Batchelor equation

Deviating from expectations, the ZIF-8-phosphonium dispersions both demonstrated a sharp decrease in viscosity upon introduction to the ZIF-8 solid (Figure (3.2.3)). However, as mass loading further increased, the dispersions showed contrasting trends. While the ZIF-8-$[P_{1,4,4,4}][NTf_2]$ dispersion viscosity steeply increased 120% from 2.5% to 10% loading, the ZIF-8-$[P_{6,6,6,14}][NTf_2]$ dispersion showed a shallow viscosity decrease of 10%.

A study by Servantie et al [32] modelled shear thinning and thickening effects in nanoparticles dispersed in a Lennard-Jones fluid by evaluating the viscosity changes as a function of shear rate and nanoparticle volume fraction. The results showed that at low volume fractions (ϕ= 0.08, 0.16, 0.25), dispersion viscosity decreased with shear rate and volume fraction due to the nanoparticles arranging themselves in a layered microstructure longitudinal to the flow direction. This arrangement causes the nanoparticles to follow flow paths that avoid collisions, reducing the system resistance to flow, and reducing the dispersion viscosity. Shear thinning could begin to explain the viscosity drop at 2.5% mass loading ($\phi = 0.03$) for the phosphonium-IL dispersions. As proposed by *Hoffman et al.*[33], the order-disorder transition was explored, which could be a basis explanation for the dual viscosity trend observed in ZIF-8-$[P_{1,4,4,4}][NTf_2]$ as mass loading increased. The detailed transition occurred from the ordered shear thinning regime to disordered shear thickening regime, resulting in increased collisions, frictional interactions and hence increased viscosity as volume fraction increased. The literature showed that increased dispersion volume fraction caused the critical shear rate to decrease, transitioning to the thickening regime at lower shear rates. There is potential that the applied shear rate by the Opentrons may have exceeded the critical value for 5.0% ZIF-8-$[P_{1,4,4,4}][NTf_2]$ mass loading and shifted the regime to shear thickening, causing the viscosity to further increase with mass loading.

The switch in viscosity regimes was not observed for ZIF-8-$[P_{6,6,6,14}][NTf_2]$ as the viscosity trend continued decreasing as mass loading increased. *Servantie et al* [32] commented on the potential for shear-thickening behaviour to be suppressed by strong particle-particle interactions in the dispersion causing shear stress to increase. As backed by Brown et al [34]and Singh et al [35], this can narrow or eliminate the range of shear rates at which shear thickening can occur and maintain the thinning regime. Alternatively, if the ZIF-8 particles are sufficiently stabilized, the system may resist forming the disordered structure required for shear thickening and maintain the shear thinning behaviour. The $[P_{6,6,6,14}]^+$ cation may provide steric stabilisation [36] from the bulky 14 carbon alkyl chain, preventing aggregation and viscosity increase.

A limitation of comparing the Servantie et al model results to the tested dispersion results is that ILs cannot be classified as Lennard-Jones liquids as their interactions are more complex. Also, given that a drop in viscosity was not observed across the other IL functional groups, the wider solution is likely due to specific intermolecular forces between the phosphonium ILs that are not widely detailed in the literature.

3.3. SiO_2-IL viscosity trends

The results of the silica dispersion viscosity testing failed to meet their initial purpose as a neutral comparison to the ZIF-8 dispersion data. The first batch of testing yielded the SiO_2-[EMIM][NTf_2], SiO_2-[BMIM][NTf_2], SiO_2-[HMIM][NTf_2] and SiO_2-[Bpy][NTf_2] data. From Figures 3.3.1 and 3.3.2it is evident that the silica synthesis did not produce a stable dispersion as two layers can be seen.

Identical methodology to the ZIF-8 dispersion synthesis was followed but even after 10 mins of sonication, the layers quickly separated out. It was proposed that the aspirated layer had negligible amounts of silica dispersed and predominantly contained neat IL . This was supported by the near-constant IL viscosity values as mass loading increased (See figures (3.3.1),(3.3.2)). The SiO_2-[HMIM][NTf_2] and SiO_2-[EMIM][NTf_2] 10% loading viscosity tests were repeated immediately after 10

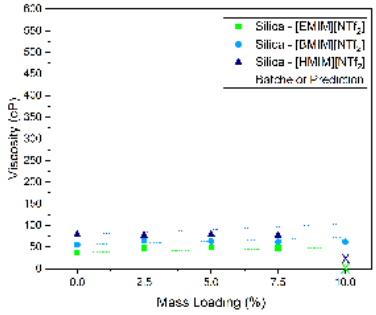


Figure 3.3.1: Viscosity (cP) against mass loading(%) trend for SiO_2-[EMIM][NTf_2] , SiO_2-[BMIM][NTf_2], SiO_2-[HMIM][NTf_2] dispersions with comparison to theoretical predictions derived from the Batchelor equation

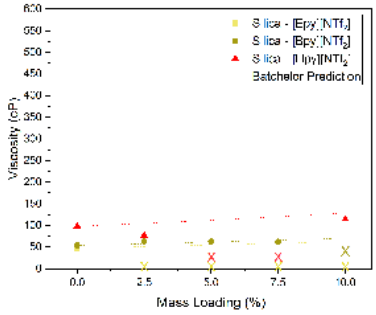


Figure 3.3.2: Viscosity (cP) against mass loading(%) trend for SiO_2-[Epy][NTf_2] , SiO_2-[Bpy][NTf_2], SiO_2-[Hpy][NTf_2] dispersions with comparison to theoretical predictions derived from the Batchelor equation

mins of sonication. However, the pipette failed to smoothly dispense the dispersions due to solid settling blocking the exit and were recorded as failed tests. To reduce settling time, the subsequent tested source vials were sonicated and immediately tested individually. This was to check whether the previous failed testings were solely due to poor sonication or due to the silica dispersion characteristics being incompatible with the testing methodology. SiO_2-[Hpy][NTf_2] was first tested but yielded sparse results. It was hypothesised that the most viscous IL SiO_2-[$P_{6,6,6,14}$][NTf_2] would be the most likely to maintain a stable silica dispersion but the tests similarly failed. This supported the hypothesis that the silica dispersion had poor compatibility with the methodology and no viscosity testing for SiO_2-[Epy][NTf_2] and SiO_2-[$P_{1,4,4,4}$][NTf_2] was carried out. This nullified the intended comparison between porous and non-porous solid dispersion effects on viscosity.

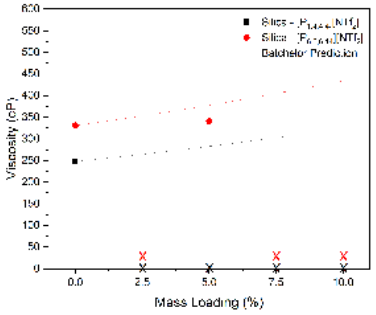


Figure 3.3.3: Viscosity (cP) against mass loading(%) trend for SiO_2-[$P_{6,6,6,14}$][NTf_2],SiO_2-[$P_{1,4,4,4}$][NTf_2] dispersions with comparison to theoretical predictions derived from the Batchelor equation

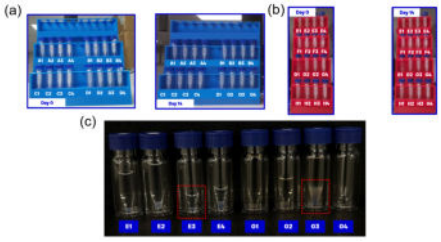


Figure 3.4.1: Dispersion testing results over two weeks. Blue: ZIF-8 Vials A) ZIF-8-[EMIM][NTf_2], B) ZIF-8-[BMIM][NTf_2], C) ZIF-8-[HMIM][NTf_2], D) ZIF-8-[Bpy][NTf_2] Red:E) ZIF-8-[Hpy][NTf_2], F) ZIF-8- [$P_{6,6,6,14}$][NTf_2] and G) ZIF-8-[$P_{1,4,4,4}$][NTf_2], H) ZIF-8-[Epy][NTf_2] 1-4 represents 2.5-10% wt%

3.4. Dispersion stability testing

Figure 3.4.1(a)-(b) shows that over the 14 day dispersion stability testing period, every ZIF-8-IL dispersion remained homogeneous. As mentioned above, the SiO2-IL dispersions were unable to maintain homogeneity 3.4.1(c). The particle size of the silica nanoparticles (150nm) is noteworthy and may have played a role in the instability of the silica dispersions as literature details examples of smaller 7nm [37] silica nanoparticle dispersions. Given the neutrality of SiO_2, increased particle-fluid interactions through hydrogen bonding or other coulomb forces would not be observed. There is potential that, at a larger particle size, the particle-fluid interactions were insufficient to overcome gravitational forces causing solid settling.

3.5. Repeatability of results

For all the dispersions, three repeat tests were carried out. These mass dispense measurements were averaged before being substituted into the calibration curve equation to increase reliability. The first round of viscosity testing was consistently successful. However, as the dispersion volume level decreased with repeats, the Opentrons began to inconsistently aspi-

rate the expected 450μl of dispersion. In those cases, the volume would be manually pipetted by the operator and the loaded pipette tip would be placed back into the Opentrons to carry out the dispense step. For failed repeats, when all the aspirated liquid was dispensed, the dispense mass data for that specific repeat was removed from the average. This means that certain data points represent only one or two successful measurements in place of three. In most cases, an average of three data points were taken. This data is presented in Appendix C.

4. Conclusion

The initial aims of this study regarding the synthesis, viscosity testing and trend exploration of ZIF-8-IL dispersions at varied mass loadings were successfully achieved.

The results showed an increase in viscosity with mass loading for ZIF-8 dispersions in the selected imidazolium and pyridinium-based ionic liquids. Deviations from theoretical expectations could be due to solid surface aggregation disrupting particle-fluid intermolecular forces. The phosphonium-based ZIF-8-IL dispersions showed conflicting trends where the ZIF-8-[$P_{6,6,6,14}$][NTf_2] experienced a viscosity decrease with mass loading while ZIF-8-[$P_{1,4,4,4}$][NTf_2] experienced a trend transition, at 5% loading , from an initial viscosity decrease to increase. These were attributed to the impact of particle-liquid interactions on the shear thinning and thickening regimes of the dispersions. Further investigation is necessary to support these hypotheses. The ZIF-8-IL dispersions all remained stable over a 14-day observation, but the silica-IL dispersions were unable to maintain homogeneity so could not act as the neutral trend comparison as intended.

Overall, this research provides a useful basis for understanding how mass loading impacts the viscosity and stability of ZIF-8-IL dispersions and begins to fill the gap in literature about the relationship with respect to type III porous liquid applications.

5. Outlook

To compare the findings in this paper, the non-porous dispersion tests could be repeated using a smaller SiO_2 particle size. This would provide further insight into the impact of porosity towards the deviations seen in the ZIF-8-IL viscosity trends. The data would also remain relevant as a neutral comparison for future porous solid dispersion analysis. The observed results of this paper could be further supported by investigating alternative microporous solid frameworks, such as CC3 or ZIF-67, and comparing their dispersion viscosity trends under identical conditions. More complex prediction models, such as the prolate ellipsoid model, take particle-particle interactions into account and could provide more detailed theoretical representations of the viscosity-mass loading trends. Investigations into the effect of sonication technique on dispersion viscosity could also provide useful insight into the extent of its impact on the results observed. CO_2 uptake measurements could be performed on the dispersions detailed in this report to quantify the impact of mass loading on their gas adsorption capacity. This data is critical to confirm the tested dispersions as type III porous liquids and recommend them for further research. Low material volume had poor compatibility with the high-throughput automation techniques. Hence, a constant excess of material would be required for successful scale-up of this methodology. The economic implications of this could be further investigated as type III porous liquids become closer to adoption in industry.

Acknowledgements

Thank you to Omar Al-Miqdadi, the Petit and Greenway research groups.

References

[1] N. O'Reilly, N. Giri, and S. L. James, "Porous liquids," *Chemistry - A European Journal*, vol. 13, no. 11, pp. 3020–3025, 2007.

[2] X. Li, D. Wang, Z. He, F. Su, N. Zhang, Y. Xin, H. Wang, X. Tian, Y. Zheng, D. Yao, and M. Li, "Zeolitic imidazolate frameworks-based porous liquids with low viscosity for co2 and toluene uptakes," *Chemical Engineering Journal*, vol. 417, p. 129239, 2021.

[3] B. D. Egleston, A. Mroz, K. E. Jelfs, and R. L. Greenaway, "Porous liquids Äì the future is looking emptier," *Chem. Sci.*, vol. 13, pp. 5042–5054, 2022.

[4] W. Shan, P. F. Fulvio, L. Kong, J. A. Schott, C.-L. Do-Thanh, T. Tian, X. Hu, S. M. Mahurin, H. Xing, and S. Dai, "New class of type iii porous liquids: A promising platform for rational adjustment of gas sorption behavior," *ACS Applied Materials & Interfaces*, vol. 10, no. 1, pp. 32–36, 2018.

[5] Y. Zhang, X. Hu, S. M. Mahurin, and S. Dai, "Type iii porous liquids: A promising platform for gas separation," *ACS Applied Materials & Interfaces*, vol. 16, no. 5, pp. 1234–1240, 2024.

[6] K. Zou, S. Xie, M. Jiang, P. Wang, T. Ning, T. Lu, L. Tan, H. Li, Y. Zhou, W. Wang, and L. Li, "Insights into the precursor specific surface area for engineering co-free ni-rich cathodes with tailorable properties," *Chemical Engineering Journal*, vol. 483, p. 149189, 2024.

[7] F. Zhang, F. Yang, J. Huang, B. G. Sumpter, and R. Qiao, "Thermodynamics and kinetics of gas storage in porous liquids," *The Journal of Physical Chemistry B*, vol. 120, no. 29, pp. 7195–7200, 2016. PMID: 27379463.

[8] L. Ma, C. J. E. Haynes, A. B. Grommet, A. Walczak, C. C. Parkins, C. M. Doherty, L. Longley, A. Tron, A. R. Stefankiewicz, T. D. Bennett, and J. R. Nitschke, "Coordination cages as permanently porous ionic liquids," *Nature Chemistry*, vol. 12, no. 3, pp. 270–275, 2020.

[9] R. Kumar, P. Dhasaiyan, P. M. Naveenkumar, and K. P. Sharma, "A solvent-free porous liquid comprising hollow nanorodÄìpolymer surfactant conjugates," *Nanoscale Adv.*, vol. 1, pp. 4067–4075, 2019.

[10] A. Knebel, A. Bavykina, S. J. Datta, L. Sundermann, L. Garzon-Tovar, Y. Lebedev, S. Durini, R. Ahmad, S. M. Kozlov, G. Shterk, M. Karunakaran, I. D. Carja, D. Simic, I. Weilert, M. Klüppel, U. Giese, L. Cavallo, M. Rueping, M. Eddaoudi, J. Caro, and J. Gascon, "Solution processable metal–organic frameworks for mixed matrix membranes using porous liquids," *Nature Materials*, vol. 19, no. 12, pp. 1346–1353, 2020.

[11] R. J. Kearsey, B. Alston, M. E. Briggs, R. L. Greenaway, and A. I. Cooper, "Accelerated robotic discovery of type ii porous liquidsÄ†Ä†electronic supplementary information (esi) available: Detailed synthetic procedures, experimental details and measurements (pdf). see doi: 10.1039/c9sc03316e.," *Chemical Science*, vol. 10, no. 41, pp. 9454–9465, 2019.

[12] Y. Zhang, X. Hu, S. M. Mahurin, and S. Dai, "Transforming porous silica nanoparticles into porous liquids with permanent porosity," *ACS Omega*, vol. 7, no. 7, pp. 5687–5697, 2022.

[13] T. D. Bennett, F.-X. Coudert, S. L. James, and A. I. Cooper, "The changing state of porous materials," *Nature Materials*, vol. 20, no. 9, pp. 1179–1187, 2021.

[14] H. Mahdavi, S. J. D. Smith, X. Mulet, and M. R. Hill, "Practical considerations in the design and use of porous liquids," *Mater. Horiz.*, vol. 9, pp. 1577–1601, 2022.

[15] B. W. Soh, A. Chitre, W. Y. Lee, D. Bash, J. N. Kumar, and K. Hippalgaonkar, "Automated pipetting robot for proxy high-throughput viscometry of newtonian fluids," *Digital Discovery*, vol. 2, pp. 481–488, 2023.

[16] A. Phan, C. J. Doonan, F. J. Uribe-Romo, C. B. Knobler, M. OÄôKeeffe, and O. M. Yaghi, "Synthesis, structure, and carbon dioxide capture properties of zeolitic imidazolate frameworks," *Accounts of Chemical Research*, vol. 43, no. 1, pp. 58–67, 2010. PMID: 19877580.

[17] M. Costa Gomes, L. Pison, C. Červinka, and A. Padua, "Porous ionic liquids or liquid metal-organic frameworks?," *Angew Chem Int Ed Engl*, vol. 57, pp. 11909–11912, 2018.

[18] S. Liu, J. Liu, X. Hou, T. Xu, J. Tong, J. Zhang, B. Ye, and B. Liu, "Porous liquid: A stable zif-8 colloid in ionic liquid with permanent porosity," *Langmuir*, vol. 34, pp. 3654–3660, 2018.

[19] T. Sato, G. Masuda, and K. Takagi, "Electrochemical properties of novel ionic liquids for electric double layer capacitor applications," *Electrochimica Acta*, vol. 49, no. 21, pp. 3603–3611, 2004.

[20] Z.-B. Zhou, H. Matsumoto, and K. Tatsumi, "A new class of hydrophobic ionic liquids: Trialkyl(2-methoxyethyl)ammonium perfluoroethyltrifluoroborate," *Chemistry Letters*, vol. 33, pp. 886–887, 06 2004.

[21] J. Cravillon, S. Munzer, S.-J. Lohmeier, A. Feldhoff, K. Huber, and M. Wiebcke, "Rapid room-temperature synthesis and characterization of nanocrystals of a prototypical zeolitic imidazolate framework," *Chemistry of Materials - CHEM MATER*, vol. 21, 04 2009.

[22] G. K. Batchelor, "The effect of brownian motion on the bulk stress in a suspension of spherical particles," *Journal of Fluid Mechanics*, vol. 83, no. 1, p. 97–117, 1977.

[23] J. Cravillon, S. Münzer, S.-J. Lohmeier, A. Feldhoff, K. Huber, and M. Wiebcke, "Rapid room-temperature synthesis and characterization of nanocrystals of a prototypical zeolitic imidazolate framework," *Chemistry of Materials*, vol. 21, no. 8, pp. 1410–1412, 2009.

[24] H. Bux, A. Feldhoff, J. Cravillon, M. Wiebcke, and J. Caro, "Kinetics of zif-8 thermal decomposition in inert, oxidizing, and reducing environments," *The Journal of Physical Chemistry C*, vol. 120, no. 15, pp. 8112–8121, 2016.

[25] A. Wishard and B. C. Gibb, "Dynamic light scattering - an all-purpose guide for the supramolecular chemist," *Supramolecular Chemistry*, vol. 31, no. 9, pp. 608–615, 2019. Epub 2019 Jun 20.

[26] N. Giri, M. G. D. Pópolo, G. Melaugh, R. L. Greenaway, K. Rätzke, M. J., Ben, S. L. James, and S. L. James, "Porous liquids: The next frontier in gas separation," *Nature*, vol. 527, no. 7578, pp. 216–220, 2015.

[27] H. Li, W. Shan, L. Kong, X. Hu, S. M. Mahurin, H. Xing, and S. Dai, "Porous liquid: A stable zif-8 colloid in ionic liquid with permanent porosity," *Langmuir*, vol. 34, no. 15, pp. 4727–4731, 2018.

[28] K. Liu, Z. Wang, L. Shi, S. Jungsuttiwong, and S. Yuan, "Ionic liquids for high performance lithium metal batteries," *Journal of Energy Chemistry*, vol. 59, pp. 320–333, 2021.

[29] D. A. Corrigan, "The catalysis of the oxygen evolution reaction by iron impurities in thin film nickel oxide electrodes," *Journal of The Electrochemical Society*, vol. 134, no. 2, pp. 377–384, 1987.

[30] A. P. Abbott, "Application of hole theory to the viscosity of ionic and molecular liquids," *ChemPhysChem*, vol. 5, no. 8, pp. 1242–1246, 2004.

[31] J. M. Crosthwaite, M. J. Muldoon, J. K. Dixon, J. L. Anderson, and J. F. Brennecke, "Phase transition and decomposition temperatures, heat capacities and viscosities of pyridinium ionic liquids," *The Journal of Chemical Thermodynamics*, vol. 37, no. 6, pp. 559–568, 2005. Ionic Liquids.

[32] E. Küçüksönmez and J. Servantie, "Shear thinning and thickening in dispersions of spherical nanoparticles," *Phys. Rev. E*, vol. 102, p. 012604, Jul 2020.

[33] R. L. Hoffman, "Explanations for the cause of shear thickening in concentrated colloidal suspensions," *Journal of Rheology*, vol. 42, no. 1, pp. 111–123, 1998.

[34] E. Brown and H. M. Jaeger, "Generality of shear thickening in dense suspensions," *Nature Materials*, vol. 9, pp. 220–224, 2010.

[35] A. Singh, S. Pednekar, J. Chun, M. M. Denn, and J. F. Morris, "From yielding to shear jamming in a cohesive frictional suspension," *Physical Review Letters*, vol. 122, no. 9, p. 098004, 2019.

[36] R. Hunter, "Colloidal dispersions," in *Colloid Science*, pp. 1–72, Elsevier, 2023.

[37] T. Wang, M. Ni, Z. Luo, C. Shou, and K. Cen, "Viscosity and aggregation structure of nanocolloidal dispersions," *Chinese Science Bulletin*, vol. 57, no. 27, pp. 3644–3651, 2012.

Investigation of first-principles surrogate models for Protein crystallisation

Asan Masruri & Mateus Porfirio
Department of Chemical Engineering, Imperial College London, U.K.

Abstract
This thesis presents the development of a surrogate mechanistic–data-driven model designed to address the challenges in scaling up protein crystallisation processes for biotherapeutic production. It integrates a parametric Method of Moments (MoM) simulator with Long Short-Term Memory (LSTM) networks, enabling accurate prediction of crystallisation concentration profiles under varying kinetic parameters. Using Latin Hypercube sampling, a broad range of mechanistic scenarios was generated, and the resulting data were pre-processed, normalised, and structured for model training. Two modelling approaches were investigated: a forecasting strategy that uses dense early-stage observations to predict the remaining profiles, and an interpolation strategy that relies on sparse, evenly spaced samples. Hyperparameter optimisation with Optuna and rigorous k-fold cross-validation ensured robust, generalisable models. While forecasting achieved near-perfect accuracy, it proved operationally unrealistic due to the requirement for intensive early sampling. In contrast, an hourly-based interpolation approach offered strong predictive performance and greater practicality, accurately capturing complex system behaviour without sacrificing scalability. This calibrated surrogate model therefore provides a valuable framework to inform and guide process optimisation. Moreover, it can be readily adapted to more complex scenarios, including seeded or template-induced crystallisation, offering a sustainable pathway to meet the growing demand for large-molecule therapeutics.

1. Introduction

In the past few years, protein-based therapeutics have evolved to become vital in the pharmaceutical industry because of their efficacy in treating chronic diseases, such as rheumatoid arthritis, diabetes, and others with an estimated market size of 566 billion USD by 2030 [1]. Downstream purification is usually done via chromatography which, although operating at high purities, is very expensive and incapable of a scale-up that could reach market demand to supply the required therapeutics.

Crystallisation has been explored as an innovative solution. Current use for separating small-molecule active pharmaceutical ingredients (APIs), and templated nucleation becoming of interest to achieve quick and easily controllable large-molecule crystallisation, presents an alternative to efficiently purify therapeutics due to superior scale-up potential by improving process volumes rather than product titres [2], [3], [4], [5] .

Scaling up a protein crystallisation process, however, does not come without its challenges due to complex crystallisation behaviour and unknown early-stage nucleation mechanisms. Proteins specifically are difficult to crystallise as they are large, fold and have many sources of interactions. Incomplete characterisation of large-molecule crystallisation due to this system complexity and other inaccuracies challenges systematic and evidence-based decision making in the industrial scope for large-molecule crystallisation scale-up as predictions must be correct or the design will be for the wrong process. To ensure in depth mechanistic understanding can be found in these crystallisation systems; researchers proposed the use of an initial integrated mechanistic-data-driven surrogate model on an unseeded crystallisation process to model a realistic scale-up to be transferred to seeded and even template-induced processes required to purify these large molecule therapeutics. Transferred learning can reduce the number of experiments performed for new templates and inform the performance of untested templates. This research will focus on the investigation of an in-silico/in-vitro surrogate model of an unseeded lysozyme crystallisation process. Combining both parametric and non-parametric models serves as a foundation for constructing and validating the model. The parametric model will simulate experiments under varying conditions and this data will be used to train and evaluate the neural network. Following this, the model will be optimised to ensure minimal loss for the purpose of implementing transferred learning to enable scale-up of various protein crystallisation systems.

2. Background

2.1 Crystallisation basics

A crystal is a solid structure that consists of an ordered, regular array of molecules organised into a 3D lattice with periodic patterns of molecule arrangements [6]. Crystallisation is the process by which a dissolved molecule in solution forms or is incorporated into these solid crystalline structures. Nucleation refers to the birth of small crystalline nuclei from the liquid solution. Nuclei that

overcome the kinetic barrier of the phase transition, i.e for a first-order phase transition a solution becomes a solid crystal, grow and do not re-dissolve [7]. Nucleation types are divided into primary and



secondary classes: Primary homogeneous nucleation occurs spontaneously in solution with no surface; primary heterogeneous nucleation requires a foreign surface to be observed; secondary nucleation occurs on a larger crystal of the same solute [8].

Figure 1: Nucleation pathways

Classical Nucleation Theory (CNT) describes the mechanism for primary homogeneous nucleation of crystals [9], [10] shown in Eq (1). In the case of a crystal embryo the overall change in Gibb's free energy will decrease due to nuclei formation, as shown by the first term, and increase due to the presence of a new phase in solution, shown by the second term.

$$\Delta G = \frac{4}{3}\pi r^3 + 4\pi r^2 \gamma$$

Eq 1: Classical Nucleation Theory (CNT)

2.2 Heterogeneous and templated nucleation

Heterogeneous nucleation is a subset of primary nucleation driven by the phase boundary between the solute and solid surface where a solute crystallises in contact with a foreign surface. The phenomenon is generally considered undesirable and typically occurs on reactor walls or impellers [11], [12], [13]. Heterogeneous surfaces can act as templates with lower free energy demand where tuning and purposeful selection can help control various process features of crystallisation systems such as observed polymorphs or kinetic rate [14]. Studies show when mixing silica nanoparticles with a lysozyme solution of moderate supersaturation (S≤10 molecules firstly absorb onto the silica surface and proceed to aggregate to form crystals, greatly enhancing nucleation rates in the system; with addition of lysozyme-saturated silica nanoparticles further reducing nucleation induction time [15], [16].

2.3 Computational modelling of crystallisation systems

Crystallisation models are used to calculate process conditions and their changes during the operation of a crystalliser unit. Formulating a model is no easy task, as the non-linear mathematical structure of the model forms a hyperbolic integro–differential problem that requires careful application of numerical solution methods to precisely solve with reasonable computational expenditure [17], [18].

$$\frac{\partial n}{\partial t} + \frac{\partial (Gn)}{\partial L} + D(L) - B(L) + n\frac{dlnV}{dt} + \sum_k \frac{n_k Q_k}{V} = 0$$

BC: $\quad n^o = \frac{B_o}{G(L = L_O)} \qquad\qquad n = \frac{dN}{dL}$

Eq 2: Population Balance Equation

Mass and energy balances ensure that conservation laws remain valid in the crystallisation model. However, as particle properties are not conserved in the same way, a population balance equation (PBE), shown in Eq (2), is used to describe changes in crystal nucleation and growth [19]. Another hurdle, as such partial differential equations (PDEs) are challenging to solve accurately [20].

3. Hybrid Model

The hybrid model integrates the strengths of High-Fidelity (HF) and non-parametric models, combining the physical insight provided by mechanistic models with the flexibility of non-parametric approaches to improve process predictions during computational modelling.

3.1. High Fidelity models.

High-Fidelity (HF) models comprise a system of partial differential-algebraic equations that rigorously incorporate detailed kinetic mechanisms governing crystallisation processes. Numerical schemes solve these models and determine solution properties such as concentration profiles, crystal size distribution (CSD), and the number and dimensions of nucleated and grown crystals. They can be used for process optimisation [21]; process control [22]; or to investigate mechanistic knowledge [23]. Several solution methods were developed to solve the PBE, each with specific advantages and disadvantages. This section focuses on two contrasting approaches: the Method of Moments (MoM) and the Finite Volume (FV) classes.

3.2. MoM Class

The Method of Moments (MoM) simplifies the complexity of the full crystal size distribution by reducing it to a finite set of key quantities, known as moments. These moments correspond to physical properties of the crystals, such as their number, volume, or surface area. By reducing the PBE into a system of ordinary differential equations (ODEs), MoM significantly reduces the computational effort

required, enabling faster and more efficient solutions compared to solving the full integro-differential equation. Nevertheless, while MoM is computationally efficient, it has inherent limitations. It struggles to accurately capture the full dynamics of crystal growth, particularly in systems involving large molecules or complex crystallisation behaviours, where the size distribution evolves in highly non-linear ways. This limitation makes it less suited for scenarios demanding detailed resolution of crystal size dynamics.

3.3. FV Class

In contrast to Method Of Moments, the Finite Volume method involves a discretisation of the domain of interest. FV differs from other methods in that it ensures the conservation of the intrinsic property in the balance for each volume element, this is most commonly mass. High resolution finite volume (HRFV) methods were developed to specifically reduce numerical diffusion and oscillations found in the presence of non-smooth solutions due to the stiffness of a problem. This limitation is addressed by introducing a flux limiter function dependent on the ratio of successive population density gradients across 3 adjacent bins.

3.4. Non-Parametric Models

Non-parametric models instead are derived and determined exclusively from measurement data where parameters are fitted to an arbitrary structure, and they have no physical meaning [24]. With no requirement to understand the inherent complexity of the system, the applicability of these models in these systems where non-linear behaviours are observed cannot be overstated. Several types of non-parametric models exist, differing on how the input and output states are correlated. It was decided to focus on only Neural Networks (NN) due to their ability to model nonlinear behaviour without prior knowledge and their fast computational speed once trained.

3.5. Neural Networks

Inspired by the structure and the intricacy of the biological neural networks found in the human brain, artificial neural networks (ANNs) attempt to emulate how biological neurons process and transmit information [25], [26]. ANNs consist of interconnected layers of nodes, or 'neurons', where each connection carries a weight that determines the strength of the interaction. By adjusting these weights during training, ANNs are capable of learning complex patterns and relationships within data.

3.6. Recurrent Neural Networks

As neural networks evolved, certain limitations naturally emerged. From speech recognition to process systems engineering, what the standard neural network lacked is the capability of capturing temporal dependencies over varying time scales. This limitation led to the development of Recurrent Neural Networks (RNNs).

Recurrent Neural Networks extended the capabilities of ANNs by introducing a feedback connection, enabling the model to retain and utilise information across time steps. In addition, unlike Feedforward networks, which process inputs independently, RNNs maintain a hidden state that acts as a form of memory, allowing the network to model dynamic, time-dependent behaviours.

3.7. Long Short-Term Memory Neural Networks

While RNNs offer advantages for sequential data modelling, they are not without limitations. During the training period of standard RNNs, issues such as the vanishing gradient and exploding gradient arise, hindering the learning capability of long-term dependencies [27]. These challenges prompted the development of more advanced architectures such as Long Short-Term memory (LSTM) and Gated Recurrent Units (GRUs). In the context of this research, LSTMs will be the main focal point due to their enhanced performance on longer sequences allowing the capability of managing temporal dependencies.

3.8. Optimising the LSTM

Due to NNs being extremely sensitive, its critical to tune the hyperparameters of the model. Hyperparameters refer to the variables that do not change while the NN is being trained. There are several hyperparameters that can be adjusted to improve the predictive capabilities of the LSTM:

Hidden size: refers to the number of hidden units (neurons) in each layer of a neural network.
Number of layers: represents the depth of the neural network, indicating how many stacked layers are included in the architecture
Batch size: defines the number of training samples processed simultaneously during one forward and backward pass
Learning rate: controls the step size for updating model weights during training.

4. Methods

4.1 Criteria for the creation of the surrogate model

	MoM	FV
Pros	- Reduces CSD to key moments, simplifying the system. - Set of ODEs are quick and easy to solve - Computational strain is reduced	- Ensures conservation of intrinsic properties like mass and energy. - Reduced numerical diffusion and oscillations - Manages stiffness and non-smooth solutions
Cons	- Limited ability to capture full dynamics of crystallisation - Overlooks some information regarding CSD	- Complex implementation - Computationally expensive particularly for large-scale problems

Table 1: Comparison of High-Fidelity Models

	LSTM	GRU
Pros	- Captures long-term dependencies in sequential data - Manages vanishing gradient issues effectively - More expressive due to additional gates - Suitable for complex sequences and long-term patterns	- Faster training due to simpler architecture with fewer parameters - Performs comparably to LSTMs, especially with smaller datasets - Computationally more efficient, requiring less memory - Easier to tune due to fewer hyperparameters
Cons	- Computationally expensive due to complex architecture - Longer training time as there are more parameters - Can lead to overfitting with smaller datasets	- Struggles with extremely long-term dependencies. - Less expressive than LSTM in complex scenarios. - May not always generalise well for highly complex sequences.

Table 2: Comparison of Non-Parametric Models

Comparison of combinations of HF and non-parametric models was approached systematically as displayed in table 1 and 2.

For MoM vs FV, these comparisons highlight a trade-off between computational efficiency and accuracy in concentration profiles. For LSTM vs GRU, while both effectively handle sequential data, LSTMs perform better at capturing long-term dependencies due to their state-based architecture.
To leverage these strengths, a surrogate model integrating the computational efficiency of MoM with the predictive power of LSTMs is proposed.

4.2 Parameter Estimation

With the MoM class simulations provided and set up the next step was to determine which loss calculation allows minimal difference when comparing experimental kinetic parameters against simulation ones. The parameters are nucleation pre-exponential constant (A_j), surface energy (γ ($mJ.m^2$)), growth rate pre-exponential constant (A_g), the growth rate constant (g), and initial concentration (IC (mol/m^3)) The mean squared error (MSE), maximum likelihood estimate (MLE), and mean absolute error (MAE) were calculated as loss functions using the scikit.learn library, as seen in equation 3, 4 and 5.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \check{y}_i)^2$$

Eq 3: Mean Squared Error

$$MLE = \prod_{i=1}^{n} P(y_i|\theta)$$

Eq 4: Maximum Likelihood Estimation

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \check{y}_i|$$

Eq 5: Mean Absolute Error

After obtaining the values, an optimisation function was set up to minimise each of these losses and observe the set of simulation kinetic parameters that match that profile. MSE gave the closest values towards the experimental parameters. Therefore, MSE will be the main calculation for validation loss during the LSTMs training period.

Kinetic Parameter	Acceptable Range	Nominal Value
A_j	10-60	37.7
A_g	0.15-2.5	0.682
g	1-4	2.4
γ	0.15-1.2	0.881
ic	12-20	15

Table 3: Kinetic parameters and their ranges

4.3 Building the Surrogate Model

Having confirmed the combination of HF and non-parametric models and determining which loss calculation should be used for the training period; it was then time to create the model. The Methodology is laid out below in figure 1.

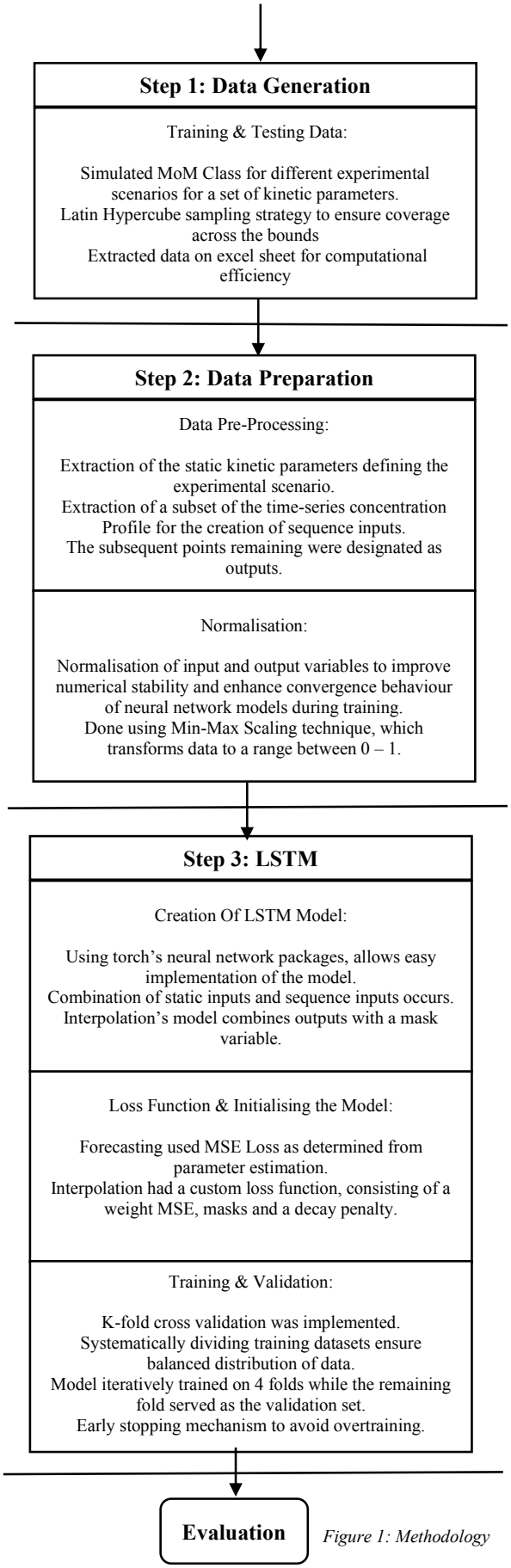


Figure 1: Methodology

4.4 Varying Sequence inputs

To test how effective the surrogate is, the model must be pushed to its limits revealing its potential issues and shortcomings. One way to do this is to essentially withhold information from the model to evaluate its predictive performance by varying sequence inputs.
Sequence inputs are a unique feature of LSTMs. Utilising this, more operationally realistic models were developed that would adhere to industrial workflow, allowing for the consistent agglomeration of data to further train the model. Three main variations were investigated: Multi-step time series forecasting, ran.dom time series interpolation, and hourly time series interpolation.
The forecasting model fed the initial 5 timepoints of the simulated concentration profile to the LSTM and tasked it with predicting the remainder of the profile. Random and hourly interpolation, on the other hand, fed 5 random concentrations and 5 hourly concentrations to the LSTM respectively. The LSTM was then tasked with predicting the concentration profile between these 'true' values.

4.5 Interpolations Data Preparation:

During the pre-processing step for interpolation, to simulate sparse data acquisition, a custom function was implemented. For each model, two sampling strategies were applied: random interpolation, where five points were randomly selected across the profile, and hourly interpolation, where eight points were selected at hourly intervals.
To structure the data appropriately, the observed points were extracted based on the model. A binary mask was simultaneously generated to explicitly indicate the observed and unobserved positions, with zeros representing known points and ones denoting missing values. The unobserved concentration points were extracted separately as the ground truth for model predictions.
This systematic approach allowed the separation of time-invariant static parameters and dynamic sequence inputs, enabling the model to predict the evolution of concentration profiles based on the corresponding sequence points over the defined time horizon.

4.6 Normalisation and conversion to Tensors.

Three separate Min-Max scalers from the scikit-learn pre-processing package were applied to the main components of the dataset stated within the Data preparation. The static kinetic parameters, sequence inputs, and target outputs were normalised each using a unique scaler. As masks are already binary variables, they do not require normalising.

Following normalisation, the three component datasets were converted into PyTorch tensors using the torch package. These three tensors were combined into one dataset for forecasting and masks were converted into tensors and combined for interpolation. This step enabled seamless integration into the LSTM training pipeline. By scaling all features and targets to a comparable range, the normalisation process ensured that the model could learn effectively without bias toward variables with larger magnitudes.

4.7 Creation of an LSTM class:

Two LSTM-based architectures were developed to perform forecasting and interpolation of concentration profiles. Both architectures utilised static kinetic parameters alongside sequential concentration inputs, enabling the models to capture both time-invariant system properties and the temporal evolution of the concentration profiles.
The theoretical foundation of LSTMs, which manage both short-term and long-term memory via gating mechanisms, is directly utilised using Torch's Neural Network package.

For the forecasting model, static kinetic parameters were first processed through a fully connected layer with a ReLU activation function to transform and extract fixed features. The sequence input points, consisting of the initial data of the concentration profile, were passed through an LSTM layer consisting of 2-layer depth. Combined with the processed static features, it is passed through a fully connected layer that generates the predicted concentration profile. Creating a Forward propagation enables the model to process input data, apply transformations, combine features, and generate predictions, forming the basis for the learning process. The sequential concentration inputs are passed through the LSTM layer and combined with the static features allowing the model to learn the temporal relationships within the sequence. The integrated features are then used to generate predictions ensuring the shape of the output aligns with the targe profile format. Interpolations model follows the same procedure however with an extra variable known as masks, which required adjusting. The predicted output was constrained using a binary mask to retain observed values while updating the unobserved positions.

4.8 Loss function and Optimiser

With the LSTM architecture implemented and the tensor dataset prepared, the next step was to define the loss function and optimiser to train the model. For forecasting, a standard (MSE) loss function was used to evaluate the validation loss as directed by the parameter estimation.

For the interpolation model, a more complex custom loss function was developed to address the additional challenges posed by sparse observations and the need for smooth transitions between predicted values.

The weighted MSE was introduced to emphasise the importance of specific observed points within the concentration profile. By assigning a weight factor of 3 to the observed positions, the loss calculation prioritised accuracy at these critical points. This ensures that the model focuses on correctly predicting known values, which are sparse but significant for reconstructing the full profile.
To manage the sparsity of the data, the mask was applied to exclude unobserved points from contributing to the loss calculation. This approach aligns the loss function with the constraints of the data, ensuring that irrelevant regions do not influence the model's optimisation process.
To enforce smooth transitions between consecutive predictions, a decay penalty was incorporated. The penalty was calculated as the mean squared difference between consecutive time steps, scaled by a regularisation factor of λ=0.005. The final custom loss function combines the weighted MSE and the decay penalty to form a single objective:

4.9 K-fold cross Validation During Training Period.

To mitigate overfitting and ensure robust model performance, K-fold cross-validation was implemented with k=5. This approach systematically divides the training dataset into multiple subsets, enabling both training and validation to be conducted iteratively across all data points.
The training dataset, consisting of the post-processed experiments, was first divided into 10 equally sized subsections to ensure a balanced distribution of data. These subsections were then allocated across the five folds, with each fold representing one-fifth of the total data. This process ensured that each fold contained samples representative of all subsections, maintaining consistency and avoiding bias during model training and evaluation.
This systematic approach provided a rigorous evaluation of the model's performance while reducing the risk of overfitting to any particular subset of data.

4.10 Testing Data and Model Evaluation

The testing data was prepared using the same preprocessing steps as the training data to ensure consistency. The model was switched to evaluation mode using model.eval() in PyTorch to disable dropout and batch normalisation layers.

The model was run on each experiment in the testing dataset to generate predictions for the concentration profiles. These predictions were then inverse transformed to return them to their original scale for comparison. The predicted profiles were plotted against the ground truth to visually assess performance, and the corresponding calculated loss function as a quantitative metric.
This process ensured a rigorous evaluation of the model's ability to generalise to unseen data, providing both qualitative and quantitative assessments of its predictive accuracy.

4.11 Hyperparameter Optimisation

To improve the performance of the LSTM model, the Optuna hyperparameter optimisation framework was implemented. Optuna utilises a Bayesian optimisation approach with a dynamic search space, enabling efficient exploration of hyperparameter combinations to minimise the validation loss.
Within the framework, an objective function was defined, in this case the training and validation of the LSTM model. The objective function received suggested hyperparameter values from Optuna, trained and evaluated the model, and returned the corresponding loss function for each trial respective to the model. To guide the optimisation process, hyperparameters were defined within specific bounds using the trial.suggest functions provided by Optuna.

Parameter	Acceptable range
Hidden size	**10 - 200**
Number of layers	**1 – 4**
Lr	**1.10^{-5} – 1.10^{-2}**
Batch size	**16, 32, 64**
Number of input points	**1 - 30**
Training experiments	**256, 512, 1024, 2048, 4096, 8192, 16384**

Table 4: Hyperparameters and their ranges

Optuna then iteratively explored the search space in the minimise direction to identify the combination of hyperparameters that yielded the lowest validation loss. By leveraging Optuna's adaptive search strategy, the optimisation process efficiently balanced exploration and exploitation, enabling the identification of an optimal set of hyperparameters to maximise model performance.

5. Results & Discussion

5.1 Hyperparameter Optimisation results:

Parameter	Forecasting	Interpolation
Hidden size	**26**	**50**
Number of layers	**2**	**2**
Lr	**0.0055**	**0.0036**
Batch size	**32**	**32**
Number of input points	**27**	**8**
Training experiments	**16384**	**8192**
Optimal Loss	**6.10^{-5}**	**0.0044**

Table 5: Optuna optimal parameters

Table 5 shows Optuna's optimised hyperparameters and parameters. The important factor to consider here is the fact that the optimal loss for forecasting was roughly 2 orders of magnitude lower than that of interpolation suggesting it is the superior model in terms of predictions.

Forecasting required a smaller hidden size (26) compared to interpolation (50). This suggests that the forecasting model can efficiently learn the underlying patterns/ relationships within the data with a smaller number of LSTM units. This is likely because forecasting uses dense initial samples as the initial inputs, enabling the model to capture these temporal dependencies with fewer units. In contrast, Interpolation involves few spread-out points, requiring a larger hidden size to capture the subtle and complex relationships and fill gaps between observations emphasising the need for a higher number of units to approximate the missing concentration points accurately.

Interpolation's slightly lower learning rate indicates its sensitivity to large weight updates when working with little data. A higher learning rate in this case would lead to convergence on oversimplified profiles. For forecasting the higher learning rate reflects the model's ability to converge quickly due to the richer input data available at the start of the prediction process and the relatively simple profile towards the end of the process.
Forecasting heavily relies on an extended sequence of initial observations to make long-term predictions, whereas interpolation operates under constrained inputs reflecting the challenges of the forecasting model in terms of its practicality as frequent early-stage sampling is both difficult and potentially compromises the results obtained.

Training experiments had different effects for both models. Forecasting showed that as we increased the number of training experiments, the model learns more effectively to forecast the rest of the concentration profile, which requires exposure to a

wider variety of training experiments to ensure a reliable performance. On the other hand, Interpolation had an interesting effect. Too little training data, the model undertrains too much data, the model over trains creating a convex profile. This highlights that there should be more focus on creating an impactful training set as the model relies more on localised patterns and benefits less from large-scale datasets.

5.2 Hyperparameter Importance:

Optuna's visualisation tools allow for the plotting of parameter importance, leading to the analysis of the search process and deduction of relative influences of each parameter.
With these hyperparameters and their importances laid out they were implemented into the model to test the improvement of the model from its previous unoptimised parameters.

5.2.1 Forecasting:

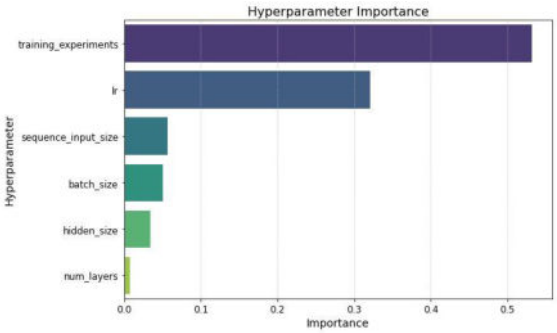


Figure 2: Forecasting Hyperparameter importance

Figure 2 helps visualise the importance number of training experiments has on validation loss when changed; followed closely by learning rate with the other parameters, such as hidden size and batch size, lagging behind significantly .This finding reflects the importance of larger datasets for the model to accurately predict long-term concentration profiles, as an increase in the training experiments will allow the model to generalise more efficiently.

5.2.2 Interpolation:

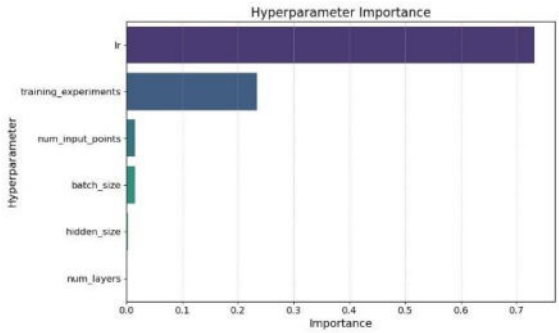


Figure 3: Interpolation Hyperparameter importance

Figure 3 shows that learning rate is king when considering Interpolation with training experiments lagging relatively far behind with all other parameters being seemingly insignificant. The sensitivity of learning rate highlights the challenges posed by sparse data distributions and underscores the need for careful tuning to prevent premature convergence. Hence why an optimal learning rate is crucial for this model.

5.3 Final Simulation Results:

After obtaining the final parameters and new models etc. an evaluation on different experiments was run. This section will focus on experiment 421 where the kinetic params are 30.06, 0.65, 1.98, 2.46 and 19.73 for A_j, γ, A_g, g and IC respectively.

5.3.1 Forecasting

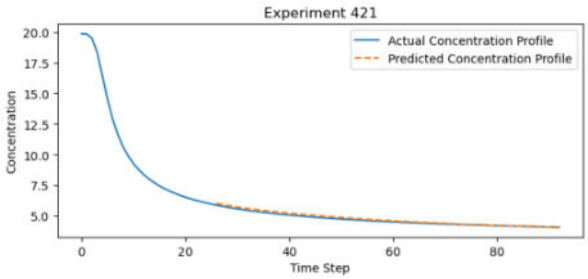


Figure 4: Forecasting model final prediction

Figure 4 shows the final result of the forecasting model boasting a loss of 6.10^{-5}. This allows for a simple visualisation of how the density of the initial inputs allows for the near perfect predictions of the remaining concentration profile. However, it also reveals how these same inputs allow the model to ignore the initial, non-linear stages of the profile; critical to capturing the complex and unknown mechanistics of crystallisation.
While its accuracy is undeniable, real-world considerations limit its practicality for the future improvement of this model. Collecting a sample takes time making consecutive sampling during the early stages of a crystallisation process operationally unrealistic. Furthermore, frequent sampling in the early stages can introduce pertubations that may compromise the stability and consistency of the crystallisation process [28].

5.3.2 Random Interpolation

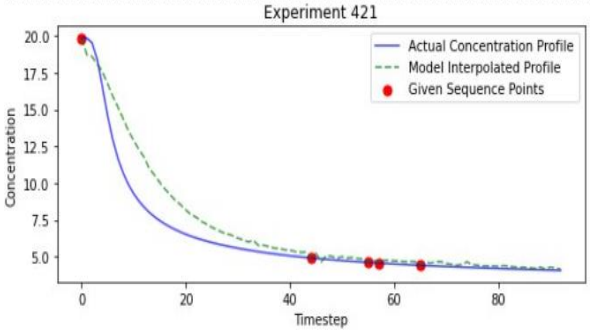


Figure 5: Random Interpolation model final prediction

From figure 5 we can see that random interpolation exhibits significant weakness. When the random points are distributed further down the concentration profile, the model struggled with reconstructing the earlier concentration profile effectively. While the model serves as a useful baseline for assessing the capabilities of the model, it lacks the accuracy and utility that is desired. When assessing practicality, random sampling does not make sense in a modern working environment and the model was deemed to be unrealistic and not very useful but was the baseline for the creation of the hourly interpolation model.

5.3.3 Hourly based interpolation

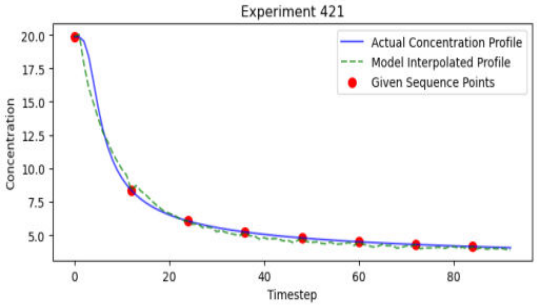


Figure 6: Hourly Interpolation model final predictions

Evolved from random interpolation, hourly interpolation shows strong resilience when predicting the initial stages of the concentration profile and competes with forecasting at the later timesteps. This outcome highlights the resilience of the hourly-based sampling strategy, which mirrors current industrial practices such as in-process checks (IPC) [29] for crystallisation systems. The periodic nature of hourly observations ensures that the model receives a well-distributed representation of the concentration profile, enabling it to accurately interpolate missing values without requiring extended initial data. Moreover, the collection and utilisation of future experimental data on seeded and unseeded crystallisations to be implemented into the training period will allow the model to constantly improve over time.

5.4 Final Comparison of Models:

While forecasting achieved a significantly lower loss, this must be contextualised. The model presents itself as the most suitable for an unseeded crystallisation process, where we are able to model the system using the MoM class. However, the large amount of training experiments coupled with the dense input data required for the model to achieve a higher accuracy is an idealistic scenario and does not reflect real-world implications, where data sampling limitations often make forecasting impractical for a seeded or template induced processes where experimental data is scarce. Interpolation, by contrast, operates with sparse, random input data, making optimisation a far more challenging task. The emphasis on learning rate and hidden size highlights the models struggle to balance both learning speed and understanding of the intricate relationships within the data. However, by using the optimal parameters with an hourly spaced model mitigates this inaccuracy as evenly distributed samples allow for a more effective capture of the temporal dynamics.
By understanding the relationships between hyperparameters, model complexity, and the data's structure, the optimised interpolation model emerges as a flexible and realistic solution for industrial crystallisation processes. Being a great candidate for implementation to match the predicted increase in demand.

6. Conclusion

In this study, an integrated mechanistic-data-driven LSTM-based model combined with a Method of Moments class was developed to address the increase in demand of large molecule biotherapeutics. The decision to proceed with interpolation over forecasting was driven by practical considerations from an industrial perspective, allowing the model to reconstruct missing values providing a more immediate and actionable outcome. A novel hourly-based interpolation model was created by combining mechanistic insights from the Method of Moments (MoM) with the predictive power of Long Short-term memory networks. The model utilised sparse, evenly spaced observations at hourly intervals, simulating realistic sampling strategies currently used. By integrating static kinetic parameters with sequential inputs, the surrogate approach successfully captured the temporal dependencies and underlying dynamics of the crystallisation process.

The model performed admirably, achieving minimal loss for an unseeded crystallisation process. This result highlights the model's ability to

effectively simulate crystal nucleation and growth behaviours, even within complex and highly nonlinear systems. The combination of mechanistic understanding and data-driven learning allowed for accurate predictions of concentration profiles, validating the model's capability in handling these systems.

Furthermore, the high-performance of the developed model positions it as a reliable tool for testing more advanced crystallisation processes. Its framework can be readily extended to seeded or template-induced crystallisation systems, where additional complexities arise due to external influences on crystal growth and nucleation. Moreover, if implemented in industry, the consistent generation of experimental data will allow the model to constantly train and learn, consolidating its knowledge and deepening its understanding of various crystallisation processes. This flexibility enhances the model's applicability across various industrial crystallisation processes, ensuring its relevance in process optimisation and scale-up scenarios as introduced in this paper.

In summary, this research successfully developed a robust and flexible surrogate that balances physical mechanistic principles with the predictive capabilities of deep learning. The model addresses the critical challenge of interpolating missing concentration data, and providing accurate and actionable results for crystallisation processes facilitating scale-up, industrial implementation and a reduction in experiments performed on new templates or informing untested template performance bridging the gap in the demand for large molecule biological therapeutics.

Future work will explore its validation in seeded and template-induced systems, along with expanding its generalisability to other large-molecule and template induced crystallisation frameworks with the ability to eventually model optimal templates for different processes. All models have room to be fine tuned and different combinations of high-fidelity and non-parametric models to form new hybrids capable of transforming the pharmaceutical industry will be investigated.

References

1. Yadav et al. (2023) Protein-Based Therapeutics, Springer Nature, 23-47.
2. Lemmer et al. (2022) Cryst. Growth Technol., IntechOpen.
3. Hatcher et al. (2020) CrystEngComm, 22, 43, 7475-7489.
4. Chen et al. (2020a) Handb. Contin. Crystallization, RSC, 372-392.
5. Dos Santos et al. (2017) Biotechnol. Adv., 35, 1, 41-50.
6. Roberts et al. (2017) NATO Sci. Peace Secur. A Chem. Biol., Springer Netherlands.
7. Oxtoby (1998) Acc. Chem. Res., 31, 2, 91-97.
8. Xu et al. (2020) Ind. Eng. Chem. Res., 59, 41, 18335-18356.
9. Becker & Döring (1935) Ann. Phys., 416, 8, 719-752.
10. Volmer & Weber (1926) Z. Phys. Chem., 119, 1, 277-301.
11. Akrap et al. (2012) Phys. Rev. B, 86, 23, 235207.
12. Liang et al. (2004) Cryst. Growth Des., 4, 5, 1039-1044.
13. Yazdanpanah et al. (2020) Mol. Pharmaceutics, 17, 5, 1624-1634.
14. Artusio et al. (2018) Org. Process Res. Dev., 22, 9, 1104-1114.
15. Weichsel et al. (2017) Anal. Chem., 89, 22, 12066-12073.
16. Chen et al. (2021) Chem. Eng. Sci., 230, 116171.
17. Braatz et al. (2001) Annu. Rev. Control, 25, 1, 87-99.
18. Casas-Orozco et al. (2023) Processes, 11, 2, 234.
19. Randolph & Larson (1988) Theory of Particulate Processes, Academic Press, 2nd ed.
20. LeVeque (2002) Finite Volume Methods for Hyperbolic Problems, Cambridge Univ. Press.
21. Aamir et al. (2009) Ind. Eng. Chem. Res., 48, 20, 9230-9238.
22. Mesbah et al. (2009) J. Process Control, 19, 3, 527-537.
23. Li et al. (1995) AIChE J., 41, 10, 2344-2355.
24. Von Stosch et al. (2014) Comput. Chem. Eng., 60, 86-101.
25. McCulloch & Pitts (1943) Bull. Math. Biophys., 5, 4, 115-133.
26. Rosenblatt (1958) Psychol. Rev., 65, 6, 386-408.
27. Katte-Bangayya et al. (2018) J. Pharm. Sci., 107, 4, 960-971.
28. Braatz et al. (2002) Handb. Ind. Crystallization, Butterworth-Heinemann, 2nd ed., 353-374.
29. Pharmaguddu (2021) Pharmaguddu.

Exploring the Growth and Fluorescent Protein Expression of Encapsulated Pichia with Varying Concentrations of Methanol

Pratham Agrawal and Niam Shah
Department of Chemical Engineering, Imperial College London, U.K.

Abstract This study investigates the growth and fluorescent protein expression of Pichia pastoris strains (Wild Type, GFP, and mCrimson) in varying methanol concentrations, with an emphasis on the encapsulation of the GFP strain in alginate hydrogel beads. Pichia, known for its high cell-density growth and methanol-inducible AOX1 promoter, was analysed for optimal methanol concentrations to maximise fluorescent protein expression. A liquid culture experiment was conducted to evaluate growth and fluorescence over a 48-hour induction period under methanol concentrations ranging from 0% to 2%. Results showed that optimal GFP expression peaked at 1% methanol, while mCrimson performed best at 0.5%, highlighting strain-specific responses. Subsequently, Pichia strains were encapsulated in alginate beads to enhance operational stability and fluorescence detection. Encapsulation experiments demonstrated a downward trend at 2% methanol and a variability in readings on Day 2. Multiple T-tests further indicated high variability, implying that higher methanol concentrations may not strongly induce gene expression as anticipated. While gene expression was observed, the lack of growth underlines the need for biological repeats to improve data accuracy. This research underscores the potential of Pichia encapsulation for biotechnological applications, offering insights into optimising conditions for protein production and environmental sensing. Future work includes refining encapsulation techniques and expanding their application scope to real-world scenarios, including wastewater treatment.

Keywords: Pichia, Wild Type, GFP, mCrimson, Encapsulation, Methanol, Fluorescence

1 Introduction

Pichia pastoris, more commonly known as Pichia, is a genus of yeast and more specifically a methylotrophic yeast due to its well-known property of utilising methanol as a sole carbon source. This is attributed to a strong, methanol-inducible and tightly regulated promoter (AOX1) that leads to the secretion of high amounts of heterologous proteins with lower amounts of endogenous proteins [1]. Previous research has shown a mixed feed of glycerol and methanol has worked best to express recombinant proteins and fully induce the AOX1 promoter [2]. On top of this, Pichia is widely used for the production of recombinant proteins due to its unique ability to grow high cell densities upwards of 100g/L dry cell mass [3]. With these characteristics, Pichia is used in a large variety of industrial and research applications such as the production of pharmaceutical proteins, vaccines and research reagents. Currently, Pichia is used to express antigens for vaccines against modern diseases such as COVID-19 [4], produce industrial enzymes - more specifically, food-grade enzymes [5] - and also used in wastewater treatment to remove synthetic dyes [6]. Although these are ongoing research, its vast applications have led to a growing interest in protein production through methods including genetic engineering in the healthcare industry for the last two decades [7].

This paper looks to explore the potential of controlling Pichia growth using varying concentrations of methanol to produce fluorescent proteins. By investigating the fluorescence at various concentrations, predictions were made for the fluorescence levels expressed with Pichia encapsulated in alginate hydrogel beads before being tested in an experiment.

2 Background

Yeast, such as Pichia, possess chromosomes as well as plasmids which are simply smaller, circular extra chromosomal DNA. A well-documented property of bacteria and Pichia is their ability to share plasmid information through a process called conjugation which is a form of horizontal gene transfer [8]. This allows them to share and exchange beneficial genetic information with neighbouring cells when entering a new environment. However, Pichia are also known to also remove plasmid through cell division in a process called plasmid curing [9]. This occurs if a plasmid does not provide a selective advantage in a given environment and so is lost over successive generations to conserve energy [10]. In order to ensure the yeast does not get rid of the fluorescent gene expression, an antibiotic resistance gene is added upstream in the plasmid as a selection marker. For this research, the antibiotic chosen was Zeocin and Pichia was then cultured in a medium containing Zeocin ensuring that it retains the plasmid and hence the necessary gene expression. Along with this, a fluorescent protein gene, either GFP or mCrimson, was inserted in the plasmid so that the yeast can use the nutrients in the medium to produce either of the fluorescent proteins.

The genetically modified Pichia is then encapsulated in alginate beads to protect it from environmental factors such as pH fluctuations and toxic compounds due to its real-life applications such as wastewater treatment [11] to detect methanol concentrations.

Currently, there has been some initial research into the conditions needed to entrap Pichia in calcium alginate gels, looking at biomass concentration at different phases of Pichia growth [12] as well as successfully integrating the GFP gene expression into Pichia to yield an expression level of over 40% of all proteins [13]. Furthermore, alginate encapsulation has been used to improve the operational stability of Pichia by producing protein of high yield, high purity and high activity [14]. These properties would be particularly useful in the wastewater treatment industry as the expression for fluorescent proteins can be more easily detected. On top of this, alginate beads have been developed that encapsulated GFP-expressing Live Vaccine Strain (LVS) [15]. Whilst there are papers accomplishing the ability to encapsulate various biopolymers in alginate beads, they do not combine Pichia's valuable fast-growing properties with the benefits achieved by being entrapped in alginate beads [16] such as extended longevity, controlled drug delivery and the production of proteins in a confined space [17].

3 Methods

All cultures, recipes and preparations were conducted in sterile conditions, unless otherwise stated, due to the nature of biological substances used and the high risk of contamination. This was ensured by using a Bunsen burner to create a sterile radius of 10cm [18]. On top of this, filter sterilisation and autoclaving were used for solutions and equipment respectively. Unless otherwise stated, all water used to make solution was distilled water due to it not containing any impurities. *Section 3.1* specifies recipes [19] used throughout the entire research project with stage specific preparations detailed in their respective sections.

3.1 Preparation of Recipes

3.1.1 10X Yeast Nitrogen Base

Yeast Nitrogen Base (YNB) solution was prepared by using 134g of YNB, with Ammonium Sulphate and without amino acids, in 1L of water resulting in a 13.4% YNB concentration. The solution was then heated and magnetically stirred to dissolve YNB completely in water. Using 0.2µm porous membrane filters and a syringe, the solution was filter sterilised, and then stored at 4°C. It should be noted that vacuum sterilisation can also be used but is typically a process saved for larger volumes.

3.1.2 500X Biotin

0.02% Biotin solution was prepared by simply dissolving 20mg of Biotin in 100mL of water. This was also filter sterilised and stored at 4°C.

3.1.3 10X Glycerol

10% Glycerol solution was prepared by mixing 100mL of Glycerol with 900mL of water. This was sent off to be autoclaved after which it was stored at room temperature.

3.1.4 1M Potassium Phosphate Buffer

This buffer was prepared by combining 132mL of 1M Dipotassium Phosphate and 868mL of 1M Monopotassium Phosphate. It was crucial to ensure the pH of this buffer is 6.0 ± 0.1 and since this was not the case initially, Potassium Hydroxide was added as required to achieve the desired pH. This was tested using a pH probe which was recently standardised. The buffer solution was then sent off to be autoclaved after which it was stored at room temperature.

3.1.5 Yeast Extract Peptone Dextrose Medium

Yeast Extract Peptone Dextrose (YPD) medium was prepared by dissolving 10g of yeast extract and 20g of peptone in 900mL of water. This was dissolved using a magnetic stirrer on a heated plate and autoclaved after which 100mL of 20% Dextrose solution was added. The liquid medium was also stored at room temperature.

3.1.6 Yeast Extract Peptone Dextrose Agar

YPD agar was prepared by dissolving 5g of yeast extract, 91.9g of Sorbitol and 10g of Peptone in 450mL of water. This was then autoclaved after which 50mL of 20% Dextrose was added. The new solution was left to naturally cool to ~55°C when 500µL of 100mg/mL Zeocin was mixed in. The plates were stored at 4°C in the dark due to Zeocin being light sensitive and degrading rapidly [20].

3.1.9 Buffered Methanol-Complex Medium

Buffered Methanol-Complex Medium (BMMY) was prepared by dissolving 10g of yeast extract and 20g of Peptone in 795mL of water. This was then sent to autoclave. After, 100mL of Potassium Phosphate buffer, 100mL of YNB, 2mL of Biotin and 100mL of Glycerol was mixed at room temperature as well as 100% methanol to reach a final concentration of 0.5-1%.

3.2 Preparation of Agar Plates

Before pouring plates, it was crucial to create a sterile environment to work in which involved using a Bunsen burner as well as wiping down surfaces with ethanol. After this, the bottle containing YPD agar was heated in a microwave to liquidise it as it solidifies at room temperature. It was then sterilised by running the bottle opening under a flame and, using a 'PIPETBOY', 20mL was poured into each sterile agar plate ensuring the lid was kept as close as possible to avoid airborne contamination. This process had to be conducted quickly to avoid the agar medium from solidifying again. The plates were wrapped in foil to stop light entering and stacked upside down at 4°C to prevent condensation from dripping onto the agar.

Three types of Pichia were used in this experiment: Wild Type, GFP and mCrimson with the latter two being fluorescent protein genes. For these two, YPD agar plates with Zeocin were used as the strains contained a Zeocin selection marker. YPD agar plates without Zeocin were also prepared for the Wild Type Pichia to grow in. To streak the different types of Pichia onto the agar plates, plastic inoculating loops were used due to their smaller diameter allowing for easier streaking. This process was also conducted with a Bunsen burner. A quadrant streaking method was used due to it growing large dense cultures with single colonies.

3.3 Liquid Culture Experiment

Once the streaking of the plates had been completed, they were left in a static incubator at an optimal temperature of 30°C [21] over a weekend to allow for colonies to grow as much as possible. For any plates not used, perhaps kept for later experiments, these were wrapped in foil and placed in a fridge. *Figure 1* shows the subsequent results.



Figure 1. Petri dish containing Wild Type Pichia and its growth over a weekend. Streaking patterns that were performed previously can also be seen.

3.3.1 Culture Stage

The first steps of the experiment involved using the previously prepared starter culture of YPD medium and biomass culture of BMMY. To prepare the starter culture, 6 falcon tubes were gathered: 2 for each strain of Pichia. 5µL of Zeocin was deposited onto the side of four tubes (for the GFP and mCrimson strains). Then, 5mL of YPD medium was added to all 6 tubes making sure to wash any Zeocin down, if present, to ensure mixing.

Pichia grows in very high cell density, so it was important to identify singular colonies to pick out. Using a pipette, 2 colonies from each strain were picked and ejected into each falcon tube depending on the presence of Zeocin. This process was all completed in the presence of a flame for sterility purposes. All falcon tubes were lightly screwed and covered in autoclave tape before being placed in a shaking incubator at 250 rpm, also at 30°C. The plates containing the colonies of the different Pichia strains were wrapped in parafilm to prevent microbial contamination and placed in a refrigerator until required for further use. After a day, the biomass culture of BMMY was added; 10mL in each falcon tube. The total volume occupied by the cultures of each 50mL falcon tube was always between 10-30%, as per guidance when attempting to induce expression [19].

3.3.2 Centrifugation Stage

A minimal media, named Buffered Minimal Methanol (BMM) was then prepared the following day consisting of 1M Potassium Phosphate Buffer of pH 6, 13.4% YNB and 0.02% Biotin. To create 50mL of BMM, the following was added together in a falcon tube under a flame; 40mL of distilled water, 5mL of YNB, 5mL of potassium phosphate buffer and 100µL of Biotin (methanol to be added as necessary). This was then repeated to make another batch of BMM but with 50µL of Zeocin for the GFP and mCrimson strains. In total, 40mL of BMM without Zeocin and 80mL of BMM with Zeocin was required. The final concentration of BMM should come to 100mM of potassium phosphate buffer, 1.34% YNB and $4x10^{-5}$% Biotin.

The Pichia was retrieved from the incubator and pelleted via centrifugation at 3000xg for 5 minutes making sure all falcon tubes were equally balanced in the centrifuge. Before adding 20mL of BMM to each tube (with the appropriate antibiotics – BMM with Zeocin for GFP and mCrimson, BMM without Zeocin for Wild Type), the supernatant was removed and disposed of under sterile conditions. The solution was then gently resuspended using a 'PIPETBOY' and stripette. 10mL of each cell suspension was transferred into a new falcon tube; 12 in total - 4 for each strain which led to finally adding the appropriate volume of methanol for a desired final concentration. For each strain, 4 different concentrations were tested which were 0%, 0.5%, 1% and 2%, which is further explained in *section 4.1.3*. Since the total volume was 10mL, 100µL

of methanol was added for 1% for example and so on. Gradual introduction of methanol was crucial to allow the culture to adapt to growth on it. Adding methanol too rapidly could be lethal to the cells [2]. For safety purposes, the methanol was added in a fume cupboard without a flame due to it being highly flammable [22]. The tubes were then returned to the incubator for overnight growth. Sufficient aeration is required during methanol induction for efficient protein expression in Pichia [19]. This represented Day 0.

3.3.3 Fluorescence Measurements

Just before the centrifugation in *section 3.3.2*, 250µL aliquots, including technical repeats, of each strain were taken and put on a transparent spheroid 96-well plate. BMMY was also added to the plate in separate wells as a blank. Note that sides and corners of the plate were avoided where possible to limit evaporation of samples. A CLARIOstar (BMG Labtech) plate reader, as seen in *figure 2* was used to take optical density (OD_{600}) absorbance readings as well as GFP and mCrimson fluorescent readings to examine protein expression.



Figure 2. The set up of the CLARIOstar plate reader, where the 96-well plate is placed and inserted before beginning OD, GFP and mCrimson readings using the adjacent computer system.

3.3.4 Induction Stage

On Day 1, more samples were taken from the tubes, now from each concentration of methanol and each strain of Pichia – including technical repeats. These were placed in a similar transparent plate in order to track cell growth and fluorescent protein expression that occurred overnight. The same protocol for testing OD and fluorescent protein expression was applied as in *section 3.3.3*. All Pichia cultures were then spiked with the same methanol concentrations respectively as the day before and 50µL of BMM was added to the cultures with 0% methanol in order to balance the total volume. Overnight, the previously added methanol could have evaporated, due to its volatility, or the Pichia has metabolised the methanol, so more methanol was added to keep the desired level of induction every 24 hours. These steps were repeated for Day 2 signalling the end of the experiment ensuring a sufficient growth period.

3.4 Encapsulation Experiment

To set up the encapsulation experiment, the culture stage in *section 3.3.1* was repeated using new colonies of the 3 Pichia strains.

3.4.1 Additional Solutions Preparation

A solution of 10mM Tris was prepared from a stock solution of 1M Tris buffer and dilution with distilled water to make up a total volume of 1L. A 400mL solution of Calcium Chloride was made by combining the previously prepared Tris and 2.32g of powdered Calcium Chloride. In a falcon tube, a 2% alginate solution was prepared with 10mM Tris and then filter sterilised into a new tube.

Next, 500mL of a Krebs–Ringer N-(2-hydroxyethyl) piperazine-N′-ethanesulfonic acid (HEPES) buffer (KRH buffer) was prepared with 20mM HEPES, 135mM Sodium Chloride, 5mM Potassium Chloride and 0.4mM Dipotassium Phosphate. This was vacuum filtered using a filtration cup and pump and then supplemented with previously prepared sterile solutions of 1mM Magnesium Sulphate and 1mM Calcium Chloride.

3.4.2 Dilution and Centrifugation Stage

The cultures were retrieved from the incubator and under sterile conditions, 250µL aliquots were taken and placed onto a similar transparent 96-well plate used in *section 3.3.3*. Technical repeats were accounted for as well as BMMY as a blank before OD readings were taken and recorded of the Pichia cultures.

The concentration of Pichia culture required was $7.2\text{x}10^7$ cells mL^{-1}, and in order to calculate the dilution or concentration required, using literature, a conversion of 1 $OD_{600} \approx 5\text{x}10^7$ cells mL^{-1} was used [19]. A sample calculation is provided below:

OD Reading: 5.295

BMMY Blank OD Reading: 0.099

Adjusted OD Reading:

$$5.295 - 0.099 = 5.196 \quad (1)$$

Concentration:

$$5.196 \times 5 \times 10^7 = 2.598 \times 10^8 \frac{cell}{mL} \quad (2)$$

New volume of Pichia culture required (per mL):

$$\frac{7.2 \times 10^7}{2.598 \times 10^8} = 0.277mL = 277\ \mu L \quad (3)$$

Volume of BMM minimal media required:

$$1000 - 277 = 723\ \mu L \quad (4)$$

The cells were then pelleted via centrifugation in the same conditions as *section 3.3.2* before resuspension with 20mL of fresh BMM minimal media (prepared as required using same method in *section*

3.3.2). Likewise, the suitable BMM was added to each culture in terms of the presence of Zeocin.

3.4.3 Formation of Alginate Beads

In sterile 1.5mL Eppendorf tubes, the 2% alginate solution and Pichia cultures were mixed in a 1:3 volume ratio, specifically 250µL of Pichia and 750µL of alginate, to create the desired cell concentration of $1.8x10^7$ cells mL^{-1} hence why solutions of Pichia cultures were made up to final concentrations of $7.2x10^7$ cells mL^{-1} as seen in *equations 3-4*. This mixture was drawn into a 1 mL syringe, and a sterile 30G- blunt end needle was added to the tip prior to the bead formation. Simultaneously, 3 bottles (per Pichia strain) of 100mL Calcium Chloride-Tris solution (also vacuum filtered) each were gently agitated with a magnetic stirrer and a sterile magnetic bead. Drops of the alginate-Pichia mixtures were delicately added from a height of 1 cm into the respective Calcium Chloride solutions to form the alginate hydrogel beads, which were then left to incubate 30 minutes to ensure crosslinking. After 30 minutes, any excess Calcium Chloride was drained from the bottle and the remaining solution along with the transparent beads were transferred to separate petri dishes, where they were incubated for 1 minute in 10 mL of KRH buffer.

3.4.4 Fluorescence Measurements

To prepare for bead extraction and fluorescence measurements, whilst waiting for crosslinking, a 96-well plate was prepared by adding the 4 different concentrations of methanol being tested for each strain, including technical repeats (as seen in *figure 3).*

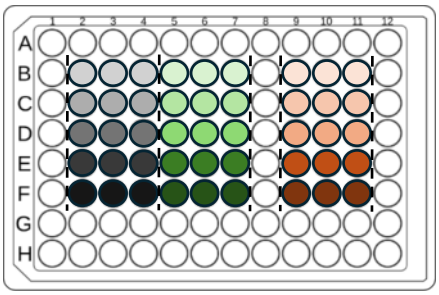


Figure 3. Example of the layout of a 96-well plate used in the encapsulation experiment. Wild Type (grey) is depicted from columns 2-4, GFP (green) in columns 5-7 and mCrimson (red) in columns 9-11. The gradient for each colour represents the methanol concentration beginning at 0% for the lighter shades progressing to 0.5%, 1%, 2% and 3% for the darkest shade. 3% methanol was tested too for further verification of results.

Respective concentrations of methanol were added with 200µL of BMM was placed in the appropriate wells taking into account whether Zeocin was required or not.

Once the beads were ready to be extracted, under a flame a cell strainer was used to isolate them and along with a pipette to transfer 1 bead to each well in the plate. This specific plate was advantageous as the spherical indent in each well helped the bead settle near the centre for uniformity while fluorescence was measured.

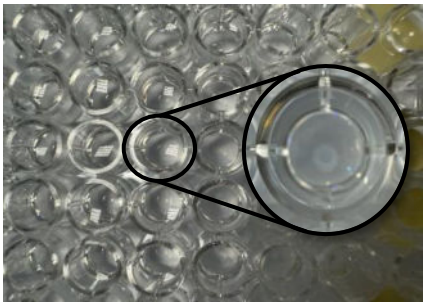


Figure 4. Close-up of the 96-well plate once the beads had been added to the BMM (and methanol). Initially they had been transparent in formation but then adjusted to a white/grey colour to indicate growth of the Pichia cells.

This was then taken to the plate reader to measure regular readings every 30 minutes for 24 hours for GFP and mCrimson expression. Initially, 20x20 matrix scanning was implemented for the readings to separate each well into sections and analyse the growth as accurately as possible, which indicated Day 0.

3.4.5 Induction Stage

After 24 hours, the data was retrieved from the reader. The plate was then returned, and another cycle of readings was initiated, only for GFP and mCrimson expression. On Day 2, a final set of readings were recorded after another top up of methanol.

4 Results and Discussion

4.1 Liquid Culture Experiment

4.1.1 OD Data

In order to use the raw data from the plate readings to show growth of the Pichia strains, some preliminary analysis was required. The background OD from BMMY, used as the blank, was totalled and averaged before being subtracted from the data obtained for each strain. These data points were also averaged since 3 technical repeats were recorded for every sample. Repeat calculations were carried out for each concentration of methanol for Day 1 and 2.

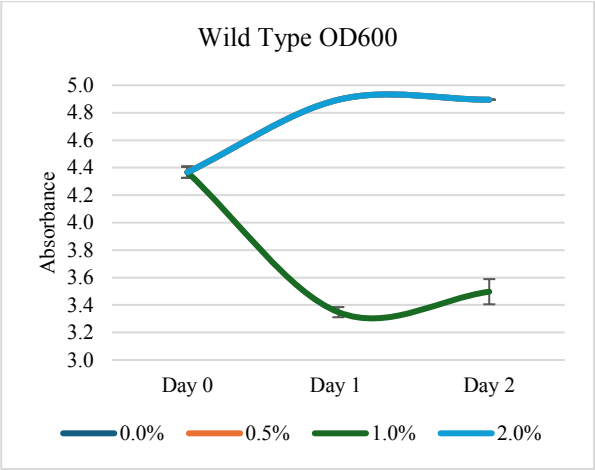


Figure 5. Line graph for Wild Type OD_{600} absorbance throughout the induction period. Some concentrations are hidden beneath other lines – can be better seen in *figure 5*.

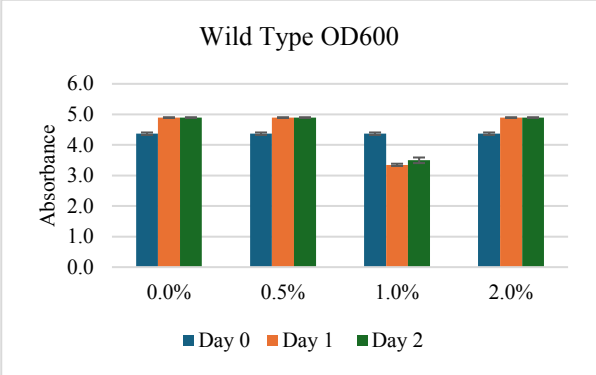


Figure 6. Bar chart representing the same Wild Type OD_{600} data to better visualise absorbance per concentration over the entire experiment.

The Wild Type OD_{600} readings in *figure 5* certainly shows overall growth during the induction period for all concentrations of methanol except for 1%. Despite this, on Day 1, the sample with 1% does begin to grow, increasing its optical density by the end of Day 2.

It is not a coincidence that all other concentrations of methanol and their respective samples have the same values by the end of the experiment. It is believed that this is the cause of the BMM media used and its medium carrying capacity limiting the growth of the Pichia. Carrying capacity refers to the average maximum number of cells that can occupy a particular habitat without permanently impairing the productive capacity of that habitat [23]. This can be easily observed and is supported by the uniform data in *figure 9* in which the maximum, and only, absorbance value is 4.895.

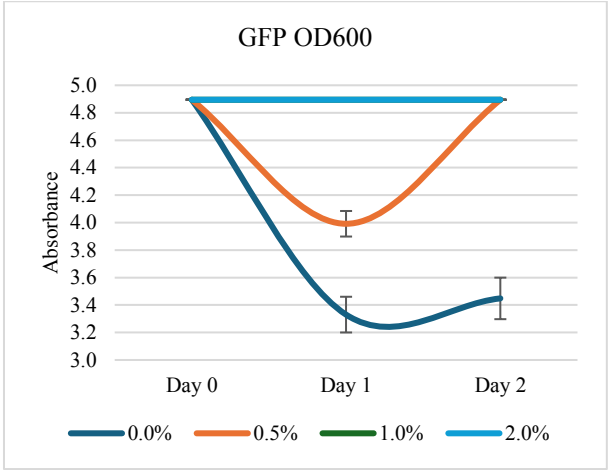


Figure 7. Line graph for GFP OD_{600} absorbance during the induction period – again hidden concentrations are better seen in *figure 8*.

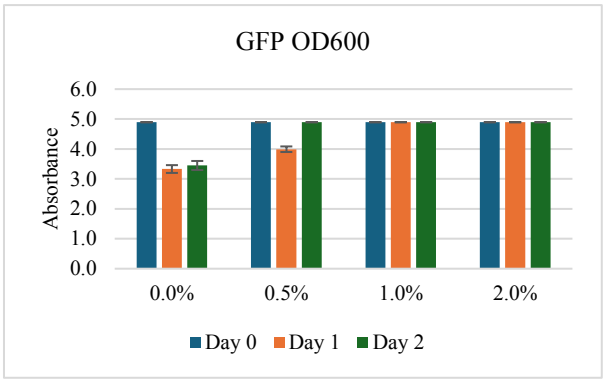


Figure 8. Bar chart representing the same GFP OD_{600} data to better visualise absorbance over the entire experiment. Here, the 1% and 2% concentrations are clearly redundant towards any conclusion.

Samples that contained the GFP strain came out with a different set of results. Higher concentrations of methanol did not seem to have any effect on the growth of GFP samples, by keeping at a consistent level, further revealed in *figure 8*. It is evident that the introduction of 0.5% methanol certainly did induce growth as absorbance increased after Day 1, far more than the control without any methanol. However, it is clear that this required at least 24 hours for any growth to begin, depicted in *figure 7* perhaps suggesting the sample was going through a delayed exponential phase on Day 1.

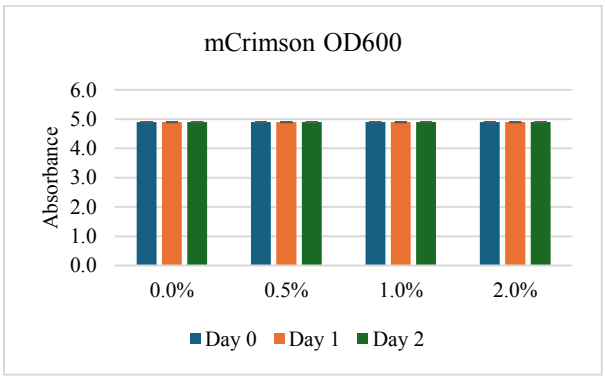


Figure 9. Bar chart representing mCrimson OD_{600} absorbance

throughout the induction period plainly displaying effects of medium carrying capacity.

Samples that contain the mCrimson strain show no sign of growth at any level of methanol concentration as depicted in *figure 9*. When comparing to the standard growth curve, it is possible that due to recording of only end point readings, intermittent growth was missed. The data recorded could cover the lag phase portion before exponential phase begins. Whilst this may be one conclusion, another possibility is, similar to GFP and Wild Type, the BMM has reached its carrying capacity limiting any further growth. This can also explain why the final data points for all methanol concentrations over the entire experiment are identical as well as no standard deviation error bars.

4.1.2 Fluorescence Data

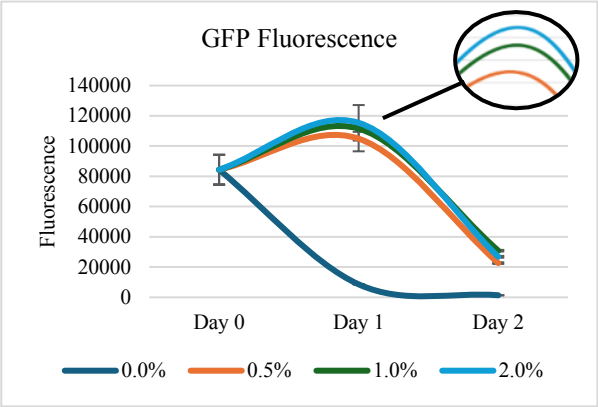


Figure 10. Line graph representing GFP fluorescence throughout the induction period. A zoom in shows 2% reaching the highest fluorescent peak although *figure 11* helps determine this may not be the case.

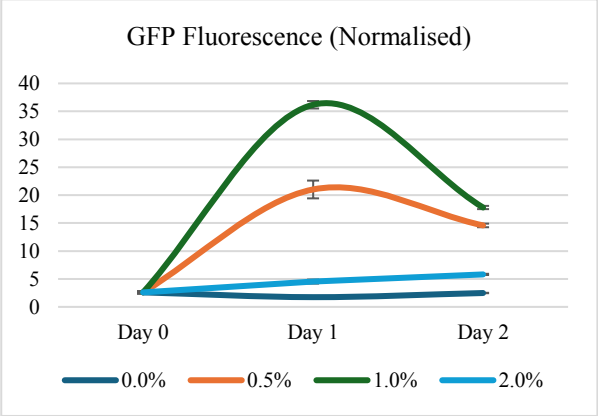


Figure 11. GFP fluorescence curves normalised (divided) against Wild Type fluorescence per methanol concentration for the same induction period.

The fluorescence data for GFP overall shows a clear trend of a peak in fluorescence on Day 1 except for the sample with 0% methanol concentration as seen in *figure 10*. At 0%, the AOX1 promoter would not have been activated due to no methanol being present in the sample which would lead to no expression of GFP as shown by *figure 10* reaching close to zero for that sample. The miniscule readings could suggest some background fluorescence being detected even after having subtracted the Wild Type value to negate this. The rest of the samples show a pattern of a higher peak with an increase in methanol concentration consistent with general understanding of a higher inducer concentration leading to a higher protein gene expression [24]. But with overlapping standard deviation error bars, this data could be delicate in assuming any significance, further confirmed with T-tests and conclusions made from *figure 11*. This phase of the curve resembles the exponential growth of a Pichia growth curve and the curve after Day 1 resembling the death phase shown by the decline in fluorescence.

Having also normalised the curve to get more uniform data, it allows for easier comparisons to be made as it is all relative to one another. Normalisation was completed by dividing the Wild Type values with fluorescence rather that the difference between the two. *Figure 11* depicts a different story with the sample containing 1% methanol experiencing the highest peak perhaps suggesting this concentration being the most optimal to induce GFP expression despite having a slightly smaller peak in *figure 10*.

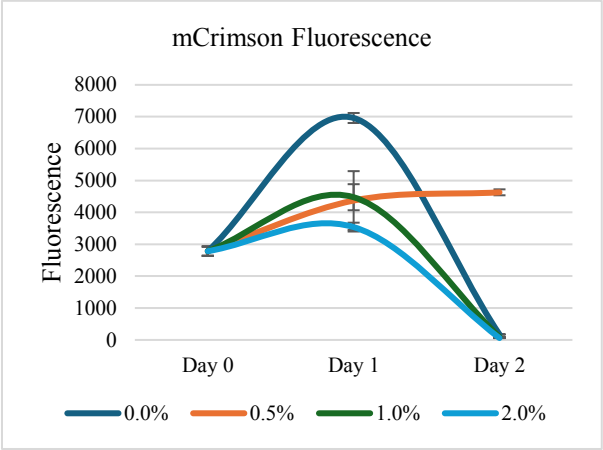


Figure 12. Line graph representing mCrimson fluorescence throughout the induction period. 0.5% stands out here and appears to continue beyond Day 2.

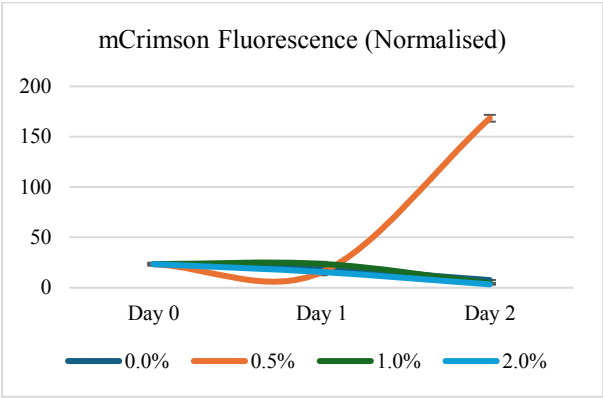


Figure 13. mCrimson fluorescence curves also normalised against

Wild Type showing that 0.5% methanol was indeed most effective in drawing protein expression with an exponential curve.

Fluorescence data was also recorded for the mCrimson strain depicting stark differences between the different samples. The 1% and 2% concentrations showed little expression with the latter being the smallest increase. Eventually by the end of the induction period, both had reached negligible fluorescence. The 0.5% curve shows that this may be the most optimal concentration of methanol to induce mCrimson expression as it does not begin to decline. However, the 0% values do seem to be an anomaly in the data set. For there to be any increase in fluorescence, there must be some addition of methanol due to the promoter. As there was no methanol added, the data appears to have discrepancies and does not align as expected.

The normalised fluorescence data for the mCrimson strain in *figure 13* clearly shows that actually most concentrations of methanol did not induce much expression and tend towards 0 as the experiment progressed. The 0.5% curve, on the other hand, exponentially rises and extrapolation would suggest continued efficacy beyond Day 2. This supports the findings from the absolute values in *figure 12*.

4.1.3 Experimental Analysis

Before the beginning of both experiments in this research project, whilst it was known that methanol would be added to the Pichia to induce growth, it was important to know how much to add and which final concentrations should be tested. It was determined that a range between 0.5% and a maximum of 3% would be used as final concentrations of methanol when growing the Pichia cultures in liquid medium [19]. Anything above just a 5% concentration would be toxic for cell viability and a minimum of 0.5% methanol was required to express recombinant proteins [25] so it was crucial to be precise in how much methanol was added to the cultures. Previous studies of similar fluorescent proteins revealed that increments of 0.5-1% were used [26] [27] and that a 48-hour induction period was most optimal for the growth of Pichia [28]. Whilst these may not have been the same proteins, it was still appropriate to use this data as a comparison. Thus, final concentrations of 0.5%, 1% and 2% were selected along with 0% as a control.

Due to the elapsed interval between the placement of the order and final delivery of a methanol assay kit, this was not used in the final experiment. Whilst this did not affect the results of the Pichia growth, it did mean that the amount of methanol in the samples on Days 1 and 2 were not able to be analysed in terms of the methanol content. This would have verified how much methanol has been used for each sample and supported any conclusion that any or all of the Pichia strains had indeed grown, along with fluorescent protein expression, in the induction period aided by the regular addition of methanol.

4.2 Encapsulation Experiment.

4.2.1 Fluorescence Data

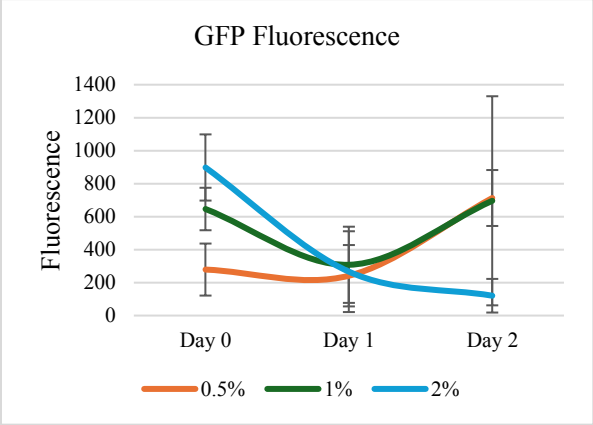


Figure 14. Line graph representing GFP fluorescence for encapsulated Pichia – note smaller values as these readings are relative due to alginate beads. Long overlapping error bars support high invariability and statistical insignificance found with T-tests in *table 1*.

Due to complications in the laboratory as well as time constraints, only GFP encapsulation data was obtained in the end. *Figure 14* shows that 2% methanol may have been too much to add to the sample as the fluorescence is on a downward trend. However, after Day 1, the other concentrations show an increase in gene expression once induced by methanol and allowed time. Since the data is normalised against the 0% concentration, this was not shown on the graph. In contrast to previous days, Day 2 data suggests high variability from the error bars, possibly affecting the precision of the readings taken.

Table 1. T-test p-values calculated for each methanol concentration each day relative to the respective previous concentrations. Values reported less than 0.05 are statistically significant compared to the control data set used [29].

	Day 0	**Day 1**	**Day 2**
0.5%	0.043	0.066	0.008
1%	0.001	0.754	0.961
2%	0.027	0.867	0.232

Table 1 contains the p-values obtained from multiple T-tests conducted to authenticate whether the data showed statistical significance. It can be concluded that there is a high degree of variability and that this proves that the increased concentrations of methanol do not have the effect of inducing gene expression as strongly as expected. To confirm this, a biological set of repeats would be useful in conjunction with the technical repeats performed in order to obtain a more accurate set of data.

Upon visual inspection, the beads appeared to be intact, and the media was still clear, indicating that

the Pichia dd not escape. This suggests that whilst there was some gene expression, there was no growth of Pichia.



Figure 15a. A close up of an encapsulated alginate bead using a microscope once strained from the Calcium Chloride-Tris solution. **Figure 15b.** The same bead with a higher magnifying power. **Figure 15c.** A zoom in on *figure 15b* in which the longer lines depict the Pichia strains and the dots depict methanol droplets.

4.2.2 Experimental Analysis

There were evident inconsistencies within the fluorescence data for GFP in the encapsulation experiment. This would likely be caused by minute differences in the volume of BMM and/or methanol in each well after using a pipette, which would affect the growth of Pichia, as well as the natural differences in measurement that the CLARIOstar plate reader would record per reading. However, another reason for this could be the size of the bead formed. When creating droplets to fall into the Calcium Chloride-Tris solution, another piece of equipment could be used such as Echo Acoustic Handlers, which use soundwaves to dispense droplets with extreme precision and reproducibility. This will ensure that the droplets are of the same size with almost negligible difference leading to a more uniform and accurate set of fluorescent measurements [30]. For ease, positive displacement pipettes can also be used to reduce inaccuracies by eliminating the air gap between a piston and the liquid resulting in the exact volume of liquid being aspirated is dispensed, with minimal accuracies [31]. These pipettes would also be advantageous to use when topping up methanol for the samples as they unaffected by highly viscous liquids due to the lack of air cushion so will not impact the amount dispensed [32].

5 Conclusion

From the liquid culture experiment, a number of conclusions can be drawn. For GFP, 1% seemed to be the better concentration to detect growth and fluorescence. Despite carrying capacity limiting the growth of Pichia, the fluorescence data shows that it was the most optimal in generating protein expression. This ultimately led to higher fluorescence readings supported by being normalised to adjust for any mishaps in the methodology applied. Disregarding the anomalous results for the 0% methanol samples for mCrimson, it is evident that 0.5% was most effective in going on to eventually produce the highest protein expression. The induction period was also not long enough to observe the full extent of growth for this sample as it was still in the exponential phase before reaching a stationary or death phase.

Whilst there may not have been a full set of results for all Pichia strains in the encapsulation experiment, there can be some clear inferences drawn. For GFP, there was some gene expression induced by the addition of methanol with both 0.5% and 1% being the most effective. But statistical analysis suggests that the data may not be as significant as expected, possibly hinting at the conclusion that Pichia may not be able to generate gene expression while encapsulated. Further repeats and variety of data would be required to confirm this. After inspecting the beads, there was also no growth, which can be tested further for verification.

For future outlook, this research can take a step forward by further analysing the cellular growth of Pichia in encapsulation. Once confined in the bead, OD readings cannot be taken and even GFP or mCrimson readings are only relative to each other. To find the true value of these readings and OD_{600}, EDTA can be introduced to dissolve and liquidise the beads allowing for more accurate measurements of optical density. Using the OD cell correlation for Pichia (see *section 3.4.2)* the number of cells can be determined to establish if there was, or how much growth occurred. These results would be able to be supported and verified by the trends observed in the previous liquid culture experiment. This research can be applied to a variety of biotechnological industries such as the wastewater industry. This would allow for real-time detection of pollutants such as methanol in water bodies through fluorescence measurements.

6 Acknowledgements

The authors would like to thank Dr Masue Marbiah and the Polizzi Lab at Imperial College London for their immense help and advice throughout the duration of this research project.

7 References

[1] Mastropietro, G., Aw, R. and Polizzi, K.M. (2021). Expression of proteins in Pichia pastoris. Methods in Enzymology, [online] 660, pp.53–80.
[2] Pichia Fermentation Process Guidelines Overview Introduction. (n.d.).
[3] Heidebrecht, A. and Scheibel, T. (2013). Chapter Four - Recombinant Production of Spider Silk Proteins. [online] ScienceDirect.

[4] Miladys Limonta-Fernández, Glay Chinea-Santiago et al. (2022). An engineered SARS-CoV-2 receptor-binding domain produced in Pichia pastoris as a candidate vaccine antigen. New biotechnology, 72, pp.11–21.
[5] Juturu, V. and Wu, J.C. (2017). Heterologous Protein Expression in Pichia pastoris: Latest Research Progress and Applications. ChemBioChem, [online] 19(1), pp.7–21.
[6] Saravanan, P., Kumaran, S. et al (2021). Bioremediation of synthetic textile dyes using live yeast Pichia pastoris. Environmental Technology & Innovation, 22, p.101442.
[7] Vuree, S., Prajapati, G.K. and Singh, S. (2020). Recombinant Pichia pastoris and its applications in healthcare industry. [online] Elsevier, pp.219–229.
[8] Haavisto, V. (2023). Plasmids and the Spread of Antibiotic Resistance Genes. [online] American Society for Microbiology.
[9] Letchumanan, V., Chan, K.-G. and Lee, L.-H. (2015). An insight of traditional plasmid curing in Vibrio species. Frontiers in Microbiology, 6.
[10] Roy, S., Naha, S., Rao, A. and Basu, S. (2021). CRISPR-Cas system, antibiotic resistance and virulence in bacteria: Through a common lens. Progress in Molecular Biology and Translational Science, pp.123–174.
[11] Tsilo, P.H., Basson, A.K., Ntombela, Z.G., Maliehe, T.S. and Pullabhotla, V.S.R.R. (2022). Production and Characterization of a Bioflocculant from Pichia kudriavzevii MH545928.1 and Its Application in Wastewater Treatment. International Journal of Environmental Research and Public Health, 19(6), p.3148.
[12] Carvajal-Vallejos, P.K., Planas, A. and Barbera, E. (2007). Immobilized Pichia pastoris in alginate beads for continuous culture. Afinidad -Barcelona-, [online] 64(529), pp.384–389.
[13] Lenassi Zupan, A. (2004). High expression of green fluorescent protein in Pichia pastoris leads to formation of fluorescent particles. Journal of Biotechnology, 109(1-2), pp.115.
[14] Yin, W., Wang, X., et al. (2022). Encapsulating IM7-Displaying Yeast Cells in Calcium Alginate Beads for One-Step Protein Purification and Multienzyme Biocatalysis. Frontiers in bioengineering and biotechnology, [online] 10.
[15] Catanzaro, Kelly & Lahmers, Kevin & Allen, Irving & Inzana, Thomas. (2022). Alginate microencapsulation of an attenuated O-antigen mutant of Francisella tularensis LVS as a model for a vaccine delivery vehicle. PLOS ONE. 17. e0259807. 10.
[16] Flamminii, F., Di Mattia, C.D. et al. (2020). Structuring alginate beads with different biopolymers for the development of functional ingredients loaded with olive leaves phenolic extract. Food Hydrocolloids, 108, p.105849.
[17] Lee, K.Y. and Mooney, D.J. (2012). Alginate: Properties and biomedical applications. Progress in Polymer Science, [online] 37(1), pp.106–126.
[18] www.labxchange.org. (2021). LabXchange. [online]
[19] Life Technologies 01.2014 (2014). Pichia Expression Kit For expression of recombinant proteins in Pichia pastoris. [online]
[20] Haon, M., Grisel, S., Navarro, D. et al. (2015). Recombinant protein production facility for fungal biomass-degrading enzymes using the yeast Pichia pastoris. Frontiers in Microbiology, 6.
[21] Liu, W.-C., Gong, T. et al. (2016). Scaling-up Fermentation of Pichia pastoris to demonstration-scale using new methanol-feeding strategy and increased air pressure instead of pure oxygen supplement. Scientific Reports, [online] 6.
[22] The National Institute for Occupational Safety and Health (2019). CDC - The Emergency Response Safety and Health Database: Systemic Agent: METHANOL - NIOSH. [online] CDC.
[23] McLachlan, A. (2006). Coastal Zone Management. Elsevier eBooks, pp.303–323.
[24] Atroshenko, D.L., Sergeev. et al. (2024). Additivities for Soluble Recombinant Protein Expression in Cytoplasm of Escherichia coli. Fermentation, [online] 10(3), p.120.
[25] Karbalaei, M., Rezaee, S.A. and Farsiani, H. (2020). Pichia pastoris: A highly successful expression system for optimal synthesis of heterologous proteins. Journal of Cellular Physiology, [online] 235(9), pp.5867–5881.
[26] Dharmarathna, C.S., Kumar, S., et al. (2023). AOX1 promoter-driven expression of yeast-enhanced green fluorescent protein in Pichia pastoris. The Microbe, [online] 1, p.100012.
[27] Naseem, M.U., Tajti, G., et al. (2021). Optimization of Pichia pastoris Expression System for High-Level Production of Margatoxin. Frontiers in Pharmacology, 12.
[28] Santoso, Adi; Herawati, Neng; and Rubiana, Yana (2012) "Effect of Methanol Induction and Incubation Time on Expression of Human Erythropoietin in Methylotropic Yeast Pichia Pastoris," Makara Journal of Technology: Vol. 16: Iss. 1, Article 5.
[29] University of Southampton (2020). T test | Practical Applications of Statistics in the Social Sciences | University of Southampton. [online] Southampton.ac.uk.
[30] www.mybeckman.uk. (n.d.). Echo Acoustic Technology. [online]
[31] BNO Team (2024). Positive Displacement Pipette – Principle, Procedure, Applications, Advantages. Biology Notes Online.
[32] Henderson, T. (2017). How Positive Displacement Pipettes Work: A Guide for Precision Liquid Handling. [online] Lab Manager.

Effect of Agitation on Isothermal Paracetamol Crystal Regeneration

Helena Rigg Franco and Alice Woodcock

Department of Chemical Engineering, Imperial College London, U.K.

Abstract

Crystal regeneration in paracetamol-ethanol systems is crucial for pharmaceutical applications; yet the role of agitation on crystal regeneration remains unclear. This study investigates the effect of varying agitation speeds (0 to 200 rpm) on crystal regeneration at a supersaturation ratio of 1.1. Crystal growth was monitored through regular imaging, revealing that agitation significantly influenced secondary nucleation, reducing its onset time (from 1.75 hours at 100 rpm to 0.75 hours at 200 rpm) and its extent, thereby hindering growth. Growth rates prior to secondary nucleation were higher for agitated systems (on average 0.4588 pixel/min) compared to non-agitated ones (0.1969 pixel/min), though no clear trend was observed with increasing agitation speed, suggesting that growth is governed by surface integration. Lower supersaturation levels were found to be more suitable for this investigation, albeit at the expense of longer experimental duration. These findings highlight the nuanced role of agitation in crystallisation, with future studies needed to explore a wider range of agitation speeds and increased experimental repetitions to refine these insights.

Keywords: Agitation, Crystal Regeneration, Crystal Growth, Nucleation, Paracetamol.

Introduction and Background

Crystallisation is a fundamental process present within the chemical and pharmaceutical industries, used to create highly ordered and structured arrangements of a solid from solute atoms, ions or molecules. Crystallisation occurs in two main steps: nucleation, primary or secondary, and crystal growth. Primary nucleation is the initial formation of crystal nuclei in a supersaturated solution without the presence of existing crystals. Secondary nucleation occurs in the presence of existing crystals.

Primary nucleation takes place in supersaturated solutions, where solute molecules spontaneously form stable nuclei, driven by the thermodynamic need to minimise free energy. This can occur as homogeneous nucleation, forming without external surfaces, or as heterogeneous nucleation, where foreign surfaces reduce the activation energy required for nucleus formation [1].

Secondary nucleation occurs through mechanisms such as initial breeding, where crystal fragments act as seeds for further growth; contact nucleation, triggered by collisions between particles or surfaces; and shear breeding, where mechanical forces like shear stress break crystals into smaller fragments which become secondary nuclei [2].

Crystallisation occurs when a solution becomes supersaturated, having more solute than present at saturation, making it thermodynamically unstable. The metastable zone lies between the supersaturation limit and the solubility curve, where no spontaneous nucleation occurs unless crystal seeds are present. The metastable zone width (MSZW) is the temperature range of this zone during cooling. Within the metastable zone, well-defined crystals grow rather than forming fine precipitates, allowing for controlled growth. The MSZW depends on factors such as cooling rate, agitation, impurities and solution properties [3].

Crystal growth is a complex process which can be simplified into two distinct stages: bulk diffusion/mass transfer, and surface integration. In the bulk diffusion stage, solute particles travel through the molecular boundary layer towards the crystal surface. During surface integration, these solute molecules are then incorporated into the crystal lattice [4]. These two stages are visualised in Figure 1.

Surface integration involves four main steps. First, solute molecules diffuse through the solution to reach growth sites like kinks on the crystal surface. Next, desolvation occurs, where the solute molecules detach from the solvent molecules. The desolvated molecules then adsorb onto the crystal surface, and finally, they are incorporated into the

lattice at favourable sites, contributing to crystal growth [5].

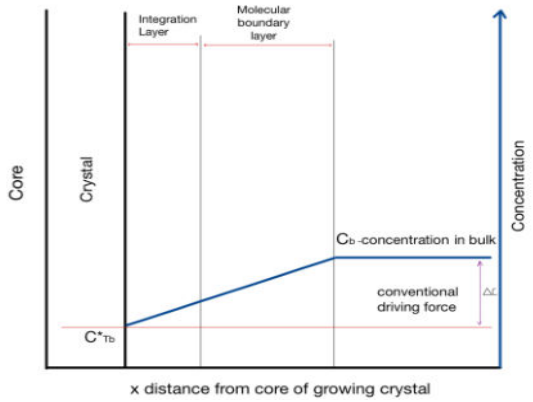


Figure 1 - Linearised profiles of conventional crystallisation driving forces, displaying the layers of crystallisation

Crystal regeneration is a new phenomenon where a cleaved crystal is observed to regrow back to its original shape pre breakage. Although it is a new area of research it has been found that the internal plane (010) is the preferred cleavage plane for crystal regeneration. This is unexpected as according to morphology laws it has the lowest attachment energy. It was expected to grow the slowest, but experiments showed it exhibited much faster growth compared to other facets and around 3 times faster than growth before breakage [6] [7].

The effects of agitation on crystal regeneration have been previously investigated, but only briefly, and it was observed to have quickened this process. However, it was determined that much more research was necessary to obtain definitive conclusions regarding this [6]. Lastly, cooling crystallisation was chosen for this experiment over evaporative due to it being widely used in industrial crystallisers.

This area of research has importance in industry as mechanical breakage often occurs within industrial crystallisers. Especially along the cleavage planes, because they are the weakest in the crystal structure. This research covering growth after cleavage helps to attain better understanding of crystal behaviour post breakage. This is useful for applications in the pharmaceutical industry: it could help improve crystalliser performance and downstream processes such as milling and tableting. This research can contribute to the development of more accurate process models, enabling better predictions of crystal behaviour within crystallisers. This, in turn, could lead to improved control and optimisation of crystallisation processes [7].

Aims

The objectives of this study were to investigate the impact of agitation on secondary nucleation, as well as its effect on growth rate. Thirdly, we aimed to briefly investigate the influence of supersaturation ratios with agitation present.

Methodology

Materials

Ethanol Absolute (C_2H_5OH, CAS-No:64-17-5, >99.8% purity) was purchased from VWR Chemicals. Paracetamol (Acetaminophen) powder ($CH_3CONHC_6H_4OH$, CAS-No: 103-90-2, >98% purity) was purchased from Sigma-Aldrich and meets USP testing specifications. Reagents' structural formulae can be seen in Figure 2. All reagents were used without further purification as they were of analytical grade.

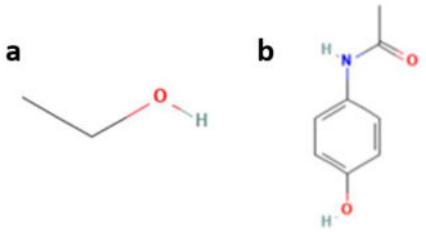


Figure 2 - structural formula of (a) ethanol and (b) paracetamol Sourced: [8] and [9]

Preparation of crystal seeds

The first step was to prepare a saturated paracetamol solution.100 ml of ethanol and excess paracetamol powder were mixed in a glass bottle. A magnetic stirrer was used to mix the solution for 24 hours at roughly 400 rpm. The solution was then taken off the magnetic stirrer plate for an hour to allow the excess paracetamol powder to settle at the bottom of the glass bottle before the solution at the top could be used to grow crystals.

To grow seeds, 10 ml of the saturated solution was filtered through cellulose acetate membrane filters (pore size 2 µm) using a 10 ml plastic syringe into a small beaker. The top then was covered by parafilm and using sharp tweezers some holes were punctured into the parafilm to allow for some level of controlled evaporation. The beaker was then left to sit in the fume hood overnight, checking on it daily till the seeds were ready to harvest. Using

tweezers, promising seeds were picked out, washed with a very small amount of ethanol then stored.

The process of growing seeds into larger crystals began by placing a seed into a small vial, adding 2 ml of the saturated solution via cellulose acetate membrane filters (pore size 2 μm). Then the vials were covered with parafilm, and two holes were poked in the film with tweezers. The vials were then left in the fume hood until the crystals had reached an appropriate size to be used for the experiment (0.7 ± 0.2 cm for this investigation).

When preparing and growing the crystal seeds, variability of the crystal habit was noticed, making it difficult to consistently grow suitable crystals (the desired morphology can be seen in Figure 3(c)). This is because the crystal seed habit is highly influenced by the seeding conditions. An example of this can be seen in Figure 3 which displays the different morphologies/habits of paracetamol crystals grown at different supersaturation ratios. The supersaturation was not controlled; therefore, it was inevitable that our seeds would vary in shape. Some crystals of the shape seen in Figure 3(b) were also used. To address this a plate with 24 wells was used instead of the individual vials to efficiently maximise our chances of having desirable crystals available. The seeds were also placed in the centre away from the walls to prevent any deformations and checked on regularly.

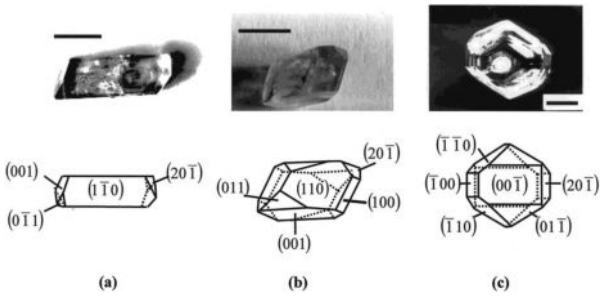


Figure 3 - Morphology of paracetamol crystals grown at (a) low, (b) medium and (c) high supersaturation. showing the observed forms of the crystals Source: [10]

Crystal Cutting

The crystals were cleaved by applying a small amount of pressure parallel to facet (010) (as can be seen in Figure 4) with a scalpel. This cut the crystal into two exposing the internal preferred facet for the crystal regeneration. This study focused only on the regeneration of paracetamol crystals cleaved in this way.

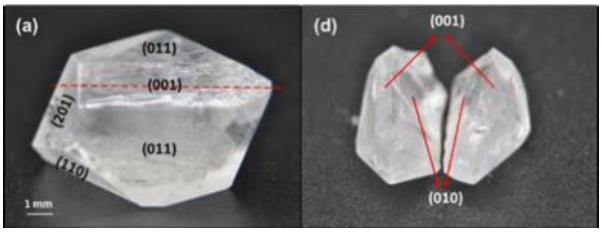


Figure 4 - Macroscopic crystal of monoclinic paracetamol (a) whole crystal, (d) cleaved crystal. (010). Source: [7]

Crystal Regeneration

Another saturated paracetamol solution was prepared by mixing excess paracetamol powder with approximately 100 mL ethanol in a glass bottle and stirring at 400 rpm and 30°C, using a jacketed vessel and water bath for 24 hours. The mixture was then left to settle for 1 hour to separate excess powder from the saturated solution. 60 mL of the saturated solution was filtered through cellulose acetate membrane filters (pore size 2 μm) into a screw-cap glass bottle. The filtered solution was then equilibrated at 30°C for 2 hours in the jacketed vessel. After this, the temperature was reduced (e.g., to 25.5°C for a supersaturation of 1.1) and left for 2 hours, stirring at 0 or specified rpms (e.g., 100–200).

When deciding the range of agitations implemented, a couple of factors had to be taken into consideration. At 200 rpm, excessive secondary nucleation was observed which made it difficult for image analysis to be carried out. Thus, this was chosen as our maximum agitation. 100 rpm was chosen to be the minimum agitation possible due to the limitations of the magnetic stirrer plate. The dial for the plate was found to be uncalibrated. To address this, the magnetic stirrer was videoed and manually counted to ensure it was at the desired speed; if not, it was adjusted until this was the case. As the experiments were being carried out at relatively low agitation speeds, a test was conducted to ensure mixing was occurring efficiently. This was done using water and dye, with a glass rod in the solution to imitate the experiment conditions. This test confirmed proper mixing was occurring.

The supersaturation ratio of the solution was achieved by the temperature the solution was cooled to. To achieve our chosen supersaturation ratio (S = 1.1) the temperature required was calculated using solubility data of paracetamol at 30˚C and Eq.1.

$$S = C/C^* \qquad \text{Eq. 1}$$

where C is the concentration and C^* is the saturation concentration.

The cleaved crystal was then immobilised by glueing it onto a 1 mm soda lime glass capillary with Loctite super glue (cyanoacrylate adhesive) and then allowed to cure for 2–3 hours. The capillary was passed through a pre-drilled lid and secured with blue tack. This new lid was placed on the cooled solution.

The crystal was regrown at the set temperature for 24 hours at a time. Growth was recorded with a high resolution DSLR camera in a lightbox within a fume hood. The camera was set to take a photo every 30 minutes using a timer and the growth was checked at least daily to determine if the crystal had finished regenerating. At the experiment's end, the crystal was transferred to a new vial for analysis and the solution was returned to the original bottle for reuse.

Quantifying Experimental Results

Image analysis was an integral part of quantifying the experimental results. Videos were created from the photos for a visual aid to determine when growth had ceased and when secondary nucleation started. The time at which secondary nucleation occurred was defined to be the point when the solution appeared to have turned cloudy, indicating the presence of secondary nuclei in the solution [11]. However, it must be noted that secondary nucleation would have occurred earlier than was visually apparent, as initially nuclei are too small to be captured on camera.

MATLAB was used for image processing by writing code that utilised MATLAB's built in functions. There were some initial challenges as the code incorrectly detected the liquid level and glass rod as part of the crystal when using the raw images. This was resolved by adding a manual crop feature using the **`imcrop`** function. A rotation feature that allowed input of a specific angle of rotation anticlockwise for the images was also implemented. This was critical for accurate crystal measurements and analysis to ensure the crystal surface was correctly orientated, so that it was parallel to the x-axis. The length of the crystal was found at each time interval using the **`regionprops`** function, which calculates properties such as bounding boxes and centroids, these features can be seen in Figure 5. These measurements, recorded in pixels, were visualised through plots and organised into tables with **`uitable`**.

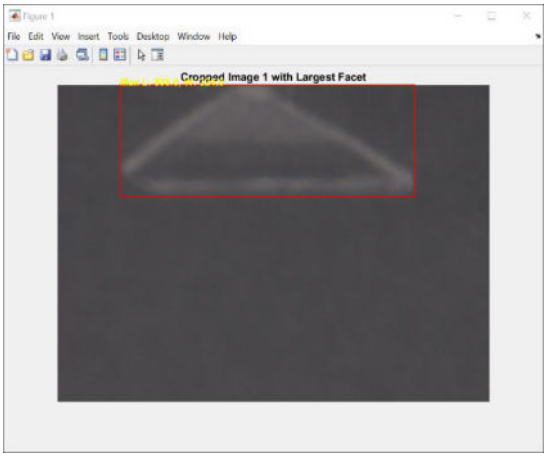


Figure 5 - Example shot of rotated and manually cropped image on MATLAB and the region (red box) which it measures.

Data was extracted from MATLAB into excel to calculate the growth rates. This was done by plotting the length of each crystal against time; the gradient of the best fit lines gave the growth rate.

Results and Discussion

General Crystal Habit

Figure 6 illustrates the final regenerated habit of some of the paracetamol crystals, displaying their shape and overall clarity following the regeneration process. Looking at Figure 6(a) this crystal's habit aligned with our predictions from literature as the cleaved crystal grew back to its shape pre-breakage. However, some crystals did not follow this behaviour and grew back to unexpected habits like in Figure 6(b). It was also observed that the regenerated half of some of these crystals exhibited reduced clarity compared to the rest of the crystal as in Figure 6(a).

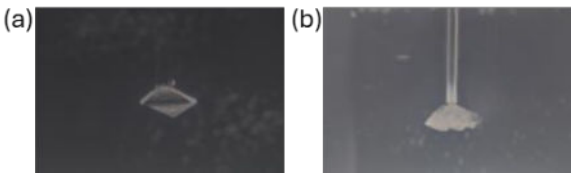


Figure 6 - Photos displaying crystal regeneration at agitation speed of 100 rpm, at time (a) 31 hours and (b) 29 hours

Upon cleavage, mechanical stress is experienced by the crystal which can lead to strain in the crystal material. Strain can manifest within the crystal in the form of defects such as dislocations [12]. Consequently, upon crystal growth it is likely that

a perfect lattice structure would not be obtained. This led to both this cloudy appearance of the regenerated section and unexpected crystal habits. It is important to refer to impurities here: one example of an impurity that is present is the glue that was used to secure the crystal onto the glass rod. If present, impurities in the solution could potentially change crystal habit [13].

To fully understand the extent of these defects in the crystals and impurities in the system more advanced imaging and other characterisation techniques should be used.

Secondary Nucleation

It was observed that secondary nucleation did not occur in non-agitated runs. This was not the same for the agitated cases showing us that the introduction of agitation induced secondary nucleation for this supersaturation ratio and within this time frame. Table 1 shows us that for all agitations, the average time for nucleation to occur was less than 2 hours. It also shows that increasing the agitation speed reduced the average time taken for the onset of secondary nucleation: it decreased from 1.75 hours at 100 rpm to 1 at 150 rpm and then 0.75 at 200 rpm.

Table 1 – Average time taken for secondary nucleation to occur at each agitation speed.

Agitation Speed (rpm)	Av. Time Taken for Secondary Nucleation to Occur (hours)
0	-
100	1.75
150	1.00
200	0.75

As noted before, the onset of secondary nucleation was identified by observing changes in cloudiness of the solution [11]. In some cases, the presence of nucleation was observed not only by the presence of secondary nuclei in solution (cloudiness) but by the observed growth of seeds on the crystal, glass rod, and walls of the vessel. This can be seen in Figure 7 where at 150 rpm seeds are growing on the surface of the crystal and at 200 rpm they are growing on the surface, glass rod and walls.

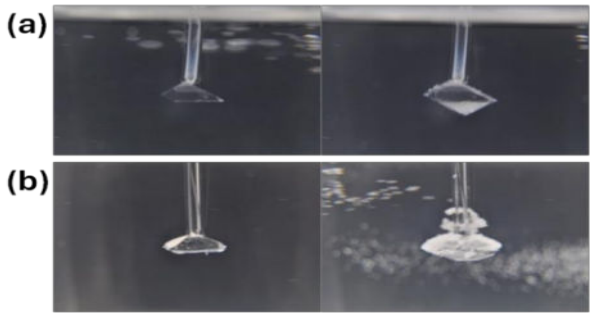


Figure 7 - Photos displaying extents of secondary nucleation at different agitation speeds of (a) 150 rpm, and (b) 200 rpm. Photos at time 0 hours and (a) 19.5 hours, (b) 30 hours.

As previously mentioned, the cleavage process can cause defects on plane (010), increasing the likelihood of small fragments being present on this surface. Higher agitation rates introduce higher fluid velocities within the solution, leading to increased fluid shear. Through greater shear forces being exerted on the crystal, micro fragments may be removed from the regenerating crystal surface and moved into the bulk solution [2]. These fragments would then be considered secondary nuclei and would begin to grow in solution and/or deposit on crystal surface/rod as seen in Figure 7. This is an example of the mechanism of shear breeding.

Secondly, higher fluid velocities within the solution, due to agitation, will increase the collision frequency between these fragments and the glass rod, vessel walls, and crystal, and between the fragments themselves. This process, known as contact nucleation, leads to an increased formation of secondary nuclei due to these collisions. From literature it was found that at a chosen level of supersaturation, the number of observed nuclei increased with rising shear (rpm) [2]. This validates our results: higher agitation rates increase the level of secondary nucleation.

The presence of crystals on the glass rod, vessel walls, and the crystal itself can be attributed to heterogeneous (primary) nucleation and/or secondary nucleation. The glass rod and vessel walls provide surfaces that lower the energy barrier for nucleation, facilitating heterogeneous nucleation in the supersaturated solution [1].

It is unclear whether these crystals have formed from heterogenous nucleation or mechanisms of secondary nucleation, as these processes can occur simultaneously with similar outcomes, and it is challenging to observe the exact moment or location of nucleation.

There was also inevitably human error involved in a few aspects; for example, the presence of impurities. The use of glue on the crystal may have introduced impurities into the solution and there was a chance that other impurities may have been present. This would have interfered with the level of secondary nucleation that was observed. It could have amplified it by increasing contact nucleation. Impurities also provide additional surfaces within the solution facilitating heterogenous nucleation.

Also, the determination of the time taken for secondary nucleation to occur can be considered relatively subjective as it involves identifying a slight colour change due to cloudiness of solution which can be hard to recognise.

Growth Rates

Table 2 – Length growth rate before secondary nucleation at each agitation speed.

Agitation Speed (rpm)	Growth Rate before Secondary Nucleation (pixel/min)
0	0.1969
	-
100	0.3000
	0.5352
150	0.1667
	0.6167
200	0.3833
	-

As previously mentioned, secondary nucleation occurred early: within the first few hours of this experiment for all agitation speeds. This complicates comparisons of growth rates beyond this point, as the extent of secondary nucleation remains unclear and is difficult to quantify. Consequently, it cannot be assumed that the supersaturation level was consistent across all experiments. Thus, valid comparisons are not able to be made. Therefore, only comparisons of growth rates prior to the onset of secondary nucleation were made. The growth rates observed after secondary nucleation began can only provide insights into how secondary nucleation may have influenced them.

Firstly, Table 2 shows that in general, agitated runs had higher growth rates compared to the non-agitated ones. There is one exception at 150 rpm where the initial growth rate calculated was 0.1667 pixels/min and the non-agitated growth rate was 0.1969 pixels/min. This data point was therefore treated as an anomaly.

These higher growth rates are the result of continuous fluid motion introduced through the presence of agitation. Therefore, allowing for continuous flux of solute molecules towards the crystal surface. Agitation also reduces the thickness of the molecular boundary layer. Both factors combined contribute to the enhanced mass transfer of solute to the crystal surface therefore increasing the growth rate.

There was no clear trend observed between increasing agitation speed and the growth rates prior to secondary nucleation. It would be expected that the growth rate would increase with agitation speed due to enhanced mass transfer. However, as this was not the case it was concluded that diffusion/mass transfer was no longer the limiting step in the crystal growth process, and instead it was governed by surface integration.

In crystal regeneration, agitation primarily enhances the mass transfer of solute molecules to the crystal surface. However, it has minimal direct influence on stages in surface integration such as desolvation, adsorption, and incorporation, as these steps are governed by molecular interactions and the thermodynamics at the crystal surface. These processes depend on intrinsic crystal properties rather than on the external agitation conditions.

When looking at Figures 8-11, which display the length of crystal over time, the growth rate of the crystal after secondary nucleation occurs can be observed. For all agitation speeds, the presence of secondary nucleation hindered the growth rate of the crystals. This is because the secondary nuclei are now competing with the regenerating crystal for solute. From literature, it was found that at higher agitation speeds, supersaturation decreases more rapidly [14]. This is due to increased secondary nucleation, which creates additional crystal surfaces for solute deposition. This validates the idea proposed that the presence of secondary nucleation decreases crystal growth rate.

Conventionally in crystallisation, nucleation occurs first, followed by growth. However, in this system, both processes happen simultaneously. The regenerating crystal, along with secondary and potentially heterogeneous nuclei, are present, allowing crystal growth to occur alongside new nucleation. This simultaneous occurrence of nucleation and growth highlights the complexity of the system, where both processes happen in parallel rather than sequentially.

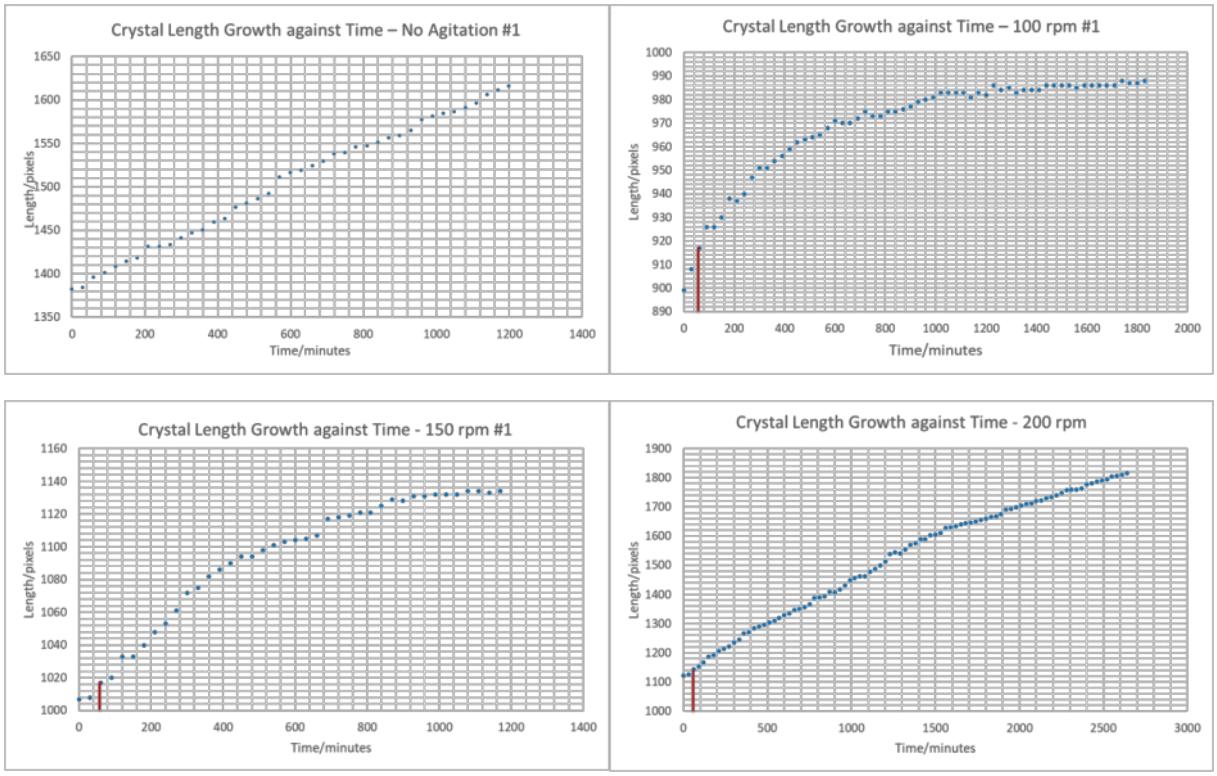


Figure 8-11 - Length against time for first run at 0 rpm (8, top-left), 100 rpm (9, top-right), 150 rpm (10, bottom-left), and 200 rpm (11, bottom-right). The red line indicating when secondary nucleation began.

Several experimental limitations also impacted the reliability and accuracy of the results. First, the insufficient number of data points for growth rate calculations reduced the accuracy of the measurements. Tackling this would require reducing the interval time between photos to increase data collected before secondary nucleation begins.

Secondly, issues with image processing and photo quality, compounded by barriers such as glass and water and the limitations of the camera, affected the precision of measurements. Improvements to the MATLAB script for detecting crystal edges and higher-quality imaging equipment could mitigate these issues. The current setup's lack of sensitivity may have also prevented the detection of subtle changes, potentially masking correlations.

Supersaturation Ratios

The supersaturations ratios were varied while an agitation speed of 100 rpm was kept constant.

From image analysis it is evident that at a lower supersaturation of 1.05, regeneration takes significantly longer compared to a supersaturation of 1.1 (average time of 29 hours at 100 rpm). The regeneration experiment was run for 35.5 hours at the lower supersaturation and the crystal was not complete in its regrowth. When comparing to 1.1 it can be seen that this growth is evidently slower.

At lower supersaturations, crystal growth is slower due to the reduced concentration of available solute. With fewer solute particles in the solution, growth depends on the deposition of these particles onto the crystal surface. The low solute concentration limits the diffusion rate to the regenerating crystal, thereby reducing the overall growth rate.

At the higher supersaturation of 1.15, the process was rapid, but there was a disruptive level of secondary nucleation observed. This is evident in Figure 12, where the solution rapidly becomes very cloudy suggesting the presence of crystal seeds in solution. Crystals can be observed on the surface of the crystal as well as the glass rod. Although growth rates could not be calculated due to the high level of secondary nucleation, it can be inferred that at higher supersaturation, the regeneration process accelerates, as the solute availability is larger. The increased solute concentration promotes faster deposition onto the regenerating crystal surface but also promotes a higher rate of secondary nucleation. Although regeneration is quicker, the risk of data analysis being disrupted by excessive secondary nucleation increases, as the solution becomes more prone to forming new

crystal seeds due to the higher levels of supersaturation.

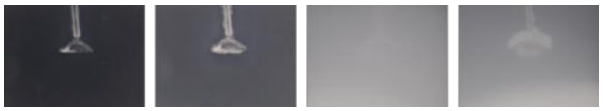


Figure 12 - Photos showing disruptive amount of secondary nucleation in the experiment run at a supersaturation ratio of 1.15, taken over 19 hours.

Through carrying out this experiment at ratios of 1.05 and 1.15, the choice to run our investigation at 1.1 was validated. It was an optimal balance of experiment duration and minimal secondary nucleation, ensuring reliable results.

The onset time for secondary nucleation was very short, limiting the observation window during which only crystal growth occurred. Using lower supersaturation levels would extend this period, allowing for clearer measurements of growth rates without interference from secondary nucleation. A lower supersaturation would also allow for greater agitation rates to be studied qualitatively.

Conclusions

Overall, this study highlighted the critical role of operating parameters in industrial crystallisation processes, particularly in industries such as pharmaceuticals and specialty chemicals, where precise control over crystal properties is essential. Agitation was found to induce secondary nucleation, reducing the time to its onset and increasing its extent at higher agitation rates. The onset of secondary nucleation hindered crystal growth, disrupting the regeneration process. While agitation enhanced initial growth rates before the onset of secondary nucleation, no clear trend was observed when increasing agitation. This lack of trend suggests that growth rates were governed by surface integration limitations beyond a certain threshold, indicating that growth rates become independent of further agitation.

Understanding this threshold is essential for designing more efficient crystallisation reactors, enabling consistent product quality while minimising energy consumption. However, the lack of accurate models for surface integration-dominated growth remains a challenge, as crystal growth is a complex process often modelled using empirical correlations that fail to fully represent the actual physical phenomena [4].

Secondary nucleation, the dominant nucleation mechanism in industrial crystallisers, is crucial as for controlling crystal formation and process efficiency [2]. Understanding its role in crystal regeneration, particularly the impact of shear forces, aids in optimising crystalliser design and improving product quality in industries like pharmaceuticals and food processing.

Additionally, this study highlights that lower supersaturation ratios would be more suitable for future investigations, as they would extend the observation window before secondary nucleation occurs, facilitating more precise data collection and analysis.

Outlook

Future studies could expand the range of agitation speeds and increase data collection by imaging at smaller intervals, enabling for more decisive results. Further development of the MATLAB code to incorporate facet specific growth measurements could prove beneficial, however this would depend on the improvement of the image quality. By using other characterisation techniques such as Time-of-Flight Secondary Ion Mass Spectrometry (TOF-SIMS) the extent of a crystal's defects can be further explored and quantified. It might also be insightful to explore the effects of agitation on the regeneration of crystals cleaved to different percentages as it could provide awareness into optimising conditions.

In the long-term, models could be developed to simulate crystal behaviour post mechanical breakage, assessing how agitation-induced changes in crystal size and shape influence downstream properties like dissolution rates, bioavailability, and tablet compressibility. It could also enable better scaling from laboratory to industrial processes. Research into scaling up crystal regeneration, including testing different types of agitation, could be valuable for industrial applications. The study could also be extended to encompass other active pharmaceutical ingredients (APIs) to see if the agitation-crystallisation relationships for paracetamol apply more broadly. Finally, studying systems with multiple crystals would better reflect industrial conditions and provide insight into crystal-crystal interactions. These advancements would improve understanding of crystallisation, especially its relationship with agitation and enhance the study's relevance in pharmaceutical manufacturing.

Acknowledgements

We would like to sincerely thank Isha Bade for her invaluable support, guidance and constructive feedback throughout the course of this study. We also extend our gratitude the Heng Research Group.

References

[1] n.d, "Tec-Science," 2018. [Online]. Available: https://www.tec-science.com/material-science/solidification-of-metals/heterogeneous-nucleation/. [Accessed December 2024].

[2] S. Agrawal and A. Paterson, "Secondary Nucleation: Mechanisms and Models," 2015.

[3] Y. Zhang, "6 - Modeling and simulation of batch and continuous crystallization processes," in *Modelling of Chemical Process Systems*, S. A. Imtiaz, Ed., Elsevier, 2023, pp. 181-202.

[4] A. Borissova, "Driving Force of Crystallization Based," 2019.

[5] M. Mazotti, "Rate-Controlled Separations in Fine Chemistry," 2021. [Online]. Available: https://ethz.ch/content/dam/ethz/special-interest/mavt/process-engineering/separation-processes-laboratory-dam/documents/education/RCS_2021/Crystallization/2021_Predicting_crystal_morphology.pdf.

[6] I. Bade and J. Heng, "Effect of Agitation during Cooling Crystallization on the Unique Regeneration Phenomenon Observed in Paracetamol Crystals Post Breakage," 2023.

[7] I. Bade, V. Verma, I. Rosbottom and J. Y. Y. Heng, "Crystal regeneration – a unique growth phenomenon observed in organic crystals post breakage," Royal Society of Chemistry, 2023.

[8] N. C. f. B. Information, "PubChem Compound Summary for CID 70, Ethanol.," 2024. [Online]. Available: https://pubchem.ncbi.nlm.nih.gov/compound/Ethanol. [Accessed 12 2024].

[9] N. C. f. B. Information, "PubChem Compound Summary for CID 1983, Acetaminophen.," 2024. [Online]. Available: https://pubchem.ncbi.nlm.nih.gov/compound/Acetaminophen. [Accessed 12 2024].

[10] R. Ristic, S. Finnie, D. B. Sheen and J. N. Sherwood, "Macro- and Micromorphology of Monoclinic Paracetamol Grown from Pure Aqueous Solution," Septemper 2001. [Online]. Available: https://www.researchgate.net/figure/Macromorphology-of-paracetamol-crystals-grown-at-a-low-b-medium-and-c-high_fig1_231629174. [Accessed 12 2024].

[11] A. Hussein, "Chapter 4 - Principles of Flow Assurance Solids Formation Mechanisms," in *Essentials of Flow Assurance Solids in Oil and Gas Operations*, Gulf Professional Publishing, 2023, pp. 143-197.

[12] J. Morris, in *A Survey of Materials Science IV. Mechanical Properties*, 2008, pp. 76-107.

[13] M. Nosan, "Understanding Crystallographic Defects: Imperfections in Crystalline Structures," 2024.

[14] P. J. Frawley, N. A. Mitchell, C. T. Ó'Ciardhá and K. W. Hutton, "The effects of supersaturation, temperature, agitation and seed surface area on the secondary nucleation of paracetamol in ethanol solutions," *Chemical Engineering Science,* vol. 75, pp. 183-197, 2012.

Author Index

Linked to paper number

Abdalla, Mariam 39
Adeyemi, John 3
Agrawal, Pratham 73
Ahadi, Bellal 9
Ahmed, Saarah 16
Akinnola, Oluwakemi 33
Alami, Sounni Karim 46
Al Khalifa, Hamad 44
Ali, Abdullah 68
Ali, Nasem 53
Arif, Danyar 12
Ates, Ayce I 30
Azeem, Zain 58
Bavaharan, Bhargavi 65
Beddows, Isabelle 64
Bell, Emma 18
Bentley, Isobel 59
Bin Nor Azam, Kamarul Izzat 40
Cambridge-Lau, Reuben 17
Cerullo, Elise 47
Chan, Kar Jan 62
Charlton, Thomas 67
Chen, Bide 67
Chen, Boren 27
Chen, Tachun 45
Chen, Xin 39
Cheng, Xinyu 21
Christie, Leonidas 56
Cho, Yongbin 13
Chukwuemeka, Mitchel 25
Corbett, Hannah 29
Coy, Hannah 48
Dam, Irene 52
Deinde, Olusola 71
Dolbel, Thomas 13
Echeverría de Encio, Álvaro 54
Elyamani, Sohaib 26
Fung, Chun Wai 57
Gao, Yingshan 66
Ghafir, Yasmine 52
Goh, Braxton 5
Goh, David 36
Gregory, Billy 35
Gylyjova, Aylar 43
Hajdu, Gergely 61
Harper, Shannon 4
He, Yuqi 31
Hill, Zach 22
Hong, Charmaine 17
Hyndman, Sarah 14
Kantaria, Pranay 11
Ho, William 56
Jawaharlal, Aromal 69
Ji, Sooa 7
Jiao, Shunhua 45
Jin, Yi 63
Jonnalagadda, Soumya 63
Khan, Emaad 26
Kim, Myungchan 6
Kitching-Williams, Angus 20
Kkantis, Giorgos 8
Kong, Zhuocheng 49
Lai, Ka Chun 55
Lal, Taran 8
Lau, Chris Z W 69
Lee, Donggyu 42
Leong, Weng Kit 15
Lewis, Miles 19
Li, Bingqi 31
Lim, Ashley 64
Limbachia, Rahul 32
Liu, Zewen 15
Liu, Ziyang 66
Low Yi Wei, Jonathan 62
Lu, Guanhao 27
Luna Rueda, Ari 60
Ma, Wenqing 41
Mahendren,Sharven 3
Mao, Yujie 41
Martin, Frederick 20
Masruri, Asan 72
Mauri, Chiara 51
Mazauric, Jeremy 28
McLoughlin, Amy 48
Mee, Peter 18
Minsun, Kwak 33
Murugan, Gavin 70
Nagpal, Stutya 9
Najib, Danyal 25
Narongchartsopon, Suphapacha 38
Neseyif, Hadi 11
Nies, Alexander 54

Ng, Jake	58
Ng, Wi Rong	50
Okezie-Akwiwu, Chelsea	71
Okorogheye, David	19
Olaleye, Hephzibah	34
Ong, David	50
Penn, Daniel	14
Popovich, Paul J.	37
Porfirio, Mateus	72
Punnarujwong, Rada	4
Rigg Franco, Helena	74
Rowley, Francesca	43
Ruzickova, Hana	16
Samuel, Joshua	65
Shah, Niam	73
Shaik, Siraj	53
Singh Lalli, Juproop	36
Sim, John P	30
Srikumar, Shiam	28
Stathogiannopoulos, Dimitrios	37
Sturman, Callum	35
Tai, Louisa	55
Tan, Yew Jueen	38
Tan, Xiang Hong	44
Tang, Bohui (Robert)	10
Tang, Hiu Chung	57
Tay, Xuanyong	5
Thillairajah, Thagshanen	70
Torres Quiros, Claudia	51
Trane, Julia	22
Utting, Macey	29
Wang, Jasmine	2
Wang, Qianxiang	24
Wang, Yicong	24
Wang, Yuhan	23
Wijaya, Colleen Handriani	10
Wong, Jacky	68
Woodcock, Alice	74
Wootton, Callum	34
Worthington, Jack	32
Xie, Jing	21
Xu, Yuan	1
Xu, Zhenyu	1
Zakharov, Yaroslav	47
Zhang, Yihao	23
Zhou, Kelvin	12
Zhou, Yijie	49
Zou, Yujie	61
Zulkifli , Suzulyana Binti	40

Supervisor Index

Linked to paper number

Cabral, João	8, 15, 59
Chachuat, Benoit	45, 57, 64
Chen, Rongjun	4, 24
Ceroni, Francesca	53
del Rio Chanona, Antonio	56, 60, 69
Eslava, Salvador	25, 26, 62
Fennell, Paul	23, 40, 44
Galindo, Amparo	58
Hallett, Jason	5, 48
Hankin, Anna	1, 10, 19
Hawkes, Adam	32, 47, 55
Hellgardt, Klaus	13, 42, 50, 67
Heng, Jerry	17, 49, 74
Jackson, George	20
Krishnan, J	22
Kontoravdi, Cleo	16, 36
Luckham, Paul	18
Markides, Christos	29, 33
Matar, Omar K.	7, 54
Mercangoz, Mehmet	3, 27, 68
Petit, Camille	37, 43, 71
Pini, Ronny	9, 52
Papathanasiou, Maria	51, 65, 72
Polizzi, Karen	41, 70, 73
Rinaldi, Roberto	11, 39
Shah, Nilay	31, 34, 66
Shmool, Talia	14
Song, Qilei	35
Tighe, Chris	6
Titirici, Magda	2, 30, 46, 61
Trusler, Martin	21, 28
Xu, Yun	38, 63