Chemical Process Design Approaches Leveraging Process Modularization, Intensification, and Large Language Models

by

Chinmoy Basak Mukta

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
December 14, 2024

Keywords: process intensification, process modularization, optimization, large language models

Approved by

Mario R. Eden, Chair, Joe T. and Billie Carole McMillan Professor of Chemical Engineering, Auburn University
Selen Cremaschi, Co-chair, B. Redd & Susan W. Redd Endowed Eminent Scholar Chair Professor of Chemical Engineering, Auburn University
Sushil Adhikari, Professor of Biosystem Engineering, Auburn University
Zhihua Jiang, Associate Professor of Chemical Engineering, Auburn University
Chris Kieslich, Assistant Professor of Chemical Engineering, Auburn University

Abstract

This dissertation investigates critical aspects of chemical engineering, including reactor intensification, process expansion, distributed natural gas desulfurization, and large language models. The key findings of this work highlight the complex interplay among reactor intensification, separation design, production costs, and environmental considerations, emphasizing the necessity for tailored approaches. Economic analysis confirmed triazine-based absorption and SourCat^{TM} to be cost-effective desulfurization methods. Moreover, In one of our study we underscored the limitations of general-purpose language models and advocated for the use of domain-specific training data to enhance reliability in process design applications.

As a case study, the first study discusses the critical effects of reactor intensification on downstream processes and the impact of different process expansion approaches on the cost and environmental effects of ethylene oxide production. In this study, the production of ethylene oxide was simulated at a production capacity of 100 kt/yr using a conventional reactor and reaction kinetics. Subsequently, the impact of reactor intensification was investigated using microfibrous-entrapped catalysts on separation section design. Second, three expansion approaches, namely, the Brownfield, retrofitting, and Greenfield processes, were explored with an emphasis on the implications of their cost and environmental impact. Despite offering cost reduction, the limited separation capacity of the Brownfield process limits the potential increase in production achievable through reactor intensification. Although retrofitting reduces production costs, it introduces complexities at a capacity of >130 kt/yr, necessitating additional separation columns. The Greenfield process, with the lowest carbon emissions among the three investigated expansion approaches, is cost-effective only at production rates above 160 kt/yr. The retrofitted process

involves a moderate increase in carbon emissions, whereas the Brownfield process leads to a substantial increase in carbon emissions with capacity expansion. These findings emphasize the complex interplay among reactor intensification, separation train design, production costs, environmental considerations, and capacity limitations in chemical processes, underscoring the importance of formulating tailored approaches to suit specific requirements and carbon emission goals.

The second section discusses natural gas as a reliable, clean, and abundant energy source for heat and electricity generation. A vital aspect of using natural gas as an energy source is the removal of water, CO_2, and H_2S before pipeline transmission. Desulfurization is a crucial and routinely used purification step for removing H_2S from natural gas in conventional reservoirs. With increasing energy demand, stranded/distributed natural gas resources, which remain unused for economic reasons (small scale and remote from the market), are being increasingly considered at present. Investigating low-cost distributed natural gas desulfurization is essential as a preprocessing step for utilizing stranded gas. In this study, three desulfurization processes, namely, triazine-based absorption, a liquid redox system (LOCAT®), and solid-bed oxidation (SourCat™), were analyzed to study their economics and capability to remove H_2S from small-scale distributed natural gas resources. Besides considering the impact of modularizing these processes, which typically use standardized equipment sizes owing to the transmission difficulties of stranded gas resources, the economics of the modularized processes were compared with those of traditional-design processes. Each process was simulated using Aspen Plus at different sulfur concentrations (500–2500 ppm) and natural gas capacities (1–100,000 MSCFD), and the corresponding desulfurization costs were estimated. In addition, a sensitivity analysis was used to identify the parameters influencing the desulfurization cost. The results revealed that the triazine-based

absorption process is economical for processing capacities lower than 500 MSCFD, whereas SourCat™ is cost-effective for processing capacities between 500 and 100,000 MSCFD. Furthermore, the raw material, pump utility, and sorbent costs substantially impact the triazine-based absorption, LOCAT®, and SourCat™ processes, respectively. An analysis of the applicability of modularized desulfurization units in gas fields is expected to provide valuable insights into the quantitative assessment of the modularization effectiveness and its potential application limitations.

In our third study explored the usefulness of the large language models (LLMs)in process design context-based evaluation of, specialized metrics to assess chemical process design-specific query evaluation, problem-solving abilities, and correct terminology usage. This work highlights the general-purpose language model limitations of using general training sets to convey process engineering concepts, emphasizing the need for training on design diagrams such as process flow diagrams. The general-purpose model hallucinates in a specific process design task by either over- or under-assuming. This research reiterates the age-old phenomenon of "garbage in, garbage out" in system modeling, demonstrating that models trained on general language data exhibit lower accuracy in specialized contexts. Each engineering term was heavily contextualized to address the issues of ambiguity and polysemy in process design terminology to reduce model hallucinations. Subsequently, words and phrases were transformed into vector representations using techniques like word-to-vector. By calculating the Euclidean distances between these word vectors, terms that were close in meaning were identified within the process design context. Clustering algorithms, such as k-means, group these vectors into clusters of contextually similar terms, effectively disambiguating polysemous words based on their usage in process design documents. Graphs were then used to find relationships and contextualize technical understanding using process design

diagrams such as flowsheets, flow diagrams, and graphs. Furthermore, maintaining consistent distances and clustering while training the language model ensured that the language model learned specifically in the specific flowsheet context and provided consistent answers to specific questions, thereby enhancing proficiency and maintaining consistency for each query. By analyzing the impact of training data quality on LLM performance, the integration of domain-specific texts was found to enhance model reliability. These findings suggest that targeted training approaches, such as the strategy described in this work, for different types of training data other than text are essential for developing LLMs capable of effectively operating within technical fields, such as process design engineering.

Acknowledgments

First, I express my deepest appreciation and gratitude to my advisors, Dr. Mario Eden and Dr. Selen Cremaschi, for their support, mentorship, and guidance. They generously shared their knowledge and experiences, helping me navigate through challenges and contributing significantly to my academic growth. I am truly grateful for their patience, encouragement, and belief in my abilities throughout this challenging journey.

I would also like to thank Dr. Chris Kieslich for his valuable advice and constructive feedback, which improved my research considerably. Additionally, I thank Dr. Soledad Peresin for serving as my university reader and providing valuable input and guidance. I also thank my colleagues at Auburn University for valuable shared experiences and intellectual discussions. Their support and inspiration motivated me to strive for excellence, and I am grateful for the privilege of working alongside them.

I want to express my heartfelt appreciation to my father, Loknath Basak, who has gone above and beyond to uplift a boy from a small village, providing opportunities and opening doors to a brighter future. I am fortunate to have collaborated with Dr. Jae Jim Shim, Dr. M. A. A. Shoukat Choudhury, and Dr. Salem Monem, who opened the doors to academia for me. Their mentorship has been instrumental in shaping my journey and fueling my passion for research and scholarship.

I want to thank my partner, Protima Paul, for her patience and unwavering emotional support during the challenging times I faced throughout my journey. Her love, understanding, and

encouragement were excellent sources of strength, and I am profoundly grateful for her presence in my life. Finally, I thank my mother for her selfless love and support throughout this process.

Table of Contents

List of Tables

List of Figures

Chapter 1 – Introduction

This dissertation investigates critical aspects of chemical engineering, including reactor intensification, process expansion, distributed natural gas desulfurization, and large language models. The key findings highlight the complex interplay between reactor intensification, separation design, production costs, and environmental considerations, emphasizing the need for tailored approaches for optimal process design. Economic analysis reveals that triazine-based absorption and SourCat^{TM} are cost-effective desulfurization methods. This study also underscores the limitations of general-purpose language models and advocates for domain-specific training data to enhance reliability in chemical engineering applications.

This study examines the effects of reactor intensification and process expansion on downstream processes, along with their environmental impact and economic viability using ethylene oxide (EO) production as a paradigmatic case study. An initial simulation of EO production was conducted using a conventional reactor with a 100 kt/yr production capacity, followed by a rigorous assessment of reactor intensification via microfibrous-entrapped catalysts to elucidate its implications on the separation section. Furthermore, three expansion strategies—Brownfield, retrofitting, and Greenfield—are scrutinized for their cost-effectiveness and environmental implications, revealing distinct advantages and limitations. Notably, Brownfield expansion offers cost advantages but encounters capacity limitations in separation, whereas retrofitting reduces production costs but requires additional separation units above 130 kt/yr. Conversely, Greenfield processing is the most cost-effective approach at 160 kt/yr and exhibits the lowest carbon emissions.

This study analyzes the intricate interplay between reactor intensification, separation design, and carbon emission goals, underscoring the need for tailored strategies to meet the economic and environmental objectives of chemical processes.

Furthermore, this study investigates the economic desulfurization of stranded natural gas resources to address the pressing need for alternative energy sources. The economic feasibilities of triazine-based absorption, LOCAT® liquid redox, and SourCat™ solid-bed oxidation were evaluated across diverse sulfur concentrations and gas capacities, considering modularization for overcoming the challenges associated with small-scale deployment. Simulation results indicated that triazine absorption is the most cost-effective method for lower capacities (<500 MSCFD), whereas SourCat™ excels in the 500–100,000 MSCFD range. Sensitivity analysis indicated that the raw material, utility, and sorbent costs are the primary economic drivers of the process, providing valuable insights into the cost implications of modularizing the desulfurization of distributed gas sources.

This study also probes the proficiency of large language models (LLMs) in chemical engineering tasks, particularly in process design. The limitations of general-purpose language models in conveying technical concepts are accurately highlighted, reaffirming the "garbage in, garbage out" principle in system modeling. By leveraging word embeddings and clustering, the ambiguity in process terminology can be mitigated, thereby enhancing the precision of language models in technical contexts. Proximity graphs were used to examine relationships within the process design documents, ensuring correct terminology and contextual consistency. These findings advocate domain-specific training data to enhance LLM reliability and consistency in specialized fields, such as chemical engineering, suggesting that tailored language models can significantly augment the accuracy and relevance of process-engineering tasks.

Chapter 2 – Literature Review

2.1 Modular Processes

Modular process units, pre-engineered and pre-tested, are integral to chemical plants; they optimize functions such as natural gas processing and ammonia production through scalable and economical processes that defy traditional power-law cost escalations (Allen et al., 2018; Palys et al., 2018). Besides enhancing operational efficiency and safety because of their transportability and robust design, these units offer substantial flexibility and economic benefits across various applications, including water purification and petrochemical manufacturing (Huang et al., 2020; Al-Fadhli et al., 2019; Baldea et al., 2017; Yang ; Mokhatab et al. 2013 and You, 2018).

.

2.1.1 Modularized Desulfurization

The development of economically modularized desulfurization processes has enabled the utilization of stranded natural gas. Recent studies on distributed modular processes (Chen and Grossmann, 2019; Shao and Zavala, 2020) suggest that numbering standardized modules may provide advantages, such as off-site fabrication, swift transportation, deployment of machinery, expediting experimentation, and a better understanding of the processes. Consequently, the technological costs of such processes are lower than those of conventional processes. Bhosekar et al. (2021) proposed modularization for considering the effects of uncertainty in the supply chain and studied the design and utilization of such modular units using a mixed-integer two-stage stochastic programming model. The results indicate that modularization provides adaptability for

managing market volatility. Lara et al. (2019) studied the design and planning of manufacturing networks by comparing distributed and conventional centralized chemical processes using generalized disjunctive programming.

2.1.2 Stranded Natural Gas

The desulfurization cost of stranded natural gas depends on the selected process, design, and operating conditions. The production rate of stranded natural gas varies with the location, geology, age of the gas well, and market conditions, such as prices (Wood et al. 2008. Wagner et al. 2002). For example, the production rate of gas wells in the Appalachian Basin lies within 1.1–26.3 MMSCFD, and the H_2S concentration may vary within 1–60,000 ppm, typically declining over time (Zavala-Araiza et al., 2015). However, the selection of the desulfurization process and its optimum design and operating conditions depend on the feed flow specifications, which are related to the production rate and H_2S composition of the stranded gas resources (Thomas et al. 2003; Calel et al. 2020). Owing to the wide range of production rates and sulfur compositions, a systematic approach is key to selecting the most economical desulfurization process for a specific feed.

In this study, process simulation and techno-economic analyses were used to evaluate the potential of three processes, namely, triazine-based absorption (Schulz, 2012), liquid redox system (LOCAT®) (Yong et al., 2014), and solid-bed oxidation (SourCat™) (Mukta et al., 2022; Xu et al., 2020), for the desulfurization of sour gas produced from stranded resources. The analysis considered the impact of process modularization by restricting the equipment sizes that allowed truck transportation and identifying the number of standardized pieces of equipment required for each process. Technoeconomic evaluations were used to ascertain the most cost-effective desulfurization process for varying flow rates and sulfur concentrations of natural gas using both

modular and traditional setups. An overview of the three desulfurization processes is provided in Section 2. Section 3 presents the methodology used to analyze the process simulation, technoeconomic study, and sensitivity analysis (SA). The results are presented and discussed in Section 4, and the conclusions are presented in Section 5.

2.2 Overview of Three Desulfurization Processes

2.2.1 Triazine-based Absorption

Triazine-based absorption is a commonly used H_2S removal method that does not require solvent (triazine) regeneration because the products are readily biodegradable. The process flow diagram is shown in Figure 1 (Schulz, 2012). Three reactions (R1, R2, and R3) occur in the absorption vessel. The primary reaction is R1, and the product of R2 forms solid concrete inside the vessel and may cause plugging (Chakraborty et al., 2017). Here, the inlet triazine flow rate was maintained at a 20% excess to avoid solid formation (Subramaniam et al., 2018). The inlet sour gas requires a relatively dry stream because the extra water intensifies the side reaction of triazine with water, which causes additional solvent loss (Subramaniam et al., 2018). Here, dry sour gas with a 52 wt% triazine solution was fed into an absorber to absorb the sulfur component within 10–70 ℃ (Schulz, 2012). The final gas phase comprises a sweet gas with a sulfur content lower than 4 ppm, while the liquid phase reacts with the disposed triazine (Schulz, 2012).

$$C_9H_{21}N_3O_3 + 3H_2S \rightarrow C_3H_6O_3 + 3C_2H_7NO \quad \text{(R1)}$$

$$C_9H_{21}N_3O_3 + H_2S \rightarrow C_7H_{16}N_2O_2S + C_2H_7NO \quad \text{(R2)}$$

$$C_7H_{16}N_2O_2S + H_2S \rightarrow C_5H_{11}NOS_2 + C_2H_7NO \quad \text{(R3)}$$

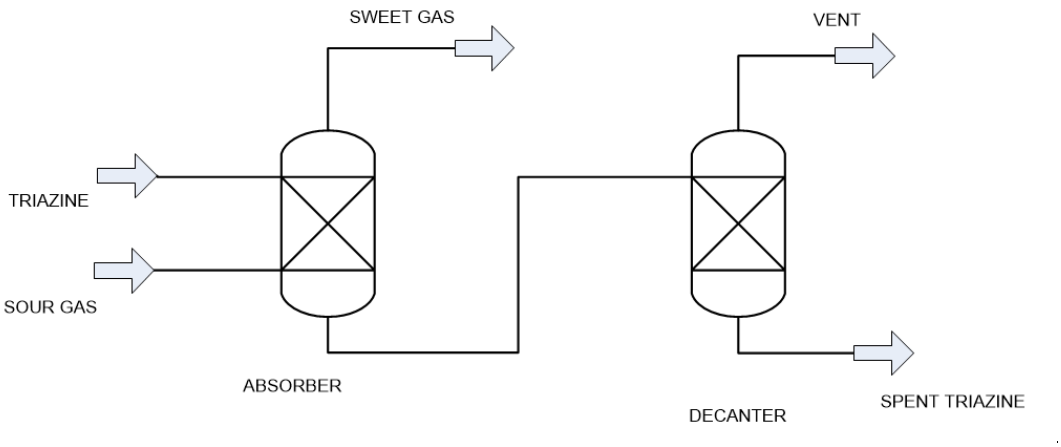

Figure 1. Triazine-based absorption process flow diagram.

2.2.2 Liquid Redox System

LOCAT® processes utilize an environmentally clean liquid oxidation catalyst (for example, the use of chelated iron is reported by Yong et al., 2014) to convert H_2S into elemental sulfur, and the catalyst can be regenerated by reacting with oxygen. These reactions are labeled R4 and R5. Here, a small amount of sodium was added to control the solution pH between 8 and 9 to maintain H_2S solubility and ferric ion stability in the system (Echt et al., 2017). The inlet aqueous catalyst concentration remained within 250–3000 ppm (Judd, B., Eng, 2003), and the inlet Fe^{3+} molar flow rate was four times higher than that of H_2S (Kohl and Nielsen, 1997). Figure 2 shows the LOCAT® process flow diagram. The dry sour gas and catalyst were fed into an absorber to generate elemental sulfur in the temperature and pressure ranges of 27–55 ℃ and 1–34 atm, respectively (Echt et al., 2017). The top product of the absorber was sweet gas. The bottom liquid, containing sulfur and ferrous ions, was fed into the regeneration column. Ferrous ions were regenerated in the column using oxygen, and solid sulfur was separated using a filter. The liquid catalyst regenerated with a fresh makeup catalyst was recycled back into the absorber. The mass fraction of the sulfur cake in this process varies within 30–90%, depending on the filter type (Echt et al., 2017).

$$2Fe^{3+} + HS^{-} \rightarrow 2Fe^{2+} + S + H^{+} \quad \text{(R4)}$$

$$2Fe^{2+} + \tfrac{1}{2}O_2 + H_2O \rightarrow 2Fe^{3+} + 2OH^{-} \quad \text{(R5)}$$

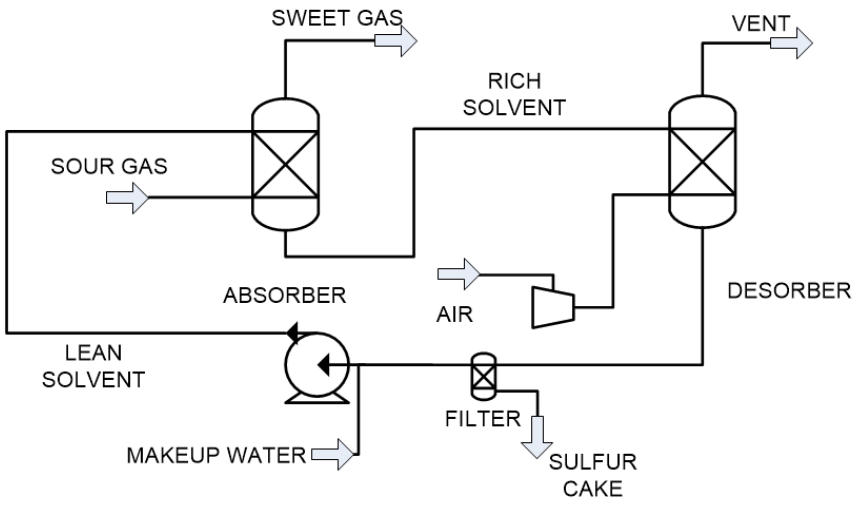

Figure 2. LOCAT® process flow diagram.

2.2.3 Chemical Oxidation

The SourCat™ process involves H_2S conversion into elemental sulfur using a solid-bed catalyst. The primary and side reactions are R6 and R7, respectively. (Mukta et al., 2022; Xu et al., 2020). Figure 3 shows the process flow diagram of SourCat™. Unlike the triazine-based absorption and LOCAT® processes, the wet sour gas can be fed into the reactor in this process. Here, sour gas and oxygen were heated up to 200 ℃ at 115 psi, and the mixed gas was fed into a solid-bed reactor to generate elemental sulfur. This reaction is highly exothermic and requires precise temperature control in the reactor, particularly while being conducted at a large scale. After the reaction, elemental sulfur was washed with water, and a filter was used to separate the sulfur compounds. Because of the side products and remaining H_2S in the gas stream, a sorbent bed was used to eliminate these contaminants. The makeup water replenished the lost water.

$$\frac{1}{2}O_2 + H_2S \rightarrow H_2O + S \quad \text{(R6)}$$

$$\frac{3}{2}O_2 + H_2S \rightarrow H_2O + SO_2 \quad \text{(R7)}$$

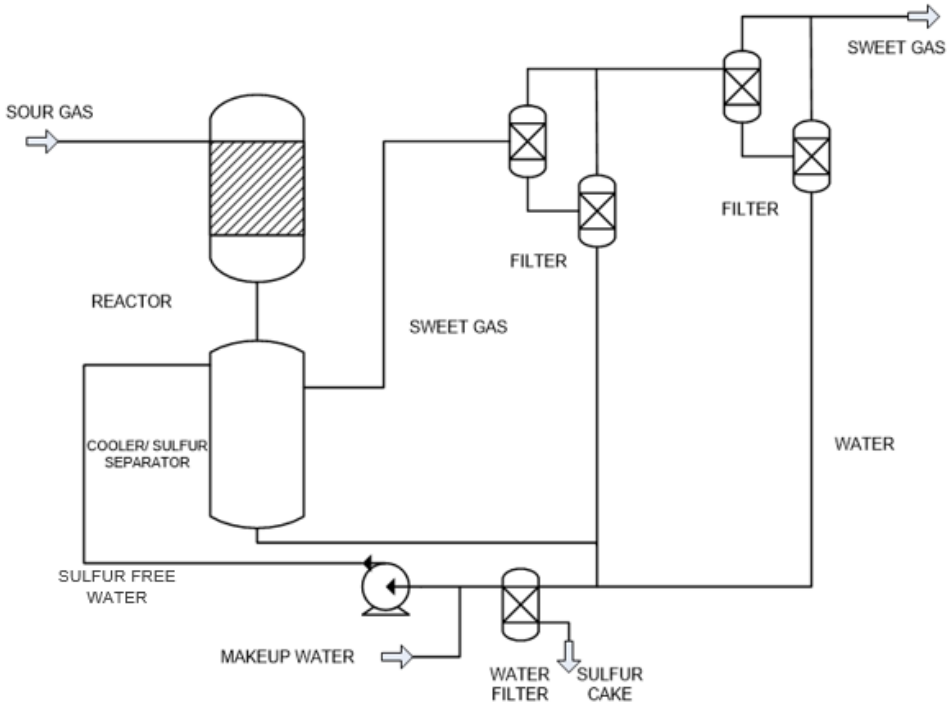

Figure 3. SourCat™ process flow diagram.

2.3 Methodology

Comparing two technologies can be challenging, particularly when they involve proprietary or trade-secret methods. The engineering level required to assess each technology increases the complexity of the evaluation process in several industries, including chemical processing. Such assessments often require a thorough understanding of specialized engineering principles and the application of rigorous analytical techniques to ensure accurate comparisons and decision-making.

2.3.1 Process Simulation

The three selected desulfurization processes were simulated using the Aspen Plus V11 software. To compare the economic performance of the processes at different natural gas flow rates and sulfur concentrations, the gas flow rate was varied from 1 MSCFD to 100 MMSCFD, and the H_2S concentration was varied from 500 to 2000 ppm. In all simulations, the sweet gas H_2S concentration was set to ≤20 ppm. The process design requirements for different feed flow conditions were different. Therefore, the Aspen Simulation Workbook was used to evaluate the simulations at different feed flow rates and analyze the associated operating conditions. The Aspen Simulation Workbook uses the ActiveX connection between Aspen Plus and Microsoft Excel and automates data collection at the respective inlet flow specifications. The process-specific simulation inputs are explained in *Sections 3.1.2* and *3.1.3*.

2.3.1.1 Triazine-based Absorption

A flooded-system-based triazine absorption was simulated in this study, as shown in Figure 4. Diluted MEA triazine (52 wt% 1,3,5-tris-(2-hydroxyethyl) hexahydro-s-triazine), a water solution (S1), and dry sour natural gas (S2) were fed into absorber T-1, which was simulated using a RadFrac unit with two equilibrium stages to remove the sulfur component through reactions R1–3 (Subramaniam et al. 2018). Reaction kinetics data reported by Bakke and Buhaug (2004) were used to estimate the product compositions, and the liquid–gas mixture (S5) was separated within flash vessel V-1, where sweet gas S-6 was collected from the top and the spent scavenger S-7 was collected from the bottom. In this simulation, the process operating temperature and pressure were set to 80 and 300 psia, respectively, for all streams and units. ELECNRTL was used as the thermodynamic package because it can process the ionic reactions R1–3 inside absorber T-1.

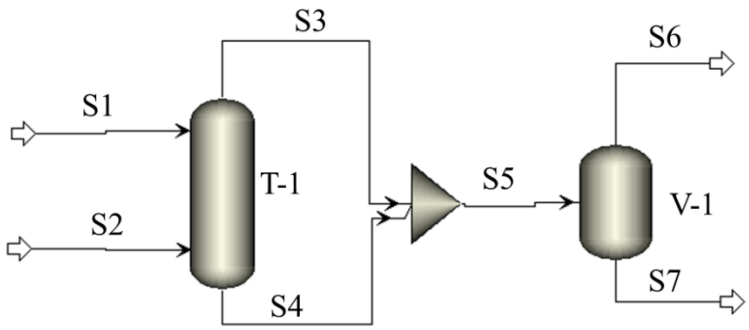

Figure 4. Process simulation flowsheet of the MEA-triazine process.

2.3.1.2 Liquid Redox System

The LOCAT® simulation process flowsheet is given in Figure 5. Owing to the ionic reactions R4 and R5, the thermodynamic package ELECNRTL was used for analysis. In this process, the regenerated absorbent (S10-2) with an iron content of 250 ppm and sour natural gas (S1) were fed into the absorption column T-1 to remove the sulfur component in the sour natural gas by converting H_2S into elemental sulfur via reaction R4. Sweet natural gas (S2) was produced at the top of the column. The rich solution S3 containing elemental sulfur and ferrous ions was fed into the desorption column T-2 from the top. Under atmospheric conditions, air (S4) was compressed and cooled in the compressor C-1 and heat exchanger H-1 before entering the desorption column T-2 from the bottom. In column T-2, the absorbent was regenerated by oxidizing ferrous ions to ferric ions via reaction R5. The regenerated solution S6 containing elemental sulfur was filtered (F-1) to separate solid sulfur (S7). The liquid solution S10 and makeup solution S9 were pumped (P-1) and heated (H-2) before being recycled back to the absorption column T-1. A RadFrac unit was used in the simulation to model the absorption and desorption columns. The reaction kinetic model of Wubs and Beenackers (1994) was used to

estimate the reaction products inside the column at 43 ℃ and 7.8 atm. In addition, owing to the rapid reaction, five equilibrium stages were observed in columns T-1 and T-2. Filter F-1 was modeled using a separator, where all solid sulfur was removed with a 0.01% loss of the liquid solution.

2.3.1.3 Chemical Oxidation

Figure 6 shows a flowchart of the chemical oxidation process. The inlet natural gas (S2) at 95 ℃ and 7.83 atm and air (S1) at 25 ℃ and 1 atm passed through the heat exchangers H-1 and C-1 to heat the natural gas to 210 ℃ and compress the air to 7.82 atm. The mixed gas stream at 210 ℃ and 7.82 atm was fed into reactor R-1, where elemental sulfur was generated via reactions R6 and R7. The reactor was simulated as the stoichiometric conversion reactor R1. After the reaction, a subsequent flash vessel (V-1), operating at 34 ℃ and 7.83 atm, was used to both cool and separate the gas stream S5 by mixing the hot reactor outlet S4 with the recycled water stream S11. The S-containing water stream (S6) was removed from the bottom of the vessel V-1, and the gas stream S5 was collected from the top of the vessel. To ensure complete removal of impurities from the gas stream and the safe operation of V-1, the gas stream S5 and recycled water stream S12 were fed into the second flash vessel V-2 at 34 ℃ and 7.83 atm to separate the remaining elemental sulfur (S7). Subsequently, the sorbent bed vessel V-3 (at 31.9 ℃ and 3.4 atm) was used to further separate impurities, such as SO_2 (S15), produced during side reactions occurring in the reactor. After these separations, stream S14 became a sweet natural-gas product. The sulfur-containing water streams S6 and S7 were pumped from vessels V-1 and V-2 and fed into a filter (F-1) to remove solid sulfur. The resultant sulfur cake was discarded using stream S8-1, and freshwater was circulated back to vessels V-1 and V-2. The Peng–Robinson package was used to

simulate the SourCat™ process, and a separator unit operation was used to model the filter (F-1) and adsorber (V-3).

After process analysis, the sulfur-removal vessel (V-1) was identified as the critical equipment dictating the overall capital and operating costs (Figure 7). This equipment removes the sulfur produced in the reactor and cools its effluent. Here, circulating water removed hot sulfur from the reactor effluent gas while cooling the reactor effluent. Figure 7 shows the regression model generated using the experimental data shared by IntraMicron, Inc., which was used to determine the water flow rate required for sulfur removal from the effluent gas at various H_2S concentrations and natural gas flow rates. A worst-case scenario was considered to ensure that the reactor effluent was adequately cooled by simulating an adiabatic reactor with various water flow rates in the condenser. Moreover, to determine the possible condenser outlet temperature, a worst-case scenario design was considered for the condenser with a specific flow rate and H_2S concentration to ensure that the reactor outlet cooling requirements were satisfied.

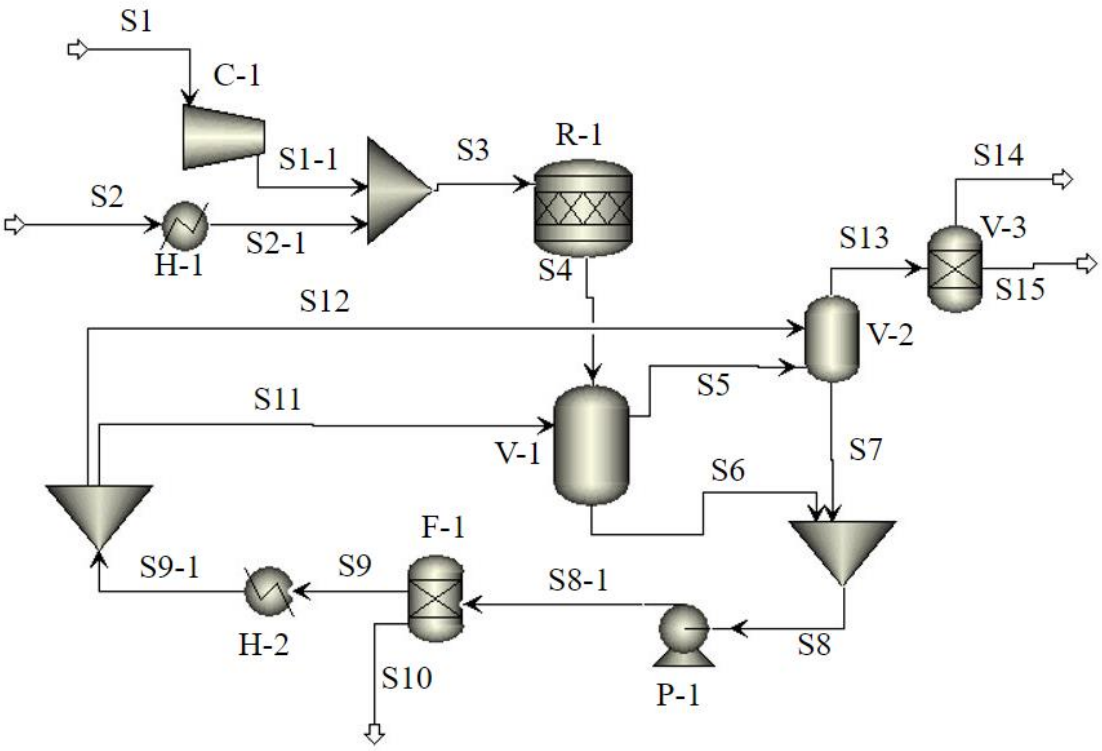

Figure 5. Process simulation flowsheet of the SourCat™ process.

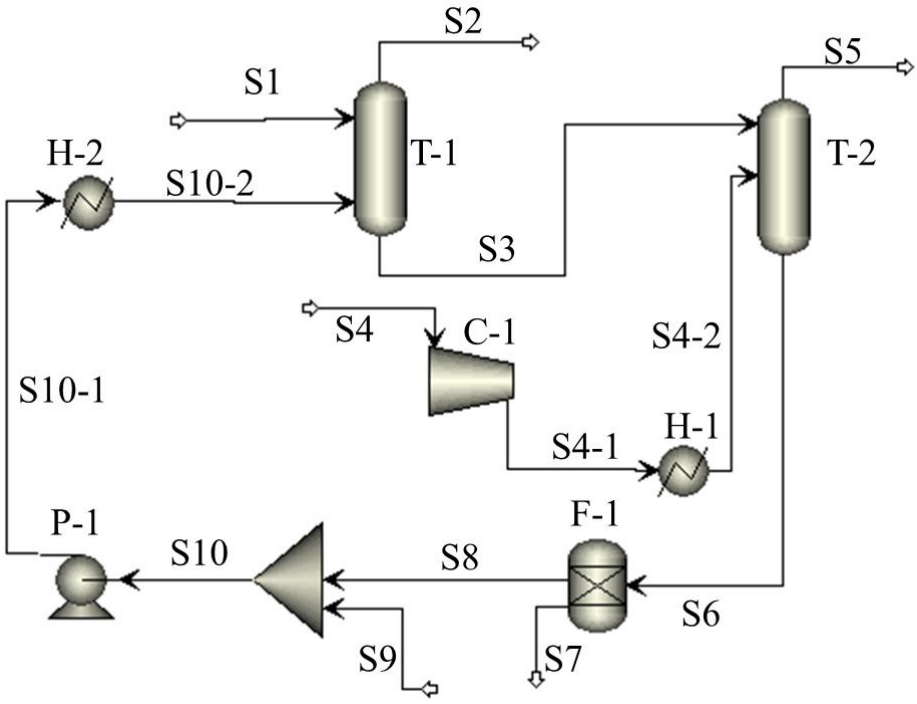

Figure 6. Process simulation flowsheet of the LOCAT® process.

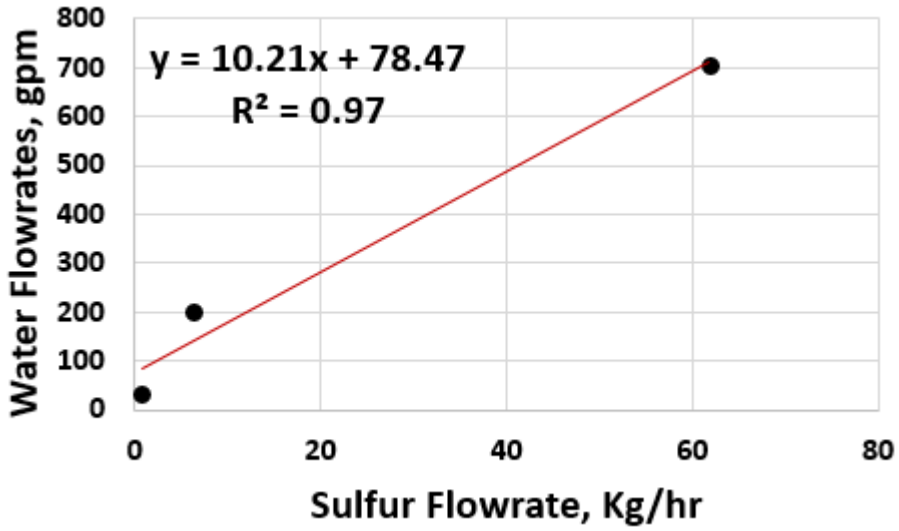

Figure 7. SourCat™ process condenser sizing experimental data.

2.3.2 Sizing

Equipment design parameters, such as the inlet/outlet flow rate and heat duty, were extracted from the converged Aspen Plus process simulation models and transferred to Microsoft Excel using the Aspen Simulation Workbook. The equipment shown in Figures 4, 5, and 7 were sized using the equations listed in Table 1.

Table 1. Equipment sizing criteria (GPSA Engineering Data Book, 2004)

Equipment	Sizing equations	Details
Reactor	$V_R = \frac{F\, G_{\text{natural}}}{\rho_{cat}}$	V_R.: Reactor Volume, m^3
	$D_R = 2\sqrt{\frac{Fv}{\pi\, Gv}}$	$G_{natural}$: Natural Gas Volume Flow Rate, MSCFD
	$H_R = \frac{V_R}{\pi (R_R)^2}$	D_R: Reactor Diameter, m
		F_V: Volumetric Flow Rate, m^3/s

			G_V: Gas Velocity, m/s F: Linear Correlation Factor ρ_{cat}: Catalyst Density, kg/m H_R: Reactor Height, m R_R: Reactor Radius, m
	Condenser	$H_{con} = \frac{R\ F_{outlet}}{1000}$ $D_{con} = \sqrt{\frac{4V_C}{\pi \mathrm{H_{con}}}}$	H_{con} : Height of the Condenser V_c: Condenser Volume, m^3 F_{outlet}: Reactor Outlet Flow Rate, L/min R: Residence Time, min D_{con} : Diameter of the Condenser, m
	Sorbent Bed	$V_{sorb} = \frac{\mathrm{M_{H_2S,outlet}}}{J_{sorb} \times \rho_{sorb}}$	V_{sorb}: Sorbent Bed Volume, m^3 $\mathrm{M_{H_2S,outlet}}$: H_2S Mass Flow Rate After Reaction, kg/h J_{sorb} : Sorbent Capacity, g H_2S Adsorbed/g Sorbent ρ_{sorb}: Sorbent Density, kg/m^3
or	Heat Exchanger Air Cooler	$A_{He} = \frac{UA}{U_d}$	A_{He}: Area of the Heat Exchanger, m^2 UA: Heat Transfer Coefficient Multiplied by the Area of the Process Simulator, W/K U_d: Heat Transfer Coefficient from Exchanger Design and Rating Software, $W/(m^2 \cdot K)$
Pumps	Compressor and	P	P: Power Requirements, kW
	Columns	$H_{col} = N_{col} h$	H_{col}: Total Height of the Column, m h : Distance Between Each Stage (Minimum Number of Stages was used as the Separation is Easy, with Equilibrium Stages), m N_{col}: Column, Number of Stages

2.3.3 Cost Estimation

The purchase cost of each piece of equipment was calculated based on the standard cost-correlation factors proposed by Turton et al. (2008), which are provided in the Supplementary Material. The general cost-calculation correlations can be expressed by Eqs. (1) and (2), as follows:

$$C_p^o = K_1 + K_2\, log_{10}(A) + K_3[log_{10}(A)]^2 \quad (1)$$

$$C_{BM} = C_p^o F_{BM} = C_p^o(B_1 + B_2 F_M F_P) \quad (2)$$

In Eqs. (1) and (2), C_p^o is the equipment base cost; C_{BM} is the module cost; K_1, K_2, K_3, B_1, B_2, and B_2 are the cost correlation factors; F_M is the material factor; F_P is the pressure factor; and A corresponds to the respective equipment unit. The cost-correlation factors are listed in Table S4.

The desulfurization costs ($DeCost_A$ and $DeCost_B$) are defined as the sweetening cost for one MSCF of sour natural gas before and after the payback time; this includes the capital expenditure and operating costs, as shown by Eqs. (3) and (4), respectively. The capital cost was calculated using the equipment size (*Section 3.2.1*) to determine the C_{BM} of each piece of equipment. The operating costs, including raw material and utility costs (C_r and C_u, respectively), were calculated based on the heat duty and inlet/outlet flow rate estimates of the Aspen Plus process simulation. Using Eq. (4), the raw material and utility costs were used to calculate the overall operating costs. The standard cost calculation coefficients reported by Turton et al. (2008) (listed in Tables S4 and S5) were used for capital and operating cost calculations. The desulfurization cost equations with $DeCost_A$ and without $DeCost_B$ payback, respectively, are provided in Eqs. (5) and (6). In Eqs. (5) and (4), the annual interest rate r and payoff time t are 10% and three years, respectively.

$$CAPEX = \sum C_{BM} \tag{3}$$

$$OPEX = \sum C_u + \Sigma C_r \tag{4}$$

$$DeCost_A = \frac{CAPEX \times \frac{r(1+r)^t}{(1+r)^t - 1} + OPEX}{C_{cap}} \tag{5}$$

$$DeCost_B = \frac{OPEX}{C_{cap}} \tag{6}$$

where C_{cap} denotes the natural gas flow rate (MSCF/yr).

2.3.4 Process Modularization

This study used a simple modularization criterion to limit the size of the principal equipment in each process to enable facile transportation. Each piece of equipment was sized without any limitations according to the process requirements for conventional plants. Here, the equipment size for modularized plants was constrained based on federal transportation guidelines (Morris, 2003). The algorithm used to determine the equipment size for the modularized plant is shown in Figure 8 (Shao and Zavala, 2020). First, critical equipment requiring sizing to meet federal transportation guidelines was identified. Second, data for the critical equipment were collected from the Aspen Plus process simulation and used for sizing (*Section 3.2.1*). Third, each critical equipment dimension was checked against federal transportation guideline limitations, i.e., the modularization criteria shown in Figure 8. If the equipment failed to satisfy the criteria, the number of necessary pieces of equipment (N_{eq} in Figure 8) was increased by one, and the inlet flow rate was equally divided among the units. Each piece of equipment was repetitively resized, and its dimensions were checked against the Modularization Criteria. A modularized plant design was accepted when these criteria were met.

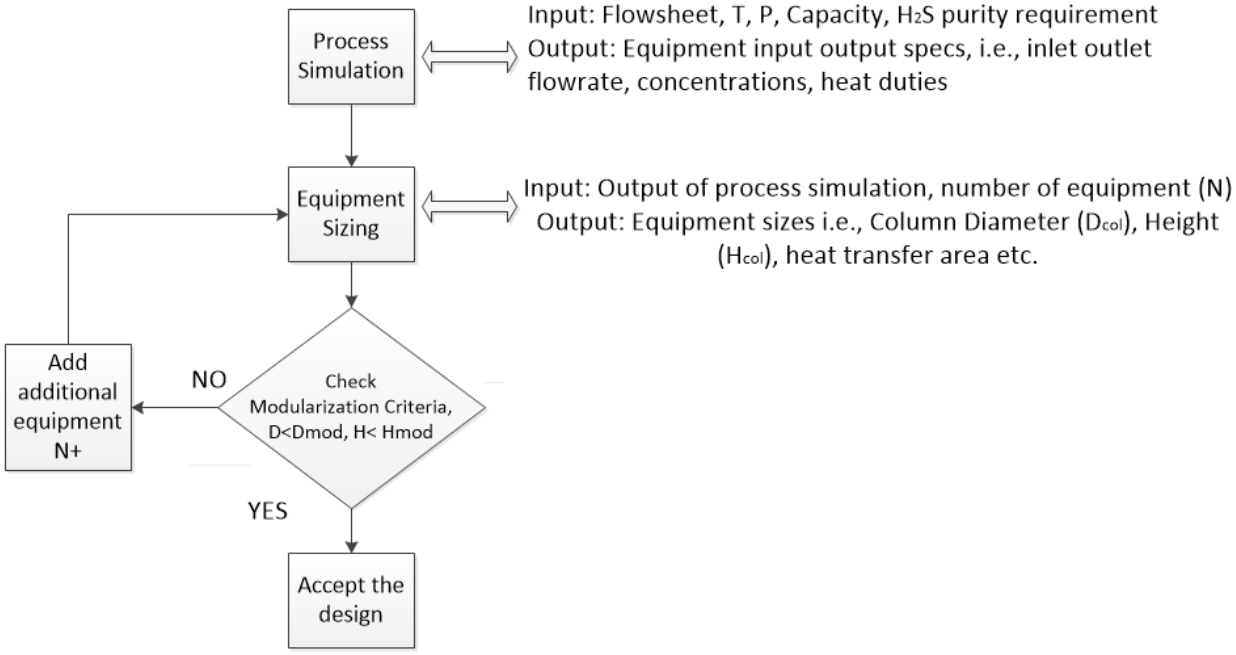

Figure 8. Strategy used for modularized plant design.

The process and economic parameters relevant to the capital and operating costs are included in the SA desulfurization cost. Tables 2–4 summarize these parameters, their nominal values, and feasible ranges for the triazine, LOCAT®, and SourCat™ processes. SA includes the gas flow rate and H_2S concentration of the sour gas for all three processes. The operating temperature and pressure were considered for the triazine and LOCAT® processes along with parameters controlling the utility costs. As a result, a total of 6, 7, and 8 parameters were considered for the triazine, LOCAT®, and SourCat™ processes, respectively, which consequently required 140, 160, and 180 desulfurization cost calculations, respectively.

Table 2. Potential parameters for the triazine-based absorption process

Parameters	Nominal Value	Range
Temperature	80 °F	50–160 °F
Pressure	300 psi	300–600 psi
Gas flow rate	100 MSCFD	100–100000 MSCFD
H_2S concentration	500 ppm	500–2000 ppm
Triazine concentration	52 wt%	20–80 wt%
80 wt% triazine scavenger price	$2/kg	$1–2.3/kg

Table 3. Potential parameters for the LOCAT® process

Parameters	Nominal Value	Range
Pressure	14.7 psi	14.7–600 psi
Temperature	80 °F	60–110 °F
EDTA–FeNa concentration	250 ppm	250–3000 ppm
Sulfur cake water fraction	50 wt%	50–90 wt%
EDTA–FeNa price	$2.15	$1.6–2.36/kg
Gas flow rate	100 MSCFD	100–100000 MSCFD
H_2S concentration	500 ppm	500–2000 ppm

Table 4. Potential parameters for the SourCat™ process

Parameters	Nominal Value	Range
Gas flow rate	100 MSCFD	100–100000 MSCFD

H_2S concentration	500 ppm	500–2000 ppm
Conversion of SO_2	1.96%	0–20%
Conversion of Sulfur	96%	80–98%
V-3 split fraction	99%	80–100%
Sorbent price	$2/lb	$2–5/lb
Sorbent capacity	0.03 g H_2S/g sorbent	0.03–0.1 g H_2S/g sorbent
Catalyst price	$26/lb	$20–30/lb

2.4. Results and Discussion

2.4.1 Validation of the process simulation and cost estimation

The process simulation and cost estimation model proposed in this study were first validated by comparing the calculated results with reference values. Table 5 lists the reference operating costs of the triazine and LOCAT® processes. The simulation results, including stream tables for the triazine process operated at 1 MMSCF and 100 ppm, as well as the LOCAT® and SourCat™ processes operated at 100 MMSCF and 1000 ppm, are included in the Supplementary Material. Tables 5–7 list the calculated equipment sizes and purchase costs as well as the utility costs of the three processes based on the heat duties and liquid/vapor volume flow rates estimated from the simulation results. For the triazine-based absorption process, the raw material cost of triazine dominates the overall expenses, possibly because the process relies on chemical absorption, which requires a single vessel/column for adsorption without the need for additional utility or adsorbent regeneration. Thus, the absorbent price is a critical factor that significantly affects the cost and applicability of the desulfurization process. At feed conditions of 1 MMSCFD and 100 ppm, calculations indicated a cost of 0.51 $/MSCF based on the full operating cost, which

is lower than the reference cost. This result can be attributed to the consideration of only the utility and raw material costs in the economic analysis reported in this study in the Figure 9. For the LOCAT® process, the calculated operating (utility and raw material) cost was 966 $/ST, which is within the reference range. Thus, the developed simulation and cost estimation models can provide robust cost estimation results for the triazine and LOCAT® processes.

Table 5. Reference costs for the triazine-based absorption and LOCAT® processes (Speight 2018 et al.; Subero et al.2004)

Parameter	**Triazine-Based Absorption**	**LOCAT®**
H_2S concentration	100 ppm	1000 ppm
Gas flow rate	1 MMSCFD	0.01–200 MMSCFD
Operating cost	1 $/MSCF (Personal Communication)	$250–1000 Per Standard Ton Range (Chemicals And Electricity) (Echt et al., 2017; Judd and Eng, 2003)

Table 5. Equipment sizes and costs of the triazine-based absorption process

LOCAT® **Equipment Name**	**Dimensions**	**SourCat™** **Equipment Name**	**Size**	**Triazine-Based Process Equipment Names**	**Dimensions**
Absorption Tower T-1 (Diameter, m)	5.09	Air Compressor C-1 kW	32	Absorption Tower T-1 (Diameter, m)	1.24
Desorption Tower T-2 (Diameter, m)	5.08	Reactor R-1 (Volume, m^3)	30.72		
Air Compressor C-1 MW	2.89	Vassel V-1 (Volume, m^3)	332.00		
Filter F-1		Vassel V-2 (Volume, m^3)	43.02		
Pump P-1 MW	3.78	Filter-1	-		
		Heat Exchanger H-1 (cal-sec-K) 0	16600.0		
		Pump-1	81.29		

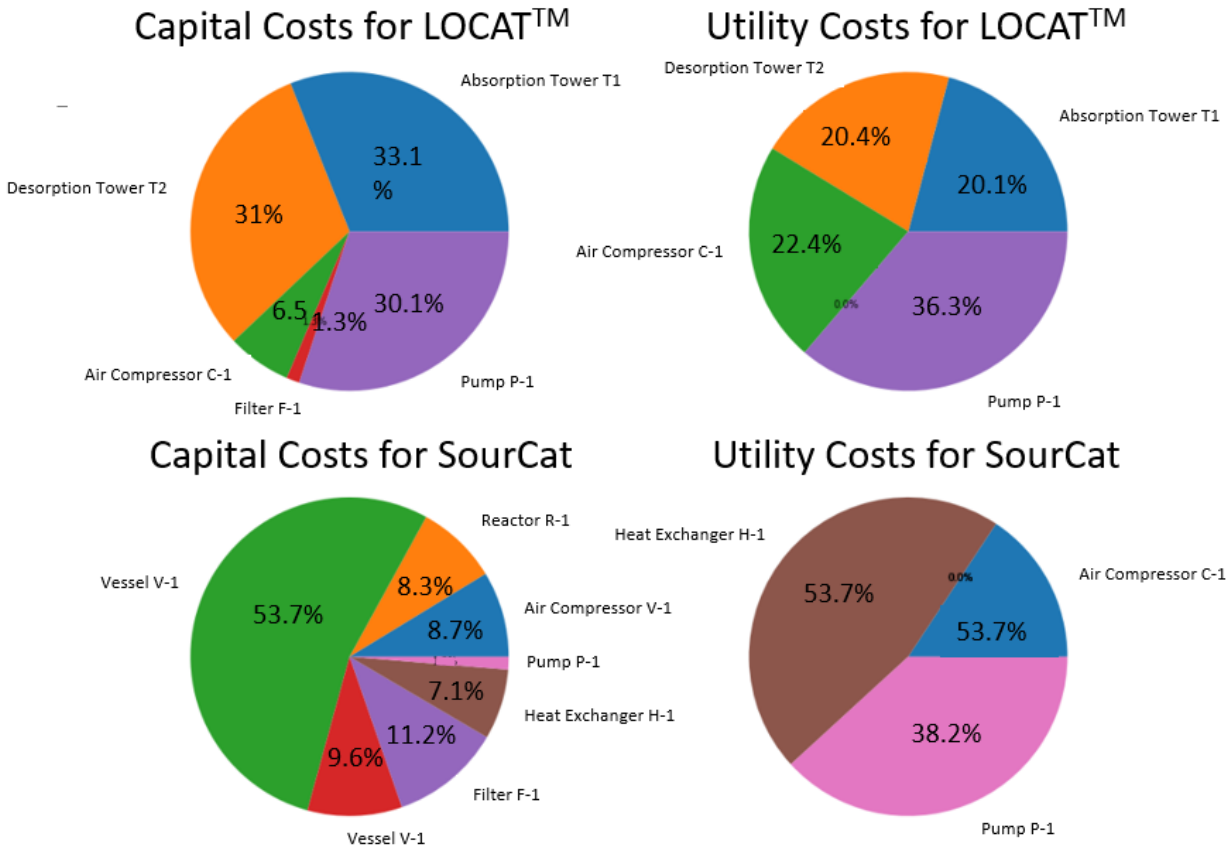

Figure 9. Equipment costs of the SourCat™ and LOCAT® processes.

2.4.2 Process Economic Analysis

Understanding how the desulfurization cost changes at different flow rates and H_2S concentrations for the three processes is critical for selecting a cost-effective desulfurization process for a specific flow rate and H_2S concentration. Here, the desulfurization costs before and after payback of the process equipment at different gas flow rates and H_2S concentrations were calculated for all three processes. Figure 10(a) and (b) show how the costs before and after payback change with the H_2S concentration at a sour-gas flow rate of 10,000 MSCFD. As expected, for all three processes, the desulfurization cost increases with increasing concentration of H_2S in the sour gas. Higher concentrations of H_2S require higher amounts of triazine to react with sulfur and remove it. Therefore, doubling the H_2S concentration results in a doubling of the amount of triazine required (and hence, its cost). Thus, although the total volume of gas requiring treatment increases, the concentration of H_2S remains constant. The chemical cost increases in this case, but this increment is less steep than that observed in the first scenario because the concentration of H_2S remains unchanged. Utility costs are likely to increase significantly owing to the increased volume of gas to be processed. The primary factor remains the need for effective treatment. Higher H_2S concentrations require more triazine to maintain the same level of sulfur removal, and an increased flow rate requires extensive processing, which leads to higher utility and disposal costs.

In the LOCAT® process, higher H_2S concentrations require more solvent to effectively absorb sulfur, resulting in an increase in the solvent procurement cost. Therefore, a significant increase in H_2S concentration may necessitate a substantial increase in solvent usage. Utility costs may also increase because of the need for additional solvent recovery and processing. Disposal costs could increase if the process generates large amounts of waste products or if the solvent requires special handling. At higher flow rates, more gas passes through the system; however, the

sulfur concentration remains constant. Although the amount of solvent required may increase, leading to higher procurement costs, the increment is lesser than that observed in the first scenario. Utility costs are likely to rise because of the higher volume of processed gas. Moreover, disposal costs increase with larger waste volumes. As a result, the operational costs related to raw materials and utility costs are expected to increase with increasing H_2S concentration. Both plots indicate that the cost for the SourCat™ process is the lowest among the three processes analyzed at a sour-gas flow rate of 10,000 MSCFD over a wide range of H_2S concentrations. During analysis, data related to the triazine process were excluded from Figure 11 (c) and (d) because the cost associated with this process is significantly higher at high flow rates than the costs of the LOCAT® and SourCat™ processes, primarily due to the proportional relationship between the amount of triazine required and the H_2S concentration in the gas stream. Thus, significant amounts of triazine are required at high flow rates, leading to an increase in the operational costs related to chemical procurement, utility, and disposal. In contrast, the LOCAT® and SourCat™ processes show cost-effective scalability at high flow rates, making them more suitable for inclusion in the comparative analysis shown in Figure 10 (c) and (d).

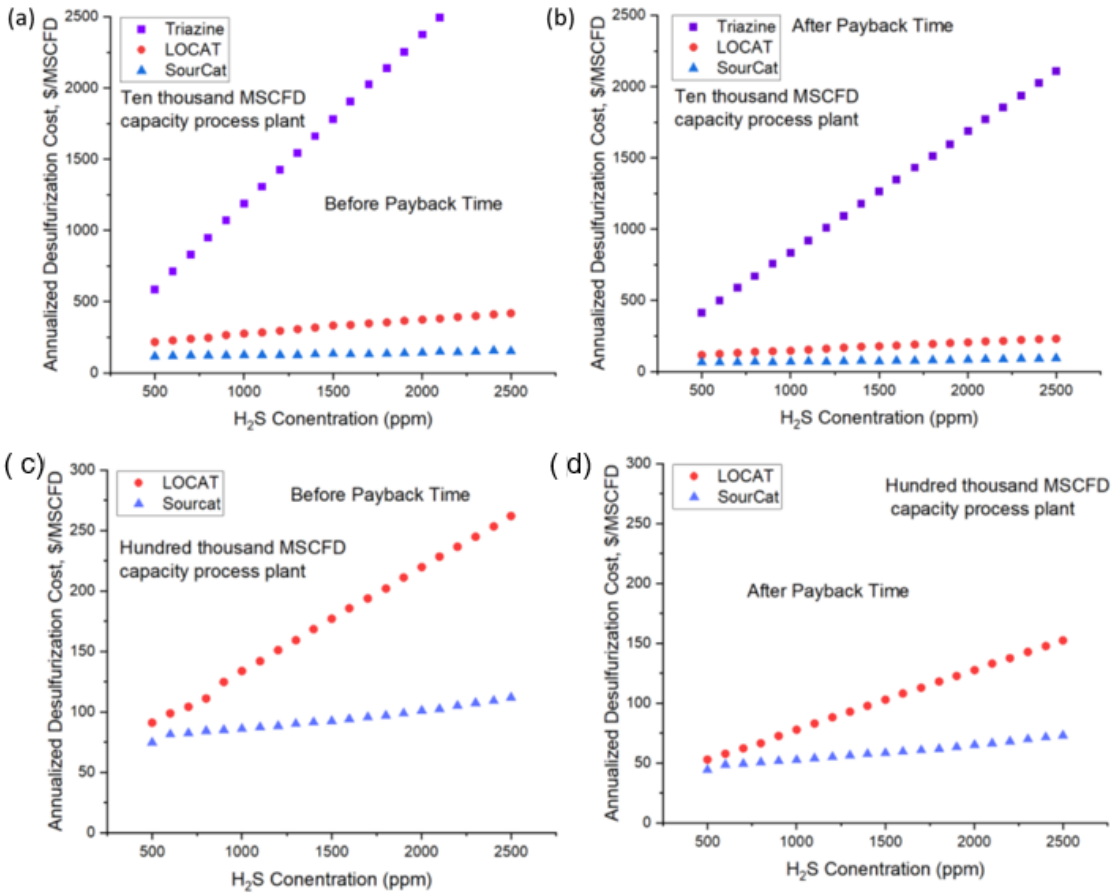

Figure 10. Desulfurization cost at a sour-gas flow rate of 10,000 MSCFD for triazine-based absorption and the LOCAT® and SourCat™ processes (a) before and (b) after the payback time. Desulfurization cost at a sour-gas flow rate of 100,000 MSCFD for triazine-based absorption and the LOCAT® and SourCat™ processes (c) before and (d) after the payback time.

The results shown in Figure 10(a)–(d) indicate that the SourCat™ process is the most cost-effective solution at high flow rates because it maintains the lowest costs (among all three tested methods) both before and after the payback period. This trend is observed across a wide range of H_2S concentrations, indicating the adaptability and advantages of the SourCat™ process under varied operating conditions. Figure 10(c) and (d) show that at low H_2S concentrations, the total costs of the LOCAT® process are approximately the same as those of the SourCat™ process. Although the LOCAT® process offers some competitive advantages over the SourCat™ process

at low H_2S concentrations, the process performance (in terms of cost-effectiveness) decreases significantly with increasing H_2S concentration. Thus, the LOCAT® process is might only suitable under certain scenarios.

Figure 12 shows a heat map comparing the cost-effectiveness of different desulfurization processes under varying operating conditions, focusing on different flow rates and H_2S compositions. This heatmap can enable an identification of the most cost-competitive process for specific source-gas conditions. The color gradations in the heatmap represent the relative cost efficiency of each process at various gas flowrates and H_2S concentrations. At high natural gas flow rates, specifically above 600 MSCFD, and over a wide range of H_2S concentrations, the SourCat™ process is the most cost-effective option. Conversely, at low natural gas flowrates, specifically under 600 MSCFD, and low H_2S concentrations, the triazine-based absorption process is the most economical. In addition, the heatmaps in Figure 12(c) and (d) indicate that under certain conditions, specifically at a high flow rate of ~100,000 MSCFD and low H_2S concentration of 500–600 ppm, the LOCAT® process can be cost-competitive. However, over a wide range of gas flow rates and H_2S concentrations, the SourCat™ process is the most cost-effective. Data related to the triazine process were excluded from Figure 10(c) and (d) because the operational costs of

this process at high flowrates were significantly higher than those of the other two processes.

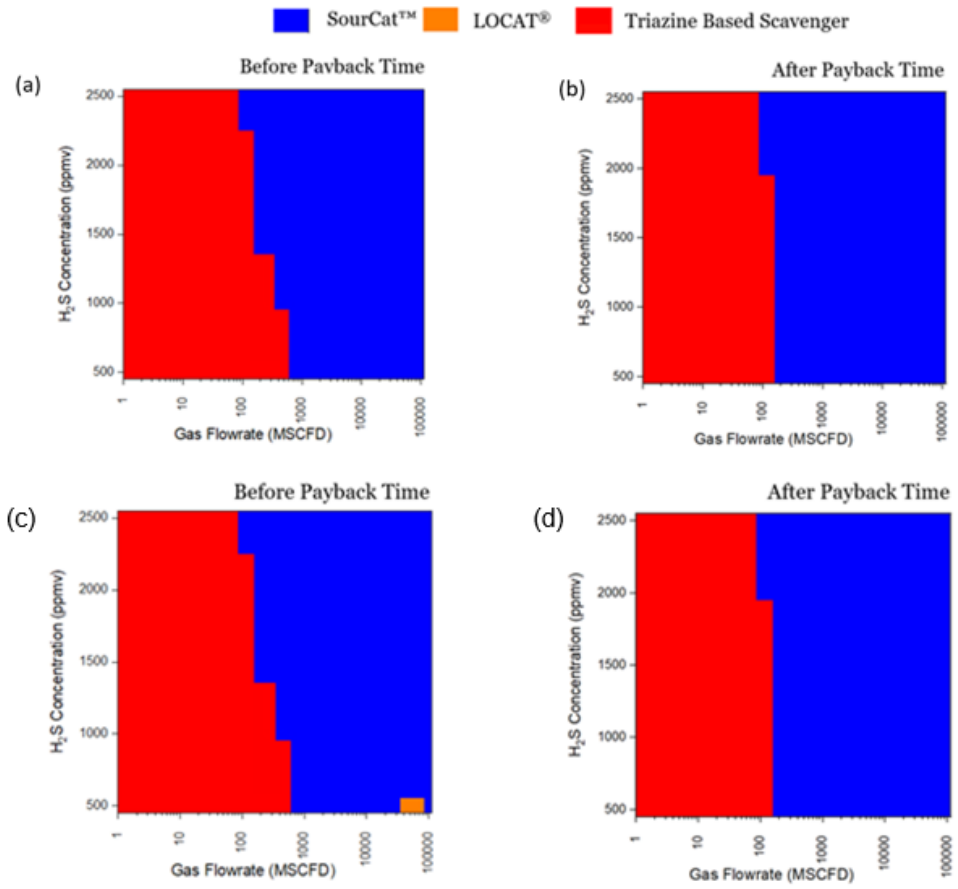

Figure 11. Desulfurization cost heatmaps of (a, b) conventional and (c, d) modularized processes.

2.5 Conclusion

In this study, techno-economic and sensitivity analyses were conducted for three small-scale natural gas desulfurization technologies. The techno-economic analysis results enabled the identification of a cost-effective operating region, while the SA results obtained using MOAT provided an understanding of the overall impact of the specific process and economic parameters on the desulfurization cost.

The technology with the lowest sulfur removal cost was identified by comparing the desulfurization costs of the three analyzed processes over a wide range of gas flowrates and H_2S concentrations for modularized and traditional configurations. The new SourCat™ process was confirmed to be the most cost-effective and economically favorable process for flow rates above 100 MSCFD, regardless of the H_2S concentration. In contrast, triazine-based absorption processes were identified as most suitable for small-capacity and low-flow rate systems. Within the flow rate and concentration ranges analyzed in this study, the LOCAT® process was only promising at high flow rates and low H_2S concentrations in a modularized configuration. For medium- to larger-size systems, the flexibility and scalability of the SourCat™ process outperformed the other tested technologies. SA results indicated that the parameters determining the cost of the triazine solvent material contributed the most to the desulfurization cost of the triazine-based absorption process. The desulfurization cost of the LOCAT® process was the most sensitive to parameters affecting the pump utility cost, while that of the SourCat™ process was determined by parameters relevant to the sorbent cost.

Future work will involve the selection of desulfurization techniques for gas fields using the developed economic models. Modularized techniques that can be easily transported and operated at various locations are expected to facilitate the utilization of stranded or flare gases. Studies on the applicability of modularized units for desulfurization in gas fields are expected to provide valuable insights into quantitative aspects of the modularization effectiveness and the potential limitations associated with the implementation of modularized techniques.

Chapter 3 – Techno-economic Study of the Effect of Ethylene Oxide Reactor Intensification on Downstream Separation Processes

This study discusses the critical effects of reactor intensification on downstream processes and the impact of different process expansion approaches on the total cost of the process and its environmental impact. First, a chemical processing system for separating ethylene oxide and carbon dioxide was simulated, revealing the changes in reactor intensification with the number of stages in the separation column and the influence of different parameters on the separation efficiency. Second, three expansion approaches, namely, the Brownfield, retrofitting, and Greenfield processes, were investigated with particular emphasis on the implications of their cost and carbon emission impacts. The Brownfield process offers cost reduction; however, its limited separation capacity hinders the attainment of the required reactor capacity. Retrofitting reduces production costs but introduces intricacies at high capacities (>130 kt/yr), necessitating additional separation columns. In contrast, the Greenfield process shows cost-effectiveness only at a production rate of ≥160 kt/yr along with the lowest carbon emissions. The retrofitted process involves a moderate increase in carbon emissions, whereas the Brownfield process involves a significant increase in carbon emissions with capacity expansion. These findings emphasize the interplay among production costs, emission considerations, and capacity limitations in traditional chemical processes, underscoring the necessity of designing specific approaches to suit specific production requirements and carbon emission goals.

3.1 Introduction

Ethylene oxide (EO), a high-volume chemical intermediate produced via highly exothermic partial oxidation reactions, is a key chemical used in various products, including plastics, detergents, and solvents. Its production is one of the most energy-intensive and inefficient processes in the chemical industry (Baharudin et al., 2021). Even with numerous studies aimed at increasing efficiency, current state-of-the-art EO processes produce large amounts of waste CO_2, typically through the conversion of 20–25% raw ethylene with low per-pass conversion and selectivity (Cai et al., 2022).

Similar to most chemical processes, reactors are the core of the EO process. A significant portion of the final overall cost and emissions are attributed to the upstream and downstream operations of the reactor, such as separation and compression, as determined by the reactor operating window and reaction effluent. Hence, the reactor design should consider process-level objectives related to economics and CO_2 emission metrics. Moreover, as the chemical industry accounts for ~9% of global greenhouse gas emissions, advancements should focus on carbon efficiency improvements (Peschel et al., 2011). Chemical engineering methods designed on the basis of the interdependence of reactor and process designs are expected to encourage the development of high-efficiency processes that are less wasteful.

Here, the synergy between the catalytic reaction, heat exchange, and membrane separation was investigated through an analytical approach using characteristic times and Damköhler numbers to enhance the exothermic reaction using dimethyl ether synthesis from CO_2-rich feedstock as the reference. This methodology highlights the optimal operating conditions and reveals that the proposed catalyst distribution in the reactor shows similar outcomes as a uniform distribution with a low catalyst mass and hotspot intensity [5]. The multilevel reactor design was

extended to optimal EO reactors by incorporating reaction routes, process intensification concepts, and heat and mass transfer criteria. A membrane reactor with an advanced cooling strategy was proposed using this approach to enhance the selectivity of ethylene epoxidation by ~3% compared with the reference case (Peschel et al., 2011). Furthermore, a new method for identifying the best reaction route by simultaneously optimizing the reaction and design parameters was introduced, which considers various factors, such as kinetics and process separation parameters. The O_2-based EO process reduces the potential cost, utility, and CO_2 emissions, exhibiting high potential to expedite the development of economical and sustainable chemical processes. Process retrofitting was developed using process intensification technologies to identify and evaluate optimal process intensification solutions from a database, addressing complex bottlenecks in the chemical industry, and its effectiveness was confirmed for EO production (Baharudin et al., 2021).

Advancements in process intensification in multifunctional reactor designs were reviewed, with particular emphasis on strategies to enhance chemical-species transport processes, integrate reactors with other operations, and address challenges in energy efficiency, space optimization, and cost-effectiveness. Synergistic effects can improve the chemical reaction efficiency, overcome thermodynamic constraints, and ensure safe and optimal productivity, highlighting the importance of novel concepts such as reactive distillation, reactive adsorption, and the design of suitable control systems (Baharudin et al., 2021). The key considerations for system advancement include heat management, catalyst loading, material choice, and tailored reactor designs (Peschel et al., 2012).

A packed bed reactor with orifice plates (PBR@OP) was designed for Prileschajew epoxidation, which exhibited better mass/heat transfer and reaction characteristics than standard reactors. PBR@OP increases the efficiency, safety, and cost-effectiveness of the process,

facilitating continuous epoxidation and outperforming stirred-tank reactors in terms of transfer rates and speed [2]. Biodiesel has emerged as a viable alternative energy source with advancements in production technologies addressing challenges such as mass transfer limitations and scalability. Ongoing research is focusing on optimizing existing methods, designing hybrid reactors, and using environmental assessment tools to measure the economic and ecological impacts of the process to ensure sustainable biodiesel production (Behloul et al., 2007). Luyben identified the “snowball” phenomenon in certain recycling processes, whereby small changes result in large alterations to the recycle flow rate. Subsequently, a mathematical analysis examined various kinetic systems and demonstrated the effects of different control structures on the severity of the snowball effect (Luyben et al., 1994.

Localized hotspots can originate from flow channeling inside the reactor, causing some tubes to operate under runaway conditions, leading to the complete oxidation of ethylene in the ignited tubes, which is undesirable and results in a reduction in the overall selectivity of the process (Baharudin et al., 2021). Conductive heat transfer to the active catalytic sites is restricted by negligible point contact between the catalyst pellets. Effective heat conduction is only possible for catalysts located immediately next to the internal wall of the reformer tube and not for catalyst pellets at the tube center (Baharudin et al., 2021). The impact of design and operational variables on the autothermal operation of type III reactors has previously revealed the optimal conditions for homogeneity as well as the requisite heat and mass transfer gradients. However, the focus on the steady-state, one-dimensional modeling, and physical property variations are among the limitations of this study. Nonetheless, the results of these studies pave the way for research that can provide deep insights into the reactor behavior and catalyst development for high-temperature and short-contact-time catalytic reactions.

Catalyst ignition in partial oxidation processes refers to the spontaneous combustion or heating of the catalyst during chemical reactions, which can increase CO_2 emissions (Luyben et al., 2019). In particular, heat is released when catalysts are ignited, which may initiate unintended reactions, generating CO_2 and other emissions. Thus, managing catalyst ignition is crucial for optimizing industrial processes and minimizing their environmental impact.

Newly developed microfibrous-entrapped catalysts (MFECs), nonwoven microfibrous metal meshes composed of nickel, steel, or copper, improve the thermal management of exothermic reactions by rapidly removing the reaction heat from catalyst-filled tubes through enhanced conduction (instead of convection) inside reactors (Sheng et al. 2013). Moreover, MFECs show high ignition prevention capacity and the ability to mitigate hotspots. The benefits of MFEC-based methods have been experimentally confirmed in various exothermic reaction systems, including Fischer–Tropsch synthesis systems (Choudhury et al., 2020). Moreover, MFECs with large surface areas and void spaces can facilitate a uniform flow distribution along the reactor. These properties of MFECs facilitate EO production, particularly in terms of increased per-pass ethylene conversion, hotspot prevention, enhanced process safety, stable performance, and extended catalytic activity (Luyben et al., 1994, Baharudin et al., 2021, Rozovskii et al., 2007).

Capacity increments and the multiscale production of chemical processes can be implemented through the Greenfield, Brownfield, and retrofitting modes. The Greenfield approach, which refers to building a new facility from the ground up on unused land, offers high flexibility in designing the plant layout, enabling the incorporation of the latest technology and optimizing energy, material flow, and adherence to environmental regulations. Brownfield expansion refers to building or expanding existing facilities. This method is typically more rapid and less expensive than the Greenfield process because most of the required infrastructure may

already be in place in the existing process plant. However, the design of a plant based on a previously developed plant is associated with several drawbacks and constraints in terms of design and operation. Retrofitting involves modifying existing facilities to enhance their performance or adapting them to meet new requirements.

This study investigated the critical effects of reactor intensification on downstream processes and the impact of different process intensification approaches on the cost and environmental effects of the process. First, a chemical processing system for separating EO and CO_2 was simulated, focusing on the number of stages in the separation column and the influence of the parameters on separation efficiency. Second, the Brownfield, retrofitting, and Greenfield approaches were analyzed.

3.2 Overview of the EO Process

3.2.1 Feed Conditioning and Conversion

The EO reactor, labeled R-101, comprises a multitubular structure with a packed plug flow system. An Ag-based catalyst with excellent selectivity and efficiency was selected for this procedure [14]. Considering the standard industrial-scale operational conditions for EO production, a surplus of ethylene was used. The resulting ethylene and EO conversions at the reactor exit were ~8% and 2% for EO, respectively. Figure 1 shows schematics of the EO process.

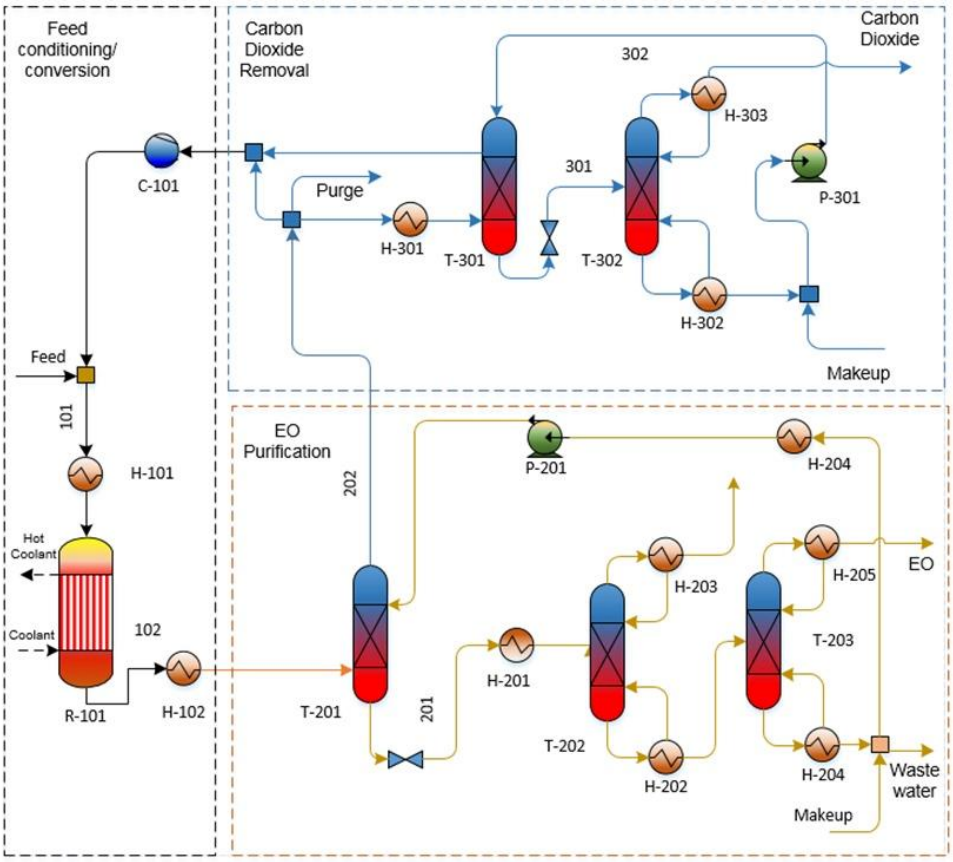

Figure 12. Process flow diagram of the EO process.

3.2.2 Refinement and Purification of EO

EO was purified by dissolving it in water under an increased pressure of 20 bar inside a countercurrent column labeled T-201. The water flow rate was adjusted to ensure complete reclamation of the dilute EO. The gaseous flow exiting the absorber with reduced EO content was divided into three distinct portions: one directed toward the CO_2 extraction segment and two recycled directly back to the reactor. This recycling mechanism is crucial because it curtails separation expenses. A minute segment (<1%) of the gas stream was discharged to avoid the accumulation of impurities. The remaining dissolved gases were then released into a T-202 column.

Finally, EO underwent desorption under diminished pressure in a T-203 column, resulting in a product with 99% purity.

3.2.3 CO_2 Extraction

CO_2 was effectively filtered using a combination of physical and chemical immersion in a warm potassium carbonate solution diluted to 30% in water. This process is unsuitable for systems containing monoethanolamines because of the potential risk of forming stable EO complexes. Here, the absorption process was modeled using the reaction kinetics reported by Kothandaraman. The absorption column was designed with numerous separation layers and operated at 20 bar, whereas the desorption column was operated at a reduced pressure of 1 bar. The purity of the extracted CO_2 stream was maintained at 95% to minimize solvent waste.

Four separation systems—acetone/chloroform, acetone/methanol, benzene/cyclohexane, and methanol/methyl acetate—and their potential solvents were selected from a review paper by Gerbaud et al. (2019). Table 2 provides an example of the property calculation and ranking results for the acetone/chloroform mixture for an equimolar composition in the inlet stream. Five solvents, dimethyl sulfoxide (DMSO), ethylene glycol (EG), chlorobenzene, o-xylene, and benzene, were used to separate the acetone/chloroform mixtures. The FSES model ranked the five solvents in the following order: DMSO > EG > chlorobenzene > o-xylene > benzene. The optimization results confirmed that DMSO was the best solvent for acetone/chloroform separation. However, contradictory to the optimization results, the FSES model predicted that o-xylene should outperform chlorobenzene. This contradiction possibly occurs because the system contains a separation boundary, which must be approximated using the residue curves calculated using the process simulation software. The solvent flow rate of o-xylene was 1.22 times larger than

chlorobenzene but 1.61 times larger than the derivative-free optimization (DFO) results, indicating that the o-xylene system shows a higher reboiler duty than predicted.

3.3 Methodology

3.3.1 Reactor Simulation

Catalyst ignition in partial oxidation processes refers to the spontaneous combustion or heating of the catalyst during chemical reactions, which can increase CO_2 emissions in the commodity industry (Luyben et al., 1994). Therefore, MFECs are considered prominent systems for transporting excess heat and mitigating excess emissions. Figure 2 shows a base-case Aspen Plus diagram of the EO process. Numerical simulation of a chemical reactor system, which is a common approach in chemical engineering and process modeling, relies on mathematical models to predict the behavior of chemical reactions under various conditions. The overall thermal performance of the EO reactor was assessed by examining key factors, including reaction rates, temperature changes along the reactor, heat transfer coefficients, and the pressure drop, as they are interrelated and significantly influence the reactor efficiency and safety. The simulation code incorporated the principles of reaction rates (Eq. (1)), reactor heat transfer (Eq. 2 and 3), and fluid dynamics, i.e., the friction factor (Eq. 4 and 5):

$$r = y_E y_{O_2} k P^2 k_0 e^{-\frac{E}{RT}} \quad (1)$$

$$\frac{dT}{dW} = \frac{\lambda r}{C_{P_E} F_E + C_{P_O} F_O + C_{P_{EO}} F_{EO}} \quad (2)$$

$$\frac{1}{U} = \frac{1}{h_w} + \frac{R_t B_i + 3}{R_t B_i + 4} \quad (3)$$

$$\frac{dP}{dW} = \frac{-f L_R \rho V^2}{D_p W_{cat}} \quad (4)$$

$$f = \frac{(1-\varepsilon)}{\varepsilon^3}\left[1.75 + 150\frac{(1-\varepsilon)}{Re}\right] \quad (5)$$

where r is the specific reaction rate, P is the total pressure, y_E is the ethylene concentration, y_{EO} is the EO concentration, W_{cat} is the total weight of the catalyst in the reactor, λ is the heat of the reaction, F_j is the molar flow rate of component j, C_{pj} is the molar heat capacity of component j, U is the overall heat transfer resistance, h_w is the heat transfer resistance at the wall, R_t is the tube diameter, λ_r is the effective radial thermal conductivity, B_i is the tube Biot number, T is the temperature along the reactor, L_R is the length of the reactor, ρ is the gas density, V is the superficial gas velocity, and D_p is the particle diameter of the catalyst.

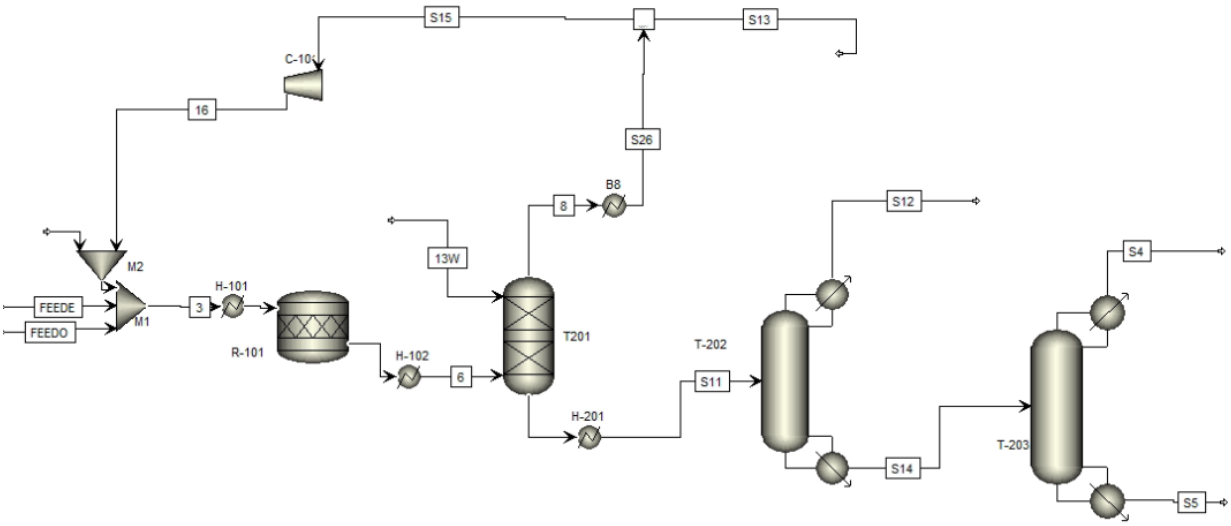

Figure 13. Base-case Aspen plus diagram of the EO process.

The reaction rate (Eq. 1) determines the rate at which the reactants are converted into products. Higher reaction rates are associated with enhanced heat generation, which affects the thermal profile of the reactor. The temperature variations (Eq. 2) within the reactor affect the reaction kinetics. Although higher temperatures can accelerate reactions, they may result in unwanted byproducts or catalyst ignition/deactivation. Therefore, monitoring the temperature profile is critical for optimizing the reaction conditions to ensure the desired reactions. Efficient heat transfer (Eq. 3) is essential for managing the temperature within the reactor. The heat transfer coefficient determines the rate at which heat is removed or added. A high heat-transfer coefficient can facilitate temperature control, whereas a low heat-transfer coefficient can result in temperature gradients and hotspots. Pressure drop (Eq. 4 and 5) occurs because of fluid flow resistance within the reactor, thereby affecting the overall performance of the reactor, including the flow rates and temperature distribution. High-pressure drops can increase energy consumption, thereby affecting the reactor efficiency. These phenomena during reactor simulation are related to the reaction rates,

which often increase with temperature. However, excessively high temperatures can lead to catalyst ignition or low reaction rates, i.e., low selectivity. Temperature changes affect the rate of heat generation, which can cause a thermal runaway if not managed by adequate cooling. Similarly, the pressure drop affects the flow rate, which can affect the heat transfer and temperature profiles in the system.

3.3.2 Process Simulation

In this study, a conventional O_2-based process was selected for producing EO, which offers superior efficiency in EO formation compared with air-based processes (Peschel et al., 2012. Based on the established EO capacities, an annual product flow rate of 100 kt with an EO purity of 99.1% was used as the benchmark. The UNIFAC group contribution method was used to model both the reactor section and EO feed conditioning and conversion phases to investigate the intricacies of this process. The accuracy of the proposed model was confirmed by juxtaposing the simulation results of this study with actual data from process plants to ensure alignment and validation. For the EO absorption and purification stages, SR-POLAR and experimentally derived CPA models based on a previous study were incorporated into calculations (Kapteijn et al., 2011). CO_2 separation was represented using the electrolyte nonrandom two-liquid model paired with the Redlich–Kwong equation for the vapor phase. Henry's coefficients were used to account for gas solubility, thereby making the modeling process holistic in nature.

The simultaneous simulation and optimization workflow involves combining the simulation of a chemical reactor process with optimization techniques to ensure the best results in terms of efficiency, cost, and environmental impact. Figure 14 shows schematics of the inputs and outputs of the EO process simulation and optimization. The workflow inputs included reactor specifications, plant capacity throughput, and base-case design data, which were used to create a

simulation model of the chemical reactor process. The simulation model computes various processes and performance metrics. Simultaneously, optimization algorithms were applied to the simulation model to determine the optimum values for the design parameters, operating conditions, and capacity utilization to minimize costs and reduce CO_2 emissions while meeting the desired plant capacity.

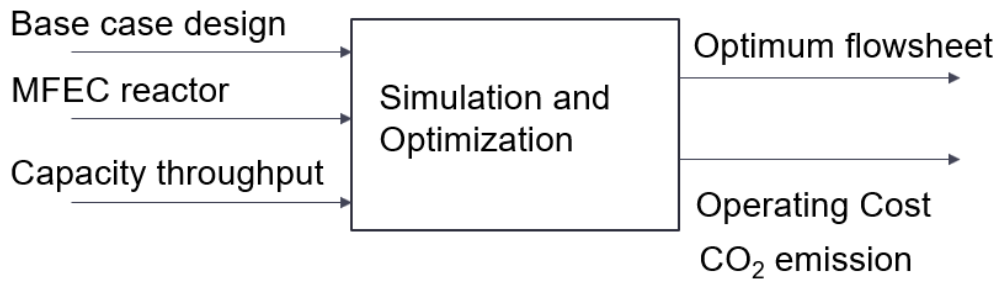

Figure 14. Workflow of the simulation and optimization of the EO process (Mukta et al. 2023).

DFO in Figure 15 is a technique used to solve optimization problems without an explicit mathematical expression or derivatives for the objective function. DFO relies on specific algorithms that iteratively improve the value of the objective function using sets of decision variables. The objective function is considered to be a black-box model, where the initial decision variable values are provided and the objective function values are obtained, which are used by the DFO algorithm to generate new decision variable values. This process is repeated until the termination criterion, which is often based on the maximum number of black-box model evaluations, is satisfied.

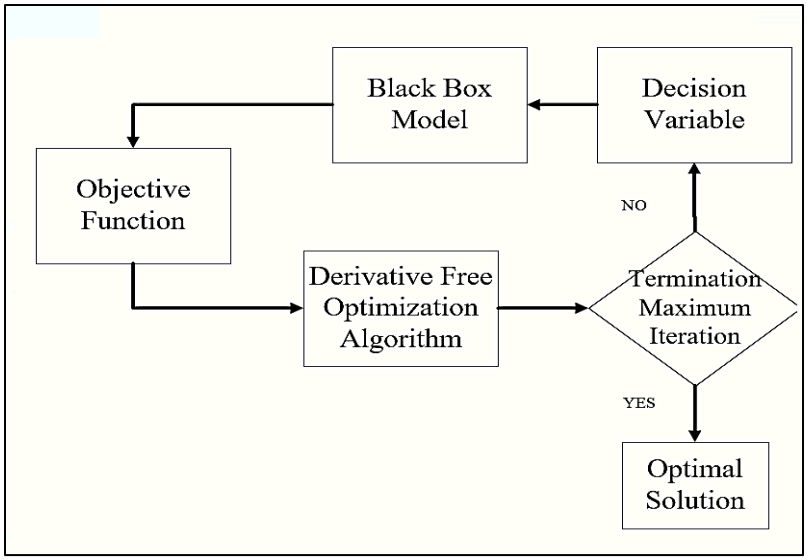

Figure 15. Flow diagram of the DFO algorithm.

In this study, a DFO algorithm suite called the radial basis function was used, where a 7a surrogate model-based gradient-free search method was used to select the optimal design and operating variables to minimize the operating cost. The objective functions (i.e., annualized costs) were calculated using the following optimization functions [Eqs. (6)–(13)]:

3.3.3 Process Economics

Table 1 outlines the various capacity increment options for chemical plants, focusing on the effects of downstream separation. The following four approaches were compared: Greenfield, Retrofit 1, Retrofit 2, and Brownfield. Strategies for plant modification or expansion with the associated capital cost implications and utility system requirements characterize each approach.

Table 6. Different expansion options and their capital cost consideration in chemical plants

Expansion Option	Description	Capital Cost Consideration	Utility Systems
Greenfield	A new plant on newly developed land	Capital cost calculations based on the grassroots equipment cost	New utility systems

Retrofit 1	Replacing the reactor or other equipment	Capital cost calculations based on modular equipment cost	Existing and new utility systems are being installed
Retrofit 2	Replacing reactor		Existing utility systems
Brownfield	Replacing the catalyst in the existing reactor		

The Greenfield approach, which involves the construction of a new plant on newly developed land, is characterized by capital-cost calculations based on grassroots equipment costs. This approach requires the installation of new utility systems to support the newly constructed plant infrastructure. The Retrofit 1 approach involves the replacement of the reactor or other significant equipment at the existing plant. Here, the capital cost for retrofitting was calculated based on modular equipment costs. In addition to existing utility systems, the option of installing new utility systems was utilized, thereby providing flexibility in upgrading the operational capability of the plant. Similar to Retrofit 1, the Retrofit 2 approach focused on modifying the existing plant by replacing the reactor. The capital cost is based on modular equipment costs, indicating a targeted investment in the need-based process unit. This retrofit option relies exclusively on existing utility systems, involving a more controlled approach (in terms of utility upgrades) than the Retrofit 1 method. The Brownfield approach involves replacing the catalyst in an existing process-plant reactor without changing the reactor shell. As fresh capital infusion in an existing process plant is difficult, this strategy, a minimal approach to increase capacity, focuses on enhancing the efficiency of existing equipment rather than replacing or adding major components. The capital-cost implications and utility system requirements are not explicitly mentioned for this option, implying a lower impact on both

compared with the other strategies.

$$min: TAC = \frac{i(i+1)^n}{(i+1)^n - 1} \times IC + AUC \quad (6)$$

$$st.\ IC = \sum_j Cost_j(q_{s,j}) \quad (7)$$

$$AUC = 24 \times 300 \times \sum_j Utility_j(q_{o,j}) \quad (8)$$

$$q = \Theta\ (x,\ z,\ p(x)) \quad (9)$$

$$\sum_j Utility_j \leq Utility_0 \quad (10)$$

$$x_{product} \geq purity \quad (11)$$

$$x \in X \quad (12)$$

$$z \in Z \quad (13)$$

3.3.4 Catalyst Performance

Effective heat removal from an exothermic reactor is a critical factor for enhancing reactor efficiency and overall process performance. In exothermic reactions, heat is generated as a byproduct of chemical reactions. Inefficient heat removal can cause several technical challenges that negatively affect the reactor efficiency. First, inadequate heat removal can increase the temperature within the reactor catalyst bed spatiotemporally, accelerating side reactions that produce unwanted CO_2, thereby reducing the selectivity and potentially causing thermal runaway with an uncontrollable increase in the ignition rate. This is a safety hazard that affects the yield of

the desired products. Second, localized high temperatures can degrade the catalyst used in the reactor, reducing its overall activity and lifespan. In terms of economics, catalyst deactivation can increase operating costs owing to frequent catalyst replacement or regeneration and excessive downtime. Finally, elevated temperatures can cause reactor corrosion and plug-flow reactor structural damage over time, necessitating expensive equipment maintenance and repair.

3.4 Results and Discussion

Effective heat removal techniques, such as the use of heat exchangers and cooling jackets in exothermic EO reactors, can help maintain optimal reaction conditions. However, maintaining such control in a conventional reactor is challenging. Precise temperature control results in reactions with high selectivity, the formation of few unwanted byproducts (such as CO_2), and high yields, ultimately increasing the reactor efficiency. Moreover, temperature modulation affects the reactor and subsequent separation equipment. Therefore, efficient heat management in reactors is fundamental for optimizing the exothermic reactions in EO process plants.

Spatial variation along a reactor is a fundamental parameter in chemical engineering and process modeling and plays a crucial role in enabling the analysis of the reaction kinetics and reactant/product concentration profiles. Figure 16 shows the effects of the EO reactor on different parameters of the downstream processes. The x-axis in the plots represents the reactor length (feet), signifying the spatial position along the EO reactor, where various properties, such as reaction conversions, pressure drop, heat transfer, and temperature, were observed and analyzed. Temperature profiles are important in chemical reactor analyses because they affect the reaction rates and equilibrium conversions (Fogler, 2015). In the MFEC reactor, the reactor temperature remained constant, whereas, in the conventional reactor, it fluctuated.

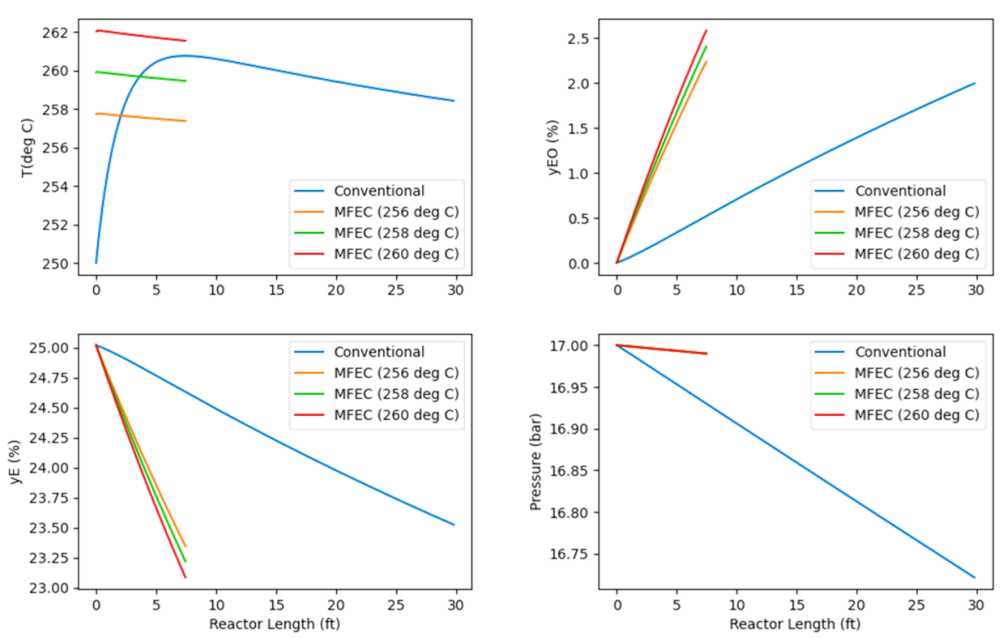

Figure 16. Effect of the EO reactor on downstream separation processes.

Here, the molar fraction profiles were evaluated to monitor the progress of the forward reactions and to identify the extent of reactant consumption and EO production rates, conversion, and yield. Pressure profiles are important in reactor design and safety considerations because they can affect reaction rates and system stability (Levenspiel, 1999).

The trends of the lines in these plots represent the evolution of various properties with reactor operation. Each line corresponds to a different set of input parameters, and the effects of varying the parameters on the reactor, such as the inlet reactant concentration, can be assessed by comparing the graphs.

4.1 Base-case simulation and results at different ignition rates

Table 2 summarizes data related to the requirements and parameters of various columns during EO chemical processing, specifically focusing on those involved in EO production and CO_2

separation. The changes in the requirements with the ignition percentage inside the reactor were elucidated in this study, and the effects on the number of stages in the columns and specific operational parameters of each column were adjusted. For the EO absorption column T-201, the number of stages remained constant for all 15 scenarios analyzed. However, the solvent flow rates varied from 188 to 209 t/h, suggesting that adjustments to the solvent flow can affect the efficiency of EO absorption. For the EO purifier column T-202, the number of stages varied from 14 to 18, and the number of feed stages varied from 5 to 9, indicating adjustments in the feed location. The pressure was maintained at 2 bars. These deviations highlight the importance of finely optimizing EO separation and purification processes to meet specific production rate and purity requirements. For the EO purifier column T-203, the number of stages varied significantly from 65 to 84, and the number of feed stages ranged from 25 to 50, varying with the feed stage placement. The reboiler duty also fluctuated from 29 to 31 MW, whereas a constant pressure of 1 bar was maintained throughout. These variations underscore the importance of customization for optimal EO purification. For the CO_2 absorption column T-301, the solvent flow rate increased from 21 to 48 t/h, indicating that adjustments to the T-301 column stage are critical for efficient CO_2 absorption. For the CO_2 desorption column T-302, the number of stages varied from 9 to 20 with three to four feed stages, underscoring the importance of flexibility for optimizing CO_2 desorption. The data listed in the table emphasize the importance of flexibility and customization in chemical processing systems to accommodate varying production requirements and achieve efficient separation. The choice of parameters, such as the solvent flow rate and number of stages, directly affects the efficiency of the separation columns. These results highlight the nature of the exothermic EO process and confirm the necessity of adjustments for achieving specific ethylene conversion and production rates.

Table 7. Optimal process configuration for the EO process

	0%	2%	4%	10%	BU
No. of stages in the EO absorption column T-201	15	15	15	15	8–28
Solvent flow rates in the EO absorption column T-201 (t/h)	188	191	195	209	1000
No. of stages in the EO purifier column T-202	14	14	15	18	5–20
No. of feed stages in the EO purifier column T-202	8	9	5	7	5–20
Pressure of the EO purifier column T-202 (bar)	2	2	2	2	1–15
No. of stages in the EO purifier column T-203	65	68	70	84	35–95
No. of feed stages in the EO purifier column T-203	50	32	25	44	35–95
Reboiler duty of the EO purifier column T-203 (MW)	29	30	30	31	-
Pressure of the EO purifier column T-203 (bar)	1	1	1	1	1–15
No. of stages in the CO_2 absorption column T-301	6	5	5	5	4–7
Solvent flow rates in the CO_2 absorption column T-301 (t/h)	21	26	31	48	-
No. of stages in the CO_2 desorption column T-302	9	11	13	20	3–22
No. of feed stages in the CO_2 desorption column T-302	4	4	3	3	3–22

The Brownfield, retrofitting, and Greenfield processes represent different approaches to chemical production, with varying cost and environmental impact implications, as shown in Figure 6. The Brownfield process initially offers a cost reduction. However, the separation capacity is hindered when the reactor capacity exceeds 112 kt/yr (Figure 17). Retrofitting reduces the production costs but results in complications as the production exceeds 130 kt/yr, necessitating the incorporation of additional EO separation columns. In contrast, the Greenfield process expansion is cost-effective when production is scaled up to ≥160 kt/yr and results in the lowest carbon emissions among all three tested approaches. Carbon emissions are moderately increased in the retrofitting process owing to the changes made during retrofitting and significantly increased with increasing capacity in the Brownfield process. These findings indicate a fine balance between production costs, emission impacts, and capacity limitations in EO processes, highlighting the importance of selecting a suitable separation-section design based on specific production and purity requirements and CO_2 emission goals.

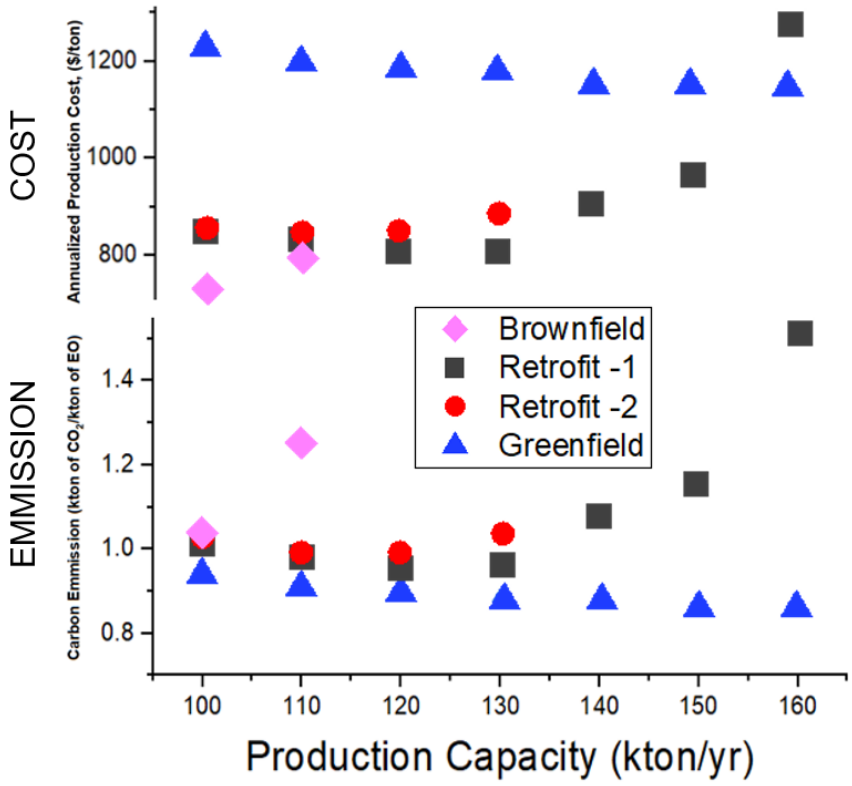

Figure 17. Effect of different modification processes on the production cost and emissions in EO separation processes with the plant capacity.

3.5 Conclusion

This study addressed two critical aspects of the chemical processing industry. First, it investigated the impact of varying EO reactor ignition conditions on EO and CO_2 separation design and efficiency. The results highlight the importance of optimizing separation processes to meet key capacity expansion requirements. Second, this study analyzed three distinct expansion approaches, namely, the Brownfield, retrofitted, and Greenfield approaches, with unique costs and emission implications. Although the Brownfield process initially reduced costs, it showed capacity limitations. Retrofitting resulted in cost savings but introduced operational difficulties at high capacities. In contrast, the Greenfield process was cost-effective at a high production rate and

showed the lowest carbon emissions among all tested alternatives. These results emphasize the necessity of carefully selecting a production expansion approach based on specific requirements and CO_2 emission objectives, highlighting the need for an exothermic EO process and tailored strategies to achieve efficiency, cost-effectiveness, and CO_2 emission goals.

Chapter 4 – Use of Large Language Models (LLMs) in Chemical Process Design

Dense feature relations in LLMs, which are trained on large amounts of data, are transforming most sectors in academia and exhibit high potential to streamline process engineering design by enabling rapid and efficient analysis, synthesis, and chemical process optimization. In process design, LLMs can excel in literature augmentation, extracting insights, and generating solutions by mining vast amounts of research data. By facilitating certain tasks, such as creating process flow diagrams, identifying alternative process flowsheets, and suggesting process options based on existing literature, LLMs can increase the speed of task completion and provide multiple potential options for conducting processes. Additionally, LLMs can interpret technical diagrams and convert process literature data into effective suggestions for researchers or engineers, enabling them to streamline design iterations and facilitate data-driven decisions in chemical engineering research.

4.1 Literature Review

The utilization of machine learning and artificial intelligence in chemical process design and engineering has been researched for several decades. Traditional machine learning approaches show promising potential in various sectors, such as catalyst design, reactor design, and flowsheet design, and leveraging techniques, such as support vector machines, random forests, and neural networks, can predict chemical properties, optimize reaction conditions, and model unit operations (Panerati et al., 2019). Despite significant advances in machine learning, developing natural language processing (NLP)-based models is critical, as traditional methods require extensive

feature engineering and struggle to capture feature relationships in large texts and engineering communications within the chemical process flowsheet. The emergence of transformer-based language models has the potential to significantly advance in the application of LLMs in process design and overall chemical engineering.

Moreover, the emergence of graph-based methods has marked a significant advancement in machine learning, as graph neural networks have enabled the representation of molecular structures (SMILE) and process flowsheets (SFILE) as graphs. This facilitates graph-based learning, which allows researchers to predict properties of either processes or molecules) and outcomes based on graph representations. Consequently, despite limitations in capturing long-range dependencies and complex processes, graph-based methods improve the local reasoning of inherent system structures.

Mann et al. proposed eSFILES for intelligent process flowsheet synthesis by combining process knowledge representations, symbolic AI techniques, and machine learning for flowsheet generation and optimization. This research might prove to be pivotal in shaping the use of LLMs in process flowsheet designs. Their framework integrates process flowsheet-specific process knowledge with symbolic AI methods, including novel flowsheet grammar and graph-based structural representations. The system's hypergraphic representation of flowsheets, with process groups as edges and streams as nodes, enables rich annotations. These key technologies might enable the curation of a large database of previously developed flowsheets that are suitable for LLM models. This combination might enable LLMs to generate meaningful flowsheets based on the literature while leveraging the power of data-driven approaches through machine learning integration. LLMs enable rich feature relations and contextual information, might facilitate the informed analysis of the traditional process design. Moreover, their potential extensibility and

ability to interface with commercial process simulators might enhance their impact on chemical and process systems engineering.

ProcessBERT, developed by Kato et al., enables the equivalence judgment of variable definitions in process models through the pretraining of process descriptions. ProcessBERT was pretrained on ChemECorpus, a dataset constructed from over 130,000 chemical engineering papers, allowing it to understand domain-specific language and concepts. The model uses the BERT architecture and is explicitly fine-tuned to compare the variable definitions in processes. The author reported that ProcessBERT outperformed several other language models, including the original BERT and SciBERT models, in the similarity-based method. This study aims to use the model for standardizing variable definitions across multiple process-model documents, which is a laborious task.

Zhang et al. introduced a molecular grammar tree transformer (G-MATT) for single-step retrosynthesis prediction, leveraging molecular grammar trees and transformer architectures to predict retrosynthesis steps directly from molecular structures traditionally based on SMILEs, which according to the autors outperforms previously reported graph-based methods. In this work, researchers used molecular grammar trees, i.e., G-MATT, which represents molecules as grammar trees rather than traditional graph structures. This representation captures the hierarchical nature of molecular structures, including local structures and functional groups, which are often overlooked in traditional SMILES-based representations. G-MATT combines powerful data-driven models with prior domain knowledge and incorporates crucial chemical information into the prediction process. G-MATT outperforms baseline retrosynthesis models. This is particularly impressive for the USPTO-50K dataset.

Similarly, polyBERT, a chemical language model developed by Kenneth and Ramprasad (Kuenneth et al, 2023) facilitates rapid polymer property prediction by treating chemical structures as a "language" and generating polymer fingerprints using the NLP technique. The transformer model was trained on a large set dataset of 80 million polymer chemical structures based on previously reported polymers, making it useful for understanding the language of polymers, i.e., polymer structures. This model generates polymer fingerprints (numerical representations of polymer structures) more efficiently (over two orders of magnitude faster) than traditional methods of polymer structure generation. PolyBERT's multitask deep neural networks enable the simultaneous prediction of multiple polymer properties by leveraging correlations within data.

Furthermore, Balhorn et al. explored the potential of generative model for autocorrecting incomplete chemical process flowsheets, achieving high accuracy on artificial datasets. Balhorn et al. proposed an innovative generative model using an autoencoder architecture to automatically identify errors in chemical process flowsheets and suggested complete corrections using the proposed model. The success of LLMs in grammatical autocorrection, which utilizes multithreaded attention to model complex relations in human language, inspired studies to enable autocorrection in flowsheets. Subsequently, a model that inputs a potentially erroneous and incomplete flowsheet and outputs suggestions for a corrected flowsheet was developed.

Current trends underscore promising future research directions in language models, including end-to-end process design, multimodal process understanding, developing interactive design assistants, enabling transfer learning for specialized domains, uncertainty quantification and risk assessment, sustainable process design, and integration with process simulation.

These developments and potential research trajectories underline the potential transformative impact of LLMs on process design practices and may pave the way for highly

advanced AI-powered tools that augment the capabilities of human engineers. Our work utilizes multithreaded attention to model the complex relations of human language, inspired by the approach of finding the same in flowsheets with a contextual understanding of the flowsheet, graphs and diagrams. Evaluating the performance of LLMs can be challenging because they often lack practical grounding in real-world applications. To bridge this gap, developing methods to evaluate the effectiveness of such models in various scenarios is crucial. For tasks such as programming, tools and contexts designed to measure the performance of LLMs in generating, debugging, and understanding code are essential. This ensures the evaluation of contemporary theoretical benchmarks, focusing on the physical outcomes and practical utility of the applied work.

4.2 Introduction to LLM

LLMs are based on a generative pretrained transformer that is predictive in nature, trained on large text-based datasets to predict the next word based on some initial text, and an extensively trained, dense, neural network-based model developed to find relations between texts over a long range. The attention mechanism in the transformer architecture using parallel processing is equipped to handle vast amounts of text data. Some key concepts used in the LLM are discussed next (Milakov et al. 2013, 2018) .

Tokenization is an important step in which text is converted into a numerical matrix based on a specific vocabulary to train a language model. Researchers have developed various transformations or encoding (numerical representations of real-world data, such as text or other types of data) models for efficient tokenization. Models trained on generalized vocabularies tend to perform unsatisfactorily in process design tasks. However, a flowsheet is essential for the process design representation of specialized vocabulary, particularly in a specialized context, and specialized vocabulary requires models to be fully trained.

Text is expressed as low-dimensional vectors, where the geometric distance between two vectors in the vector space is a projection of the relationship between two real-world objects represented by the vectors. Vectors enable compact representations of different types of data while allowing a comparison of two data objects and an analysis of the similarities or differences between data types on a numerical scale. Benefits of the vector representation of text are provided in the Results and Discussion section.

The generative capabilities of LLMs inspired from recent advancements in generative models, specifically in autoencoders and generative adversarial networks (GANs). Here, two different neural networks were first trained to create a latent representation of the data and then to predict the latent representation. Moreover, variation autoencoders have introduced probabilistic generative modeling using latent variables. GANs use competitive optimization to produce realistic data samples, and transformers integrate self-attention mechanisms, enabling parallel processing and enhancing the overall generative performance of the system.

Long short-term memory (LSTM) and gated recurrent unit (GRU) are common recurrent neural networks (RNNs), which are used for language problems, such as machine translation, text classification, text summarization, and question answering (Hochreiter et al. 1997 . RNNs sequentially process the input and output sequences to generate a sequence of hidden states based on the previous hidden state and current input (Cho et al 2004). The sequential nature of RNNs is found to be computation-intensive, and state space modeling helps parallelize the tasks in the process.

Transformers are a class of neural networks that leverage the self-attention mechanism to process sequences of tokens in parallel, enabling the effective modeling of long-term contexts, which results in significantly shorter training times and improved handling of long-term

dependencies in long-sequence tasks. (Dao et al. 2022; Devlin et al.2018)The output size of the transformer encoder is scaled linearly with the input size. The transformer comprises multiple layers, including the multi-head attention, layered norms, feedforward, linear, and softmax layers, which can be categorized as input, hidden, and output (Mikolov et al. 2013, 2018).

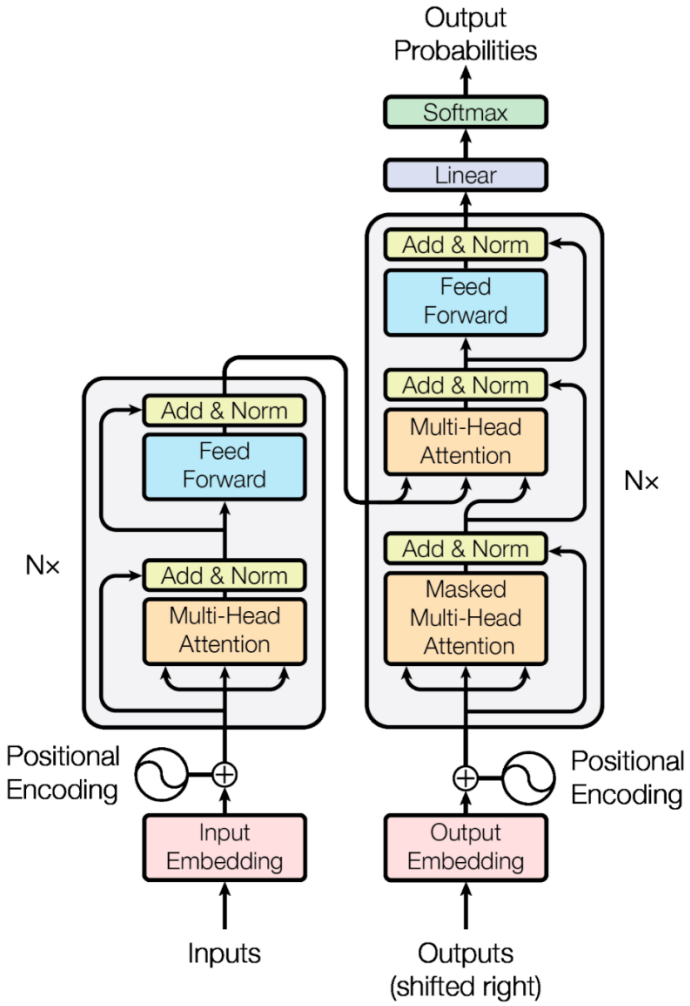

Figure 18. Transformer architecture of Large Language Models (LLMs).

As shown in Figure 18, (Dao et al. 2022; Devlin et al.2018) the input layers, such as input/output embedding, receive raw data and represent input tokens. The output layers, as exemplified by Softmax, produce the final output. Hidden layers, including the multihead attention

layer, perform core processing between the input and output layers. In machine translation models, the input embeddings represent the source language words, whereas the output embeddings represent the target language words.

Here, the attention mechanism is discussed for the following scenario: "A distillation column separates a feed into two streams (the distillate and residue) because the products need to be separated as the individual components have different boiling points." In the process of flowsheet design, understanding the relationships between different units and streams is important. In the above-mentioned scenario, the "distillation column" and "streams" are intrinsically connected because they are the outputs of column operation. In this context, self-attention can be thought of as a mechanism to mathematically determine the relationships between components and their roles in the process. For instance, it identifies how the output streams depend on the feed composition and column operation.. Matchmatically creating queries, keys, and values after linear transformation from token (as discussed numerical representation of text) are quantified using the scaled dot product of the query and key vectors (Q and K, respectively) and normalized with the softmax function to compute the attention weights, as follows:

$$\boldsymbol{Q = XW_Q} \quad \mathbf{1}$$

$$\boldsymbol{K = XW_K} \quad 2$$

$$\boldsymbol{V = XW_V} \quad 3$$

$$\mathbf{Attention\ weight\ } = \mathbf{V} \times \mathbf{Softmax}\left(\frac{\boldsymbol{Q \times K^T}}{\sqrt{\boldsymbol{d_k}}}\right) \quad 4$$

Weight matrices (Wq, Wk, and Wv) (Equeation 1-4) were used to generate the Q, K, and value (V) vectors, which were used to represent each word. Subsequently, the dot product of the Q vector of each token with the K vectors of all token in the sequence was computed by simultaneously stacking the Q, K, and V vectors for all tokens into the Q, K, and V matrices and multiplying them, as shown in Figure 3. Basic operation of attention: 1 with Q=query, K=Key, V=Value, Z=Attention, and d_k=dimension of queries and keys.

In the transformer after converting the input tokens into embedding vectors, these embeddings were fed into the multi-head attention module. Self-attention is a crucial mechanism in transformers that enables the system to focus on specific parts of the input sequence relevant to the task at hand and capture long-range dependencies (probabilities) might be effectively than traditional RNNs.

4.3 Literature Data Extraction

Various NLP techniques have been used to extract structured and unstructured data from the literature. Using various pythonic libraries such as spaCy or PDFMiner for text extraction, researcher in process design do communicate, convey and keep record through various additinal methods such as graphs, flowsheets needing extraction relevant data.

4.3.1 Preprocessing

The extracted text needed to be cleaned and preprocessed by removing irrelevant information (such as references and footnotes) and normalizing content to develop a domain-specific vocabulary representing key terms, processes, and relationships relevant to graphs and flowsheets. However, this requires complete LLM model pretraining. As pretraining is resource-intensive, the fine-tuning of existing LLMs is an attractive strategy. Clustering algorithms, such as LDA or k-means, can categorize topics, improving the ability of the LLM to understand the context within the literature.

4.3.2 Graph-Type Classification

A dataset collection comprises a dataset of various graph types (such as hierarchical, cyclic, flow, directed, and undirected) with labeled examples. Instead of text-to-number representation conversions of general text data, graph neural networks (GNNs) or graph embeddings, such as GNN, can convert graph structures into numerical representations. These conversions represent graph types in a vector format and can be used as inputs for type recognition and classification. Here, graphs generated based on flowsheet data were used to augment the dataset for training and fine-tuning the language model.

4.3.3. Graph-to-Data (Multiclass/Multi-Type Labeled Data)

This process involves the conversion of graphs representing numerical into structured formats (such as adjacency matrices and edge lists). If the graphs contain multiple labels, the labels are represented in a hierarchical or multi-dimensional format. Convolutional Neural Network CNNs are used to learn the structure and node relationships in the system. Multiclass classification models, such as hierarchical classifiers, could be applied to understand and classify data based on

multiple graph labels, which are individually represented by a unique identifier, including class and label information.

Figure 19. Graph diagram to data conversion.

In the figure 21 one example is showen a graph can be converted to numerical datasets for the use in LLMs. In annotation, different graph types are annotated based on their structures and relationships to create a dataset that teaches the LLM to interpret and convert graphs into data points.

4.3.4 Flowsheet and Flowcharts to Graphs

Pretrained models or custom classifiers are implemented to detect PFD unit operations (Figure 19) and map them onto the corresponding graph nodes. The graph representation of a PFD comprises a graph in which the components are nodes and pipes/lines are edges. Here, the directionality and flow information within the graph structure were maintained to accurately reflect

the process flow. Collecting and converting multiple PFDs into labeled graphs creates a dataset that trains the LLM to accurately interpret PFDs. Here, configurations and layouts were included to generalize the LLM across different process flows.

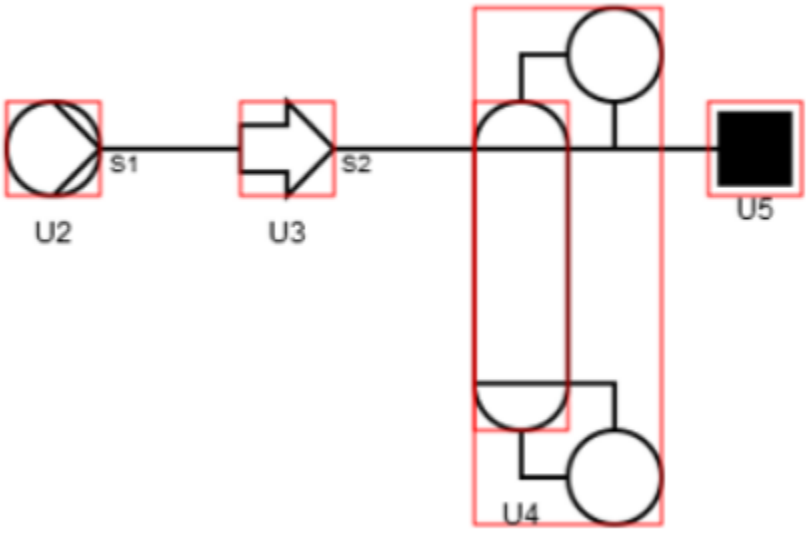

Figure 20. Process flow diagram (PFD) for the eSFILES synthesis approach.

A CNN model has been used to identify the nodes and edges in the flowcharts to detect shapes (such as circles for start/end and diamonds for decisions) and text using character recognition image processing techniques. These components are mapped to nodes and edges in a graph structure, preserving directional relationships for accurate flow representation. Each step in the flowchart corresponds to a node, and arrows or connections between shapes are translated into directed edges of the graphs. Annotating nodes and edges with information such as node type, directionality, and condition (if applicable) can improve the performance of LLMs in flowchart-

based queries. Object detection and symbol recognition techniques are used to identify the components of process flow diagrams (PFDs) (including pumps, tanks, and valves).

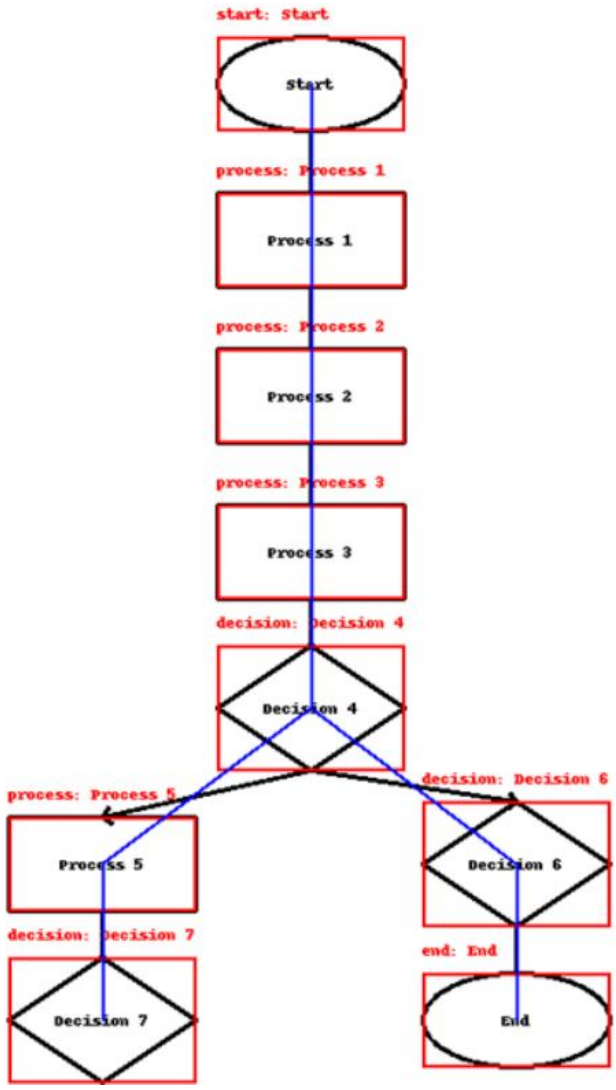

Figure 21. Process flow diagram representation using a graph structure for visualizing process steps and interconnections.

4.3.5 Data Collection

Very few open-source data journals and repositories in the process design literature contain adequate resources for data collection to test and refine LLMs are avialable. . Although journals

publish open datasets in various fields, this study focused on the process design for training LLMs. The datasets are peer-reviewed and can be used to train and test LLMs.

4.4 Processing and Training Language Model

Text preprocessing standardizes and cleans input data, removes irrelevant text, normalizes terminology, and tokenizes sentences to identify key terms such as unit operations, streams, and chemical components, providing structured input to the LLM while reducing uncertainty and noise. Model selection and hyperparameter optimization select an appropriate LLM, such as GPTs, LAMMA, Gemini, or parameter optimization, for the process design task. Here, the base model was selected with sufficient capacity and fine-tuning parameters for process-design-specific data, resulting in a tuned LLM capable of understanding relevant terminology and generating contextually accurate flowsheet designs.

The conversion of text data into vector representations using token embeddings helps context-aware decision-making, whereas fine-tuning LLMs trains the model on domain-specific datasets for flowsheet generation and task automation, using transfer learning to retain a general linguistic understanding while specializing in flowsheets.

Fine-tuning the LLM on a combined dataset involves the incorporation of prompts or queries related to each data type (such as "convert this flowchart to a graph" and "classify this graph type") into the workflow.

The LLM is tested on new examples from each category the mythology given in Figure 22, and iterations are carried out based on the model performance to adjust the preprocessing, labeling, or fine-tuning steps. This structured approach allows the LLM to handle various data types effectively and precisely understand and process the literature, graphs, flowcharts, and PFDs.

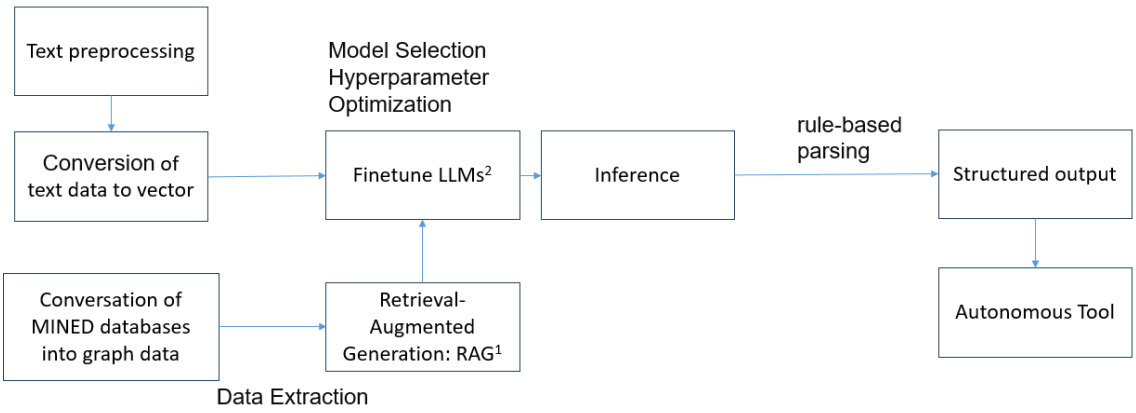

Figure 22. Language model-based autonomous tool.

Inference is used in the trained model to generate or analyze the code generated by inputting problem statements and generating structured flowsheets, including units and connections, as textual or graphical outputs, providing automated solutions for design tasks.

Rule-based generation enhances the structured generation of code elements by integrating process design rules, ensuring physical phenomena (such as mass and energy balance), whereas structured output converts LLM-generated text into machine-readable formats for further processing and integration with tools such as Pyflowsheet.

The conversion of collected databases into graph data involves depicting chemical engineering databases (chemical properties and flowsheets) as graph structures, extracting relationships, and representing them as directed graphs, hence facilitating efficient querying and reasoning during code generation. Retrieval uses relevant research data to augment the output of a model by retrieving flowsheet templates, best practices, solved examples, and matching queries with stored data using vector similarity.

Augmented generation combines retrieval-based data, i.e., obtaining data from the vector representation of the database with LLM generation, guiding LLMs to generate scripts by incorporating retrieved knowledge with high accuracy and rapidly generating reliable outputs. These steps contribute toward solving the process design and facilitate the automation of programming tools, improving the efficiency and accuracy of the system, reducing manual labor, and increasing the potential for innovation.

Here, the performance of the model was assessed across several problems, with each task targeting specific aspects of the process design and simulations. The evaluation spanned core concepts from basic unit operations to advanced process analysis and control. A detailed breakdown of the problems addressed is provided in Table 8. As the context length of both models is a limiting factor, prompts were set to create the shortest possible code to test the performance of each model.

Table 8 Performance comparison datasets for GPT and Chen_LLM

Problem	Description
Unit converter	Code to convert units (such as meters to feet, Celsius to Fahrenheit).
Mixing material balance	Calculate mass flow rates in a mixing tank with two inlet streams.
Heat capacity calculator	Compute heat capacity using empirical formulas provided.
Vapor pressure estimator	Estimate vapor pressure using the Antoine equation.
Heat exchanger design	Size a heat exchanger using the LMTD method.
Batch reactor simulation	Model concentration changes over time in a batch reactor.
Binary distillation column	Design a binary distillation column using the McCabe–Thiele method.
Pump power requirement	Calculate the power required for a pump to operate at a certain flow rate and head.

Reaction yield optimization	Optimize temperature to maximize the reaction yield in a reactor.
Multicomponent distillation	Simulate distillation of a ternary mixture using shortcut methods.
Non-ideal reactor model	Model a reactor using the axial dispersion.
Process with the recycle loop	Simulate a process flow with recycle and purge streams.
Pinch analysis tool	Perform pinch analysis for heat integration in a process.
PID control implementation	Code a PID controller for temperature control in a reactor.
Fluid flow simulation	Simulate laminar flow in a pipe using simplified Navier–Stokes equations.

- LMTD: Log mean temperature difference

4.5 Results and Conclusion

Initially, the attention mechanism within the transformer models was investigated to elucidate the interactions between tokens using the graphical representation shown in Figure 23 and 24. Tokens are conceptualized as nodes in the graph, and attention weights are represented by directed edges weighted by attention scores derived from the query–key similarity.

The attention mechanism is modeled as a graph G=(V, E) comprising tokens (V) and weighted edges (E), indicating a token-to-token probabilistic relation. Token numerical conversions are transformed into Q, K, and V vectors using the learned weight matrices.

Token interactions are quantified using the scaled dot product of the query and key vectors and normalized with the softmax function to compute attention weights. These weights determine the importance assigned to the other tokens while updating the representations.

Moreover, the weight difference and its position also affect the individual scores in the heat map, indicating the process of training transformers in LLMs.

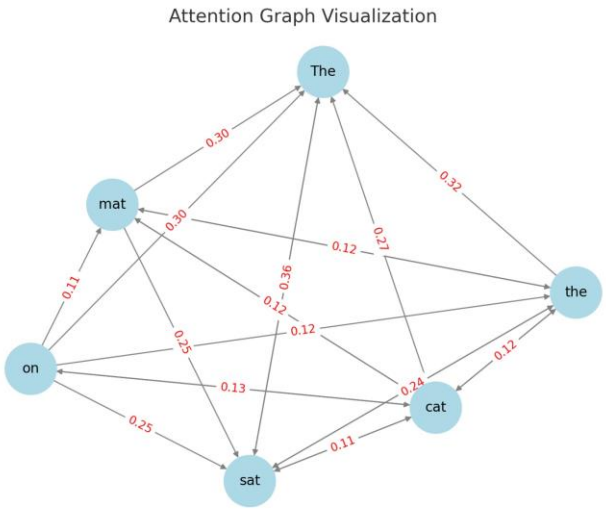

Figure 23. Attention graph visualization shows tokens as nodes with directed, weighted edges representing attention scores.

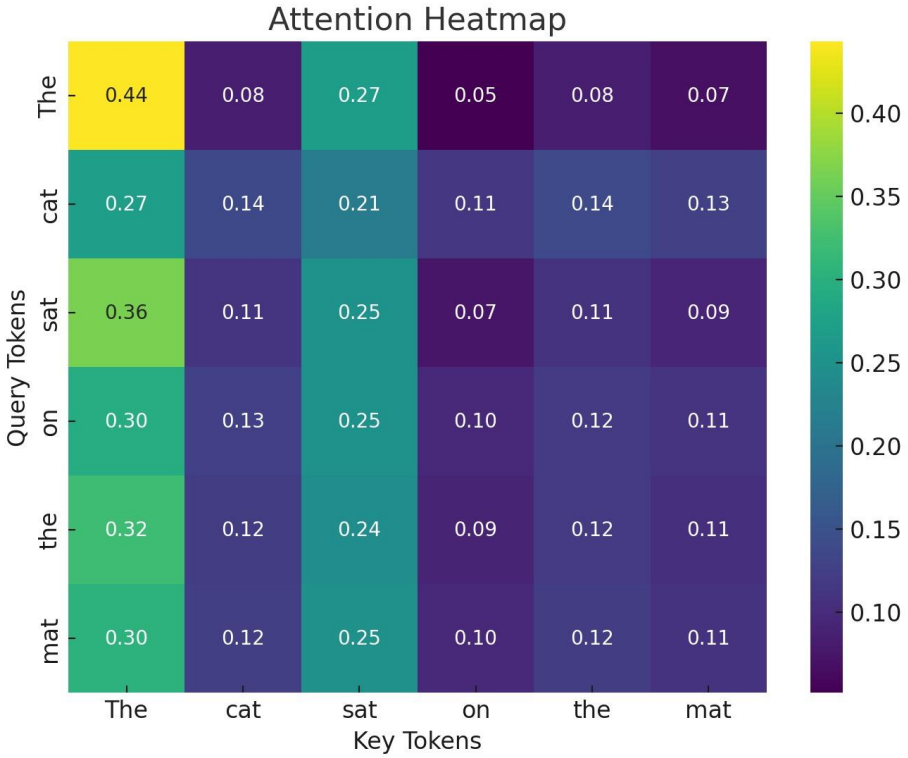

Figure 24. Attention heat map attention weights with rows representing query tokens.

Here, the attention was taught to capture long-distance feature relations, aiming to compare the performance of transformer models and RNNs in capturing feature relations within a general short text and over long-distance text, specifically in the context of a process design language. The ability to process a longer, specific text is related to the distillation column design. Here, only single-headed attention and a relatively simple RNN were tested.

The RNN performance is depicted using hidden state activations over the token range, which processes sequences sequentially. In contrast, the attention heat map of the transformer confirms the ability of the system to capture feature relations between all words simultaneously and effectively identify long-distance relationships in provided text.

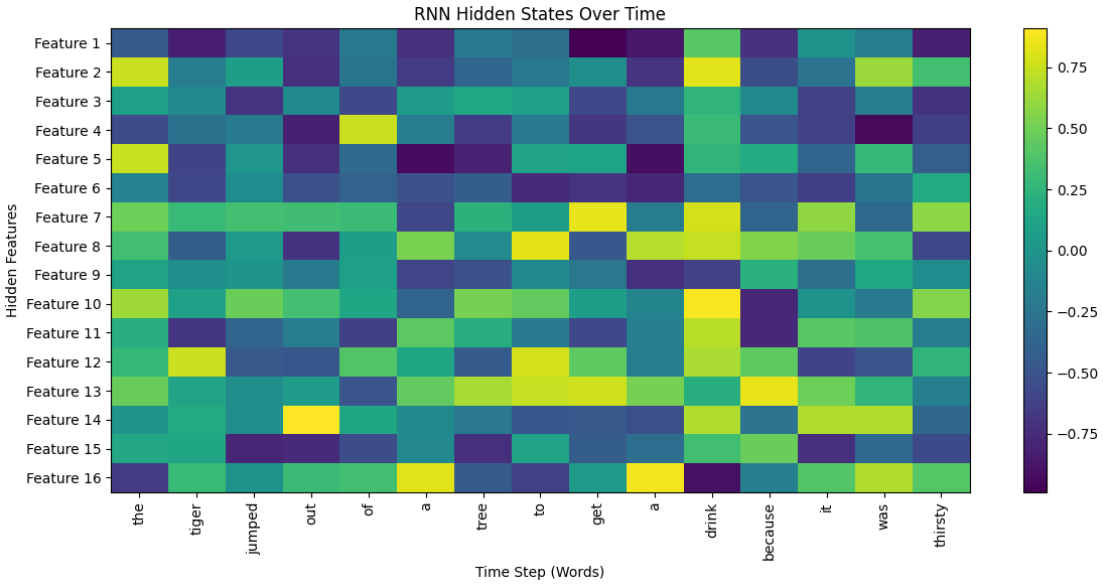

Figure 25. RNN hidden states for short text.

Figures 25and 26 show the utilization of a linear transformation of numerical data by a transformer model; the dot product effectively identifies dependencies across the entire text, precisely

identifying relationships between key terms and concepts, regardless of their positions in the sequence.

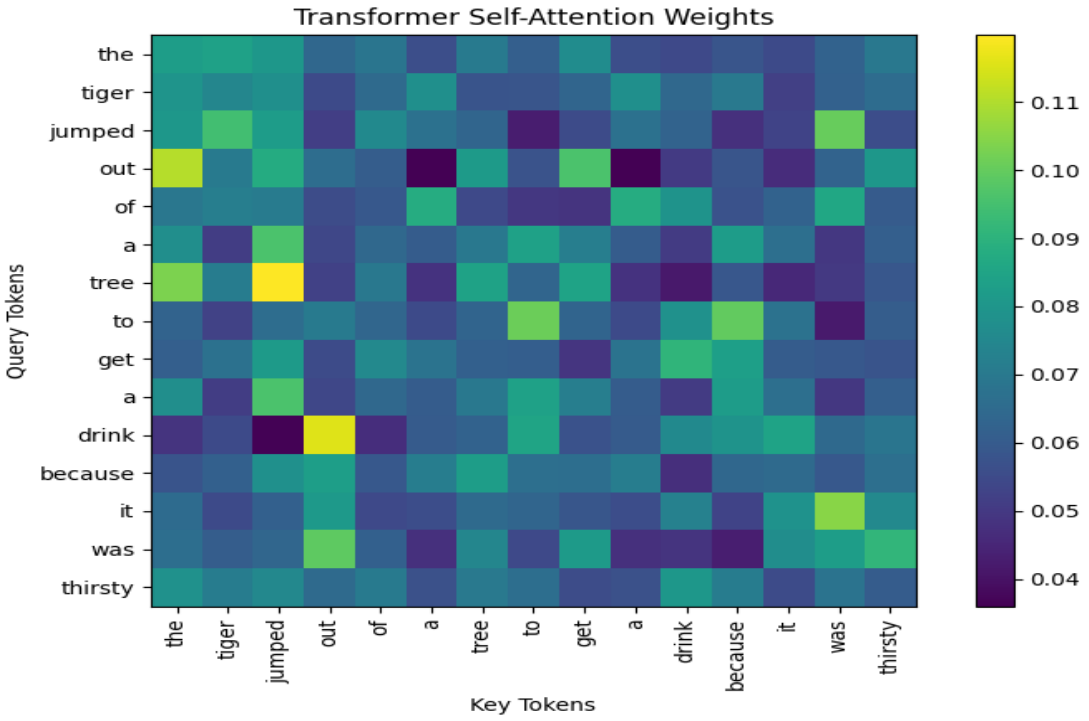

Figure 26. Transformer self-attention weights.

A long-distance text Figures 27 and 28, specifically in the context of process design language, involves processing a large section of text; transformers and RNNs process long sequences differently. The transformer model effectively captures overall weights as relationships among words across the entire text using the linear transformation of the same token and the dot product of their (attention) mechanism. In contrast, the ability of the RNN to model dependencies diminishes over longer distances because of its sequential nature.

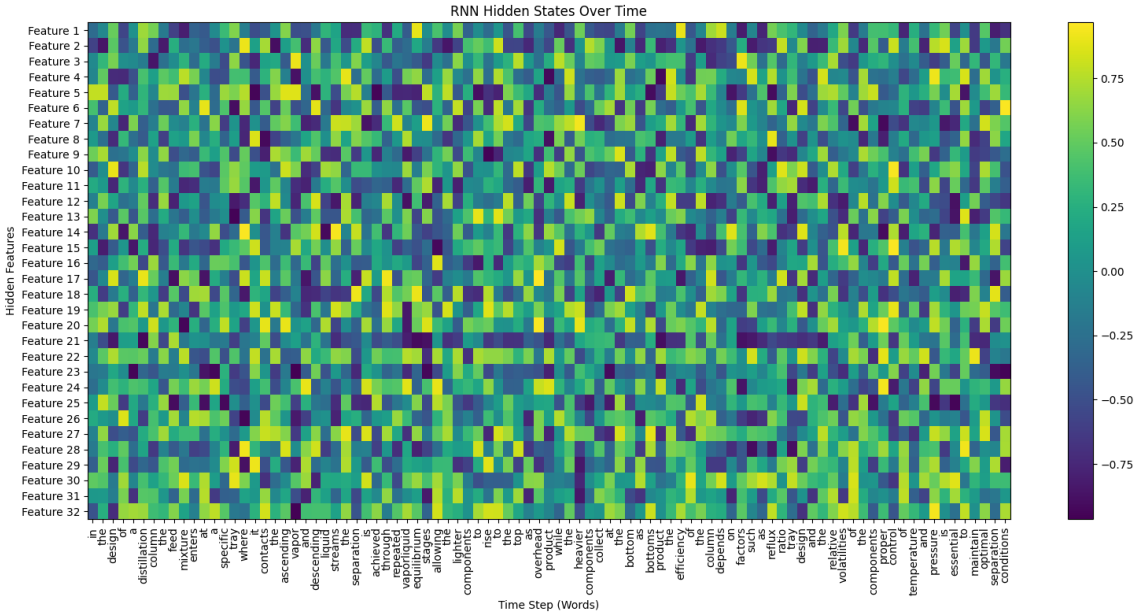

Figure 27. RNN hidden states over different text context lengths.

Therefore, to ensure the high-accuracy systematic extraction and processing of complex data from the literature, graphs, flowcharts, and PFDs using Natural Language Processing and LLMs, integrating techniques such as preprocessing, data labeling, and graph-type classification is critical.

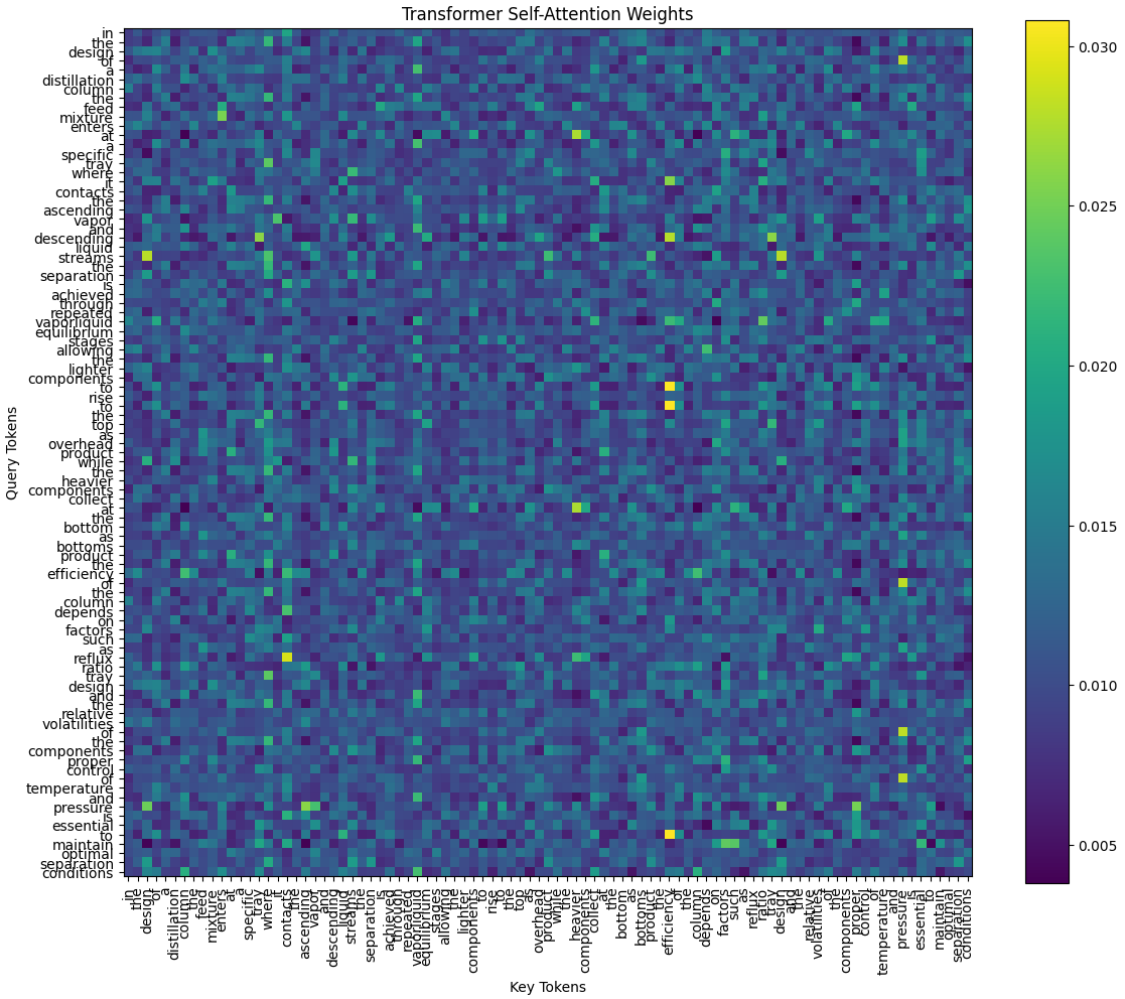

Figure 28. Transformer self-attention weights.

Subsequently, the advantages of applying the above-mentioned analysis, which motivated an in-depth analysis of the applicability of applied LLMs in training large research texts, in the process design arena were investigated. Figure 26 outlines a systematic approach for leveraging LLMs for a language-model-driven design tool coder in chemical engineering tasks.

Table 9. Performance comparison of GPT and the Chen_Model across various problems

Problems heading	**GPT_Performance**	**Chen_Model_Performance**
Unit Converter	Accurate	Accurate
Mixing Material Balance	Accurate	Accurate

Heat Capacity Calculator	Accurate	Accurate
Vapor Pressure Estimator	Accurate	Accurate
Heat Exchanger Design	Accurate	Accurate
Batch Reactor Simulation	Accurate	Accurate
Binary Distillation Column	In-Accurate	Accurate
Pump Power Requirement	In-Accurate	Accurate
Reaction Yield	In-Accurate	Accurate
Multicomponent Distillation	In-Accurate	In-Accurate
Non-Ideal Reactor Model	In-Accurate	In-Accurate
Process with Recycle Loop	In-Accurate	Accurate
Pinch Analysis	In-Accurate	Accurate
PID Control	Accurate	Stable
Fluid Flow Simulation	In-Accurate	In-Accurate

Typically, LLMs face challenges during implementing codes and exhibit inconsistencies due to variations in their reference frameworks compared with the current implementation. Moreover, the solutions proposed by LLMs during error-correction loops are frequently constrained and meet the requirements of local contexts without considering the overall goal. In modular codebases in which different sections of the code interact, an effective correction often requires an understanding of the entire system. Sometimes, the addition of new features to an existing codebase using LLMs is hindered by the same challenges as those associated with error correction. There are instances in which LLMs may enter iterations, wherein they repeatedly attempt to modify or enhance the same features without completing the task.

Chapter 5 – Conclusions and Future Work

5.1 Achievements

This study contributes significantly toward chemical engineering by analyzing the impact of reactor intensification and process expansion on ethylene oxide production, evaluating expansion strategies, identifying optimal desulfurization methods for stranded natural gas resources, and developing tailored language models for process design.

Methodologically innovative, this study involves simulation-based analysis, modularization approaches, and the application of word embeddings, clustering, and proximity graphs.

The key implications of this study include expediting the development of tailored strategies for reconciling economic and emission objectives, designing high-accuracy and relevant reactor intensification tasks, and providing insights into the cost implications of modularized desulfurization for distributed gas sources. Overall, this study advances the understanding of complex tasks in chemical processes and utilizes process modularization, intensification, and language models to minimize emissions and find efficient solutions to process-related problems.

5.2 Future Work

This study lays the groundwork for several future research trajectories, including exploring process modularization and intensification as well as investigating scalable intensification strategies.

This study is also expected to expedite advancements in distributed natural gas desulfurization, which are pivotal and involve the design of modularized desulfurization units and the evaluation of alternative desulfurization methods.

Enhancing LLMs is important and requires the compilation of domain-specific training datasets, specialized evaluation metrics, and the fine-tuning of LLMs for process design applications. Interdisciplinary investigations should focus on integrating reactor intensification and desulfurization as well as leveraging LLMs for process optimization for a better understanding of the relevant processes to drive innovation and promote the use of modularization, intensification, and LLMs for sustainability in process engineering, contributing toward the design and development of efficient and green processes.

Chapter 6 – References

Al-Fadhli, F.M., Baaqeel, H., El-Halwagi, M.M., 2019. Modular design of carbon-hydrogen-oxygen symbiosis networks over a time horizon with limited natural resources. Chem. Eng. Process. Intensif. 141 (2019) 107535.

Allen, R.C., Allaire, D., El-Halwagi, M.M. Capacity planning for modular and transportable infrastructure for shale gas production and processing. Ind. Eng. Chem. Res. 58 (2018) 5887–5897.

Rozovskii, A.Y., Kipnis, M.A., Volnina, E.A., Lin, G.I., Samokhin, P.V. Selective oxidation of CO under conditions of catalyst surface ignition, Kinetics Catal. 48 (2007) 701–710

Ba, J. L. Layer normalization. arXiv preprint (2016) arXiv:1607.06450.

Bahdanau, D., Cho, K., Bengio, Y. Neural machine translation by jointly learning to align and translate. arXiv preprint (2014) arXiv:1409.0473.

L. Baharudin, M.J. Watson, A.C. Yip, Process intensification in multifunctional reactors: A review of multi-functionality by catalytic structures, internals, operating modes, and unit integrations, Chem. Eng. Process. Process. Intensif. 168 (2021) 108561.

Balhorn, L.S., Hirtreiter, E., Luderer, L., Schweidtmann, A.M. Data augmentation for machine learning of chemical process flowsheets, Comput. Aided Chem. Eng. 52 (2023) 2011–2016.

Balhorn, L.S., Caballero, M., Schweidtmann, A.M. Toward autocorrection of chemical process flowsheets using large language models, Comput. Aided Chem. Eng. 53 (2024) 3109–3114.

Bakke, J.M., Buhaug, J.B. Hydrogen sulfide scavenging by 1,3,5-triazinanes. Comparison of the rates of reaction, Ind. Eng. Chem. Res. 43 (2004) 1962–1965.

Baldea, M., Edgar, T.F., Stanley, B.L. Modular manufacturing processes: Status, challenges, and opportunities, 63(10) (2017) 4262–4272.

Bhosekar, A., Badejo, O., Ierapetritou, M. Modular Supply chain optimization considering demand uncertainty to manage risk, AIChE J. 67 (2021) e17367.

Behloul, C. R., Commenge, J. M., & Castel, C. (2021). Influence of the synergy between reaction, heat exchange and membrane separation on the process intensification of the dimethyl ether direct synthesis from carbon dioxide and hydrogen. *Chemical Engineering and Processing-Process Intensification*, *167*, 108513.

Calel, R., Mahdavi, P. The unintended consequences of antiflaring policies—and measures for mitigation, Proc. Natl. Acad. Sci. 117 (2020) 12503–12507.

Chakraborty, S., Lehrer, S., Ramachandran, S. Effective removal of sour gases containing mercaptans in oilfield applications. Paper presented at the SPE Middle East Oil & Gas Show and Conference, Manama, Kingdom of Bahrain, (2017) https://doi.org/10.2118/183974-MS

Chen, Q., Grossmann, I.E., 2019. Economies of numbers for a modular stranded gas processing network: Modeling and optimization, in: Computer Aided Chemical Engineering. Elsevier, pp. 257–262.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.

Dao, T., Fu, D., Ermon, S., Rudra, A., Ré, C. Flashattention: Fast and memory-efficient exact attention with io-awareness. Advances in Neural Information Processing Systems, 35 (2022) 16344–16359.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. (2018)

Echt, B., Leppin, D., Mamrosh, D., Mirdadian, D., Seeger, D., Warren, B.,. Fundamentals of low-tonnage sulfur removal and recovery, in: Laurance Reid Gas Conditioning Conference (2017).

Economides, M.J., Sun, K., Subero, G. Compressed natural gas (CNG): an alternative to liquefied natural gas (LNG), SPE Prod. Oper. 21 (2006) 318–324.

F. Kapteijn, J.A. Moulijn, Structured catalysts and reactors–Perspectives for demanding applications, Catal. Today 383 (2022) 5–14.

Gan, Y., Duan, Q., Gong, W., Tong, C., Sun, Y., Chu, W., Ye, A., Miao, C., Di, Z.. A comprehensive evaluation of various sensitivity analysis methods: A case study with a hydrological model, Environ. Model. Softw. 51 (2014) 269–285.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y. Generative adversarial nets, Adv. Neural Inf. Process Syst. 27 (2014) 2672–2680.

Hochreiter, S. & Schmidhuber, J. Long short-term memory, Neural Comput. 9(8) (1997) 1735–1780.

Huang, X., Wang, D., He, B., Liu, Q., Hu, L., Jiang, G. A 3D-printed modularized purification system for rapid, high-throughput MALDI-MS analysis of small-volume biological samples, Chem. Commun. 56 (2020) 1637–1640.

Judd, B., Eng, P., Fundamentals - gas sweetening, in: Laurance Reid Gas Conditioning Conference (2003).

Kato, S., Kanegami, K., Kano, M. ProcessBERT: A pre-trained language model for judging equivalence of variable definitions in process models, IFAC-PapersOnLine 55(7) (2022) 957–962.

Khalilpour, R., Karimi, I.A. Evaluation of utilization alternatives for stranded natural gas. Energy 40 (2012) 317–328.

Kingma, D. P., & Welling, M. Auto-encoding variational Bayes. arXiv preprint (2013). arXiv:1312.6114.

Kohl, A.L., Nielsen, R., Gas purification. Elsevier. (1997)

Kuenneth, C., & Ramprasad, R. polyBERT: A chemical language model to enable fully machine-driven ultrafast polymer informatics, Nat. Commun. 14 (2023) 4099.

Lara, C.L., Bernal, D.E., Li, C., Grossmann, I.E. Global optimization algorithm for multi-period design and planning of centralized and distributed manufacturing networks, Comput. Chem. Eng. 127 (2019) 295–310.

Luyben, W. L. Snowball effects in reactor/separator processes with recycle, Ind. Eng. Chem. Res. 33(2) (1994) 299–305.

Mann, V. Domain-informed language models for process systems engineering (Master's thesis). https://doi.org/10.7916/kgkh-yj15

Mann, V., Sales-Cruz, M., Gani, R., Venkatasubramanian, V. eSFILES: Intelligent process flowsheet synthesis using process knowledge, symbolic AI, and machine learning, Comp. Chem. Eng. 181 (2024) 108505.

Milakov, M., & Gimelshein, N. (2018). Online normalizer calculation for softmax. arXiv preprint arXiv:1805.02867.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., Dean, J. Distributed representations of words and phrases and their compositionality. Adv. Neural Inf. Process Syst. 26 (2013) 3111–3119.

Mokhatab, S., Mak, J.Y., Valappil, J. V, Wood, D.A., 2013. Handbook Of Liquefied Natural Gas. Gulf Professional Publishing.

Morris, J.R., 2003. TRB Special Report: Regulation of Weights, Lengths, and Widths of Commercial Motor Vehicles. TR News.

Mukta, C.B., Cremaschi, S., Eden, M.R. Operational envelopes of cost-effective sour gas desulfurization processes, Comp. Aided Chem. Eng. 49 (2022) 1009–1014.

Mukta, C.B., Cremaschi, S., Eden, M.R. A novel approach for cost-effective and sustainable capacity expansion utilizing process intensification in chemical process industries, Comp. Aided Chem. Eng. 52 (2023) 1205–1210.

Palys, M.J., Allman, A., Daoutidis, P. Exploring the benefits of modular renewable-powered ammonia production: A supply chain optimization study. Ind. Eng. Chem. Res. 58 (2018) 5898–5908.

Panerati, J., Schnellmann, M. A., Patience, C., Beltrame, G., & Patience, G. S. Experimental methods in chemical engineering: Artificial neural networks–ANNs. The Canadian Journal of Chemical Engineering, 97(9) (2019) 2372–2382.

Peschel, A., Karst, F., Freund, H., & Sundmacher, K. (2011). Analysis and optimal design of an ethylene oxide reactor. *Chemical engineering science*, *66*(24), 6453-6469.

A. Peschel, A. Jörke, K. Sundmacher, H. Freund, Optimal reaction concept and plant wide optimization of the ethylene oxide process, Chem. Eng. J. 207–208 (2012) 656–674.

Schulz, M., 2012. Developments in H_2S scavenging Advances [WWW Document]. URL www.digitalrefining.com/article/1001502

Shao, Y., Zavala, V.M. Modularity measures: Concepts, computation, and applications to manufacturing systems, AIChE J. 66 (2020) e16965.

M. Sheng, C.F. Gonzalez, W.R. Yantz Jr., D.R. Cahela, H. Yang, D.R. Harris, B.J. Tatarchuk, Micro scale heat transfer comparison between packed beds and microfibrous entrapped catalysts, Eng. Appl. Comput. Fluid Mech. 7 (2013) 471–485.

Speight, J.G., 2018. Natural Gas: A Basic Handbook. Gulf Professional Publishing.

Subero, G., Sun, K., Deshpande, A., McLaughlin, J., Economides, M.J., 2004. A comparative study of sea-going natural gas transport, in: SPE Annual Technical Conference and Exhibition. OnePetro.

Subramaniam, R., Yasa, S., Bertrand, T., Fontenot, B., Dupuis, T.F., Hernandez, R. Advanced simulation of H_2S scavenging process with triazine at different depths of gas well, J. Nat. Gas Sci. Eng. 49 (2018) 417–427.

Taylor, M., Dawe, R.A., Thomas, S., 2003. Fire and Ice: Gas hydrate transportation-A possibility for the Caribbean region, in: SPE Latin American and Caribbean Petroleum Engineering Conference. OnePetro.

Thomas, S., Dawe, R.A. Review of ways to transport natural gas energy from countries which do not need the gas for domestic use. Energy 28 (2003) 1461–1477.

Turton, R., Bailie, R. C., Whiting, W. B., & Shaeiwitz, J. A. (2008). Analysis, synthesis and design of chemical processes. Pearson Education.

Vaswani, A. (2017). Attention is all you need. Advances in Neural Information Processing Systems.

Wagner, J. V, van Wagensveld, S., 2002. Marine transportation of compressed natural gas a viable alternative to pipeline or LNG, in: SPE Asia Pacific Oil and Gas Conference and Exhibition. OnePetro.

Wood, D., Mokhatab, S., Economides, M.J., 2008. Technology options for securing markets for remote gas, in: Proceedings of 87th Annual Convention of the Gas Processors Association (GPA).

Wong, K. Y., Ng, J. H., Chong, C. T., Lam, S. S., & Chong, W. T. (2019). Biodiesel process intensification through catalytic enhancement and emerging reactor designs: A critical review. Renewable and Sustainable Energy Reviews, 116, 109399.

Wubs, H.J., Beenackers, A.A.C.M. Kinetics of H_2S absorption into aqueous ferric solutions of EDTA and HEDTA, AIChE J. 40 (1994) 433–444.

Xu, S., Deng, Y., Webb, K., Wright, H., Dimick, P.S., Cremaschi, S., Eden, M.R. Sour gas sweetening technologies for distributed resources–a process simulation study, Comp. Aided Chem. Eng. 48 (2020) 1483–1488.

Yang, M., You, F. Modular methanol manufacturing from shale gas: Techno-economic and environmental analyses of conventional large-scale production versus small-scale distributed, modular processing, AIChE J. 64 (2018) 495–510.

Yong, Y.U., Youzhi, L.I.U., Guisheng, Q.I. Rapid regeneration of chelated iron desulfurization solution using electrochemical reactor with rotating cylindrical electrodes, Chinese J. Chem. Eng. 22 (2014) 136–140.

Zavala-Araiza, D., Allen, D.T., Harrison, M., George, F.C., Jersey, G.R. Allocating methane emissions to natural gas and oil production from shale formations. ACS Sustain. Chem. Eng. 3 (2015) 492–498.

Zhang, K., Mann, V., Venkatasubramanian, V. G-MATT: Single-step retrosynthesis prediction using a molecular grammar tree transformer, AIChE J. 70(1) (2024) e18244.