HYPERSCALE MODELING: MOLECULE, PROCESS, ENTERPRISE

André Bardow[a*], Ignacio Grossmann[b]
[a]ETH Zürich, 8092 Zürich, Switzerland
[b]Carnegie Mellon University, Pittsburgh, PA 15213, USA

Abstract

Process systems engineering relies on modeling tools to screen many alternatives at multiple scales: from the molecular, process and supply chain levels to the global level. Here, a review is presented of the currently available tools to study molecules, processes and enterprises. From this review, we derive the need for their effective integration to support the rapid transition needed for our societies towards sustainability targets.

Keywords

Process Systems Engineering, Computer-Aided Molecular Design, Conceptual process design, Enterprise-wide Optimization

Introduction

The discovery, manufacturing and distribution of chemicals and pharmaceutical products is a complex multiscale process as shown in Figure 1 (Marquardt, et al., 2000) that starts at the molecular level where new chemical or biological products must be discovered or synthesized. Subsequent steps aggregate the molecules into clusters, particles and films as single and multiphase systems that finally take the form of macroscopic mixtures. Through scale-up, the next step is to select process units that must be integrated into a process flowsheet that may operate either in a continuous or batch mode for manufacturing the products. Finally, the process flowsheet becomes part of a production site with several plants that are connected through suppliers, warehouses and distribution centers in a supply chain that ultimately defines the commercial enterprise for the production and distribution of chemical or pharmaceutical products. The recent trend toward sustainability has expanded the scope even further to global impacts (indicated by the globe added to the original Figure 1 by Marquardt et al., 2000).

The resulting need for effective modeling tools that allow the screening of many alternatives at the molecular, process and supply chain level, and ultimately at the globe level, to perform life cycle analysis of processes and products, has become a major research challenge in the area of Process Systems Engineering.

The major goal of this paper is to provide a concise state-of-the-art review of modeling tools at the molecular, process and supply chain level, identify major research challenges and new trends, as well as their multiscale integration to achieve sustainability targets.

Molecular level

Molecules provide the foundation of the chemical industry and thus our understanding of them is crucial for further innovation in the chemical industry. Our ability to understand and quantify molecular phenomena has increased dramatically in recent years using powerful tools provided by quantum mechanics, molecular simulations, fluid theory and statistical thermodynamics (van Speybroeck, et al., 2010).

Traditionally, phase equilibria have been the focus of developing predictive thermodynamic models. Here, models such as the quantum-chemistry-based COSMO-RS/SAC (Klamt 2011) along with the fluid-theory-based SAFT models (Vega and Jackson 2011) have established themselves as new standards for complex fluid mixtures. Key assets of these models are their availability in user-friendly software packages, including open-source implementations, and

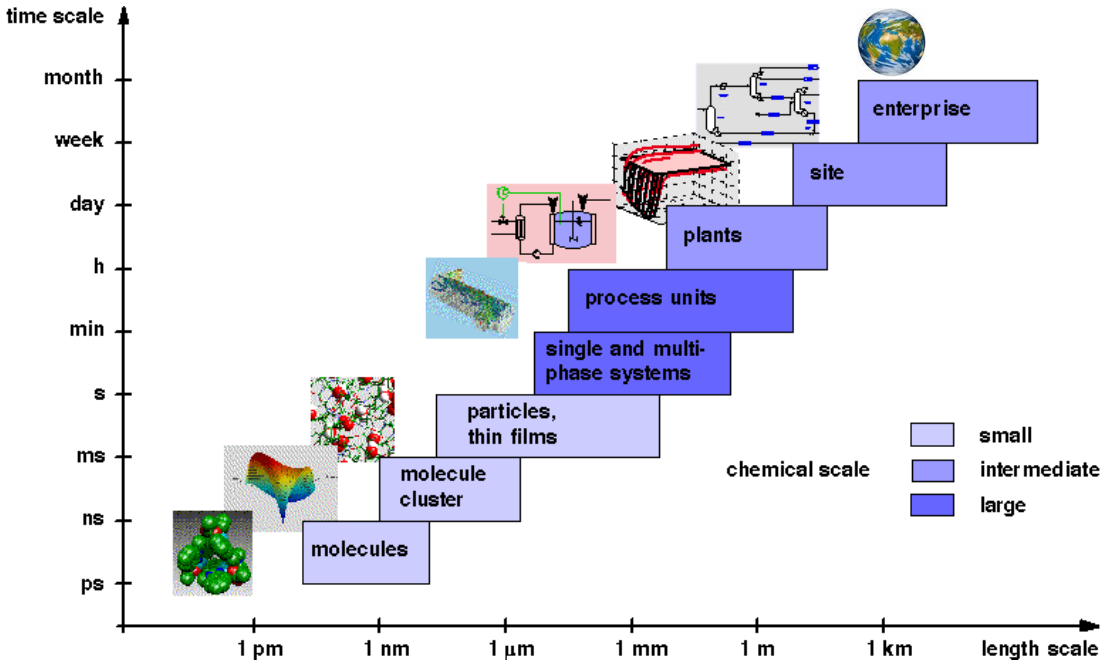

Figure 1: Multiscale modeling at the molecule, process, and enterprise levels, expanded from Marquardt, et al., (2000).

their broad applicability. Thereby, these models have become part of the chemical engineering toolbox.

An important recent development is the extension of SAFT models towards transport properties (Mairhofer and Gross 2022). This extension is based on entropy scaling: Rosenfeld (1977, 1999) found that transport properties are a univariate function of the residual entropy. This finding has been confirmed to hold for broad classes of fluids and enabled the prediction of viscosities (Lötgering-Lin et al. 2018), heat conductivities (Hopp and Gross 2017) and diffusion coefficients (Hopp et al. 2018). Thereby, both equilibrium and transport properties can now be predicted efficiently based on a single parsimoniously parameterized molecular model.

The physics-based models are complemented by data-driven approaches. While group-contribution methods and quantitative-structure property relationships (QSPRs) have a long tradition in chemical engineering (van Speybroeck et al. 2010), recent advances in machine learning have already led to new approaches towards predicting thermodynamic properties. Pioneering work by the Jirasek and Hasse groups recognized the similarity between recommender systems for movies and solvents by transferring matrix-completion methods to predict activity coefficients at infinite dilution (Jirasek et al. 2020) and Henry coefficients (Hayer et al. 2022), outperforming long-established methods like UNIFAC. Based on these encouraging results, more advanced architectures have been employed such as Graph Neural Networks (GNNs) and transformer-based models that are able to predict properties for molecules not included in the training set (Sanchez Medina et al. 2022; Qin et al. 2022; Rittig et al. 2022). These approaches achieve higher accuracies than established methods such as COSMO-RS and UNIFAC-Dortmund.

However, data-driven approaches suffer from the lack of data, in particular for multicomponent mixtures, and for transport properties. Here, the integration of experiments with modeling in coordinated learning workflows could enable major progress (Pankajakshan et al. 2019). Furthermore, hybrid modeling approaches seem promising that refine insights from physics-based models by data-driven modeling. Such an approach has recently been demonstrated by (Winter et al. 2022a; Winter et al. 2022b) for the prediction of thermodynamically consistent activity coefficients from SMILES codes. A transformer model was used which is known to require large sets of training data. To overcome the lack of data availability, synthetic data was generated using COSMO-RS. The resulting model was then refined using the limited experimental data available. This combined approach allowed to significantly improve prediction quality. Even more, integrating the NRTL equation into the machine learning framework enables thermodynamically consistent predictions. This hybrid paradigm benefit from both available knowledge and data. Overall, these improvements in thermodynamic property prediction can lead to more efficient process and product design, reducing cost and energy demands.

However, the ambitions towards sustainability require a broader set of performance measures capturing also economic, environmental and societal impact (Kleinekorte et al. 2020). While methods for social impact assessment are still in early stages of development, environmental impact assessment has been standardized using the method of life-cycle assessment (LCA) (DIN EN ISO 14040 2006). Linking molecules to their environmental impacts over the life cycle would enable the design of more sustainable processes. However, the scope of LCA to cover the full life-cycle makes it data-intensive, leading to a lack of high-

quality data sets. Thus, the practically unlimited chemical design space is only poorly represented in current LCA databases. To overcome these limitations, predictive methods have been developed (Baxevanidis et al. 2022). Physics-based models usually focus on limited parts of the life-cycle such as estimating gate-to-gate energy demands of processes (Parvatker and Eckelman 2019). In contrast, data-driven approaches try to predict cradle-to-gate impacts of chemicals directly (Wernet et al. 2008). Due to the limited training data, the accuracy is also limited. However, these predictive LCA models can also be combined with physical models in hybrid approaches to guide molecular design, e.g., in solvent selection (Fleitmann et al. 2021).

Process level

Conceptual process design is a central pillar of chemical engineering, concerning the definition, simulation, optimization, and control of chemical processes. This design task involves the synthesis of complex chemical processes through the integration of simpler unit blocks characterized by physical and chemical properties. There are two main approaches for conceptual process design: hierarchical decomposition (Douglas 1985; Siirola et al. 1971) and superstructure synthesis (Umeda et al. 1972; Chen and Grossmann 2017) through the application of mathematical programming. Superstructure synthesis is preferred for its systematic evaluation of a large space of structural alternatives (Mencarelli et al. 2020). However, the development of tools for superstructure optimization has proved to be a major challenge due to the difficulty in solving nonlinear discrete and continuous optimization problems, specifically MINLP models. To overcome these difficulties, Generalized Disjunctive Programming (GDP) (Lee and Grossmann 2000) has emerged as a modeling framework to explicitly represent the relationship between algebraic descriptions and the logical structure of a design problem, which in turn increase the robustness of the discrete/continuous optimization of superstructures.

The general form for GDP optimization models is given in Problem (GDP) (Grossmann and Trespalacios 2013).

$$
\begin{aligned}
&\min \quad z = f(x) \\
&s.t. \quad\; g(x) \le 0 \\
&\veebar_{i \in D_k} \begin{bmatrix} Y_{ki} \\ r_{ki}(x) \le 0 \end{bmatrix} && k \in K && \text{(GDP)} \\
&\veebar_{i \in D_k} Y_{ki} && k \in K \\
&\Omega(Y) = True \\
&x^{lo} \le x \le x^{up} \\
&x \in \mathbb{R}^n \\
&Y_{ki} \in \{True, False\} && k \in K,\ i \in D_k
\end{aligned}
$$

(GDP) seeks to minimize a function of the continuous variables x. $g(x)$ are the global constraints that need to be satisfied independently of the discrete decisions. (GDP) contains $k \in K$ disjunctions. Each one involves $i \in D_k$ terms, linked together by an OR operator ($\veebar$). Each disjunctive term is associated with a Boolean variable Y_{ki} and a set of inequalities $r_{ki}(x) \le 0$. One and only one Boolean variable can be $True$ ($\veebar_{i \in D_k} Y_{ki}$). When a disjunctive term is selected ($Y_{ki} = True$), the corresponding inequalities $r_{ki}(x) \le 0$ are enforced. Otherwise, the constraints are ignored. The logic relations among the Boolean variables are represented by $\Omega(Y) = True$. If $f(x)$, $g(x)$ and $r_{ki}(x)$ are convex, the model becomes a convex GDP.

There is a recent tool, Pyomo.GDP (Chen et al. 2022), that allows the formulation of models in the form of GDP, and implements several solution strategies: reformulation to MINLP (big-M of hull reformulation), logic-based outer-approximation, disjunctive branch and bound. The direct formulation of process synthesis problems through Pyomo.GDP requires a level of expertise from someone who has good knowledge of mathematical programming. This requirement has motivated the development of Pyosyn (Chen et al. 2021), a framework for superstructure-based synthesis. It makes use of the Pyosyn Graph (PSG) superstructure representation, which uniquely supports nested units (Figure 2). Nested units, together with formal interfaces defined by unit ports, improve complexity handling in the PSG and modularize modeling of larger superstructures. With PSG, Pyosyn directly models superstructure logic using Generalized Disjunctive Programming (GDP). Pyosyn incorporates logic-based decomposition algorithms tailored to process synthesis applications, including a set-covering algorithm for logic-based outer approximation, and improved variable range reduction based on disjunctive variable bounds. The use of Pyosyn has been demonstrated with a diverse range of industrially-relevant case studies that include multi-period modular facility location, synthesis of methane to syngas and methanol processes, and Kaibel column design. Pyosyn also supports optimization models for Process Intensification (PI), an area that has received recently significant attention as it helps to support the goal of sustainability in chemical processes.

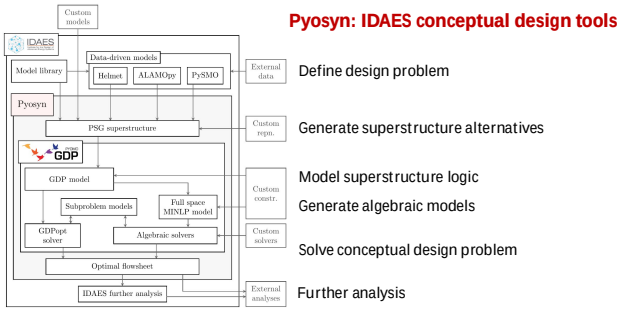

Figure 2: Architecture of Pyosyn

Supply chain level

Overview. Enterprise-wide Optimization (EWO) is concerned with the coordinated optimization of the operations of a supply chain (Shapiro 2001), namely R&D, supply of material, manufacturing and distribution of products. Process supply chains range from the ones in petroleum industry (Shah et al. 2011) to the ones in the pharmaceutical industry (Shah 2004), and include manufacturing as a major

component (Ryu and Pistikopoulos 2007). The main objectives in EWO include maximization of profits, responsiveness to customers and asset utilization, and minimization of costs, inventory levels and environmental impact. Major operational activities include planning, scheduling, real-time optimization and control (see Figure 3).

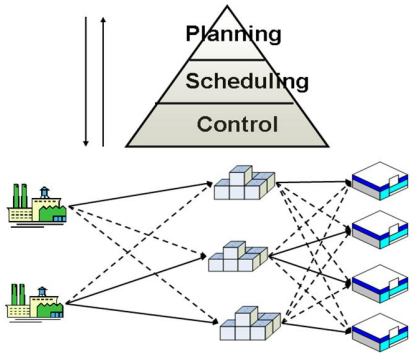

Figure 3: Major elements of Enterprise

One of the key features in EWO is integration of the information and decision making among the various functions that comprise the supply chain of the company. Integration of information is being achieved with modern IT tools (e.g. SAP and Oracle) that allow the sharing and instantaneous flow of information along the various organizations in a company. While IT tools allow many groups in an enterprise to access the same information, these tools do not provide comprehensive decision-making capabilities for optimization that account for complex trade-offs and interactions across the various functions, subsystems and levels of decision making. This means that companies are faced with the problem of deciding as to whether to develop their own in-house tools for integration, or else make use of commercial software from vendors.

To realize the full potential of transactional IT tools, the development of sophisticated decision-support tools based on mathematical programming (analytical IT tools) is needed to operate the supply chain to yield overall optimum economic performance, as well as high levels of customer satisfaction. A major challenge in EWO of process industries is the integrated and coordinated decision-making across the various functions in a company (purchasing, manufacturing, distribution, sales), across various geographically distributed organizations (vendors, facilities and markets), and across various levels of decision-making (strategic, tactical and operational) as seen in Figure 4 (Shapiro 2001).

Optimization Models. A large number of optimization problems in EWO can be described by Mixed-Integer Linear Programming (MILP) models. Examples include the optimization of production operations including planning and scheduling (Méndez et al. 2006; Pochet and Wolsey 2006), optimization of supply chains involving logistics and distribution, multiple period optimization (Grossmann 2005).

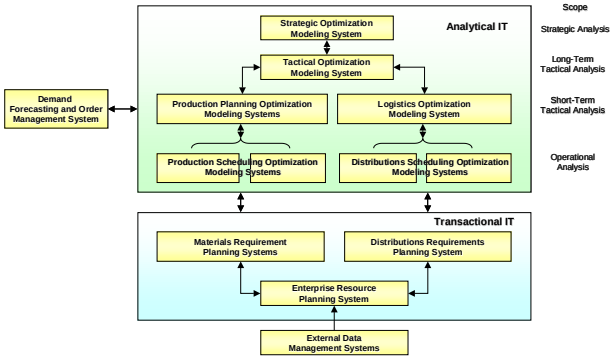

Figure 4: Transactional and Analytical IT (Tayur et al., 1999).

In the process industries, real-world problems usually lead to large-scale models, due to the size of the system under study, but also because of models that involve multiple periods. Furthermore, often new variables and equations are introduced to replace nonlinearities by piecewise linear approximations, or by performing exact linearizations (eg. product binary and continuous variables).

Generally, MILP problems can be solved using Linear Programming (LP)-based Branch & Bound (B&B) solvers (Wolsey 1998) that provide rigorous lower and upper bounds on the solution, which in turn provide information regarding the optimality of the solution. MILP solvers like CPLEX and Gurobi have implemented more sophisticated versions denoted by Branch & Cut (B&C) algorithms. These algorithms add valid inequalities, denoted by cutting planes, to the linear relaxations to reduce the size of the feasible space without eliminating any feasible integer solution. In the last 15 years, great progress has been made in algorithms and hardware, resulting in an impressive improvement of our ability to solve MILPs (Bixby and Rothberg 2007) through codes such as CPLEX, XPRESS and GUROBI. Capitalizing on theory developed during the last 20 years, off-the-shelf LP-based branch-and-bound commercial software can now solve in a few seconds MILP instances that were unsolvable 15 years ago. The current computational performance of MILP software is the result of a combination of improvements in LP solvers, pre-processing techniques for LP and MILP, linear algebra for sparse systems, cutting planes, heuristics, parallelization and faster computers (Bixby and Rothberg 2007).

Planning under uncertainty is an important problem that arises in EWO problems. Modeling and solution of this class of problems by stochastic programming directly takes uncertainty into account in terms of probability distribution functions (Birge and Louveaux 1997; Sahinidis 2004). A Stochastic Program is a mathematical program in which some of the parameters defining a problem instance are random (e.g. demand, yield). The basic idea behind stochastic programming is to make some decisions now, stage 1, and to take some corrective action (recourse) in the future, stage 2, after revelation of the uncertainty. If there are only two stages then the problem corresponds to a 2-stage stochastic program, while in a multistage stochastic program the uncertainty is revealed sequentially, i.e. in multiple stages

(periods), and the decision-maker can take corrective action over a sequence of stages. The uncertain parameters are commonly assumed to follow discrete probability distributions and a planning horizon consisting of a fixed number of decision points. Using these two assumptions, the stochastic process can be represented with scenario trees.

Given the large size of many EWO problems, the application of decomposition techniques is often required. Three common approaches in EWO problems are Lagrangean decomposition, Benders decomposition, bi-level decomposition and rolling horizon. Lagrangean Decomposition is perhaps the most common technique used for decomposing problems so as to make them tractable. An excellent review on Lagrangean Relaxation and Lagrangean Decomposition can be found in Guignard (2003). As described in the previous sections, the most common form of EWO models take the form of deterministic or stochastic MILP or MINLP models. Currently, it is possible to implement these models in commercial modeling systems since the models are expressed in equation form. Therefore, many planning, scheduling and supply chain models have been implemented in commercial modeling systems such as AIMMS, AMPL, GAMS and OPL. Furthermore, the effort involved is relatively modest. Finally, the modeling systems have capabilities of interfacing with spreadsheet, databases or graphics packages, thereby facilitating the deployment of the model as a tool with graphical user interface that can be used by non-specialist users.

Three major outstanding challenges remain to be addressed and are worth of future research efforts. In particular: 1) optimizing entire supply chains due to the very large size of the resulting model. 2) integrating control with planning and scheduling where a major barrier is the complexity and large size of the resulting MIDO optimization problem, which will require effective solution methods; 3) incorporating considerations of sustainability, which is becoming a major concern through multi-objective optimization of entire supply chain networks (Galán-Martín et al. 2018).

Perspective: Hyperscale modeling for sustainability

The previous sections discussed recent advances and challenges on the levels of molecules, processes and supply chains. Aiming for sustainability, however, requires the multiscale integration across these scales. Life-cycle thinking provides such a holistic perspective (Kleinekorte et al. 2020). Thus, our modeling and optimization tools need to be extended to cover and integrate these scales.

Integration approaches have already been successful covering the neighboring levels molecules and processes as well as process and supply chains. The simultaneous consideration of molecules and processes is the topic of Computer-Aided Molecular and Process Design (CAMPD) (Gertig et al. 2020). The CAMPD problem has for long been conceptualized (Gani 2004). However, only recently, successful solution approaches have been presented. Applications range from working fluids (Schilling et al. 2017) to solvents (Scheffczyk et al. 2018) and adsorbents (Farmahini et al. 2021). The key success factor has been the integration of powerful thermodynamic models with advanced solution algorithms (Gertig et al. 2020). Integration of predictive LCA models then allowed to design molecules that minimize environmental impacts of the full life cycle (Fleitmann et al. 2021).

The further expansion of the scope towards supply chains seems highly desirable and needed (Koberg and Longoni 2019). In particular, the required switch to renewable feedstocks and energy supply mandates the establishment of novel supply chains. Climate-mitigation targets require the transition to be rapid, reaching CO_2-neutrality mid-century. In achieving such a rapid transition to stop harming the climate, we need to avoid burden shifting towards other environmental impacts, e.g. due to increase mining activities. Thus, systems tools are needed that enable to rapidly develop environmentally benign industries. Among them is carbon capture and storage, which involves challenges at multiple scales (Bui et al. 2018).

Herein, a prominent example is our need to scale up direct air capture to provide negative emissions on the gigaton-scale (Gasser et al. 2015). This scale-up needs to be aligned with the transformation of the energy sector (Qiu et al. 2022) and the establishment of the materials supply chain, e.g., to produce the large amounts of adsorbents needed (Deutz and Bardow 2021).

The integration of the scales traditionally covered by PSE into a joint perspective could guide the way towards the fast track to a sustainable future.

Acknowledgments

AB acknowledges funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101036854.

Publication bibliography

Baxevanidis, Pantelis; Papadokonstantakis, Stavros; Kokossis, Antonis; Marcoulaki, Effie (2022): Group contribution-based LCA models to enable screening for environmentally benign novel chemicals in CAMD applications. In *AIChE Journal* 68 (3), e17544. DOI: 10.1002/aic.17544.

Birge, John R.; Louveaux, François (1997): Introduction to stochastic programming. New York, London: Springer (Springer series in operations research).

Bixby, Robert; Rothberg, Edward (2007): Progress in computational mixed integer programming—A look back from the other side of the tipping point. In *Ann Oper Res* 149 (1), pp. 37–41. DOI: 10.1007/s10479-006-0091-y.

Bui, Mai; Adjiman, Claire S.; Bardow, André; Anthony, Edward J.; Boston, Andy; Brown, Solomon et al. (2018): Carbon capture and storage (CCS): the way forward. In *Energy Environ. Sci.* 11 (5), pp. 1062–1176. DOI: 10.1039/C7EE02342A.

Chen, Qi; Grossmann, I. E. (2017): Recent Developments and Challenges in Optimization-Based Process Synthesis. In *Annual review of chemical and biomolecular engineering* 8, pp. 249–283. DOI: 10.1146/annurev-chembioeng-080615-033546.

Chen, Qi; Johnson, Emma S.; Bernal, David E.; Valentin, Romeo; Kale, Sunjeev; Bates, Johnny et al. (2022): Pyomo.GDP: an ecosystem for logic based modeling and optimization development. In *Optim Eng* 23 (1), pp. 607–642. DOI: 10.1007/s11081-021-09601-7.

Chen, Qi; Liu, Yunshan; Seastream, Grant; Siirola, John D.; Grossmann, Ignacio E. (2021): Pyosyn: A new framework for conceptual design modeling and optimization. In *Computers & Chemical Engineering* 153, p. 107414. DOI: 10.1016/j.compchemeng.2021.107414.

Deutz, Sarah; Bardow, André (2021): Life-cycle assessment of an industrial direct air capture process based on temperature–vacuum swing adsorption. In *Nat Energy* 6 (2), pp. 203–213. DOI: 10.1038/s41560-020-00771-9.

DIN EN ISO 14040 (2006): Environmental management - Life cycle assessment - Principles and framework (14040:2019).

Douglas, J. M. (1985): A hierarchical decision procedure for process synthesis. In *AIChE J.* 31 (3), pp. 353–362. DOI: 10.1002/aic.690310302.

Farmahini, Amir H.; Krishnamurthy, Shreenath; Friedrich, Daniel; Brandani, Stefano; Sarkisov, Lev (2021): Performance-Based Screening of Porous Materials for Carbon Capture. In *Chemical reviews* 121 (17), pp. 10666–10741. DOI: 10.1021/acs.chemrev.0c01266.

Fleitmann, Lorenz; Kleinekorte, Johanna; Leonhard, Kai; Bardow, André (2021): COSMO-susCAMPD: Sustainable solvents from combining computer-aided molecular and process design with predictive life cycle assessment. In *Chemical Engineering Science* 245, p. 116863. DOI: 10.1016/j.ces.2021.116863.

Galán-Martín, A.; Pozo, C.; Azapagic, A.; Grossmann, I. E.; Mac Dowell, N.; Guillén-Gosálbez, G. (2018): Time for global action: an optimised cooperative approach towards effective climate change mitigation. In *Energy Environ. Sci.* 11 (3), pp. 572–581. DOI: 10.1039/C7EE02278F.

Gani, Rafiqul (2004): Chemical product design: challenges and opportunities. In *Computers & Chemical Engineering* 28 (12), pp. 2441–2457. DOI: 10.1016/j.compchemeng.2004.08.010.

Gasser, T.; Guivarch, C.; Tachiiri, K.; Jones, C. D.; Ciais, P. (2015): Negative emissions physically needed to keep global warming below 2 °C. In *Nature communications* 6, p. 7958. DOI: 10.1038/ncomms8958.

Gertig, Christoph; Leonhard, Kai; Bardow, André (2020): Computer-aided molecular and processes design based on quantum chemistry: current status and future prospects. In *Current Opinion in Chemical Engineering* 27, pp. 89–97. DOI: 10.1016/j.coche.2019.11.007.

Grossmann, Ignacio (2005): Enterprise-wide optimization: A new frontier in process systems engineering. In *AIChE J.* 51 (7), pp. 1846–1857. DOI: 10.1002/aic.10617.

Grossmann, Ignacio E.; Trespalacios, Francisco (2013): Systematic modeling of discrete-continuous optimization models through generalized disjunctive programming. In *AIChE J.* 59 (9), pp. 3276–3295. DOI: 10.1002/aic.14088.

Guignard, Monique (2003): Lagrangean relaxation. In *Top* 11 (2), pp. 151–200. DOI: 10.1007/BF02579036.

Hayer, Nicolas; Jirasek, Fabian; Hasse, Hans (2022): Prediction of Henry's law constants by matrix completion. In *AIChE Journal* 68 (9), e17753. DOI: 10.1002/aic.17753.

Hopp, Madlen; Gross, Joachim (2017): Thermal Conductivity of Real Substances from Excess Entropy Scaling Using PCP-SAFT. In *Ind. Eng. Chem. Res.* 56 (15), pp. 4527–4538. DOI: 10.1021/acs.iecr.6b04289.

Hopp, Madlen; Mele, Julia; Gross, Joachim (2018): Self-Diffusion Coefficients from Entropy Scaling Using the PCP-SAFT Equation of State. In *Ind. Eng. Chem. Res.* 57 (38), pp. 12942–12950. DOI: 10.1021/acs.iecr.8b02406.

Jirasek, Fabian; Alves, Rodrigo A. S.; Damay, Julie; Vandermeulen, Robert A.; Bamler, Robert; Bortz, Michael et al. (2020): Machine Learning in Thermodynamics: Prediction of Activity Coefficients by Matrix Completion. In *The journal of physical chemistry letters* 11 (3), pp. 981–985. DOI: 10.1021/acs.jpclett.9b03657.

Klamt, Andreas (2011): The COSMO and COSMO-RS solvation models. In *WIREs Comput Mol Sci* 1 (5), pp. 699–709. DOI: 10.1002/wcms.56.

Kleinekorte, Johanna; Fleitmann, Lorenz; Bachmann, Marvin; Kätelhön, Arne; Barbosa-Póvoa, Ana; Assen, Niklas von der; Bardow, André (2020): Life Cycle Assessment for the Design of Chemical Processes, Products, and Supply Chains. In *Annual review of chemical and biomolecular engineering* 11, pp. 203–233. DOI: 10.1146/annurev-chembioeng-011520-075844.

Koberg, Esteban; Longoni, Annachiara (2019): A systematic review of sustainable supply chain management in global supply chains. In *Journal of Cleaner Production* 207, pp. 1084–1098. DOI: 10.1016/j.jclepro.2018.10.033.

Lee, Sangbum; Grossmann, Ignacio E. (2000): New algorithms for nonlinear generalized disjunctive programming. In *Computers & Chemical Engineering* 24 (9-10), pp. 2125–2141. DOI: 10.1016/S0098-1354(00)00581-0.

Lötgering-Lin, Oliver; Fischer, Matthias; Hopp, Madlen; Gross, Joachim (2018): Pure Substance and Mixture Viscosities Based on Entropy Scaling and an Analytic Equation of State. In *Ind. Eng. Chem. Res.* 57 (11), pp. 4095–4114. DOI: 10.1021/acs.iecr.7b04871.

Mairhofer, Jonas; Gross, Joachim (2022): Prediction and correlation of physical properties including transport and interfacial properties with the PC-SAFT equation of state. In Michael Bortz, Norbert Asprion (Eds.): Simulation and Optimization in Process Engineering: Elsevier, pp. 1–31. Available online at https://www.sciencedirect.com/science/article/pii/B9780323850438000027.

Mencarelli, Luca; Chen, Qi; Pagot, Alexandre; Grossmann, Ignacio E. (2020): A review on superstructure optimization approaches in process system engineering. In *Computers & Chemical Engineering* 136, p. 106808. DOI: 10.1016/j.compchemeng.2020.106808.

Méndez, Carlos A.; Grossmann, Ignacio E.; Harjunkoski, Iiro; Kaboré, Pousga (2006): A simultaneous optimization approach for off-line blending and scheduling of oil-refinery operations. In *Computers & Chemical Engineering* 30 (4), pp. 614–634. DOI: 10.1016/j.compchemeng.2005.11.004.

Pankajakshan, Arun; Waldron, Conor; Quaglio, Marco; Gavriilidis, Asterios; Galvanin, Federico (2019): A Multi-Objective Optimal Experimental Design Framework for Enhancing the Efficiency of Online Model Identification Platforms. In *Engineering* 5 (6), pp. 1049–1059. DOI: 10.1016/j.eng.2019.10.003.

Parvatker, Abhijeet G.; Eckelman, Matthew J. (2019): Comparative Evaluation of Chemical Life Cycle Inventory Generation Methods and Implications for Life Cycle Assessment Results. In *ACS Sustainable Chem. Eng.* 7 (1), pp. 350–367. DOI: 10.1021/acssuschemeng.8b03656.

Pochet, Yves; Wolsey, Laurence A. (2006): Production planning by mixed integer programming. New York, Berlin: Springer (Springer series in operations research and financial engineering).

Qin, Shiyi; Jiang, Shengli; Li, Jianping; Balaprakash, Prasanna; van Lehn, Reid; Zavala, Victor (2022): Capturing Molecular Interactions in Graph Neural Networks: A Case Study in Multi-Component Phase Equilibrium. Available online at https://doi.org/10.26434/chemrxiv-2022-3tq4c-v2.

Qiu, Yang; Lamers, Patrick; Daioglou, Vassilis; McQueen, Noah; Boer, Harmen-Sytze de; Harmsen, Mathijs et al. (2022): Environmental trade-offs of direct air capture technologies in climate change mitigation toward 2100. In *Nature communications* 13 (1), p. 3635. DOI: 10.1038/s41467-022-31146-1.

Rittig, Jan G.; Hicham, Karim Ben; Schweidtmann, Artur M.; Dahmen, Manuel; Mitsos, Alexander (2022): Graph Neural Networks for Temperature-Dependent Activity Coefficient Prediction of Solutes in Ionic Liquids. Available online at http://arxiv.org/pdf/2206.11776v1.

Rosenfeld, Yaakov (1977): Relation between the transport coefficients and the internal entropy of simple systems. In *Phys. Rev. A* 15 (6), pp. 2545–2549. DOI: 10.1103/PhysRevA.15.2545.

Rosenfeld, Yaakov (1999): A quasi-universal scaling law for atomic transport in simple fluids. In *Journal of Physics: Condensed Matter* 11 (28), pp. 5415–5427. DOI: 10.1088/0953-8984/11/28/303.

Ryu, Jun-hyung; Pistikopoulos, Efstratios N. (2007): Multiperiod Planning of Enterprise-wide Supply Chains Using an Operation Policy. In *Ind. Eng. Chem. Res.* 46 (24), pp. 8058–8065. DOI: 10.1021/ie070508b.

Sahinidis, Nikolaos V. (2004): Optimization under uncertainty: state-of-the-art and opportunities. In *Computers & Chemical Engineering* 28 (6-7), pp. 971–983. DOI: 10.1016/j.compchemeng.2003.09.017.

Sanchez Medina, Edgar Ivan; Linke, Steffen; Stoll, Martin; Sundmacher, Kai (2022): Graph neural networks for the prediction of infinite dilution activity coefficients. In *Digital Discovery* 1 (3), pp. 216–225. DOI: 10.1039/D1DD00037C.

Scheffczyk, J.; Schäfer, P.; Fleitmann, L.; Thien, J.; Redepenning, C.; Leonhard, K. et al. (2018): COSMO-CAMPD: a framework for integrated design of molecules and processes based on COSMO-RS. In *Mol. Syst. Des. Eng.* 3 (4), pp. 645–657. DOI: 10.1039/C7ME00125H.

Schilling, Johannes; Lampe, Matthias; Gross, Joachim; Bardow, André (2017): 1-stage CoMT-CAMD: An approach for integrated design of ORC process and working fluid using PC-SAFT. In *Chemical Engineering Science* 159, pp. 217–230. DOI: 10.1016/j.ces.2016.04.048.

Shah, Nikisha K.; Li, Zukui; Ierapetritou, Marianthi G. (2011): Petroleum Refining Operations: Key Issues, Advances, and Opportunities. In *Ind. Eng. Chem. Res.* 50 (3), pp. 1161–1170. DOI: 10.1021/ie1010004.

Shah, Nilay (2004): Pharmaceutical supply chains: key issues and strategies for optimisation. In *Computers & Chemical Engineering* 28 (6-7), pp. 929–941. DOI: 10.1016/j.compchemeng.2003.09.022.

Shapiro, Jeremy F. (2001): Modeling and IT Perspectives on Supply Chain Integration. In *Information Systems Frontiers* 3 (4), pp. 455–464. DOI: 10.1023/A:1012876804965.

Siirola, J. J.; Powers, G. J.; Rudd, D. F. (1971): Synthesis of system designs: III. Toward a process concept generator. In *AIChE J.* 17 (3), pp. 677–682. DOI: 10.1002/aic.690170334.

Umeda, Tomio; Hirai, Akira; Ichikawa, Atsunobu (1972): Synthesis of optimal processing system by an integrated approach. In *Chemical Engineering Science* 27 (4), pp. 795–804. DOI: 10.1016/0009-2509(72)85013-9.

van Speybroeck, Veronique; Gani, Rafiqul; Meier, Robert Johan (2010): The calculation of thermodynamic properties of molecules. In *Chemical Society reviews* 39 (5), pp. 1764–1779. DOI: 10.1039/b809850f.

Vega, Lourdes F.; Jackson, George (2011): 20Years of the SAFT equation of state—Recent advances and challenges. In *Fluid Phase Equilibria* 306 (1), pp. 1–3. DOI: 10.1016/j.fluid.2011.04.020.

Wernet, Gregor; Hellweg, Stefanie; Fischer, Ulrich; Papadokonstantakis, Stavros; Hungerbühler, Konrad (2008): Molecular-structure-based models of chemical inventories using neural networks. In *Environmental science & technology* 42 (17), pp. 6717–6722. DOI: 10.1021/es7022362.

Winter, Benedikt; Winter, Clemens; Esper, Timm; Schilling, Johannes; Bardow, André (2022a): SPT-NRTL: A physics-guided machine learning model to predict thermodynamically consistent activity coefficients. Edited by arXiv. Available online at http://arxiv.org/pdf/2209.04135v1.

Winter, Benedikt; Winter, Clemens; Schilling, Johannes; Bardow, André (2022b): A smile is all you need: Predicting limiting activity coefficients from SMILES with natural language processing. Edited by arXiv. Available online at http://arxiv.org/pdf/2206.07048v1.

Wolsey, Laurence A. (1998): Integer programming. New York, Chichester: John Wiley & Sons (Wiley-Interscience series in discrete mathematics and optimization).