Antonis Kokossis, Michael C. Georgiadis, Efstratios N. Pistikopoulos (Eds.)
PROCEEDINGS OF THE 33rd European Symposium on Computer Aided Process Engineering (ESCAPE33), June 18-21, 2023, Athens, Greece
 http://dx.doi.org/10.1016/B978-0-443-15274-0.50036-6

Intelligent Process Flowsheet Synthesis and Design using Extended SFILES Representation

Vipul Mann,[a] Rafiqul Gani,[b,c] Venkat Venkatasubramanian,[a]

[a]*Columbia University, New York, NY 10027, USA*
[b]*PSE for SPEED Company, Charlottenlund, DK-2920, Denmark*
[c]*Sustainable Energy and Environment Thrust, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China*

Abstract

A central problem in chemical engineering (and several related areas) is evaluating the correct sequence of unit operations, their design aspects, and the continuous optimization of their operations to efficiently convert input materials to final products. The numerous decisions to be made at each problem-solving stage renders this problem combinatorically complex. In this work, we propose a hybrid, artificial intelligence-based multi-level framework to perform fast, efficient, and reliable flowsheet design and optimization. We build upon the previously proposed SFILES-based text representation of flowsheets to incorporate additional contextual details in the extended SFILES framework using hypergraph representations. We discuss our eSFILES framework using the well-known hyrdodealkylation (HDA) process.

Keywords: process design; flowsheet prediction; artificial intelligence; computer-aided flowsheet synthesis

1. Introduction

Process flowsheet synthesis and design is a challenging task that involves identifying the correct set of unit operations and their optimal sequence that enables the conversion of input materials to desired output products, while considering energy consumption, environmental impact, safety, operability, and many more. The goal of synthesis is to identify the tasks, the operations that will perform these tasks, and their sequence to determine the flowsheet. The goal of design is to add operational and equipment details to the flowsheet so that the process can be verified, optimized, built, and operated. The size of the synthesis problem varies with the number of tasks and the number of alternative operations for each task. The complexity of the design problem is related to matching the tasks and their associated operations with design parameters so that the designed process matches the desired process specifications. The design parameters are further optimized to obtain a sustainable process design.

Model-based methods are usually used for the synthesis and design stages. Although, in principle, the synthesis, design and optimization steps could be performed simultaneously, the current practice is to solve them separately due to the size and complexity of the resulting mathematical problem. The challenge here is how does one incorporate issues, such as, economics, environmental impacts, operability, safety, and sustainability in the early stages of process synthesis and/or design (Tula et al., 2017).

Computer-aided approaches using AI-inspired methods combined with fundamental concepts offer several advantages over purely data-driven methods, which may lead to infeasible flowsheets, or purely model-based methods, which may run into difficulties in the numerical solution step due to the complexity of the models used.

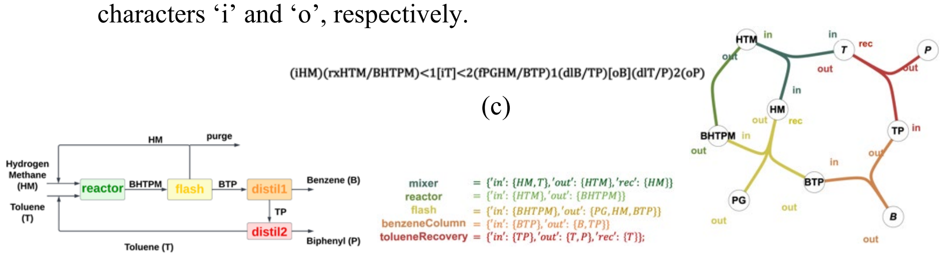
To facilitate the development of such hybrid AI-based approaches combining process knowledge with computational algorithms, an appropriate flowsheet representation is needed that is concise, complete, and accurate. The SFILES representations developed by (D'Anterroches, 2005), (Bommareddy et al., 2011), and (Tula et al., 2015) was the first step in this direction, and has been shown to have various applications, such as, flowsheet autocompletion (Vogel et al., 2022), piping and instrumentation diagram generation (Hirtreiter et al., 2022), and flowsheet pattern mining (Zhang et al., 2019). The SFILES strings represent correctly and consistently a wide range of process flowsheets, involving typical operations found in chemical and biochemical processes. Note that, like the SMILES strings for a molecule, a parser is needed to convert SFILES strings to the actual process flowsheet diagrams. As originally developed, a process flowsheet is first represented by a set of process groups, which are similar to the functional groups that represent a molecule. The process group representation is then converted to a SFILES representation. Unlike the molecular representation with SMILES, several additional details need to be considered. For example, information about the number of chemicals and their effect on the system behavior; the direction of flow-paths for reactants, products, inerts, solvents, etc. need to be tracked; and, start and end of the process need to be clearly marked. These additional issues give opportunities for a symbolic AI-based intelligent system (Venkatasubramanian & Mann, 2022) need to be incorporated to the current process synthesis-design methods such that the application of SFILES to represent process flowsheets as well as its use in computer-aided process synthesis and design can be extended.

The objective of this paper is to present a hybrid multi-level AI framework for fast, efficient, and reliable flowsheet synthesis and design taking into account concepts and theory on chemical process development, together with the knowledge and data for hundreds of process flowsheets that are known already. We present an extended SFILES representation (eSFILES) based on SFILES of process flowsheets ((D'Anterroches, 2005); (Bommareddy et al., 2011); (Tula et al., 2015)) and the annotated hypergraph representation, developed originally to study networks in organic chemistry (Mann & Venkatasubramanian, 2023). We highlight selected developments where we illustrate the hierarchical eSFILES representation framework for any chemical process flowsheet, also suitable for process synthesis, design, and innovation (Tula et al., 2015). Namely, we propose a three-level, extended SFILES framework where the lowest level contains flowsheet connectivity information which could be used to generate text-based flowsheet representations like SFILES but with additional information; the middle level contains additional details on process groups and streams providing information necessary information for process synthesis (superstructure-based optimization or process group based enumeration and test); and the top level contains process operational data in terms of design parameters for process simulation and innovation through optimization and intensification. At the core of our approach lies a hypergraph representation of flowsheet (Mann & Venkatasubramanian, 2023) that represents process groups as hyperedges and streams as nodes with annotations indicating contextual information.

2. Current state of flowsheet representation: SFILES

A text-based representation for chemical flowsheets reported by (Tula et al., 2015) is called simplified flowsheet input line entry scheme (or SFILES) analogous to the SMILES representation for molecules. Such a representation offers a concise way of representing flowsheets and could also facilitate not only storage of flowsheets, but also, enumeration, analysis, and deployment for interactive text-based algorithms process synthesis, design, and simulation of a given flowsheet. (Tula et al., 2015) developed separate rule-based algorithms for process synthesis and design, whose solutions were not added to SFILES strings. The underlying idea behind SFILES is to identify analogous process components similar to a SMILES string as follows –

- represent processes as *process groups* akin to functional groups in molecules.
- represent process flow streams akin to chemical bonds in molecules.
- represent various connections, recycle streams, and branching of streams akin to branching in molecules.
- unlike SMILES, SFILES need special unidirectional input nodes and output nodes that indicate the start and end of a process flowsheet, through special characters ‘i’ and ‘o’, respectively.



(a) (b) (d)

Figure 1: (a) HDA process flowsheet, (b) the various process groups, (c) the corresponding SFILES string, (d) HDA process represented by process groups. In the above subplots, the abbreviations refer to – mixture of hydrogen, methane (HM); mixture of benzene, hydrogen, toluene, biphenyl, methane (BHTPM); mixture of benzene, toluene, biphenyl (BTP); benzene (B); biphenyl (P); toluene (T); purge gas (purge). The hypergraph was generated using Wolfram Mathematica and algorithmic implementation would be done in Python.

We use the well-known hydrodealkylation (HDA) process to illustrate the basic concepts and the new developments. First the process flowsheet is represented by a set of process groups retrieved from the database (the process group representation is shown in Fig 1b) and the parser is used to generate the SFILES string (shown in Fig 1c). The reader is referred to (Tula et al., 2015) for detailed examples on SFILES generation.

3. Extended SFILES (eSFILES)

Hypergraph: The hypergraph representation of the HDA process is shown in Figure 1(d). While this new SFILES representation offer numerous benefits, there is scope for further improvement and incorporation of additional details that would aid in intelligent process flowsheet synthesis and design. The new information that, when incorporated, would generate additional, more complete information in the flowsheet representation such as:

- information on incoming/outgoing streams for each process group explicitly, especially when multiple streams enter or leave a process group.

- the minimum data needed to completely define a material and/or energy stream.
- information on driving force for the process groups indicating the ease/difficulty of conversion/separation.
- ability to systematically generate superstructure of flowsheet and enumerate valid alternatives aiding in the design problem.

Towards this end, we propose the extended SFILES (or eSFILES) representation that provides a framework for incorporation of this information in a hierarchical manner and further details on this are presented in the subsequent sections.

Hierarchical framework: A hierarchy is needed to efficiently organize process knowledge needed for different applications and to guide users to select, retrieve, and/or collect the data needed for a specific application. Namely, the eSFILES hierarchy has three levels of information -- first at the lower level, information needed to only represent the flowsheet (connectivity); next, at the middle-upper level, introduce additional details such as driving force and mixture composition that helps to establish the material and energy flows; and finally, the top level that contains information on composition and material/energy balance that would be need to solve a flowsheet among other applications. The eSFILES hierarchy is shown in Figure 2.

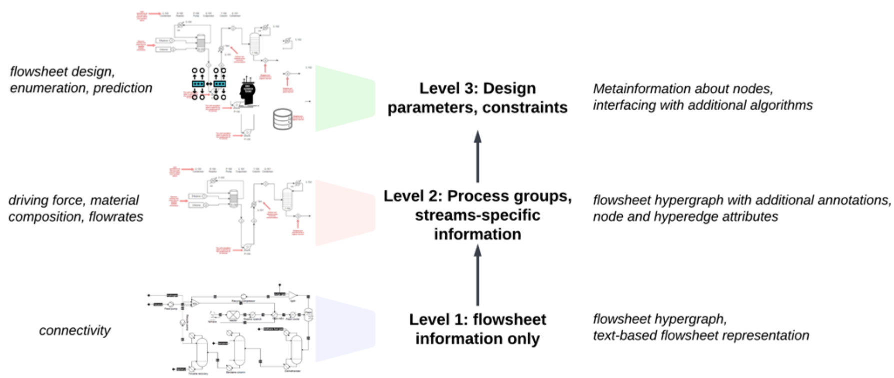
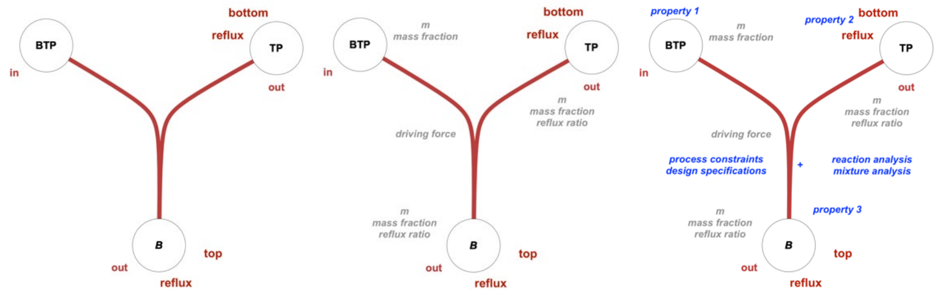


Figure 2: eSFILES framework with three levels of hierarchy

Implementation of hierarchical framework: The concept of hypergraphs is introduced to incorporate multi-level hierarchy of information with eSFILES. A hypergraph is a generalization of a graph where each edge is not limited to connecting only two vertices but could connect any number of vertices. Mathematically, a hypergraph consists of a set of hyperedges and vertices where each hyperedge consists of a non-empty subset of vertices. In addition, each vertex has a hyperedge-specific annotation indicating a predefined contextual information. The flexibility of hyperedges to connect more than one node and ability to have hyperedge-specific node annotations offers several advantages for flowsheet representation. Further details on the annotated hypergraph framework and its various applications in reaction engineering are presented in (Mann & Venkatasubramanian, 2023).

For instance, for the HDA process, the equivalent flowsheet hypergraph representation is shown in Figure 1(d) where each process group is represented as a colored hyperedge, process streams are represented as vertices, and the vertices have hyperedge-specific 'in'/'out' annotations indicating the incoming/outgoing nature of the process streams for a corresponding process groups. Moreover, recycle streams are indicated with an additional annotation 'rec' wherever applicable for the process streams. Figure 1(b) shows each process group as a hyperedge which essentially is a set of named vertices where the names (or annotations) indicate the streams' nature.

The three levels of information in the eSFILES framework is shown for the `toluene recovery' process group in Figure 3. In Figure 3(a), only the hyperedge representing the process group along with participating streams as vertices and role-specifying annotations are shown. In Figure 3(b), additional information for the entire process group (or hyperedge) such as driving force and stream (or vertex)-specific information such as mass fraction and composition is shown. Finally, information on design aspects and necessary information required for additional analyses such as mixture analysis, reaction analysis, and pure component properties are presented in Figure 3(c). Though the three levels in eSFILES are only shown for a process group brevity, the same could be done for an entire process using the same approach.



(a) (b) (c)

Figure 3: (a) Distillation column hyperedge with additional annotations. The corresponding text-based representation for this process groups is BTP(in)>dl>B(out,top,reflux)TP(out,bottom,reflux), (b) the driving force, material compositions, reflux rate, and conversion indicated as hyperedge attributes (for process groups) and hyperedge-specific node attributes (for streams) (c) the design parameters, properties, and constraints are represented as additional meta information.

Application of the eSFILES in sustainable process design: (Tula et al., 2017) proposed a 3-stages sustainable process design method consisting of 12 hierarchical steps. Stage-1 consists of steps 1-4 where the process synthesis problem is solved. Stage-2 consists of steps 5-9 where the base case process design problem is solved. Stage-3 consists of steps 10-12 where the base case design is further improved to find innovative and more sustainable alternatives. The original multi-stage method of (Tula et al., 2017) now retrieves the necessary data from the stored hypergraph. Also, different alternatives for stage-3 can easily be generated through different alternatives of the hypergraph, allowing the user to generate and analyze multiple alternatives as well as store the information for future use without having to repeat any of the previously solved steps.

4. Conclusions

The concept of hypergraph has been used to extend the original SFILES concept to multiple levels to store as well as retrieve information related to process synthesis, design, and innovation. In this way, an existing interactive multi-stage method has been made more intelligent and efficient through the adoption of an available AI method for sustainable process design. Therefore, the eSFILES-based framework for process synthesis and design incorporates a combination of artificial intelligence-based methods and well-known chemical engineering knowledge incorporated through an intelligent system facilitating fast, correct, and consistent decision-making related to process synthesis and design.

Current and future work involves the development of an extended manuscript providing more details of the eSFILES, its adoption in the multi-stage sustainable process design method, and its application in several case studies. In particular, we plan to provide rigorous examples to demonstrate the usability of eSFILES for new process flowsheets and also use them to address real challenges such as provide flowsheet alternatives, perform process intensification, and flowsheet property prediction based on the text-based flowsheet representation using a natural language framework similar to molecular property prediction (Mann et al., 2022) and chemical product design (Mann et al., 2023).

References

Bommareddy, S., Eden, M. R., & Gani, R. (2011). *Computer Aided Flowsheet Design using Group Contribution Methods* (pp. 321–325). https://doi.org/10.1016/B978-0-444-53711-9.50065-1

D'Anterroches, L. (2005). *Process flowsheet generation & design through a group contribution approach*. [CAPEC], Department of Chemical Engineering, Technical University of Denmark.

Hirtreiter, E., Balhorn, L. S., & Schweidtmann, A. M. (2022). *Towards automatic generation of Piping and Instrumentation Diagrams (P&IDs) with Artificial Intelligence*. http://arxiv.org/abs/2211.05583

Mann, V., Brito, K., Gani, R., & Venkatasubramanian, V. (2022). Hybrid, Interpretable Machine Learning for Thermodynamic Property Estimation using Grammar2vec for Molecular Representation. *Fluid Phase Equilibria*, *561*, 113531. https://doi.org/10.1016/j.fluid.2022.113531

Mann, V., Gani, R., & Venkatasubramanian, V. (2023). Group contribution-based property modeling for chemical product design: A perspective in the AI era. *Fluid Phase Equilibria*, *568*, 113734. https://doi.org/10.1016/J.FLUID.2023.113734

Mann, V., & Venkatasubramanian, V. (2023). AI-driven hypergraph network of organic chemistry: network statistics and applications in reaction classification. *Reaction Chemistry & Engineering*. https://doi.org/10.1039/D2RE00309K

Tula, A. K., Babi, D. K., Bottlaender, J., Eden, M. R., & Gani, R. (2017). A computer-aided software-tool for sustainable process synthesis-intensification. *Computers & Chemical Engineering*, *105*, 74–95. https://doi.org/10.1016/j.compchemeng.2017.01.001

Tula, A. K., Eden, M. R., & Gani, R. (2015). Process synthesis, design and analysis using a process-group contribution method. *Computers & Chemical Engineering*, *81*, 245–259. https://doi.org/10.1016/j.compchemeng.2015.04.019

Venkatasubramanian, V., & Mann, V. (2022). Artificial intelligence in reaction prediction and chemical synthesis. *Current Opinion in Chemical Engineering*, *36*, 100749. https://doi.org/10.1016/j.coche.2021.100749

Vogel, G., Balhorn, L. S., & Schweidtmann, A. M. (2022). *Learning from flowsheets: A generative transformer model for autocompletion of flowsheets*. http://arxiv.org/abs/2208.00859

Zhang, T., Sahinidis, N. v., & Siirola, J. J. (2019). Pattern recognition in chemical process flowsheets. *AIChE Journal*, *65*(2), 592–603. https://doi.org/10.1002/aic.16443