




# Computer-Aided Mixture Design Using Molecule Superstructures

**Conference Paper****Author(s):**

[Rehner, Philipp](#) ; [Schilling, Johannes](#) ; [Bardow, André](#) 

**Publication date:**

2024

**Permanent link:**

<https://doi.org/10.3929/ethz-b-000684391>

**Rights / license:**

[Creative Commons Attribution-ShareAlike 4.0 International](#)

**Originally published in:**

Systems and Control Transactions 3

# Computer-Aided Mixture Design Using Molecule Superstructures

Philipp Rehner, Johannes Schilling, André Bardow\*

ETH Zurich, Energy and Process Systems Engineering, Zurich, Switzerland

\* Corresponding Author: [abardow@ethz.ch](mailto:abardow@ethz.ch).

## ABSTRACT

Computer-aided molecular and process design (CAMPD) tries to find the best molecules together with their optimal process. If the optimization problem considers two or more components as degrees of freedom, the resulting mixture design is challenging for optimization. The quality of the solution strongly depends on the accuracy of the thermodynamic model used to predict the thermophysical properties required to determine the objective function and process constraints. Today, most molecular design methods employ thermodynamic models based on group counts, resulting in a loss of structural information of the molecule during the optimization. Here, we unlock CAMPD based on property prediction methods beyond first-order group-contribution methods by using molecule superstructures, a graph-based molecular representation of chemical families that preserves the full adjacency graph. Disjunctive programming is applied to optimize molecules from different chemical families simultaneously. The description of mixtures is enhanced with a recent parametrization of binary group/group interaction parameters. The design method is applied to determine the optimal working fluid mixture for an Organic Rankine cycle.

**Keywords:** Molecular Design, Energy Conversion, Process Design, Optimization, Exergy Efficiency

## INTRODUCTION

Computer-aided molecular and process design (CAMPD) [1] determines optimal molecules jointly with their optimal process. Evaluating molecules for their performance in a process enables a comprehensive assessment. The influence of different molecular characteristics is combined in a process-level target function that can quantify the thermodynamic, economic, or environmental performance of the process.

Molecules occur as degrees of freedom for various energy conversion processes like heat pumps [2] and organic Rankine cycles (ORC) [3-5] or separation processes like CO<sub>2</sub> capture [6]. The design space of molecules is vast but can even be extended exponentially by considering blends of multiple molecular species [7,8].

To optimize a molecule, its structure must be featurized, i.e., expressed in a machine-readable way [9]. Common features used in molecular design are group counts that indicate the number of occurrences of predefined groups in a molecule. The group counts can be used as degrees of freedom in an optimization algorithm.

Additional molecular constraints can ensure that the optimization only generates valid molecules.

The group counts are used in group-contribution methods to determine the thermophysical properties required for the process model [10]. A disadvantage of using group counts as features is that some of the structural information of the molecule is lost in the featurization. Therefore, the optimization does not necessarily allow the identification of an optimal molecule but rather a set of optimal isomers that share the same group counts.

The limitation of group counts is alleviated by maintaining structural information during the optimization. This can be achieved by optimizing the coefficients of an adjacency matrix [11]. However, the number of degrees of freedom becomes large, which is a limiting factor in the presence of expensive model evaluations that occur in an integrated molecular and process design. As an alternative, we recently introduced molecule superstructures [12]. The superstructures can encode all molecules of a particular chemical family using a significantly reduced number of binary variables.

The molecular design space superstructures offer is

smaller than first-order group-contribution methods. However, the molecular representation should not only be picked for computational convenience but also based on the expected accuracy of property prediction methods. Often, extrapolating a group contribution method to molecules significantly different from the molecules used for parameter adjustment (e.g., polyfunctional molecules) leads to low accuracies that undermine the results of a molecular design.

Evaluating a process model requires an equation of state or an equivalent model. For mixture design, the model should predict a wide range of fluids and their mixtures. Molecular equations of state like those based on statistical associating fluid theory (SAFT) [13] leverage molecular insights to describe mixtures accurately based on pure-component parameters. If a higher accuracy is required, binary interaction parameters can further enhance the description of mixtures.

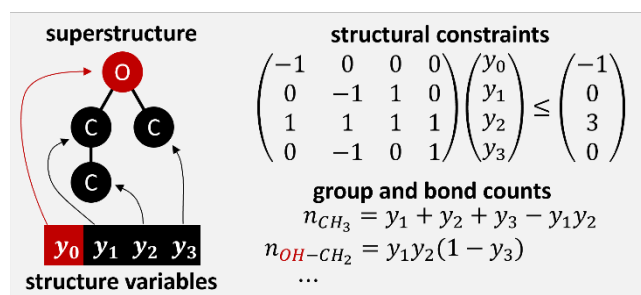
In a recent study, the authors and co-workers [14] extended the group contribution method for PCP-SAFT [15,16] by Sauer et al. [17] to mixtures by accounting for hydrogen bonding between polar molecules (aldehydes, ketones, ethers, formates, and esters) and self-associating molecules (alcohols and amines), and by parametrizing a group contribution method for binary interaction parameters. The study compares a homosegmented and a heterosegmented group-contribution method. The homosegmented approach determines PCP-SAFT parameters from group counts and group-specific parameters. In the heterosegmented approach, segments replace molecules as species in the equation of state. The heterosegmented approach was determined to be more accurate for pure components [17] and for mixtures [14], but to parametrize the model, the number of bonds between different groups is required in addition to the group counts. The molecule superstructures [12] can be used to infer both group counts and bond information.

In this work, the more accurate heterosegmented group-contribution method for PCP-SAFT is unlocked for application in an integrated molecular and process design using molecule superstructures. We demonstrate the approach in a case study determining the optimal working fluid mixture for an organic Rankine cycle (ORC). Due to the combinatorial complexity of possible working fluid mixtures, a molecular design is necessary to determine the optimal working fluid systematically.

## MOLECULE SUPERSTRUCTURES

To optimize molecules in a computer-aided molecular and process design, the structure of the molecule needs to be represented in a format that is accessible to the optimization algorithm. Molecule superstructures are a graph-based representation of a molecular family [12]. Fig. 1 shows an example of a small molecule structure. It

consists of a graph with four nodes and three edges. Every node corresponds to an atom type and a binary structure variable. Structural constraints are introduced to ensure that only valid molecules are generated (in this case water, methanol, ethanol, or dimethyl ether). The structure variables are converted to group and bond counts that are required to parametrize the equation of state. The structural constraints and group/bond counts can be derived generically for arbitrarily large superstructures. We refer to the original publication [12] for the detailed expressions.



**Figure 1:** Example for an ether/alcohol superstructure including the corresponding structural constraints and resulting group and bond counts.

Previously, optimal molecules were determined for every chemical family individually [12]. For mixtures, the number of possible combinations of chemical families becomes large. Therefore, we extend the superstructure approach to optimize any number of chemical families simultaneously using methods from disjunctive programming [18].

The superstructure converts binary structure variables  $y_k$  into molecular features used in the property prediction method. Linear inequality constraints are required to ensure that only valid molecules are found. The specific constraints are elucidated in the original publication [12]. In general, the constraints can be expressed as

$$\sum_k a_{ijk} y_k \leq b_{ij} \quad (1)$$

Here,  $i$  enumerates the chemical families and their respective molecule superstructures,  $j$  indicates different constraints, and  $k$  enumerates the structure variables  $y_k$ . The number of constraints  $N_{c,i}$  and structure variables  $N_{y,i}$  depends on the considered molecule superstructures.

To combine the different chemical families in one structure, a common set of structure variables is defined by using  $N_y = \max_i N_{y,i}$  binary variables and adding additional constraints that force the unused variables to 0.

$$\sum_{k=N_{y,i}+1}^{N_y} y_k \leq 0 \quad (2)$$

The constraints associated with each molecule superstructure can then be formally expressed as disjunctions in the optimization problem.

$$v_i [\sum_k a_{ijk} y_k \leq b_{ij}] \quad (3)$$

The Big-M method and convex hulls have been established as standard methods to convert disjunctive programs into MILPs or MINLPs [18]. The convex hull has a tighter feasible region than the Big-M formulation but requires more additional variables. Therefore, the Big-M formulation is better suited for the integrated process design task with expensive target function evaluations. Applying the Big-M method introduces additional binary variables  $c_i$ , one for each disjunction and hence chemical family. Only one disjunction can be active at the same time, leading to the additional constraint:

$$\sum_i c_i = 1 \quad (4)$$

The constraints for every molecule superstructure are rewritten by introducing the parameter  $M_{ij}$ .

$$\sum_k a_{ijk} y_k + M_{ij} c_i \leq b_{ij} + M_{ij} \quad (5)$$

In general,  $M_{ij}$  must be large enough to ensure that the constraint is turned off if  $c_i = 0$ , but as small as possible to ensure tight relaxation bounds. For the binary variables  $y_k$ , the optimal value for  $M_{ij}$  can be derived generically, as

$$M_{ij} = \sum_k \max(a_{ijk}, 0) - b_{ij} \quad (6)$$

The original publication on molecule superstructures [12] describes a generic method to infer the group and bond counts from the structure variables. The same method is used for every individual superstructure. The resulting group counts  $n_{i,\alpha}$  and bond counts  $b_{i,\alpha\beta}$  are then weighted with the chemical family indicator  $c_i$  of every superstructure to give the total group and bond counts.

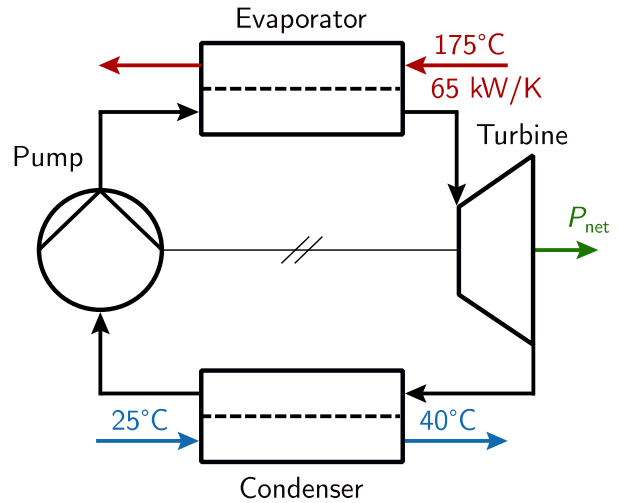
$$n_\alpha = \sum_i c_i n_{i,\alpha}, \quad b_{\alpha\beta} = \sum_i c_i b_{i,\alpha\beta} \quad (8)$$

The group and bond counts are used to parametrize the heterosegmented group-contribution method for PCP-SAFT [17], which is used to determine phase equilibria and residual properties. To calculate caloric properties, the group counts are also used in the group-contribution method for the ideal gas heat capacity by Joback and Reid [19]. The resulting MINLP is solved using the branch-and-bound implementation in Artelys Knitro [20].

## CASE STUDY: ORGANIC RANKINE CYCLE

We apply the molecular design method using molecule superstructures to the integrated design of an organic Rankine cycle (ORC) and its working fluid. ORCs can convert waste heat into power, reducing energy losses and increasing the exergetic efficiency of processes. The process flowsheet of the standard ORC used in this case study is shown in Fig. 2. Analogous to a regular Rankine cycle, the working fluid of the ORC is evaporated and potentially superheated to be then fed into a

turbine that extracts power from the working fluid. Part of the power is used to pump the working fluid that leaves the condenser back to the pressure level of the evaporator.



**Figure 2:** Flowsheet and key specifications for the ORC.

Using an organic working fluid rather than water enables the operation above atmospheric pressures, even for low-temperature heat sources. This work optimizes the working fluid mixture to maximize the power output of the ORC. The design of an actual waste heat valorization unit needs to consider additional aspects regarding safety, stability, and costs that are outside the scope of this study.

ORC processes are promising targets for mixture design since replacing pure working fluids with zeotropic mixtures leads to non-isothermal evaporation and condensation. The resulting temperature glide in the condenser and evaporator can be tuned to the respective temperature glides in the heat source and cooling mediums. A more homogeneous temperature difference across the heat exchangers reduces the exergy loss during heat transfer and, in conclusion, increases the net power output of the process.

The case study is based on the analysis of Chys et al. [21] and the design study by Schilling et al. [8]. The most important specifications are shown in Fig. 2: The heat source has an inlet temperature of 175 °C and a heat capacity rate of 65 kW/K. The cooling medium is heated from 25 °C to 40 °C. The outlet temperature of the heat source medium is only indirectly constrained via the pinch condition. The heat exchangers are modeled as isobaric. The full specification of the process is given in Tab. 1. This thermodynamic optimization aims to maximize the net power  $P_{net}$  of the cycle.

**Table 1:** ORC case study specifications from [8,21]

Component	Parameter	Value
Heat source	Inlet temperature	175 °C
	Heat capacity rate	65 kW/K
	Min. approach temp.	20 K
Turbine	Isentropic efficiency	0.65
	Min. vapor quality (outlet)	1
Pump	Isentropic efficiency	0.8
Condenser	Min. absolute pressure	1 bar
	Min. relative pressure	1e-3
Evaporator	Max. absolute pressure	50 bar
	Max. relative pressure	0.8
Cooling	Inlet temperature	25 °C
	Outlet temperature	40 °C
	Min. approach temp.	10 K

## RESULTS

As a reference, the molecule superstructures with disjunctions are first used to determine the optimal pure working fluid for the ORC. A ranking of the ten molecules that yield the highest net power output is shown in Tab. 2. The ranking consists of alkanes, alkenes, and propyne.

**Table 2:** Ranking of the best pure component working fluids based on the net power output of the ORC.

#	SMILES	Name	Net power [kW]
1	CCC	propane	472
2	C=CC	propene	435
3	C#CC	propyne	425
4	CC(C)C	isobutane	412
5	CC(C)(C)C	neopentane	409
6	CCCC	butane	399
7	C=C(C)C	isobutene	396
8	C=CCC	butene	395
9	C=CC(C)(C)C	neohexene	395
10	CCC(C)C	isopentane	394

The molecular design space also contains polar (aldehydes, ketones, ethers, formates, and esters) and self-associating molecules (alcohols and amines). However, these molecular families are unsuitable for the application due to their lower vapor pressures and higher critical temperature. The ranking is topped by the C3 hydrocarbons propane, propene, and propyne, with the highest achievable net power output being 472.25 kW using propane as the working fluid.

Higher net power outputs can be achieved by considering mixed working fluids. To obtain a ranking of working fluid candidates, we use the molecule

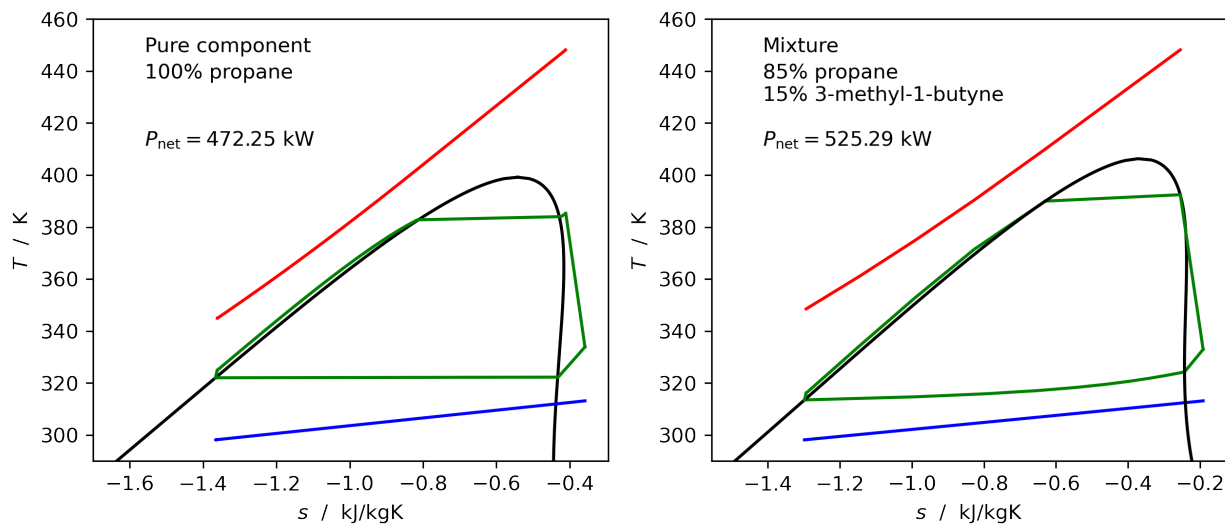
superstructure concept to find optimal additives for each of the ten best-performing pure components. The ranking is shown in Tab. 3. The ten best mixtures all consist of mainly propane or propene with traces of larger, mostly non-polar molecules. The mixtures containing one of the other well-performing fluids from the pure-component case consistently lead to lower power outputs than those based on propane or propene. This observation strongly implies that optimizing additives can determine the optimal mixture for this case study. However, both molecules must be optimized in a single optimization problem for a more comprehensive result. Schilling et al. [8] demonstrate how the approach can significantly reduce the number of function evaluations compared to a screening of mixtures using a CAMD method based on group counts. The design setup in our study allows the simultaneous optimization of both constituent molecules, but the prevalence of local optima impedes an efficient direct determination of optimal working fluid mixtures. Therefore, it was not possible yet to determine a ranking comparable to Tab. 3 by optimizing both fluids rather than an additive in a reasonable timeframe. A MINLP solution algorithm better suited for the specificity of integrated molecular and process design is required to improve the solution times and generate more robust candidate mixtures [22]. The non-convexity of the MINLP also suggests using global optimization [23], however, the expensive function evaluations that include multiple phase equilibrium calculations with the heterosegmented group-contribution method for PCP-SAFT are a limiting factor.

Fig. 3 compares the process for the optimal pure working fluid (propane) with the best mixed working fluid candidate. In the pure-component case, the temperature of the working fluid is constant during evaporation and condensation. Therefore, the optimization mainly determines process conditions in which the temperature glide of the heat source medium aligns well with the temperature during the preheating step in the evaporator. The mixed working fluid shows a temperature glide in the condenser. The resulting process conditions find a compromise between the exergy loss in the evaporator and the condenser. Consequently, the net power output of the ORC is increased by 11% compared to propane as the working fluid. This increase is partly driven by the temperature glide in the cooling medium, which is fixed to 15 K in this case study. For other process conditions, the benefit of using mixed working fluids can be small [24], and a reduced heat transfer during condensation and evaporation must be considered for an economic assessment of the process [25,8].

Choosing the correct composition of the mixture is essential for the performance of the ORC. Fig. 4 shows the optimized net power output of the ORC using the four best mixtures from Tab. 3 over the entire composition

**Table 3:** Ranking of the best working fluid mixtures found in the integrated design.

#	Component 1		Component 2		Molar composition	Net power [kW]
	SMILES	Name	SMILES	Name		
1	CCC	propane	C#CC(C)C	3-methyl-1-butyne	85% / 15%	525
2	CCC	propane	CC(C)(C)C	neopentane	92% / 8%	524
3	CCC	propane	C#CCC	1-butyne	65% / 35%	522
4	CCC	propane	CCC(C)C	isopentane	94% / 6%	519
5	C=CC	propene	CCCC	butane	83% / 17%	519
6	C=CC	propene	CCC(C)C	Isopentane	93% / 7%	518
7	CCC	propane	C=C(C)CC	2-methyl-1-butene	95% / 5%	518
8	CCC	propane	CCCCC	pentane	97% / 3%	515
9	CCC	propane	CCCO	1-methoxypropane	96% / 4%	514
10	CCC	propane	COC(C)C	2-methoxypropane	95% / 5%	512

**Figure 3:** Ts-diagram for the ORC using propane (left) and the best mixture (right) as working fluid. The phase change of the working fluid (green) is isothermal for pure components but exhibits a temperature glide for mixtures. In the optimization the temperature glide is adjusted to the temperature profiles in the heat source medium (red) and the cooling water (blue).

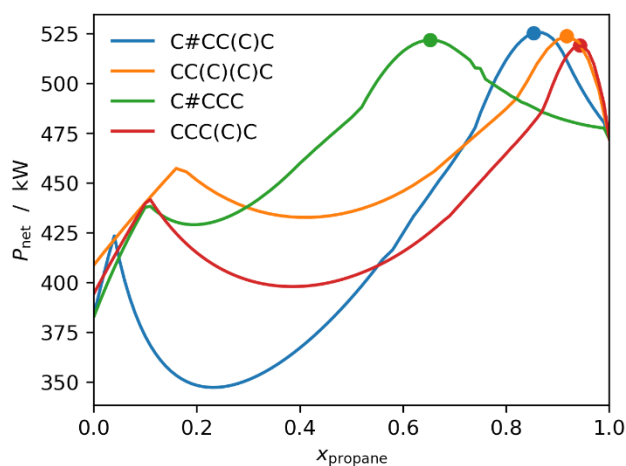
range. The solutions from the mixture design align well with the maxima in the discretized curves. In all cases, adding the longer hydrocarbon to pure propane initially increases the net power due to the establishment of the temperature glide. Further increasing the amount of additive reduces the performance, with the curves showing local minima that, in one case, are even below the net power output of the pure additive.

The net power output is strongly related to the ratio of the temperature glides in the condenser and the cooling medium [8]. Changing the composition of the mixtures affects the glide in the condenser, with the largest glide occurring in the middle of the composition range. The glide also explains the shape of the maxima in the net power: A jump of the pinch from the condenser outlet to the inlet leads to a kink in the net power (cf. Fig. 4). If the temperature profile in the condenser is shaped in a

way that the pinch is located between the inlet and the outlet, the maximum in the net power is smooth.

A strong sensitivity to variations in the composition of the working fluid can be problematic in an actual ORC process when leakage is non-negligible, and the composition can change due to the constituents' different relative volatilities. With the abundance of near-optimal mixtures in the solution space of the molecular design, the robustness with respect to composition changes can be incorporated as an additional constraint in the optimization problem.





**Figure 4:** Optimized net power output of the ORC using a propane + X mixture with the specified composition. The marks represent the optimal composition found in the molecular design.

## CONCLUSION

Molecule superstructures are used to determine the optimal working fluid mixture for an ORC. Multiple chemical families represented by their own superstructure are optimized simultaneously using disjunctive programming. Due to the full structural information available during the molecular design, the accurate heterosegmented group-contribution method for PCP-SAFT can be used as a property prediction method. The model accuracy is further enhanced by a recent parametrization of binary group/group interaction parameters [14].

The optimization determines mixtures of propane or propene with larger hydrocarbons as optimal working fluids. The top-ranked mixtures deliver comparable performances. This observation suggests that additional material properties should be considered within the optimization to reduce the molecular design space. Amongst the mixtures with the best thermodynamic properties are two that contain alkynes, components that are unsuited for long-term application as working fluids due to their low chemical stability. If predictive models are available, adding constraints for the working fluid's safety, stability, and environmental impact avoids a posteriori filtering of the results and provides a more efficient design method. Finally, economic considerations were circumvented by empirical pinch constraints instead of a direct consideration by using more detailed, rate-based process models [8]. Also, in this case, the more detailed representation by molecule superstructures could enable advanced thermodynamic models, e.g., for the required transport properties.

## ACKNOWLEDGEMENTS

PR acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 497566159. AB acknowledges support by BRIDGE as part of the project “High-Efficiency High-Temperature Heat Pumps with Temperature Glide.” We thank the Swiss National Science Foundation SNSF and Innosuisse for their support.

## REFERENCES

1. Papadopoulos AI, Tsivintzelis I, Linke P, Seferlis P. Computer-aided Molecular Design: Fundamentals, Methods, and Applications. *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering* (2018)
2. Roskosch D, Atakan B. Reverse engineering of fluid selection for thermodynamic cycles with cubic equations of state, using a compression heat pump as example. *Energy* 81:202-212 (2015)
3. Papadopoulos A, Stijepovic M, Linke P. On the systematic design and selection of optimal working fluids for Organic Rankine Cycles. *Appl Therm Eng* 30:760-769 (2010)
4. Schilling J, Lampe M, Gross J, Bardow A. 1-stage CoMT-CAMD: An approach for integrated design of ORC process and working fluid using PC-SAFT. *Chem Eng Sci* 159:217-230 (2017)
5. Bowskill D, Tropp UE, Gopinath S, Jackson G, Galindo A, Adjiman CS. Beyond a heuristic analysis: integration of process and working-fluid design for organic Rankine cycles. *Mol Syst Des Eng* 5:493-510 (2020)
6. Papadopoulos A, Badr S, Chremos A, Forte E, Zarogiannis T, Seferlis P, Papadokostantakis S, Galindo A, Jackson G, Adjiman C. Computer-aided molecular design and selection of CO<sub>2</sub> capture solvents based on thermodynamics, reactivity and sustainability. *Mol Syst Des Eng* 1:313-334 (2016)
7. Papadopoulos A, Stijepovic M, Linke P, Seferlis P, Voutetakis S. Toward optimum working fluid mixtures for organic Rankine cycles using molecular design and sensitivity analysis. *Ind Eng Chem Res* 52:12116-12133 (2013)
8. Schilling J, Entrup M, Hopp M, Gross J, Bardow A. Towards optimal mixtures of working fluids: Integrated design of processes and mixtures for Organic Rankine Cycles. *Renew Sust Energ Rev* 135:110179 (2021)
9. Austin ND, Sahinidis NV, Trahan DW. Computer-aided molecular design: An introduction and review of tools, applications, and solution techniques. *Chem Eng Res Des* 116:2-26 (2016)

10. Marrero J, Gani R. Group-contribution based estimation of pure component properties. *Fluid Phase Equilib* 183:183-208 (2001)
11. Churi N, Achenie LEK. Novel mathematical programming model for computer aided molecular design. *Ind Eng Chem Res* 35(10):3788-3794 (1996)
12. Rehner P, Schilling J, Bardow A. Molecule superstructures for computer-aided molecular and process design. *Mol Syst Des Eng* 8:488-499 (2023)
13. Chapman WG, Gubbins KE, Jackson G, Radosz M. New reference equation of state for associating liquids. *Ind Eng Chem Res* 29(8):1709-1721 (1990)
14. Rehner P, Bardow A, Gross J. Modeling mixtures with PC-SAFT: Insights from large-scale parametrization and group-contribution method for binary interaction parameters. *Int J Thermophys* 44(12), 179 (2023)
15. Gross J, Sadowski G. Perturbed-chain SAFT: An equation of state based on a perturbation theory for chain molecules. *Ind Eng Chem Res* 40(4):1244-1260 (2001)
16. Gross J, Vrabec J. An equation-of-state contribution for polar components: dipolar molecules. *AIChE J* 52(3):1194-1204 (2006)
17. Sauer E, Stavrou M, Gross J. Comparison between a homo- and a heterosegmented group contribution approach based on the perturbed-chain polar statistical associating fluid theory equation of state. *Ind Eng Chem Res* 53(38):14854-14864 (2014)
18. Grossmann I. Review of nonlinear mixed-integer and disjunctive programming techniques. *Optim Eng* 3:227-252 (2002)
19. Joback KG, Reid RC. Estimation of pure-component properties from group-contributions. *Chem Eng Commun* 57:233-243 (1987)
20. Byrd RH, Nocedal J, Waltz RA. KNITRO: An integrated package for nonlinear optimization. In: Large-Scale Nonlinear Optimization. Springer (2006)
21. Chys M, van den Broek M, Vanslambrouck B, De Paepe M. Potential of zeotropic mixtures as working fluids in organic Rankine cycles. *Energy* 44(1):623-632
22. Jonuzaj S, Gupta A, Adjiman CS. The design of optimal mixtures from atom groups using Generalized Disjunctive Programming. *Comput Chem Eng* 116:401-421 (2018)
23. Boukouvala F, Misener R, Floudas CA. Global optimization advances in Mixed-Integer Nonlinear Programming, MINLP, and Constrained Derivative-Free Optimization, CDFO. *Eur J Oper Res* 252(3): 701-727 (2016)
24. Krempeus D, Bahamonde S, van der Stelt TP, Klink W, Colonna P, De Servi CM. On mixtures as working fluids of air-cooled ORC bottoming power plants of gas turbines. *Appl Therm Eng* 236(D):121730 (2024)
25. Shah MM. A method for predicting heat transfer during boiling of mixtures in plain tubes. *Appl Therm Eng* 89:812-821 (2015)

---

© 2024 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See <https://creativecommons.org/licenses/by-sa/4.0/>

