Process Flow Sheet Generation & Design through a Group Contribution Approach

d'Anterroches, Loïc; Gani, Rafiqul

Publication date:
2006

Document Version
Publisher's PDF, also known as Version of record

Citation (APA):
d'Anterroches, L., & Gani, R. (2006). Process Flow Sheet Generation & Design through a Group Contribution Approach.

Process Flowsheet Generation & Design through a Group Contribution Approach

Loïc d'Anterroches

November 2005

Department of Chemical Engineering
Technical University of Denmark

ISBN 87-91435-35-8
Printed by Book Partner, Nørhaven Digital, Copenhagen, Denmark

Preface

This thesis is submitted as a partial fulfillment of the requirements for the Ph.D.-degree at Danmarks Tekniske Universitet (Technical University of Denmark).

The work has been carried out at Institut for Kemiteknik (Department of Chemical Engineering) from September 2002 to September 2005 under the supervision of Professor Rafiqul Gani.

I would like to thank Professor Gani for providing me with the opportunity to accomplish this work in a center close to international partners, both industrials and academics. Doctor Claude Leibovici for his support throughout the project, inspiring discussions and *l'écho.* The students and co-workers from CAPEC and the department, for good times and helping when needed.

Finally, I would like to thank my family, my friends and Pamela, for their support and understanding.

Lyngby, September 2005
Loïc d'Anterroches

Abstract

This thesis describes the development of a framework for the synthesis and design of process flowsheets using a systematic strategy for computer aided flowsheet design (CAFD). The developed framework addresses the formulation, solution and analysis of CAFD problems, based on the concept of a process group, which represents a single unit operation or a set of unit operations in the same way an atom or a group of atoms represent a molecular group. Through the definition of the process groups, their connectivity rules and their contributions to specific flowsheet properties, chemical process flowsheets are synthesized, modelled and analysed in the same way as chemical molecules are synthesized and tested for their properties. To achieve this, simple and efficient methods for process synthesis and design have been developed. The synthesis of the chemical process flowsheet alternatives is performed through a reverse flowsheet property approach, where the process group building blocks are combined to form flowsheet structures having desired (target) properties. Then the design of the most promising flowsheet alternatives is performed through a reverse simulation approach, where the design of the unit operations is back calculated from the specifications of their inlet and outlet streams inherited from the corresponding process groups. The developed framework and associated computer aided methods and tools have been tested using a series of case studies and application examples.

Resumé på dansk

Denne afhandling beskriver udviklingen af et framework til opstilling og design af proces flowsheet ved hjælp af en systematisk strategi for Computer Aided Flowsheet Design (CAFD). Det udviklede framework omfatter formulering, løsning og analyse af CAFD problemer baseret på et koncept med procesgrupper som kan repræsentere en enkelt enhedsoperation eller en række enhedsoperationer på samme måde som en molekylgruppe kan repræsentere et atom eller en gruppe af atomer. Flowsheet for kemiske processer opstilles, modelleres og analyseres ved definitionen af disse procesgrupper, deres forbindelsesregler samt deres bidrag til specifikke flowsheet egenskaber på samme måde som kemiske molekyler bliver syntetiseret og testet for deres egenskaber. Hertil er simple og effektive metoder til processyntese og design blevet udviklet. Alternative flowsheet for kemiske processer opstilles baglæns ved at kombinere procesgrupper således at der dannes flowsheet strukturer som har de ønskede egenskaber. Derefter udvælges de mest lovende flowsheetalternativer til design hvorved de enkelte enhedsoperationer beregnes baglæns udfra specifikationerne for deres ind- og udgangsstrømme svarende til procesgruppernes ind- og udgangsstrømme. Det udviklede framework med tilknyttede computerbaserede metoder og værktøjer er blevet testet på en serie case studier og anvendelseseksempler.

Contents

List of Figures

List of Tables

1

Introduction

Modelling and simulation of a chemical process flowsheet usually involve identifying the structure of the flowsheet (defined by operations involved and their sequence), deriving model equations to represent each operation, and solving the resulting total model equations according to one of various available simulation strategies. The flowsheet synthesis problem determines the type of operations and their sequence needed to achieve the conversion of raw materials to some specified set of products. The flowsheet design problem determines the optimal values for the conditions of operation and for other operation/equipment related variables for the synthesized flowsheet. It can be noted that the flowsheet modelling, synthesis and design problems are related since for generation and screening of flowsheet alternatives (synthesis/design), some form of flowsheet models are needed. Also, flowsheet models are needed for verification of the synthesis/design problem solution.

In chemical process synthesis two types of approaches exist, in the first type, one seeks to improve an existing process flowsheet (also known as the retrofit problem), while in the the second type, one seeks to find a completely new process flowsheet. The methods to address those two approaches can be classified into three categories.

The methods that employ heuristics or are knowledge based. These methods rely on a set of rules based on a combination of experience, insights and available knowledge (data). These rules are often simple but may be contradictory. The rules need careful consideration before application as the context in which they can be applied is not necessarily fully defined. The heuristic methods can be used both for retrofit or completely new designs.

The methods that employ mathematical or optimisation techniques. These methods are based on mathematical programing techniques using optimisation to determine the best process flowsheet. These methods usually involve an important preliminary step, which is to setup a mathematical superstructure which represents all possible alternatives. Therefore, identification of the optimal process flowsheet depends very much on the definition of the mathematical superstructure. These methods require efficient numerical solvers and a good knowledge of the mathematical programming techniques.

The methods that employ physical insights. This third class of methods are "hybrid methods". These methods combines the physical insights of the knowledge based methods with mathematical programming techniques to formulate and solve a collection of mathematical problems within an identified narrow search space. The physical insights help to reduce the search space and to decompose the global optimization problem into a collection of well defined inter-related sub-problems.

The method proposed in this Ph.D-project belongs to the class of the "hybrid methods" for the synthesis and design of process flowsheets including reactions and separations. The objectives are that the method should be applicable to a large range of problems without the need to employ rigorous models at each decision step. The method should support both the retrofit of existing processes and the generation of new alternatives with the ability to build and reuse knowledge over time. Finally this method should be easy to setup and use.

To achieve these objectives the framework for Computer Aided Flowsheet Design (CAFD) composed of eight main steps has been developed: the definition of the process synthesis problem; the analysis of the process synthesis problem, to refine and extract usable knowledge from the problem definition and the available knowledge bases; the selection of the process group building blocks matching with the synthesis problem – a process group represents a unit operation or a set of unit operations; the synthesis and test of the flowsheet structure alternatives – a flowsheet structure is composed of connected process groups; the ranking of the flowsheet structure alternatives using models predicting the performance of the alternatives from the contribution of the process groups in the flowsheet alternatives and the selection of the most promising alternatives; the design of the selected flowsheet structure alternatives, determining the design parameters of the unit operations underlining the process groups; the post analysis of the designed alternatives for heat integration and/or environmental impact and the final verifications.

The thesis comprises a total of seven chapters. Chapter 2 covers the theoretical background of the thesis with an overview of the current process synthesis and design methods together with computer aided molecular design. In chapter 3, the objectives of this Ph.D-project and an overview of the solution method for process synthesis and design developed in this Ph.D-project are presented. Chapter 4 presents the various methods and tools developed in this Ph.D-project for Computer Aided Flowsheet Design (CAFD). The methods and tools developed, includes a group contribution for flowsheet property estimation, a method for reverse flowsheet property prediction and a method for reverse simulation. Chapter 5 describes the computer implementation of the developed methods and tools for CAFD, together with its integration with the already available tools within CAPEC. The application of the method to a set of illustrative examples is presented in chapter 6. Finally, chapter 7 provides a summary of the thesis with some concluding remarks. The future directions of the method are also briefly discussed.

2

Theoretical Background

In this chapter, earlier works representing the state of the art in the field of product-process synthesis and design are discussed. The first section presents three main classes of methods for process synthesis and design: the methods that employ heuristics or are knowledge based; the methods that employ mathematical or optimisation techniques; and the methods that employ a combination of physical insights and mathematical programming. Special emphasis is given to methods that employ physical insights for the synthesis and design of distillation column based separation sequences, the attainable region analysis for reactor network synthesis and design, and, the synthesis of separation processes in general. The second section presents the Computer Aided Molecular Design (CAMD) framewok for the synthesis of molecular structures having desired target properties. Finally the last section discusses the strengths and the weaknesses of these methods and the reasons of the current work.

2.1 Process Synthesis and Design

The proces flowsheet synthesis and design problem is described by Hostrup[26] as:

> Given the feed and product specifications in the process, determine a flowsheet including the required tasks, appropriate equipments and solvents needed, as illustrated on figure 2.1. The flowsheet must be capable of converting input (feed streams) to output (product streams). Furthermore, determine the design of the equipments in the flowsheet and the appropriate conditions of operation. Finally, the identified solution must be analysed for verification.

In chemical process synthesis two types of approaches exist, in the first type, one seeks to improve an existing process flowsheet (also known as the retrofit problem), while in the the second type, one seeks to find a completely new process flowsheet. The methods to address those two approaches can be classified into three categories.

In order to provide a framework for process synthesis and design, process reactions, separation systems and heat/mass integration are either considered

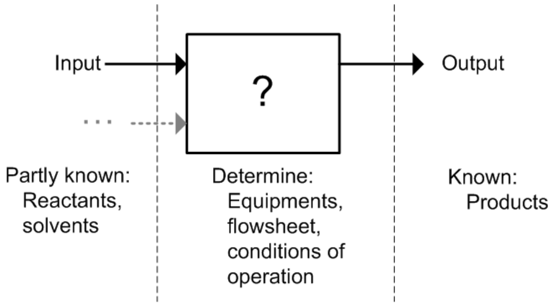


Figure 2.1: Definition of the process synthesis and design problem.

sequentially or simultaneously within the same framework. The process synthesis and design framework is based on three main classes of methods: the methods that employ heuristics or are knowledge based; the methods that employ mathematical or optimisation techniques; and, the hybrid methods combining different approaches into one method.

2.1.1 Heuristics or knowledge based methods

The methods that employ heuristics or are knowledge based rely on a set of rules based on a combination of experience, insights and available knowledge (data)[9]. Jaksland[27] show that these rules are often simple but may be contradictory. The rules need careful consideration before application as the context in which they can be applied is not necessarily fully defined. The heuristic methods can be used both for retrofit or completely new designs.

Knowledge based methods are structured around three models. First the data model, it includes all the available knowledge captured in a structured framework. Second the data mining model, it includes the procedures and rules to extract the knowledge from the data model to be applied to the current problem. Third and last, the application model, it includes the rules and methods to apply the knowledge extracted through the data mining model to the current problem. These three models[35] mimic the human approach in solving problems, where the human tends to access already available knowledge by searching for relevant information from which useful knowledge is extracted and adapted to solve the current problem.

Knowledge based methods are usually composed of sets of heuristic rules, some of which may be contradictory. For example, the rule stating that "perform the easiest separation first" and the rule stating that "remove the most plentiful component first" are contradictory when the same chemical statisfies both the rules[27]. The knowledge can also be composed of libraries of known operation tasks or processes to be used for particular purposes. These processes can involve large sets of unit operations, like an air separation unit, or smaller like a pressure swing module to separate an azeotropic mixture, and they are described as *islands* by Siirola[42]. The selection of the critical operations is made from the libraries, where each critical operation is representing an *island*. These *islands* are then interconnected to obtain the final process flowsheet. This method can be related to the case-based reasoning (CBR) methods[38]. CBR solves new problems by finding and adapting existing sucessful designs for solving new problems. Avramenko and Kraslawski[1] have designed a CBR descision supporting system for the pre-selection of the column internals in reactive separation processes. Similarities between the analysed system and available reference cases are determined to select the appropriate column internals. The similarities are based on chemical reaction description (reactants, products, etc.), process parameters (mass of catalyst, feed flow rate, etc.), catalyst properties and packing features.

The knowledge based methods can be applied in virtually all the domains of engineering, for example the automatic road extraction from aerial images[45], as long as the corresponding data models, data mining models and application models could be defined.

2.1.2 Mathematical or optimisation techniques methods

The methods that employ mathematical or optimisation techniques are based on mathematical programing techniques using optimisation to determine the best process flowsheet. These methods usually involve an important preliminary step, which is to setup a mathematical superstructure that represents all possible alternatives. Therefore, identification of the optimal process flowsheet depends very much on the definition of the mathematical superstructure. These methods require efficient numerical solvers and a good knowledge of the mathematical programming techniques. As the superstructure includes all possible inter-connections between the different unit operations and the decision variables plus the structural design parameters (like the number of plates in a column), the resulting formulation often leads to a mixed integer non-linear programming (MINLP) problem. The MINLP problem as described by Grossmann et al.[21] involved discrete linear variables (y) and continuous non linear variables (x):

$$\begin{array}{l} \min\ Z = c^T y + f(x) \\ \text{s.t.} \\ \ h(x) = 0 \\ \ By + g(x) \leq 0 \\ \ x \in X, y \in \{0,1\}^m \end{array}$$

The setup of the mathematical superstructure is often left to the experience of the user[4] or can still be an open problem e.g. for the separation of azeotropic systems of more than three components[21]. However, it is essential that the formulation of the superstructure is as comprehensive as possible, otherwise, the real optimal process may not be found as it is not available in the search space. The determination of the mathematical superstructure has been addressed by

Friedler et al.[13] with a graph-theoretic approach to express as a process graph the maximal superstructure of the synthesis problem. Shah and Kokossis[40] proposed a task based method, where tasks are representing simple to complex distillation column configurations, to generate the mathematical superstructure and the corresponding MINLP problem formulation.

If the conventional superstructure is composed of unit operations[4, 40, 13, 41], Papalexandri and Pistikopoulos[34] have proposed a superstructure composed of mass/heat exchange modules instead of conventional unit operations. As any process operation could be represented as a combination of mass and heat transfer, including the mass transfer from one specie to another in the case of reaction, the optimal process flowsheet can be represented as a network of mass and heat transfer. From this network, conventional unit operations could be identified. For example, a one feed two product distillation column could be represented with four modules, two mass/heat exchange modules representing the stripper and the rectifier sections of the column and two heat exchange modules representing the reboiler and the condenser of the column. From this approach, new types of unit operations can possibly be identified.

2.1.3 Hybrid methods

Hybrid methods combine different approaches into one. These methods combines the physical insights of the knowledge based methods with mathematical programming techniques to formulate and solve a collection of mathematical problems within an identified narrow search space. These keep the simple structure of knowledge-based approaches but replace the fixed rules with guidelines based on physical insights generated through analysis of the behaviour of the chemicals.

In this section, emphasis is on the thermodynamic insights method for the synthesis of separation processes, the driving force for synthesis and design of separation processes with distillation columns and the attainable region analysis for the synthesis and design for reactor networks.

Thermodynamic insights based flowsheet synthesis

Jaksland[27] has developed a method for the synthesis of separation processes based on thermodynamic insights. This method approaches the retrieval of the knowledge, based on the physical properties of the mixtures involved in the problem. The calculations of the indicators for each mixture provide the user with a guidance to find the matching separation task for each mixture. These indicators are ratios of physical properties, for example, a high difference in relative volatility is an indication that a separation by distillation is a feasible separation method. The method consists of two levels, each of them divided into six steps.

The first level calculates differences in pure component properties for each binary pair of components that can be generated from the components in the mixture. These property differences are the binary property ratio indicators. For each of the binary pairs the method identifies the largest property ratio, thereby also identifies the most feasible separation technique for the pair. After a screening step, the largest ratio which remains after screening is identified to indicate the first separation task to be performed and the corresponding separation technique for the flowsheet.

In the second level both pure component and mixture properties are considered in order to simultaneously sequence and select the separation tasks and techniques. By including mixture properties the second level is able to consider, more easily, separation techniques requiring mass separating agents and to give improved estimates for condition of operation. At the end of the second level a physically feasible separation flowsheet is produced together with feasible alternatives for each separation task[26].

Driving force based distillation columns design

Based on the definition of the driving force, as the difference in composition between two co-existing phases, methods exploiting the differences in the driving forces have been developed by Bek-Pedersen and Gani[3] for the synthesis and design of distillation based separation processes.

The size and the relative location of the maximum driving force is used as a design parameter for the definition of the number of plates in a column and the position of the feed location. The maximum driving force can be used not only for the design of simple distillation columns but also for thermally coupled columns[2] with multiple side draws.

By targeting the separation operation at the largest driving force, it is claimed that the near optimal separation sequence can be obtained without performing any rigorous mass or energy balance calculations.

For example, the separation sequence of a mixture of 5 components consisting of Propane, Isobutane, n-Butane, Isopentane, and n-Pentane into 5 near pure streams can be determined from the driving forces between two consecutive components (see figure 2.2). Four distillation columns are needed to achieve the desired separation. By targeting the largest driving force first, the first separation task is between Propane and Isobutane, then the n-Butane/Isopentane separation task is performed, followed by Isobutane/n-Butane and finally Isopentane/n-Pentane. The relative position of the maximum driving force between the two key components in the separation is also providing the relative position of the feed plate location in the distillation column. For example, the feed plate location of the first distillation column in the sequence performing the Propane/Isobutane separation would be at a relative position of 40% of the total number of plates (counting from the bottom of the column). 40% corresponds to the position of the maximum driving force. The total number of plates and the reflux ratio can also be obtained from a table of precalculated data based on the maximum driving force and the requested purity of the column products.

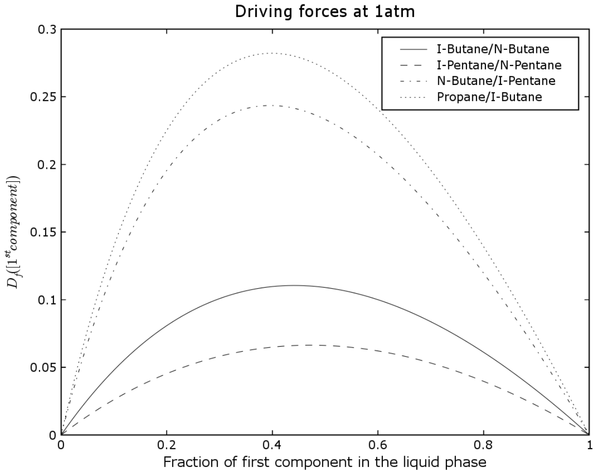


Figure 2.2: Driving forces between 5 alcanes.

Attainable region for reactor networks

Introduced by Horn[25], the Attainable Region Method for reactor network design looks at the fundamental processes occurring within the processing units. From this analysis appropriate reactor networks are identified. Fundamental processes include such phenomena as reaction, mass transfer, heat transfer and bulk mixing. The attainable region is the set of all possible outcomes that can be obtained using the allowable fundamental processes, subject to specified constraints, for a system with specified inputs.

The only element required to describe the attainable region (AR) is its boundary. The objective function is usually optimised somewhere on the boundary of the AR, as the boundary represents extremes in operation. Once the attainable region has been found the resulting boundary must be interpreted. The various surfaces that comprise the boundary represent various processing units. The arrangement of these surfaces in the boundary determine the layout of the processing units within the process. The point in the AR where the objective function is optimised then gives the optimum process specifications. For example, on figure 2.3, the attainable region, in the space of concentration of A and C, has been determined for the Trambouze reaction scheme[44]. It shows that the boundary of the AR is defined by a "bypassed" CSTR reactor between point O and A, followed by a plug flow reactor. More details on the determination of this attainable region are provided in section 4.3.4 (page 70).

The important results are that once the boundary of the attainable reagion has been constructed and interpreted both the best layout and the best

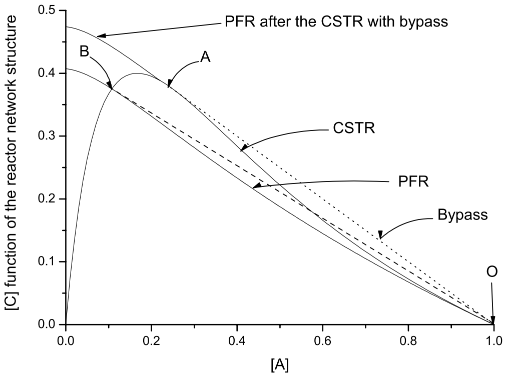


Figure 2.3: Example of an attainable region for the Trambouze reaction scheme.

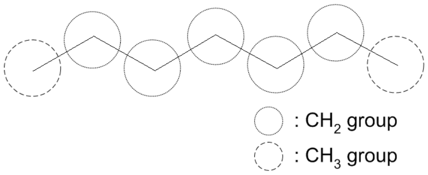
operating conditions with respect to the objective function can be determined.

2.2 Computer Aided Molecular Design

The fundamental objective of Computer Aided Molecular Design (CAMD) is to identify a component or collection of components having specific properties. The molecules are represented using descriptors (graph representing the atoms and bonds, set of molecular groups) and the properties are evaluated using methods capable of predicting properties based on the descriptors used to represent the molecules. The success of the CAMD techniques depends on the ability to rapidly predict the necessary pure component and mixture properties. As the group contribution approach[12] provides an efficient way to predict component properties, it is at the basis of most of the CAMD techniques.

2.2.1 Group contribution methods for pure component property prediction

The property estimation methods based on the group contribution approach express the properties of a chemical component in terms of functions of number of occurrences of predefined fragments (groups) in the molecule[22]. Numerous methods have been developed over the years to predict a large number of properties, from the melting point temperature[33] to the enthalpy of fusion[29]. For example, on figure 2.4, the normal-Heptane is described as composed of two types of molecular groups, 5 -CH2- groups (molecular groups with two free-attachments) and 2 -CH3 groups (molecular groups with one free-attachment). From the "a priori" regressed contributions of the molecular process groups and the Constantinou and Gani melting point temperature property model given in



hbtp

Figure 2.4: Group description of the Heptane using the Constantinou and Gani molecular group description.

equation (2.1), with n the number of different molecular groups, a_k the contribution of each type of groups and n_k the number of each type of groups. The "melting point temperature" of the normal-Heptane is predicted (see equation (2.2)), with $T_m = 175.55\ K$. This can be compared to the experimentally measured value of $T_{m_{exp}} = 182.55\ K$.

$$T_m = 102.425 \cdot ln \left(\sum_{k=1}^{n} n_k \cdot a_k \right) \tag{2.1}$$

$$T_m = 102.425 \cdot ln \left(5 \cdot 0.9246 + 2 \cdot 0.464 \right) \tag{2.2}$$

One of the issues encountered with the group contribution approaches is the difficulty to predict the difference in properties of isomers. This has been alleviated by Constantinou and Gani[5] by introducing the concept of second order molecular groups assembled from the basic set of groups (the first order groups). Later, this has been further extended to the third order by Marrero and Gani[33].

The development of a group contribution method for pure component property prediction, once the model and the molecular groups have been defined, requires the need of a large number of experimental data to regress the contributions to the model of the molecular groups. The availability of the experimental data is limiting the number of the molecular groups. In some cases, a method cannot be used to predict the property of a molecule as the molecule cannot be described with the available molecular groups. For example, the molecule in figure 2.5 cannot be described with the Constantinou and Gani molecular groups, as the acyclic Carbone-Oxygen group is not available. To overcome this limitation, Gani et al.[15] propose a correlation to determine the contribution of the missing groups.

2.2.2 The CAMD algorithms

Different approaches have been proposed for solving CAMD problems. Harper[22] group these approaches into three categories: mathematical or optimisation technique methods, stochastic optimisation methods and enumeration technique methods. In essence the mathematical or optimisation technique methods

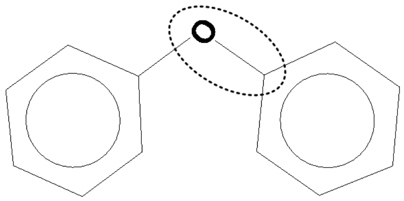


Figure 2.5: Molecule not described with the Canstantinou and Gani groups.

and the stochastic optimisation methods use the same approach as described for the process synthesis and design (see previous section 2.1.2) but applied to molecules. The mathematical programming has been used, for example, by Duvedi and Achenie[10] for the design of environmentally safe refrigerants, and the stochastic approach by Venkatasubramanian[47] for polymer design.

In the enumeration technique methods, also known as "Generate and Test" methods, first all the molecules are generated, and then, their properties are estimated and tested against the design targets. The synthesis of the molecules is based on the molecular groups and a set of rules to combine the groups to represent chemically feasible components.

Using enumeration technique based CAMD method, it is possible to generate very efficiently a large number of molecular structure alternatives[22] and the alternatives can easily be tested against the target properties using the group contribution based molecular property models. For example the complex molecule on figure 2.6 (which is one of the 20000 generated alternatives) has been generated with the following problem formulation[7]:

- generate aromatic compounds;
- generate components with double bonds;
- use a maximum of 30 groups and a maximum of 2 functional groups;
- minimum molecular weight 300 g/mol;
- minimum normal boiling point 400 K;
- minimum normal melting point 300 K.

This problem has been solved using ProCAMD a CAMD software implementation by Harper[22]. The generation of the nearly 20000 alternatives and their screening took only 0.8s (see table 2.1). It means that the generation and the analysis of one alternative took less than 1/10000s. With a different set of constraints it has been possible to generate more than 5 million alternatives in about 350s, which shows that the method scales very well.

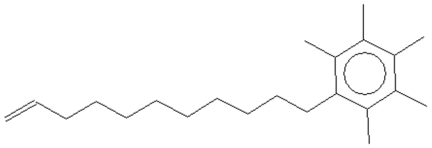


Figure 2.6: One of the large molecules synthesised using the "Generate & Test" method.

Number of compounds designed	21274
Total time used to design	0.80 s
'Screened Out' Statistics for Primary Calculations	
Normal Melting point	5574 of 21274
Molecular weight	699 of 15700

Table 2.1: Results details of CAMD problem.

2.3 Discussion

From the presented process synthesis and design methods, both for flowsheet and molecular structure, one common pattern can be found. This *synthesis pattern* is composed of three concepts, the synthesis building block, the synthesis building block connections and the model for property evaluation. In the case of CAMD, the building blocks are molecular groups, whereas for process flowsheet synthesis the building blocks could be distillation columns with their associated driving forces, mass or heat exchange modules[34], process *islands*[42] or unit operations. The building block connections are bonds for molecular groups, and, material and/or energy flow at the process level. Finally the model is evaluating a set of properties at the level of the combined building blocks, such as, the flowsheet or molecular structure. The model is a group contribution based model for molecular property in the case of CAMD or a first principle based model for the evaluation of the performance of the flowsheet.

The knowledge based and the optimisation based methods for process synthesis and design together with the thermodynamic insights method by Jaksland have in common that they need to model and simulate the process for each set of the decision variables. This is time consuming, difficult, and sometimes almost impossible to solve. If the models are oversimplified to minimize these problems, the method can finally give infeasible alternatives because of inconsistency between the simplified and the rigorous models.

Two methods for process synthesis and design are not affected by this issue: the driving force approach for distillation columns and the attainable region for reactor network synthesis and design. Using the driving force approach, first the near optimal separation sequence is determined, then each distillation column is designed. In a similar way, the attainable region analysis provides feasible alternatives, and then from the desired product specifications the optimal reactor network is chosen from the alternatives providing also the corresponding design parameters. These two methods are decoupling first the synthesis of the process flowsheet alternatives and then the design of the optimal or near optimal alternative for the desired product specifications.

For product synthesis, the CAMD method is also decoupling the synthesis of the molecular structure alternatives and their evaluation with respect to the target properties. This provides the possibility to generate a large number of molecular structure alternatives and a fast evaluation of the property targets, while having the precision and the range of application of a group contribution method.

If we could determine at the process level the three elements of the *synthesis pattern*: the building blocks; the building blocks connections and connectivity rules and a model to evaluate the performance of a given combination of building blocks, while being able to apply the CAMD techniques to those elements, it would mean that the power of CAMD could be transposed at the process flowsheet level. Building blocks could be process groups, representing a unit operation or a set of unit operation, process groups could be combined respecting connectivity rules enforcing the feasibility of the resulting flowsheet structure and a process group contribution based flowsheet property model could be used to evaluate the performance of the flowsheet structure. Finally, at a second stage, a method would be needed to determine the design parameters of the unit operations represented by the process groups of the flowsheet structure.

3

Objectives and Overview of the CAFD Framework

In this chapter, the objectives and an overview of the new framework for CAFD (Computer Aided Flowsheet Design) are presented.

From the theoretical background section, it was possible to find a common pattern among the currently available methods for process-product synthesis. This *synthesis pattern* is composed of three concepts: the synthesis building block, the synthesis building block connections and the model for property estimation. The building block is, for example, an atom, a molecular group (a set of atoms), a mass or heat transfer module[34], a process *island*[42] or a unit-operation. The building block connections can be based on bonds for atoms and molecular groups, or at the process level the connections can be based on component identities, stream composition and/or state. Finally the model is estimating a set of properties of the structures composed of a combination of building blocks (a process flowsheet, a molecular structure). This model can be based on quantum mechanics models for molecular properties, data driven models for processes or semi-empirical models for processes and/or molecules and their mixtures.

Based on this synthesis pattern, available both at the molecular level and the process level (see figure 3.1), a new synthesis framework, and its associated algorithms, methods and tools, is proposed. This framework is in particular similar to CAMD[22] and represents a successful transfer of algorithms/methods/tools from one area to another. That is, from the synthesis of chemicals to chemical process flowsheets.

3.1 Objectives of the CAFD Framework

In this section, the objectives of the new CAFD framework are stated together with a brief discussion on their scope and significance. The main objectives are to have:

O1 – A method that is applicable to a large range of problems
It is possible to find a large set of methods to solve specialized problems, but these methods are usually efficient on small, well defined problems.

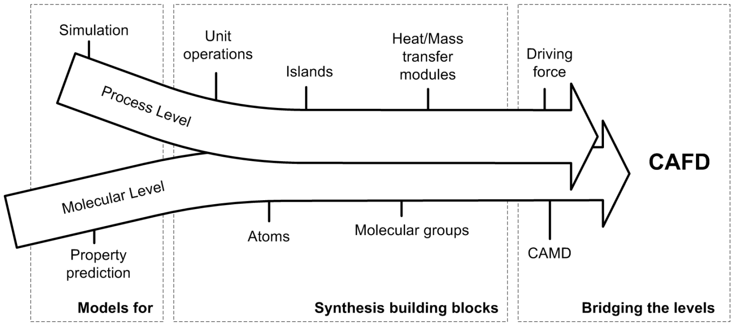


Figure 3.1: One path in the synthesis patterns

Therefore, being able to solve a large range of problems by the same method will also allow the CAFD approach to address the synthesis of complex processes involving various unit-operations.

O2 – A method that does not need to employ rigorous models at each decision step
By not having to employ rigorous models at each decision step, the generation of alternatives is faster without sacrificing the precision of the model based decisions. This will also lead to a decrease in the computational and human time for problem resolution and/or setup.

O3 – A method that supports the retrofit of existing processes and the generation of new alternatives
If completely new processes are created, nowadays, very often, existing processes are modified to satisfy new regulations or supply demands.

O4 – A method that has the ability to build and reuse knowledge over time
Knowledge is a process that evolves with its manipulation and is context dependent. If the method can propose a way to not only keep the available knowledge but also update it, the knowledge will stay accurate, consistent and enriched over a large time horizon.

O5 – A method that is easy to setup and use
A method, to be useful and also to be tested thoroughly, needs to be easy to setup so that it can effectively be used and allow constant refinement and improvements to be performed.

3.2 Definitions

First some important terms[6] frequently employed in this thesis are clarified.

Process group – A process group (PG) is the representation of a unit operation or a set of unit operations. It has at least one inlet stream or one outlet stream as connections.

Structure or flowsheet structure – A structure or a flowsheet structure is a representation of a process as an ensemble of process groups and their connections.

Flowsheet or process – A flowsheet or process provides details of the flowsheet structure and its corresponding design/operations parameters (such as, stream composition, flowrate, pressure, temperature) so that all the needed information to fully describe the process is available.

Flowsheet property – A flowsheet property is any property that provides a measure of the performance of the operations in the flowsheet and, can be expressed as a function of the contributions of each process group (unit operation or a set of unit operations of the process).

Process level retrofit – Generating and screening of design/operation alternatives for a given process considering parts of the process as fixed.

Unit operation retrofit – Generating and screening of design/operation alternatives for a given unit operation considering parts of the unit operation as fixed.

3.3 Overview of the CAFD Framework

A group-contribution (GC) based pure component property estimation of a molecule requires knowledge of the molecular chemical structure and the groups needed to uniquely represent it. The needed property is estimated from the "a priori" regressed contributions of the groups representing the molecule. If mixture properties are needed, in addition to the group contributions, the interactions between the groups contribute to the estimation of the needed property. Having the groups, their contributions and their interactions together with a set of rules to combine the groups into a molecule, allows us to synthesis/design molecules and/or mixtures. This is CAMD, computer aided molecular design (section 2.2 page 9).

Let us now imagine that each group used to represent a fraction of a molecule could also be used to represent a chemical process operation or a set of operations in a chemical process flowsheet. A process group would represent a unit operation or a set of unit operations. In the same way as the CAMD method applies connectivity rules to combine molecular groups to form chemically feasible molecular structures, process groups would have connectivity rules to combine process groups forming feasible flowsheet structures. Finally, with a process group contribution based flowsheet "property" model and the corresponding process group contributions it would be possible to predict the "property" of the flowsheet, "property" that could be an indicator of the flowsheet performance.

Having a set of process groups that represent a unit operation or a set of unit operations, and their individual contributions for a set of properties, then together with their combination rules we can perform simultaneous modeling/simulation, synthesis and design of process flowsheets. That is, exactly in the same way as group contribution based property estimation and molecular design GC-based flowsheet property estimation and flowsheet design can be carried out. If the flowsheet needs to be optimized based on maximizing or minimizing the estimated flowsheet "property", it would be quite easy to determine the optimal flowsheet by determining a set of groups that represent the flowsheet and have the desired (target) properties.

This CAFD framework, based on the group contribution method concept, is following the *synthesis patterns* with flowsheet property models, process group building blocks and process group connectivity rules.

The CAFD framework, as shown on figure 3.2 page 19 is composed of eight main steps: the definition of the problem; the analysis of the problem; the selection of the matching process group building blocks; the synthesis and test of the alternatives; the ranking of the alternatives and selection of the most promising alternatives; the design of the selected alternatives; the post analysis of the designed alternatives and the final verifications.

3.3.1 Definition of the synthesis problem

If creating a completely new process, the synthesis problem is difficult to define as hardly any information is known. Thus, it is possible to only setup the starting point of the synthesis problem. Consequently, only the terminal process groups, representing the inlet and outlet streams of the desired process, are known in the flowsheet structure that needs to be determined. On the other hand, in the case of a retrofit problem, the starting point is a backbone of already interconnected unit operations together with the available raw materials inlets and desired product outlets. Note that the retrofit of a given unit operation can also be performed during the reverse simulation step (see section 3.3.6).

3.3.2 Analysis of the synthesis problem

From the definition of the synthesis problem hardly any information is known; only the inlet and the outlet streams of the desired process are available. Analysis of the process synthesis problem is performed to further define the problem through the use of knowledge bases and physical insight methods. The analysis of the mixtures available from the specified problem provides the list of reaction and separation tasks that could be involved in the desired process.

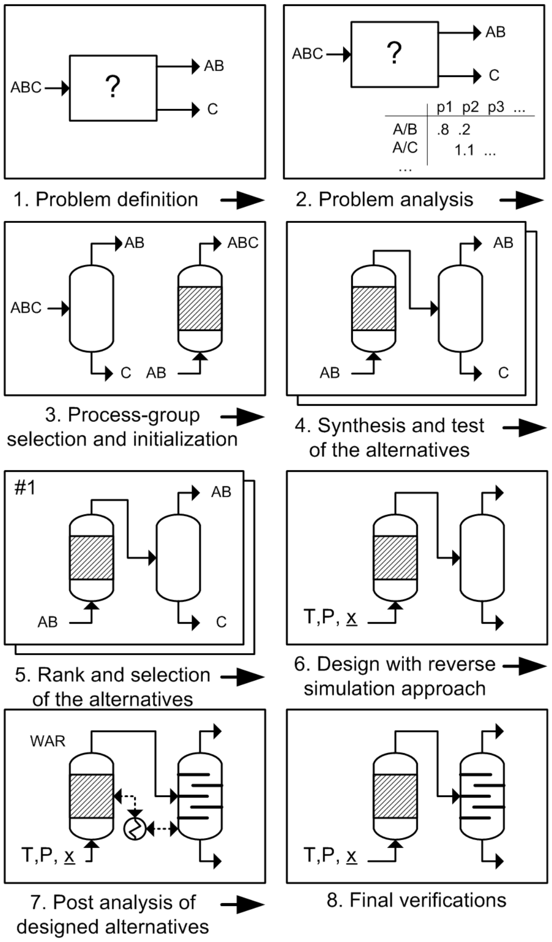


Figure 3.2: Overview of the CAFD framework

3.3.3 Process group selection and initialization

A process group, representing a unit operation or a set of unit operations, can be used with any mixture of components as long as the properties of the mixture are matched with the needed properties of the process group. For example, a distillation process group representing the separation of two components A and B can be used for any binary mixture as long as the driving force between the two components are matched with the driving force of the distillation process group.

As a process group can be used with several mixtures having matched property dependence and as the connectivity rules to combine process groups to form flowsheet structures are based on the component identities, a process group is initialized with the mixtures having matched properties. Thus, one uninitialized process group can lead to several initialized process groups, one for each set of components with matched properties. When a process group is initialized, it can be directly used in the synthesis and test of the alternatives step.

3.3.4 Synthesis and test of the alternatives

From the synthesis problem definition, the selected and initialized process groups, and, the connectivity rules, all the possible feasible alternatives can be generated and evaluated. The basis is the generate and test approach[23] for CAMD but applied now to flowsheet structures (or CAFD).

From the synthesis problem definition, the structural definition of the desired process is set. From the process group selection and initialization the list of process groups to be used in the generation of the alternative is available. It is thus possible to generate through the process group connectivity rules all the feasible flowsheet structure alternatives and test them against the structural definition.

3.3.5 Ranking and selection of the alternatives

After the synthesis of the alternatives, a ranking or benchmarking of the feasible alternatives is performed to select the most promising alternatives. To perform the screening of the alternatives to select the most promising candidates, an objective function based on the flowsheet property models is required. The objective function in this case can be a single property optimization or a minimization of a weighted sum of objectives. If the number of target properties is large and if each target property can hardly be related to another, like maximum conversion in the reactive part and minimum energy cost in the downstream separation, a non-dominated Pareto front[8] optimization method may be used.

3.3.6 Reverse simulation for process design

After the generation of the flowsheet structure alternatives, only the process groups and their connections are known. This information, as such, is only providing a block representation of the flowsheet and not the complete process model. The reverse simulation is a method to calculate from the process groups (which supply target values of the process variables) the design parameters of the process instead of calculating the process variables for assumed values of the design parameters.

For example, in the case of a process group representing a separation by distillation of a binary mixture, the process group is providing the target values of the purity of the components and the state of outlet streams at the top and the bottom of the distillation column. From these target values, the design parameters of the column are calculated. These design parameters include the number of stages, the feed stage location and the reflux ratio.

3.3.7 Post analysis

As the reverse simulation step is providing all the design parameters of the process together with the state description of the flows (pressure, temperature and individual flowrates) in all the units and streams, it is possible to perform the traditional post-analysis of the process before any rigorous simulation.

Two possible analysis are, for example, the heat integration using a pinch analysis method[31] and/or the environmental impact using the war algorithm[50]. Such post analysis step can help differentiate the most promising alternatives.

3.3.8 Final verifications

The final verification of the obtained process alternatives is a task performed by the engineer using his own experience. If a rigorous simulation is needed using a conventional process simulator, the complete set of design parameters and the initialization of the simulation is given by the reverse simulation approach reducing the time to setup the rigorous simulation.

4

CAFD Models and Methods

In this chapter, the details of the models and methods developed for the framework for Computer Aided Flowsheet synthesis and Design (CAFD) are presented. This chapter is composed of three main sections: the group contribution model for flowsheet property estimation, the reverse flowsheet property prediction method for flowsheet synthesis and the reverse simulation method for process design.

The CAFD method is composed of two reverse approaches, the reverse property prediction approach and the reverse simulation approach. The goals of those reverse approaches are respectively to generate alternative flowsheet structures matching target properties and to determine all the design parameters of each of the generated flowsheet structures.

4.1 Group Contribution for Flowsheet Properties

A similar approach as the group contribution for property prediction of molecules (see section 2.2.1 page 9 in the *Theoretical Background* chapter) has been taken to develop a group contribution method for flowsheet property prediction.

In this section the process group concept, the SFILES notation (Simplified Flowsheet Input Line Entry System) for representation of the flowsheet structure and the process group contribution based models for flowsheet property estimation are presented.

4.1.1 The process group

In this section, the process group concept and the method to represent a process flowsheet with process groups are presented.

Presentation of the process group concept

The process group concept is inspired by the molecular functional group concept. The molecular functional group concept is to represent a molecule with functional groups, each group representing an atom or a set of atoms and their connections. As shown on figure 4.1, the Bitertanol molecule, a fungicide, can be represented with 23 molecular groups using the first order functional groups of the Marrero and Gani group contribution method[33].

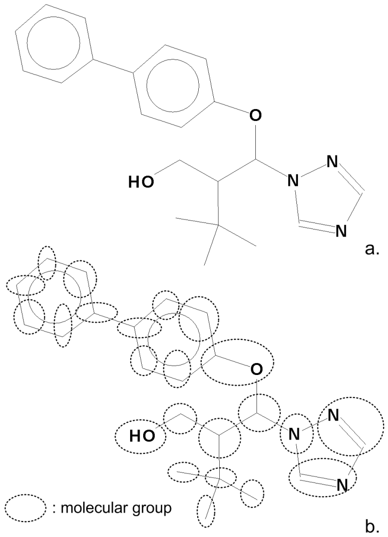


Figure 4.1: Representation of the Bitertanol fungicide (a.) with molecular groups (b.)

From the functional group description of the molecule, the corresponding group contribution model for a property (for example, the pure component normal boiling point temperature), and the contributions of each functional group to that property, it is possible to predict the required property value (for example $T_b = 779K$ for Bitertanol).

Let us now imagine that each functional group used to represent a fraction of a molecule could also be used to represent an operation in a process flowsheet. In this case, each process group may have one or more attachments to be connected to other process groups and consist of one or more unit operations. In this way, a set of process groups, representing different types of process operations, may be connected to form a flowsheet structure. The flowsheet structure being the representation of a process with process groups and their connections. For example, on figure 4.2a., a simple flowsheet composed of a reactor, then a membrane-based separation, followed by a distillation column separation and then a two distillation columns separation to separate a binary azeotropic mixture (with the addition of a makeup of solvent) could be represented with process groups.

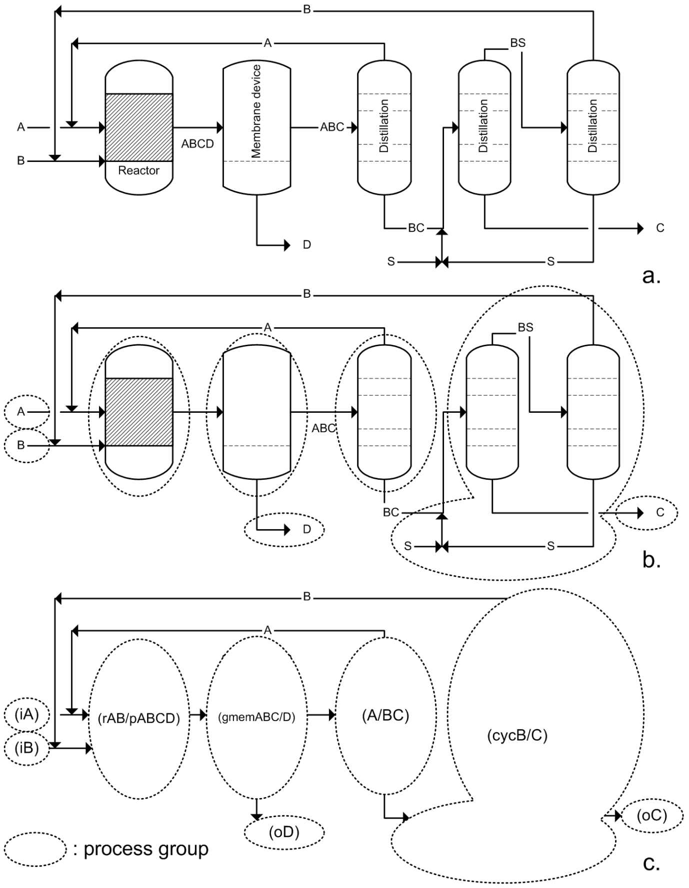


Figure 4.2: Representation of a simple flowsheet (a.) with process groups (b./c.)

Consider the process in figure 4.2a, the feed mixtures are represented by two inlet process groups (groups with one attachment), such as, `(iA)` and `(iB)`. The end-products (also groups with one attachment) are outlet process groups (a product of purity $\geq 99.5\%$), such as, `(oC)` and `(oD)`. The four intermediate process groups representing a reactor `(rAB/pABCD)`, a gas-membrane `(gmemABC/D)`, a distillation column `(A/BC)` and solvent-based separation `(cycB/C)` have at least one inlet and one oulet streams.

From the list of available process groups like `(iA)`, `(gmemABC/D)`, `(A/BC)`, etc., a feasible flowsheet structure can be created as shown in figure 4.2b-4.2c. where two components A and B are reacted to form B and C, with a downstream separation for the products and the recycle of the reactants.

As in group contribution based molecular property estimation (where the same groups may represent many molecules), the process groups are not component dependent, but component property dependent. Thus, the ability to use the same process group with different components having similar properties exist also in this case. Note, however, that the inlet and outlet streams (links) of process groups maintain a list of components present in them and that the path of a component through a process group defines the rules for combining process groups into a flowsheet. That is, process groups `(A/BC)` and `(B/C)` can be connected to form `[-(A/BC)-(B/C)-]` without knowing the identities of components A, B and C. Only when the properties of the flowsheet needs to be calculated, the identities of the chemicals (components) are needed.

Currently twelve types of process groups are available. These represent simple distillation column, solvent based azeotropic distillation, flash separation, kinetic-model based reactor, fixed conversion reactor, pressure swing distillation, polar molecular sieve based separation, molecular sieve based separation, liquid membrane based separation, gas membrane based separation, crystallization and adsorption.

Operation	**Process group example**
Distillation column	`(A/BC)`, `(ABC/DE)`
Solvent based azeotropic distillation	`(cycA/B)`
Flash separation	`(fABC/BCD)`
Kinetic-model based reactor	`(rABC/nE/pABCD)`
Fixed conversion reactor	`(rABC/nE/pABCD)`
Pressure swing distillation	`(swA/B)`
Polar molecular sieve based separation	`(pmsABC/D)`
Molecular sieve based separation	`(msABC/D)`
Liquid membrane based separation	`(lmemABC/D)`
Gas membrane based separation	`(gmemABC/D)`
Crystallization	`(crsABC/D)`
Absorption	`(abEAB/eF/EABF/EF)`

Table 4.1: Available process groups.

Method to represent a process flowsheet with process groups

In this section, a simple method to represent a process flowsheet with process groups is presented. The prerequisites are to have both the process flowsheet description and a database of available process groups. The main steps of the method are highlighted on figure 4.3 through a flow diagram.

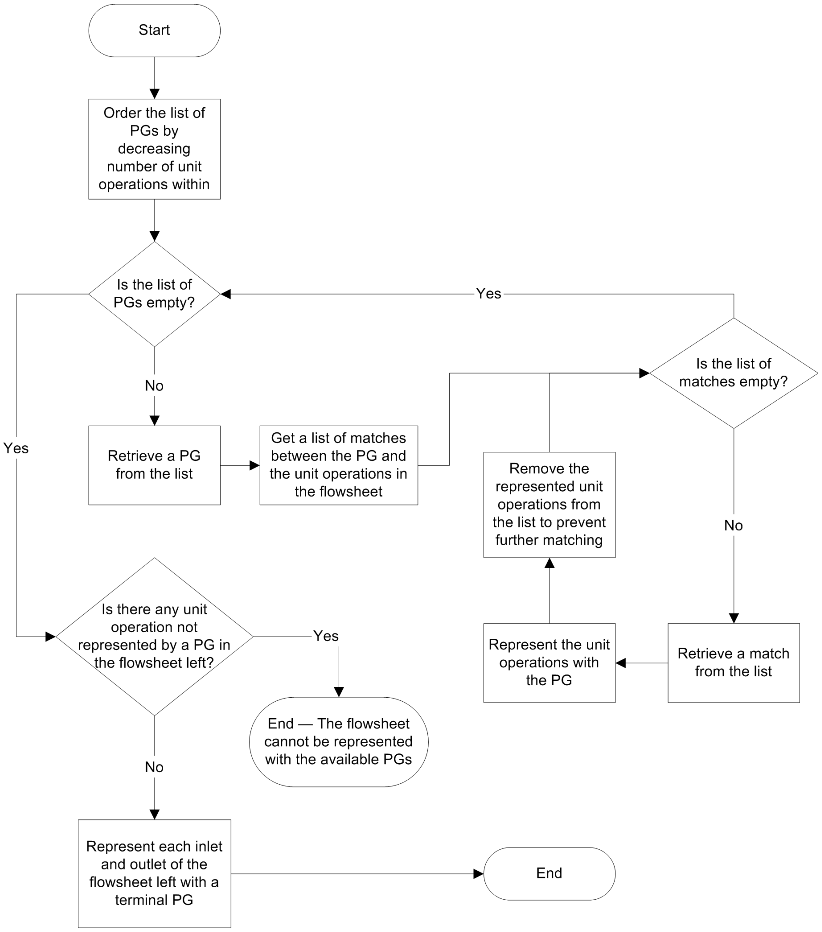


Figure 4.3: Method to represent a process flowsheet with process groups

The method for process flowsheet representation by process groups will now be applied on the process flowsheet of figure 4.2a. We start with the list of process groups given in table 4.1 and the flowsheet of figure 4.2a. According to the method (see figure 4.3), the first identified process group is the solvent based separation (`cycB/C`) as this process group is representing a set of two unit operations. Then the reactor process group (`rAB/pABCD`), the membrane process group (`gmemABC/D`) and the distillation process group (`A/BC`) are identified. At this point, all the unit operations are represented with process groups. The inlets and the outlets of the process are represented with terminal process groups, the inlets `(iA)` and `(iB)`, and, the outlets `(oC)` and `(oD)`. Note that the uptake of solvent is not represented as an inlet as it is part of the solvent based separation process group.

4.1.2 The SFILES notation of a flowsheet structure

Having a process flowsheet represented by process groups provides the possibility to employ simple notation systems for efficient transfer/storage of process group-information for the corresponding process-flowsheet. This section presents a method to uniquely describe a flowsheet structure in the form of a string. The method is based on the work by David Weininger on the SMILES[48, 49] (Simplified Molecular Input Line Entry System).

First, the SFILES notation and the similarities with the SMILES notation are presented. Then the developed SFILES specification rules are given. Finally, the method to create a unique SFILES string for any flowsheet structure is presented together with a discussion of the benefits of the generation of a unique SFILES string for a flowsheet structure.

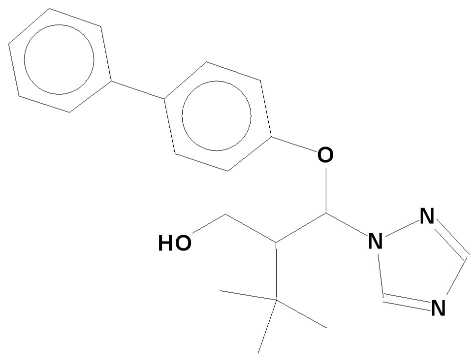
Presentation of the SFILES notation

The SFILES notation has been developed using as basis, the SMILES notation. The SMILES notation describes a molecular structure in the form of string of characters. For example, applying the SMILES notation on the Bitertanol fungicide, the following SMILES string is obtained (see figure 4.4):

`OCC(C(Oc1ccc(cc1)c2ccccc2)N3C=NC=N3)C(C)(C)C`

According to the SMILES notation, the string is read from left to right. Each letter represents an atom, with for example `C` for Carbon, `O` for Oxygen or `N` for Nitrogen. Two consecutive letters are representing a simple bonding connection between two atoms. For example, the fragment `OCC` at the start of the SMILES string represents the successive connection of an Oxygen atom to a Carbon atom, and then this Carbon atom to a second Carbon atom. An equal sign "=" signifies that the connection is double bonded connection, for example between a Carbon and a Nitrogen in the `NC=N` fragment. Branching and cycles are expressed using parenthesis and numbers. The fragment `C(C)(C)C`, at the end of the SMILES notation of the Bitertanol, represents a Carbon connected to three Carbon atoms from which two are Methyl Carbons. The cyclic structures are expressed with a number. An atom connected to another atom through a cyclic structure are both followed by the same number. For example the fragment `c2ccccc2` represent six Carbon atoms with the first connected to the last. As the atom letters are in lower cases, this also indicates that those Carbon atoms are part of an aromatic ring.

Following the same principle, applying the SFILES notation to the flowsheet structure on figure 4.5 the following string of characters is obtained:



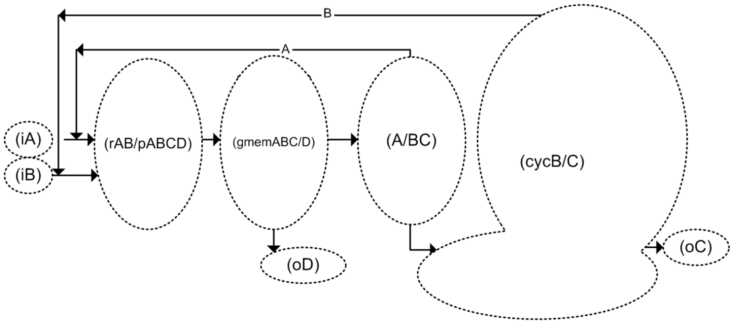
SMILES string: OCC(C(Oc1ccc(cc1)c2ccccc2)N3C=NC=N3)C(C)(C)C

Figure 4.4: SMILES string of the Bitertanol fungicide.

`(iA)(rAB/pABCD)<1<2[<(iB)](gmemABC/D)[(oD)](A/BC)1(cycB/C)2(oC)`

The SFILES string is read from left to right. The process groups are delimited by parenthesis, for example, the membrane process group `(gmemABC/D)`. Two consecutive process groups represent a connection from the first process group to the second process group. This is, for example, the case at the start of the SFILES notation with `(iA)(rAB/pABCD)` representing an inlet process group connected to a reactor process group. Branches are represented using square brackets and recycles using numbers like in the SMILES notation. The difference is that a connection between two process groups is directed. The direction, when not from the left to the right, is clarified with the "smaller than" character. This is shown with the fragment `[<(iB)]` where the process-inlet process group `(iB)` is not connected to an outlet of the reactor process group `(rAB/pABCD)` but to an inlet of this reactor process group. The recycles are indicated with numbers, one for each recycle. The number indicates the two process groups closing the recycle. For example, the number `1` indicates that one outlet of the distillation process group `(A/BC)1` is connected to the inlet of the reactor process group `(rAB/pABCD)<1`.

Having the SFILES string means that, as in the case of SMILES string, the information can be processed very easily to draw the two-dimensional process flow diagram. Also, this can become a universal way to exchange flowsheet data between flowsheet drawing tools. In the next section, the complete SFILES notation specification rules are provided.



SFILES string: (iA)(rAB/pABCD)<1<2[<(iB)](gmemABC/D)[(oD)](A/BC)1(cycB/C)2(oC)

Figure 4.5: SFILES string of a simple flowsheet structure.

The SFILES specification rules

The SFILES notation describes a flowsheet structure in the form of a string. Consequently, it is a simplification of a process flowsheet. No attempt is made to modify or add new information. A flowsheet structure is composed of process groups and connections between those process groups. The connections lead to branches expressing the different mass trains and cyclic structures expressing the recycles.

The SFILES notation for the process groups, the connections, the branching and the handling of cyclic structures are detailed below.

Rules for process groups SFILES notation. Process groups are represented using letters, the slash ”/” and start-end parenthesis characters ”(” and ”)”. The general representation of a process group is matching the following standard regular expression:

`^([a-z]*[A-Z]+(/[a-z]*[A-Z]+)*)$`

From left to right, the regular expression is explained as follows:

- `^(` – The process group starts with an opening parenthesis.

- `[a-z]*` – One optional set of lower case letters. These letters represent the type of process group, for example, `r` for a reactor or `gmem` for a gas membrane. If no type is given, the process group is set to be a distillation process group.

- `[A-Z]+` – A list of at least one upper case letters with each letter representing a component. The resulting list of component represents for example the needed reactants of a reactor process group or the overhead products of a distillation process group. The definition depends on the type of process group.

- `(/[a-z]*[A-Z]+)*` – Any number of times, a slash character `/` representing the start of another connection, a set of optional lower case letters representing the type of the connection (`[a-z]*`) and a list of at least one upper case letter (`[A-Z]+`) representing the components in the connection.
- `)$` – The process group ends with a closing parenthesis.

In the table 4.2, examples of the SFILES string of process groups are given.

SFILES string	**Description**
`(iABCD)`	An **i**nlet of 4 components A, B, C and D
`(oABC)`	An **o**utlet of 3 components A, B and C
`(ABC/D)`	A simple distillation with products of 3 components at the top and 1 at the bottom
`(rAB/pABC)`	A reactor where A and B **r**eact to form C `pABC` are the components in the reactor outlet

Table 4.2: Examples of process group syntax

Rules for describing connections. The SFILES string can be easily converted to draw a flow diagram (directed graph). The connections between the process groups are similar to the bonds between atoms in a molecular structure. The main difference is that connections are oriented, as they are material flows, whereas in a molecular structure they are mainly not oriented. In the case of the SMILES notation the directionality is not represented. For the purposes of illustration, consider 5 process groups A, B, C, D and E. They can be connected to each other using the SMILES notation to `AC[B]DE`. The corresponding non oriented graph is shown on figure 4.6, but the goal is to represent the directed graph on figure 4.7.

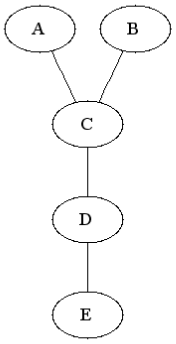


Figure 4.6: Non directed representation of a SMILES notation

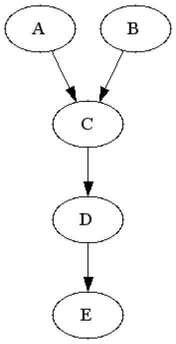


Figure 4.7: Directed representation of a SFILES notation

To have a directed graph, the connections between the process groups are oriented using the greater than ">" and smaller than "<" characters. Two process groups, A and B, connected to each other with connection coming from A and going into B would be noted as `A>B`. `A<B` would be the representation of a process group B having an outlet connection going into A.

To represent a directed graph, all the connections have to be oriented. That way, the string `AC[B]DE` becomes `A>C[<B]>D>E` which correspond to the graph on figure 4.7. As the SFILES notation is read from left to right, the simplified writing is then `AC[<B]DE`, with only the *backward* connection explicitly defined.

Rules for describing branches. The branches are represented by enclosures in square brackets. Exactly in the same way as SMILES, they can be nested or stacked, as shown for the graph on figure 4.8 for which the SFILES notation is `ABC[<J<I<K]D[<L[<K]<M]EFG`. By starting at the `H` process group, another valid SFILES notation of this graph is `HIJC[<B<A]D[<L[<K]<M]EFG`.

Rules for describing recycles. Recycles are treated in a similar way to cyclic structure in the SMILES notation (see figure 4.4 page 29 and the description page 28). The recycles are represented by breaking one connection within the recycle, and this for each recycle loop. A "smaller than" or a "greater than" character (> or <) and a digit, immediately following the process groups connected to the broken connection are used to mark the connection and its orientation. For example, the structure presented on figure 4.9 is noted `AB<1CD1E`. If considered without the orientation, it means that the process groups B and D are directly connected, but the orientation for the first digit provides also the information that the recycle loop is coming from D to B.

If a process group is present in more than one recycle, several orientation marks and digits need to be used. For example, the SFILES notation of the flowsheet on figure 4.5 (page 30) is as follows:

`(iA)(rAB/pABCD)<1<2[<(iB)](mABC/D)[<(oD)](A/BC)1(cycB/C)2(oC)`

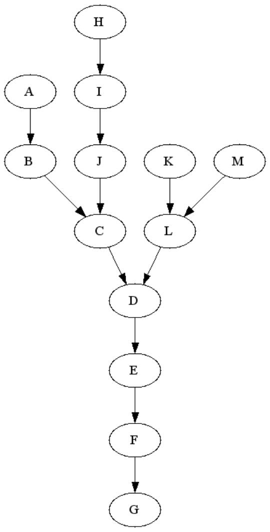


Figure 4.8: Complex branched SFILES

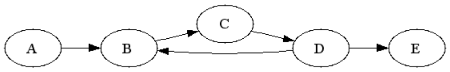


Figure 4.9: Representation of a single loop recycle SFILES

The reactor process group `(rAB/pABCD)` is part of two recycles, the recycle of A coming from the distillation process groups `(A/BC)` and the recycle of B from the solvent based separation process group `(cycB/C)`. The resulting SFILES notation for the reactor process group is `(rAB/pABCD)<1<2` with `<1` corresponding to recycle coming from the distillation `(A/BC)1` and with `<2` corresponding to recycle coming from the solvent based separation `(cycB/C)2`.

The SFILES notation can support large and complex structures like the one shown on figure 4.10. The corresponding SFILES notation of this graph is:

`OA<1E[P]F[Q]G<2MN[R]J<3K[IH3]L[S]D2CB1`

SFILES string generation algorithm

From the SFILES specification rules, it is possible to represent one flowsheet structure with two different SFILES notations. For example, the SFILES notation of the flowsheet on figure 4.5 (page 30) can be either:

`(iA)(rAB/pABCD)<1<2[<(iB)](mABC/D)[<(oD)](A/BC)1(cycB/C)2(oC)`

or

`(iB)(rAB/pABCD)<1<2[<(iA)](mABC/D)[<(oD)](A/BC)1(cycB/C)2(oC)`

Note the exchange between the two inlet process groups `(iA)` and `(iB)`.

To leverage the full benefits of the SFILES notation, it should be possible to generate automatically a *unique* SFILES string for a given flowsheet structure. In this section, the algorithm to generate the unique SFILES string from a given flowsheet structure is presented. This algorithm is based on the unique SMILES string generation algorithm[49]. The algorithm must satisfy the uniqueness of the SFILES string for a given flowsheet structure. It must also be fast to compute, while if possible to satisfy, the very subjective criteria of good aesthetic of the resulting SFILES string.

The unique SFILES string generation algorithm is composed of two steps:

Step 1. The definition of the graph invariant of the flowsheet structure. The graph invariant is a property of a graph independent of the orientation of the graph. It is used to determine the appearance order of the process groups in the SFILES notation string.

Step 2. The generation of the SFILES notation, with initial process group selection and the branching decisions based on the flowsheet structure graph invariant.

The flowsheet structure taken as an example is shown on figure 4.11. This flowsheet structure is of interest as the the two inlets `(iA)` and `(iB)` are completely symmetrical.

Step 1. The definition of the graph invariant of the flowsheet structure. The graph invariant of the flowsheet structure is a labeling of each process groups of the flowsheet structure with a unique numbers from 1 to n, with n equal to the number of process groups in the flowsheet structure. The

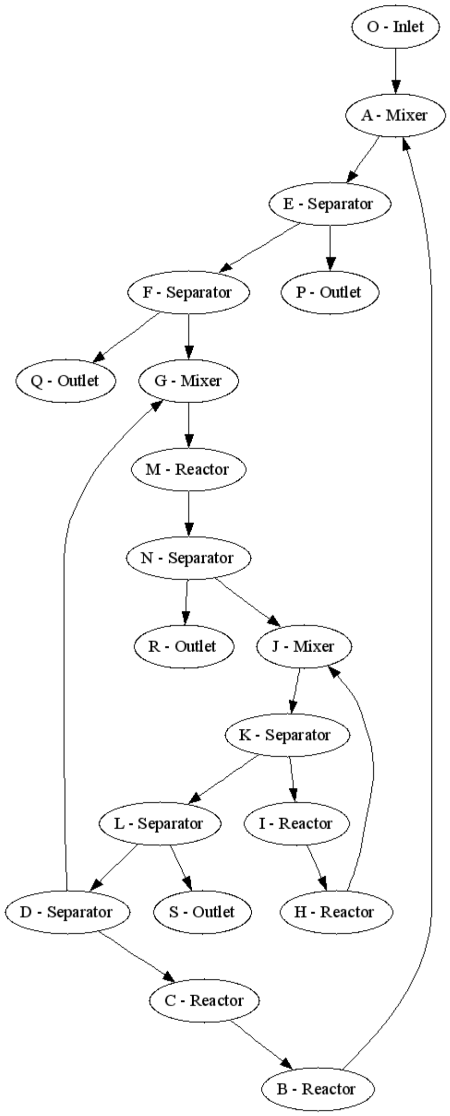


Figure 4.10: A large graph with 3 recycle loops

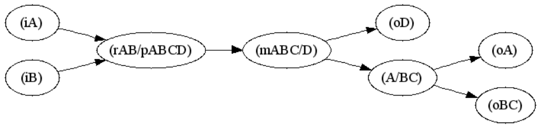


Figure 4.11: Flowsheet structure for unique SFILE generation

graph invariant is a property of a graph independent of the orientation of the graph. The procedure to generate the graph invariant of a flowsheet structure is as follows:

Step a. Generation of the initial graph invariant. The initial invariant is calculated from the four invariant rules of table 4.3. The invariant rules are associating a number to each process group, based on the nature of the process group. For the reactor process group `(rAB/pABCD)` the application of the rules gives:

Rule 1 The reactor process group is an intermediate process group, thus the number 3.

Rule 2 The reactor process group has 3 connections, 2 inlets and 1 outlet, thus the number 03. Note that this number has a leading 0.

Rule 3 The reactor process group is connected to only one process group which is not an inlet or an outlet process group, thus the number 01.

Rule 4 The SFILES notation of the process group is `(rAB/pABCD)`, which is composed of 11 characters, thus the number 11.

The concatenation of the numbers obtained for the four rules gives the initial invariant of the process group: 3030111. The initial invariants, calculated from the rules of table 4.3, for all the process groups are given in table 4.4.

Rule	Number calculated from the rule
(1)	1 (inlet), 2 (outlet), 3 (intermediate)
(2)	number of connections (01 to 99)
(3)	number of non inlet/outlet connection (00 to 99)
(4)	length of the process group name (04 to 99)

Table 4.3: Rules to generate the initial invariant of a process group.

Step b. Simplification of the graph invariant. The initial graph invariant values from the initial invariant are important for the order it provides. The invariant values are not intrinsically important but the order is. That is why the process group initial invariants:

101014-101014-3030111-3030208-2010104-3030106-2010104-2010105

corresponding to the process groups A-B-C-D-E-F-G-H can be replaced by the equivalent graph invariant:

1-1-5-6-2-4-2-3

This new set is carrying as much information as the previous one. This graph invariant is still not a valid invariant for the flowsheet structure as

Process groups	**Numbers from the rules**	**Initial invariants**	**Letter**
(iA)	1, 01, 01, 04	1010104	A
(iB)	1, 01, 01, 04	1010104	B
(rAB/pABCD)	3, 03, 01, 11	3030111	C
(mABC/D)	3, 03, 02, 08	3030208	D
(oD)	2, 01, 01, 04	2010104	F
(A/BC)	3, 03, 01, 06	3030106	E
(oA)	2, 01, 01, 04	2010104	G
(oBC)	2, 01, 01, 05	2010105	H

Table 4.4: Initial invariants for flowsheet structure on figure 4.11

the ranks are not unique, there are two PGs with rank 1 and another two PGs with rank 2. The next three steps are used to discriminate those process groups.

Step c. Rank generation from components. As the flowsheet structure graph invariant 1-1-5-6-2-4-2-3 is ranking 8 PGs with 2 pairs of PGs being at the same rank 1 and 2. These PGs are discriminated using the prime equivalent of their components.

The PGs ordered with rank 1 are `(iA)` and `(iB)`, the components are the first A and the second B. The corresponding prime numbers are 2 and 3 (See table 4.5 for a list of prime numbers). Thus the PG `(iA)` will be ordered before `(iB)`. The resulting graph invariant is:

1-2-6-7-3-5-3-4

The same procedure is applied for `(oD)` and `(oA)` of rank 3 (rank 2 before the differentiation of the PGs of rank 1). The corresponding prime numbers are 7 and 2, thus `(oA)` will be ordered before `(oD)`. The resulting graph invariant is thus:

1-2-7-8-4-6-3-5

This graph invariant is the final graph invariant (see figure 4.12) of the flowsheet structure presented in figure 4.11 (page 35) as all the PGs are numbered with a unique number between 1 and 8.

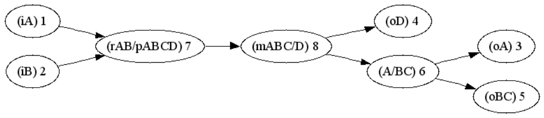


Figure 4.12: Resulting ranks for the process groups

The next steps d. and e. are performed only if the previous steps are not discriminating all the PGs in the flowsheet structure.

Step d. Rank generation from connectivity. The rank generation from connectivity is used to differentiate the nodes that are still tied. Considering the previous case without the ranking by components of step (c). The ranking is 1-1-5-6-2-4-2-3. The tied nodes can be differentiated using the rank of the neighborhood nodes. As for step (c) the equivalent prime product is used.

Considering the nodes (oD) and (oA) having both a rank of 2. The neighborhood of (oD) is (mABC/D) of rank 6 and (oA) is (A/BC) of rank 4. The corresponding prime numbers are 13 and 7. Thus (oA) will have a lower rank, and the differentiation is made:

1-1-6-7-3-5-2-4

The calculations of the prime number equivalent product is made for each node. This step is reproduced several times until no change in order is found.

Step e. Breaking the ties. Breaking ties is when it is impossible to get a strictly ordered rank set using steps (c) and (d). The method is exactly the same as the one used for breaking ties during the unique SMILES generation. This is possible as the graph is a conceptual representation of bonds and nodes. The knowledge of the physical meanings of the graph is not needed.

To break ties, all the ranks are doubled and then the first lowest rank tied node is reduced by one. The set is treated as a new invariant set and the previous step of the algorithm is repeated.

2	3	5	7	11	13
17	19	23	29	31	37
41	43	47	53	59	61
67	71	73	79	83	89
97	101	103	107	109	113
127	131	137	139	149	151
157	163	167	173	179	181
191	193	197	199	211	223
227	229	233	239	241	251
257	263	269	271	277	281
283	293	307	311	313	317
331	337	347	349	353	359
367	373	379	383	389	397

Table 4.5: List of prime numbers between 1 and 400

Step 2. Generation of the SFILES string of the flowsheet structure. The SFILES string of the flowsheet structure is generated with initial process group selection and branching decisions based on the flowsheet structure graph invariant.

Rules for the initial node selection. The lowest numbered process group is chosen as the starting point of the SFILES notation. From the definition of the initial invariant set, an inlet process group will be chosen first. This is matching with the view of a process from the raw materials (inlets) to the products (outlets). This initial process group is the starting point of the depth-first search.

Rules for the branching decisions. The branching decision is made the same way as the initial node selection, i.e. the lowest numbered process group is chosen first, with the added rule that an inlet connection to a process group has always priority to an outlet connection.

To generate the SFILES notation of a flowsheet structure, first a depth-first search adding the process groups and the connections to the output SFILES notation is performed. The starting point of the depth-first is selected by applying the initial node selection rules and the branching decisions during the search are made by following the the branching decision rules. Each time a branch is taken, a left square bracket is opened and each time a dead-end is reached, a right square bracket is added. For structures without recycles, this first step finishes the creation of the SFILES notation. However, in the case of recycle, the search will find an already visited process group. The recycle closure is known, with the previous process group and the already visited one. It is now possible to append symbols in preparation for the second step. Each pair of symbol must be unique for each recycle closure. The second step is converting the symbols into figures to ensure the ordering of the figures and the lowest numbering of the recycles by reusing the figures.

Following this algorithm, the SFILES notation of the flowsheet structure shown on figure 4.11 (page 35) is:
`(iA)(rAB/pABCD)[<(iB)](mABC/D)[(oD)](A/BC)[(oA)](oBC)`

Benefits of the SFILES notation

The SFILES notation and the ability to generate a unique SFILES string for a flowsheet structure provide real computational benefits during the synthesis of flowsheet structure alternatives. When generating a lot of flowsheet (structure) alternatives, some steps in the algorithm involve the comparison of a given flowsheet structure against a list of already generated flowsheet structures. For example, when a new complete flowsheet structure is generated, the algorithm checks if this structure has already been generated; if not it is added to the list of completed structures; if yes, it is simply discarded.

Comparing two flowsheet structures F_1 and F_2 for equality is a non-deterministic polynomial time complex problem[17], as it is equivalent to finding an isomorphism between the graph representations of the flowsheet structures F_1 and F_2.

That is why, for performance consideration, the algorithm should minimize the number of comparisons. If for example, the number of completed flowsheet alternatives is n and a new alternative is proposed. Before adding the new alternative to the list, n comparisons will be performed. The same way, if a second alternative is proposed, $n + 1$ comparisons will be performed.

By using the SFILES notation as a *signature* of a given flowsheet structure, it is possible to simply compare the signatures to know effectively if two flowsheet structures are equal. As the SFILES notation provides a character string, the computational cost of the comparison is significantly reduced. More details about the use of the SFILES notation within the computer implementation of the framework for CAFD are presented in section 5.2.3 (page 94).

SFILES notation in other fields

As presented, the SFILES notation concept is not limited to flowsheet structure representation. If a logical structure can be represented as a directed graph, and, if an initial graph invariant can be defined, the SFILES notation could be applied. This is illustrated with reaction pathways of the Ammonia synthesis reaction in section 6.5 (page 157).

4.1.3 Flowsheet property models

Having a process flowsheet represented by a set of process groups and described uniquely by a SFILES string, the next step is to predict some characteristic properties of the flowsheet. In this section, the concept of a flowsheet property model is presented together with the development of a property model of a flowsheet.

The flowsheet property model concept

Considering a process as a finite ensemble P within the universe U. The interactions between the process P and the universe U can be modeled as an exchange of information I, materials M and energy E. As shown on figure 4.13, these interactions are performed at two different boundaries, the universe/process boundary and the unit operation/process boundary.

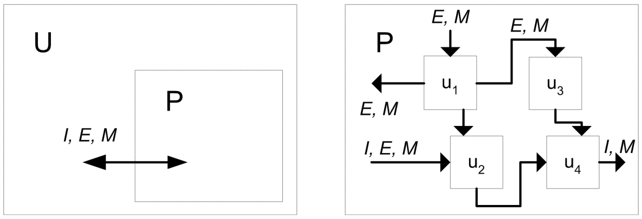


Figure 4.13: Interactions and exchanges between the universe, the process and the unit operations

The objective of a process flowsheet property is to model the impact of the process P on the universe U as the sum of the contributions of the unit operations u_i.

As a process group represents a unit operation or a set of unit operations, if a flowsheet property can be represented as a summation of process group contributions, a flowsheet group contribution based property model can be derived as shown by equation (4.1).

Equation (4.1) is an example of a generic flowsheet property model.

$$f(P) = \sum_{k=1}^{NG} pos_k \cdot a_k \tag{4.1}$$

where

$f(P)$: Flowsheet property function
NG : Number of process groups
a_k : Regressed contribution of group k
pos_k : Topology factor

A fundamental difference exists between a process group and a molecular group. Consider a simple molecule like N-Octane. This molecule can be represented as being composed of 6 CH_2 molecular groups and 2 CH_3 molecular groups. From this representation and a table of regressed parameters, it is possible to predict the critical temperature (or other pure component property) as shown by equation (4.2).

$$T_c = 231.239 \times \ln(2 \times 1.750600 + 6 \times 1.332700) = 564.71K \tag{4.2}$$

The critical temperature has been predicted using the Marrero-Gani[33] group contribution method. It is important to note that in a group contribution based method for molecular property prediction, the molecular group representation is not unique. The representation depends on the available functional groups, which corresponds to a trade off between the available experimental data and the minimum number of (molecular) functional groups that are needed to obtain reasonably accurate predictions.

On the contrary, as shown on figure 4.13 (page 40) the limits of the process group are clearly defined as they are following the limits of the underlying unit operations. This rigorous definition of the limits and the ability to quantify rigorously the impact of a unit operation on the overall process energy consumption are providing insights to simplify the regression of the individual process group contributions.

What should also be noted, compared to the case of a molecular group contribution based property model is the need for a topology factor pos_k expressing the dependence of the position of a process-group within the flowsheet structure to its property. These property models can then predict the corresponding flowsheet property like energy consumption, selectivity or risk index.

Flowsheet energy consumption index property model

As developed by Bek-Pedersen and Gani[3, 14] the design of distillation columns can be based only on the driving force between the two key components defining a separation. The driving force, D_{ij}, as defined by Bek-Pedersen and Gani is:

$$D_{ij}(x_i, y_i, \alpha_{ij}) = \frac{x_i \alpha_{ij}}{1 + x_i(\alpha_{ij} - 1)} - x_i = y_i - x_i \tag{4.3}$$

Where, x_i and y_i are mole fractions of component i in the two co-existing phases and α_{ij} is the relative separability of component i with respect to component j.

Consider now a separation process with one inlet stream and two products, which may or may not be at equilibrium. The mass balance for each component i, can be written as:

$$y_i = (R + 1)z_i - R \cdot x_i \tag{4.4}$$

Where, $R = L/V$ is the ratio of the flowrates of product 1 (liquid) and product 2 (vapor); z_i, x_i and y_i are the feed composition, the liquid product composition and the vapor product composition of component i, respectively; L and V are the liquid and vapor product flowrates.

Subtracting x_i on both sides of Eq. (4.4) and inserting Eq. (4.3), yields:

$$D_{ij}(x_i, y_i, \alpha_{ij}) = (R + 1)(z_i - x_i) \tag{4.5}$$

From Eq. (4.5), it is clear, as represented on figure 4.14, that as the driving force becomes zero, the separation $(z_i - x_i)$ must also become zero since R is ≥ 0. Therefore, for any separation involving two-phases (in equilibrium or not), the phase composition data combined with mass balance, provides a direct measure of separability.

Consider now an ideal system where α_{ij} is constant (that is, independent of composition) and the two-phases are liquid and vapor. Since in this case, energy either needs to be added or removed to create the two-phases, the amount (energy) needed to create the two phases for the purpose of separation can be obtained through an energy balance, as:

$$\frac{Q}{V} = h^V - h^L \tag{4.6}$$

Where, Q is the energy to be added (or removed), V is the amount of vapor created (or removed), and h^V and h^L are the stream enthalpies for the vapor and liquid, respectively. The assumption of constant α_{ij} implies that[30]

$$h^V - h^L \propto \frac{1}{y_i - x_i} \tag{4.7}$$

Therefore, inserting Eq. (4.7) into Eq. (4.6) and then Eq. (4.3), yields,

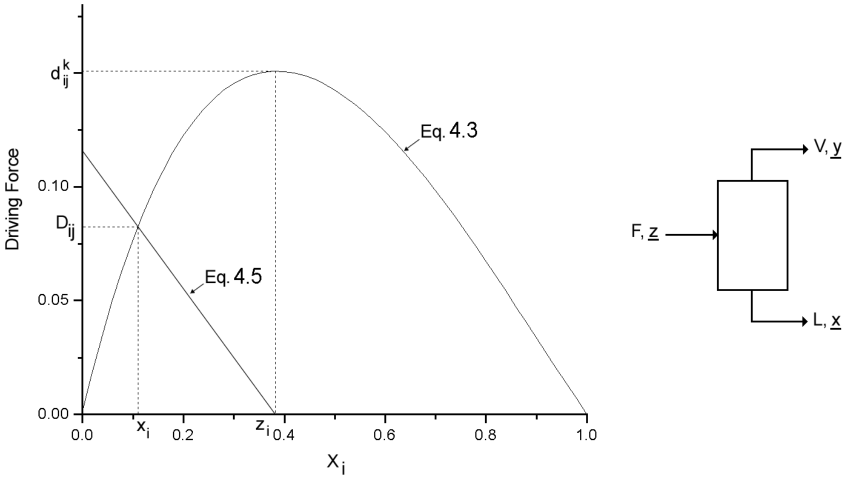


Figure 4.14: Driving force diagram

$$D_{ij}(x_i, y_i, \alpha_{ij}) = \frac{K_E}{Q/V} \tag{4.8}$$

Where, K_E is a proportionality constant. From Eq. (4.5) and (4.8), it can be noted that the driving force of component i with respect to component j is directly proportional to separability of component i from component j and inversely proportional to the energy (in the case of vapor-liquid system) added/removed to create the two-phase system. It can also be noted that Eq. (4.5) and (4.8) can be solved for given values of α_{ij}, inlet and outlet compositions without knowledge of the compounds present in the system. An important point to note is that the driving force based modeling allows the use of any property model or experimental data to obtain α_{ij}.

Two of the conclusions of the driving force model for distillation columns (and other vapor-liquid separations) is that the driving force is inversely proportional to energy consumption and that in a distillation train the easiest separation, i.e., where the maximum driving force is available, must be performed first. Bek-Pedersen and Gani[3] have shown that once the driving force for any distillation operation has been assigned, all other design parameters such as feed plate location, reflux ratio, etc., can be back calculated to obtain a near optimal design.

Based on the above theory, it is clear that any process-group contribution based model for calculating the energy consumption (as a flowsheet property) will be component independent if it is based on the driving force as an input (known) variable. Similarly, since every unit operation can be represented by its corresponding driving force, a collection of unit operations with their corresponding driving forces will represent the feasibility of the operation and the property (for example, energy consumption) of the process flowsheet.

A flowsheet property model. The objective of the property model is to predict the energy consumption performance of a unit operation (in this case, any distillation column with one feed stream and two product streams) given the process groups representing the distillation column, the driving force (related to the key compounds) and the group contributions.

The following flowsheet "property" model (4.9)[6] has been found to satisfy the objectives stated above.

$$E_x = \sum_{k=1}^{n=NG} Q_k = \sum_{k=1}^{n=NG} \frac{1+p_k}{d_{ij}^k} \times a_k \tag{4.9}$$

Where, E_x is the energy consumption performance of the flowsheet (MkJ/h for $Mmoles/h$ of feed), Q_k the energy contribution of each process-group, NG is the number of process-groups, d_{ij}^k the maximum driving force of the process-group k, a_k the contribution of the process-group k and p_k, the topology factor, is defined as

$$p_k = \sum_{i=1}^{nt} \overline{D_i} \tag{4.10}$$

Where, nt is the number of separation tasks that should be performed before the task k in the ideal case and $\overline{D_i}$ the maximum driving force of task i. For example, in a separation of a ternary mixture into two products oA and oBC where the maximum driving force of $A/B = 0.15$ and $B/C = 0.21$, $nt = 1$ (since only 1 task should have been performed before) and $\overline{D_i} = 0.21$ (the maximum driving force of this task).

Model parameter estimation. As in the development of any group contribution based method for molecular property prediction, the contributions of the individual process groups to a specific flowsheet property has been regressed through fitting experimental data. For the purposes of this work, distillation columns separating different (from binary to 8-component) mixtures into different product specifications, have been simulated through a validated rigorous distillation column model. These simulations then provided a set of "pseudo" experimental data on feasibility of separation as well as energy consumption.

For the data regression stage, since the driving force and the groups identities are known for each simulated separation operation, their corresponding process group are also known. Consequently, the process group parameter a_k can be estimated through data regression by matching the "pseudo" experimental values of E_x obtained through simulation.

The data used for regression covered a range of d_{ij}^k between 0.0168 (lower limit) and 0.15696 (upper limit). Also, it allowed the creation of process groups with 2, 3, 4, 5, 6, 7 and 8 components (note that the information on the number of components and not the component identity is used). This has resulted in a table of groups and their contributions. A partial list of process groups for distillation operations is given in table 4.6 (see appendix D.1 on page 179 for the complete list).

Calculation steps. To use the developed flowsheet property model, the following steps are needed:

Given: A separation process flowsheet, the identities of the compounds in the mixture and the product specifications for each unit operation of the flowsheet.

i) Identify the number of unit operations and the type of unit operations.

ii) Identify the separation tasks for each unit operation and calculate the corresponding d_{ij}^k for each separation task.

iii) For the identified unit operations (in this case, distillation), the product specifications and the d_{ij}^k, identify the corresponding process groups needed to represent the specified flowsheet.

iv) Retrieve the process group contributions from the parameter table (for example, table 4.6) and then apply Eq. (4.9) to estimate the flowsheet property.

4.2 Reverse Flowsheet Property Prediction Method

Having a process group based model for prediction of a characteristic property of a process flowsheet, the next step is to use this model in the reverse way to generate flowsheet alternatives described by SFILES strings and satisfying a target (desired) property value. In this section the reverse flowsheet property prediction method is presented. The reverse flowsheet property prediction problem is the core of the CAFD method and is based on the CAMD concept already presented in section 2.2 (page 9). The reverse flowsheet property prediction method is composed of the following four steps: the definition and analysis of the problem, the process group selection and initialization, the synthesis and test of the alternatives, and finally, the ranking and selection of the alternatives (see also figure 3.2 on page 19).

4.2.1 Definition and analysis of the synthesis problem

Definition of the synthesis problem

The definition of the synthesis problem is composed of two parts: i) the structural definition which is the definition of the process inlets and outlets of the desired process, and ii) the target (desired) characteristic property of the process flowsheet.

Process Group	d_{ij}^k	a_k	Light/heavy key Min. recovery
`(A/B)`	0.1570	0.01173	0.999
`(B/C)`	0.0664	0.04027	0.996
`(C/D)`	0.1092	0.01845	0.998
`(A/BC)`	0.1570	0.01016	0.999
`(AB/C)`	0.0664	0.06629	0.996
`(B/CD)`	0.0664	0.08074	0.996
`(BC/D)`	0.1092	0.01310	0.998
`(C/DE)`	0.1092	0.01476	0.998
`(A/BCD)`	0.1570	0.00808	0.999
`(AB/CD)`	0.0664	0.05101	0.996
`(ABC/D)`	0.1092	0.00911	0.998
`(B/CDE)`	0.0664	0.06426	0.996
`(BC/DE)`	0.1092	0.01113	0.998
`(BCD/E)`	0.0632	0.30529	0.997
`(A/BCDE)`	0.1570	0.01406	0.999
`(AB/CDE)`	0.0664	0.04336	0.996
`(ABC/DE)`	0.1092	0.00809	0.998
`(ABCD/E)`	0.0632	0.19277	0.997
`(B/CDEF)`	0.0664	0.02434	0.996
`(BC/DEF)`	0.1092	0.00992	0.998
`(A/BCDEF)`	0.1570	0.01262	0.999
`(AB/CDEF)`	0.0664	0.01754	0.996
`(ABC/DEF)`	0.1092	0.00830	0.998
`(ABCD/EF)`	0.0632	0.02942	0.997
`(B/CDEFGH)`	0.0664	0.01506	0.996
`(BC/DEFGH)`	0.1092	0.00823	0.998
`(BCD/EFGH)`	0.0632	0.06012	0.997
`(A/BCDEFGH)`	0.1570	0.01022	0.999
`(AB/CDEFGH)`	0.0664	0.01144	0.996
`(ABC/DEFGH)`	0.1092	0.00714	0.998
`(ABCD/EFGH)`	0.0632	0.01527	0.997

Table 4.6: Partial list of distillation process groups and their regressed contributions

Within the CAFD framework it is possible to define two different types of synthesis problem: the *new process* and the retrofit. The *new process* type is defined as looking at a completely new process alternative considering no already existing infrastructure. The retrofit at the process level is defined as looking for new process alternatives considering part of the process as being fixed. These two problems are schematically represented on figure 4.15a. for the *new process* and 4.15b. for the retrofit, respectively.

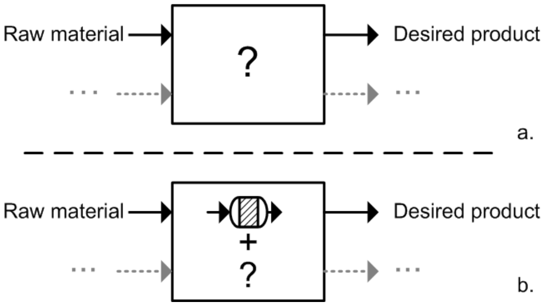


Figure 4.15: Two possible synthesis problem definitions

First part, the definition of the structural constraints. The *green field* synthesis problem has a simple structural definition, as it only requires to define the available raw materials and the desired products. In the case of the process retrofit problem, a prerequisite is the analysis of the existing process to determine which are the parts to be kept fixed for any technical or financial reasons. This defines the basic structural constraints of all the possible flowsheet alternatives to be generated.

Second part, the definition of the specifications for property targets. To be able to set a property target for the generated flowsheet structure alternatives, the corresponding flowsheet property models and the process group contributions must be available. At this point, the necessary information to proceed with the analysis of the synthesis problem is available.

Analysis of the synthesis problem

The purpose of the analysis of the synthesis problem is to gain usable knowledge from the defined problem. From the structural definition and the target properties a sequence of analysis presented in figure 4.16 is performed to obtain:

- the list of all the components (chemical species) in the problem, including the reaction intermediate or the mass transfer agents;
- if needed, the reactions to convert the raw materials into the desired products;

- the list of all the possible process operation tasks.

The three steps to be employed to gain this usable knowledge are the reaction analysis, the pure component and mixture property analysis and the identification of the feasible operation tasks.

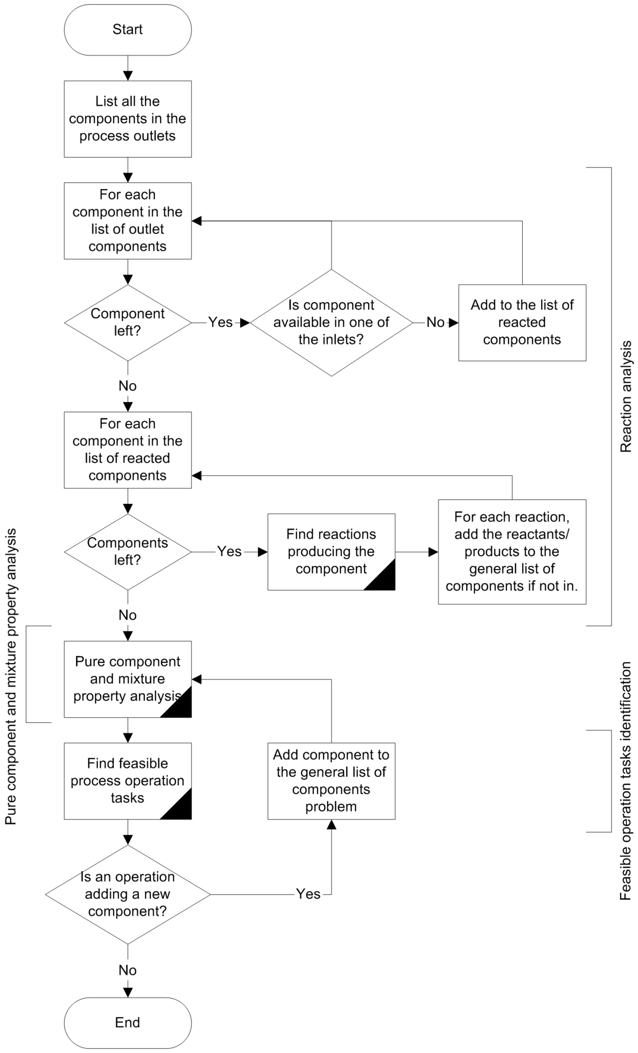


Figure 4.16: Flow diagram of the synthesis problem analysis algorithm

The reaction analysis – By identification of the components in the desired products not available as raw materials, it is possible to get the list of the components that should be formed by reaction.

For each component that should be formed by reaction, a database search is performed to find the possible reactions having those components as products. As shown on figure 4.16, this is an iterative approach as a possible reaction match may imply the need to provide reactants not currently available. These reactants can either be provided as new raw materials in the synthesis problem or as products of another reaction.

All the new reactants or by-products from the reaction matches are added to the list of components in the synthesis problem. The matching reaction information is saved for later use during the selection and initialization of the process groups (see section 4.2.2 page 50)

The pure component and mixture property analysis – The pure component and mixture property analysis is performed to obtain data (knowledge) that can be used for the identification of the feasible process operation tasks.

The pure component property analysis is performed by calculating or retrieving from a database a list of 22 pure component properties. The list of those properties is presented in table A.1 of appendix A on page 173.

The mixture property analysis is made in terms of binary pairs of components. For all the component binary pairs, analyses are made to check the presence of possible azeotropes, euctictic points or potential mass separation agents (MSA). The potential azeotropes are found using a hybrid approach of database search and calculations. The potential MSA are found by solubility analysis from the pure component property and from a list of commonly used solvents. The list of these solvents is available in tables B.1 and B.2 of appendix B on page 175.

The identification of feasible process operation tasks – The identification of the feasible process operation tasks is performed using the physical insights based method for flowsheet synthesis developed by Jaksland[28].

The physical insights method is based on the principle that for each process operation task, one or more specific pure component property can be associated. For a separation between two components and a given operation task, it is possible to assess if the operation task is feasible by comparing the ratio of the corresponding pure component properties. If the ratio is above a threshold value corresponding to this operation task and if some additional property constraints are satisfied, the operation task is considered as feasible.

For example, a crystallisation is a feasible separation task if the ratio of the melting points for the two components is greater than 1.2 and the minimum value of melting point is greater than T_min (250 K). But a distillation separation task is feasible if both the ratio of the boiling point is greater than 1.01 and no azeotropes exist between the two components.

This identification of possible process operation tasks can add new components to the problem. For example, the absorption operation task may require an entrainer not already available in the problem. If a new component is added to the problem, the pure component and mixture analysis is repeated.

4.2.2 Selection and initialization of the process groups

The selection and the initialization of the process groups is based on the analysis of the synthesis problem. Note that according to the process group concept (see section 4.1.1 page 23), the process group is component property dependent. It means that a given process group can be used for different sets of components as long as the properties are matched. The initialization of a process group is the procedure in which a process group is associated with a given set of components. This initialization is important, as a process group is component property dependent, while the connectivity of a process group is component and mixture dependent. It means that a given process group can result in more than one initialized process group, one for each set of components where the properties are matched. For example, the flash separation process group can be applied for the separation of Methane from Toluene, but also for the separation of Methane from Biphenyl.

Selection and initialization of reaction process groups

From the synthesis problem analysis, the complete list of reactions together with their specific information, such as, involved components, optional kinetic model parameters or conversion rate, are available.

For each reaction, or set of reactions if the kinetic model parameters are available a kinetic model based reaction process group is selected. If the kinetic model parameters are not known a fixed conversion reactor process group is selected. In some particular cases, a reaction process group associated with a given reaction set may be available in the database of process groups. If it is the case, this process group is selected in priority over the conventional kinetic based and fixed conversion reaction process groups.

As the corresponding reaction process groups are selected, they are initialized with the components in the reaction and the conversion or kinetic data. After initialization, each reaction process group is ready to be used for the generation of the alternatives. The details on the reaction process groups are given in appendixes D.4 and D.5 on pages 188 and 191, respectively.

Selection and initialization of separation process groups

The separation process groups are selected based on the identified separation tasks during the synthesis problem analysis. For each feasible separation task identified in the synthesis problem analysis, the corresponding process group is selected. The selected process group can be initialized with different sets of components, if each set is matching the property dependence of the process group.

For example, as shown in table 4.7, with 4 components A, B, C and D in the synthesis problem and a feasible flash separation task between B and C, the flash process group is initialized with 4 different sets of components. This is based on the assumption of ideal system, where no binary azeotropes exist and the components being ordered by decreasing relative volatility.

Components in the problem	A, B, C, D
Feasible separation task	Flash between B and C
Matching set of components	(BC), (ABC) (BCD), (ABCD)
Initialized process groups	`(fB/C)`, `(fAB/C)` `(fB/CD)`, `(fAB/CD)`

Table 4.7: Initialization of a flash process group within a 4 components synthesis problem

After initialization, each separation process group is ready to be used for the generation of the alternatives. All the initialized process groups are grouped in the pool of process groups used during the synthesis of the flowsheet structure alternatives.

4.2.3 Generation and test of the flowsheet structure alternatives

The generation of the flowsheet structure alternatives employs a combinatorial algorithm. The combinatorial algorithm generates new flowsheet structure alternatives by combining process groups according to a set of connectivity rules. The generated flowsheet structure alternatives are tested for their target property values, using the corresponding property model.

Preparation of the initial flowsheet structure working list

The generation of the flowsheet structure alternatives is performed by gradually completing flowsheet structures with process groups. While these flowsheet structure are completed, they are stored in a so-called *working list.* The initialization of the *working list* is performed with flowsheet structures representing the structural constraints of the process synthesis problem and the involved reactions.

From synthesis problem structural constraints a flowsheet structure is created. This flowsheet structure is composed of inlet and outlet process groups representing the desired process inlets and outlets. If retrofit at the process level is performed, the parts of the process being kept are translated into process group backbones and added to the flowsheet structure. A process group backbone is a collection of process groups representing a part of the flowsheet structure.

If the process synthesis problem is not involving reactions, this flowsheet structure is added to the flowsheet structure working list as the only flowsheet structure starting point. If reactions are involved, this flowsheet structure is the base

From the possible reaction paths one flowsheet structure starting point is created for each reaction path involving a unique set of reaction process groups. Each of these flowsheet structure starting points is composed of the flowsheet structure created from the synthesis problem structural constraints and the reaction process groups.

For example, considering the 4 reaction sets R_A, R_B, R_C and R_D with three possible paths:

$$R_B \rightarrow R_C \rightarrow R_D \tag{4.11}$$

$$R_C \rightarrow R_B \rightarrow R_D \tag{4.12}$$

$$R_A \rightarrow R_D \tag{4.13}$$

each reaction set is associated to a reaction process group, which can be called for the sake of simplicity `(ra)`, `(rb)`, `(rc)` and `(rd)`. The first two paths are using the same reaction process groups `(rb)`, `(rc)` and `(rd)`, the last path is using only `(ra)` and `(rd)`. As a result only two distinct sets of process groups can express the three different paths – (`(rb)`, `(rc)`, `(rd)`) and (`(ra)`, `(rd)`). Two distinct flowsheet structure starting points are created, each one being composed of the flowsheet structure created from the structural constraints augmented with an independent set of reaction process groups. The reaction process groups are added unconnected to the flowsheet structure as between two reaction process groups separation or recycle of materials may need to be performed.

Algorithm for the generation of flowsheet structure alternatives

The algorithm for the generation of the flowsheet structure alternatives is the core of the reverse property prediction approach of the the CAFD method. Going back to the *synthesis pattern* (see chapter 3), the *synthesis pattern* is composed of the building blocks, the connectivity rules between the blocks and a model to evaluate the properties of the resulting structures. In the case of the CAFD method, the synthesis building blocks are the process groups, the connectivity rules are the process groups connectivity rules and the models are the flowsheet property models. The starting points are the flowsheet structure starting points generated in the previous steps. The stop conditions are the satisfaction of the flowsheet structure valence, i.e. ensuring that all the process groups are fully connected.

The algorithm is presented on figure 4.17 (page 53). The two important steps in the algorithm are the steps tagged on the figure, *Get the list of free connections* of the flowsheet structure and *Get the list of process groups matching with the free connection from within the current structure or from the initialized process groups.*

First, the selection of the free connections in the flowsheet structure. The objective of the free connection selection is to determine the next points where the current flowsheet structure will be completed by the addition of a new process group or the connection to another process group within the flowsheet structure. The free connections in the flowsheet structure are all the process group connections not connected to another process group. They can either be process group inlets or process group outlets. In the current algorithm, only the free process group outlets are selected. This leads to a forward algorithm with the generation of the flowsheet alternatives from the raw materials to the products.

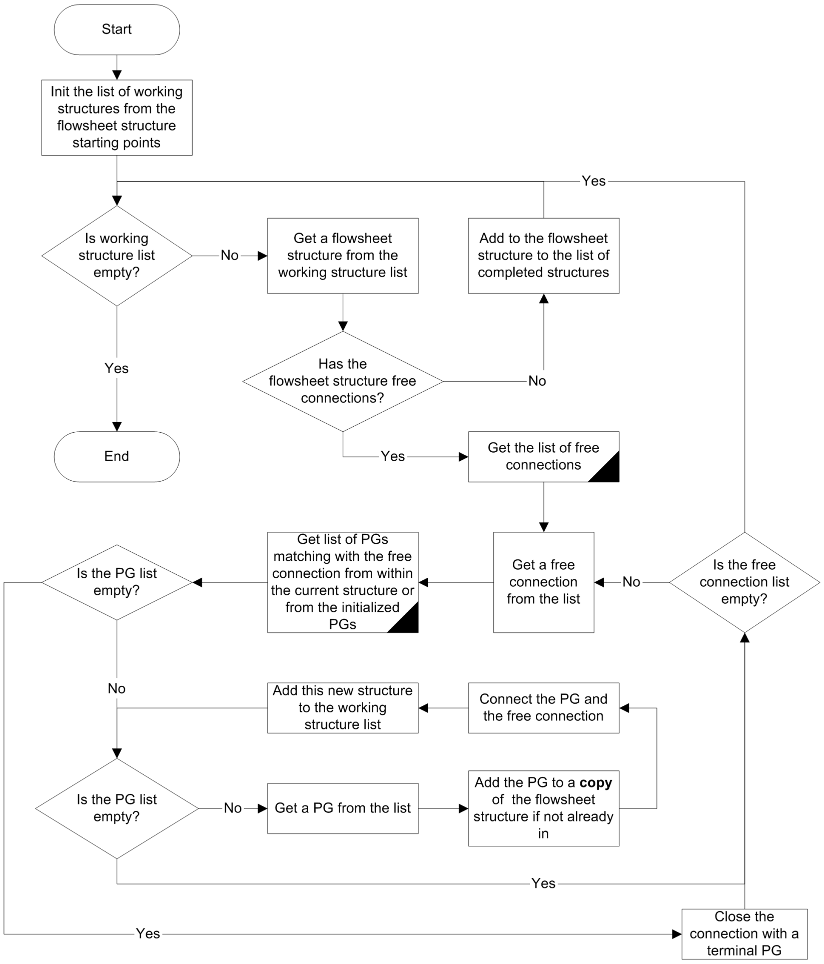


Figure 4.17: Generation of the flowsheet structure alternative algorithm

Second, the selection of the process groups matching with a free connection. The important part in the match of a free connection with a process group is the analysis of the possible material recycles. These material recycles are either for reactants or mass transfer agents.

The reaction recycles are taken into account when looking for process group matching with a free outlet connection. If a free outlet connection is containing some reactants and possibly some inerts of one of the reactor a recycle is directly created between this connection and the reactor process group. This is analog to the *Hydrogen bonding* in molecular structures, as the newly created connection has not a *formal* connection on the reactor process group side.

If the outlet connection contains some components others than the reactants of the available reactor process groups together with some reactants, this outlet connection is further processed to be able to recycle back the reactants to reactor process groups. If the free outlet connection do not contain any reactants neither requested products, it is simply closed with a process outlet process group.

The same approach as for the reactants is used to recycle the mass transfer agents. The final algorithm to generate the list of possible matching process groups and connections for a free connection is presented on figure 4.18.

Test of the flowsheet structure alternatives

The test of the flowsheet structure alternatives is composed of two parts. For each flowsheet alternative, first a test is performed to ensure that the alternative is satisfying the structural constraints, and if so, the target properties are calculated.

The verification of the structural constraints is performed on each resulting flowsheet structure alternative. For each process outlet, considering the outlet of a backbone process group as equivalent to a process inlet, it is verified that a path exists between this process outlet and at least one process inlet.

The calculation of the target properties is performed by evaluating for each flowsheet structure the corresponding flowsheet property model with the process group contributions from the process groups in the flowsheet structure. The testing also includes sub-properties (driving force, etc.).

At the end of the test step, all the remaining flowsheet structures are feasible flowsheet structures and each flowsheet structure can be evaluated against the property targets to find the optimal or to classify them in terms of the property target.

4.2.4 The ranking and selection of the alternatives

Depending on the number of target properties, the ranking of the alternatives can be on a single property optimization, on a minimizing a weighted sum of

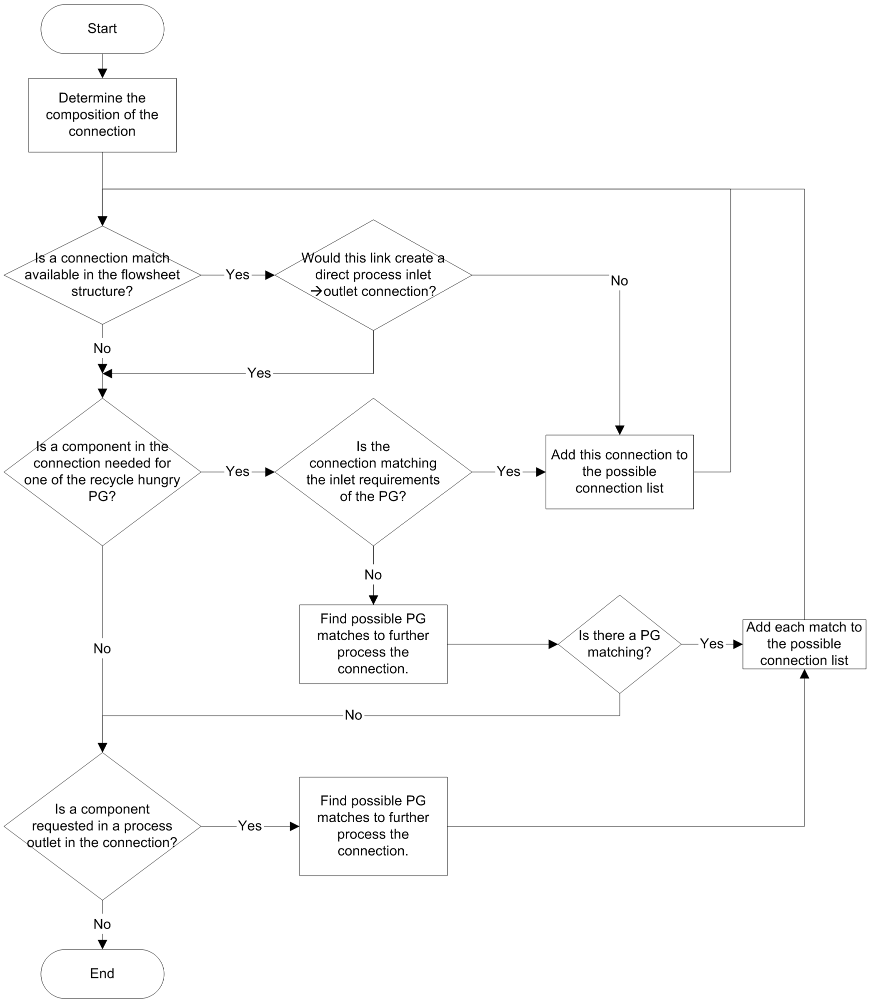


Figure 4.18: Algorithm to find the matching process groups or connections of a free connection

objectives or using a non-dominated Pareto front[8] method.

These three approaches are presented on figure 4.19. The figure is representing a set of 16 alternatives with the evaluation of two properties p_1 and p_2. The ranking of the best alternative can be done either on one property – minimization of p_2, a weighted sum of p_1 and p_2 or the set of alternatives being on the Pareto front.

Considering the weighted property minimization, it is the most commonly used approach. But in the case of figure 4.19, it is easy to remark that the objective function is sensitive to the choice of the weights for the properties. In that case, the objective function is a linear combination of the two properties with equal weights. This is often not a problem when the two properties can be linked to each other through a common criteria like the total cost of utilities. But as process synthesis often implies the optimization of "opposed" properties like environmental impact and cost, the linear combination of the two properties cannot capture the real trade-off. The selection of alternatives lying on the Pareto front ensures that the choice is performed with a correct evaluation of the associated trade-offs.

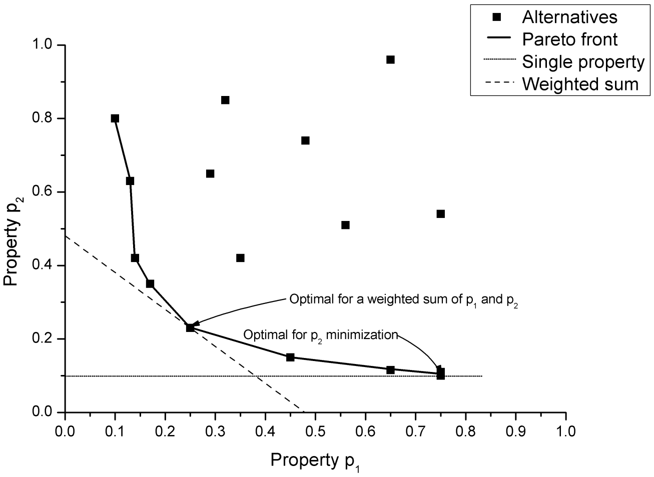


Figure 4.19: Optimization alternatives for a 2 property minimization.

4.3 Reverse Simulation Method for Process Design

Having obtained a list of feasible process flowsheets represented by process groups, described by SFILES strings, and classified by their target property values, the remaining task is to determine the flowsheet design parameters that will describe the process (the equipment parameters are not included in this study). The reverse simulation method is the process by which knowing the state variables of the inlets and outlets of a unit operation, i.e., individual flowrates, pressure and temperature, the design variables of the unit operation are calculated. The term "reverse" is used to highlight the difference with "conventional" simulation, where, knowing the inputs (individual flowrates, energy) of a unit operation and its design parameters, the outputs of the unit operation are calculated.

For example, on figure 4.20, the reverse simulation is performed on a distillation column. From the inlet and outlet stream definitions and the properties of the mixture, the number of plates, the feed plate location and reflux ratio of the column are back calculated.

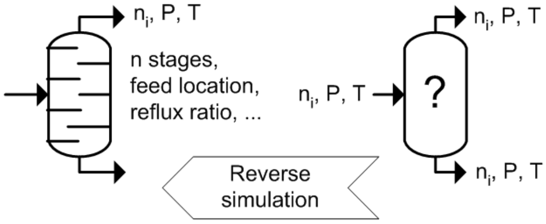


Figure 4.20: Reverse simulation overview

Using "conventional" simulation the design problem is solved by trial and error when knowing the desired targets of a given unit operation, the design parameters are modified or optimized to match those targets. As shown on figure 4.21 for a distillation column, first an initial set of design parameters are assumed, then simulation is performed and the results are compared with the desired targets. If the difference with the targets is above the acceptable error, another set of design parameters is assumed and simulation is performed again, until the results become acceptable. It is important to note, that this procedure is not guaranteed to lead to an acceptable set of design parameters. The benefits of a direct reverse simulation, is that this common trial and error approach to find the design parameters matching the targets is avoided and it is immediately known if a solution exist or not.

The reverse simulation method is a critical building block of the CAFD approach. As shown in the method overview (section 3.3 page 17), at the end of the synthesis of the alternatives, the resulting flowsheet structures are only a combination of process groups together with their connections. Thus, the need of a method to define the unit operations from the process groups and inputs and outputs of the process groups, in terms of design parameters is important.

The reverse simulation method is currently available for three types of process groups, the distillation process group, the kinetic reactor process group and for selected single stage separation process groups. The reverse simulation

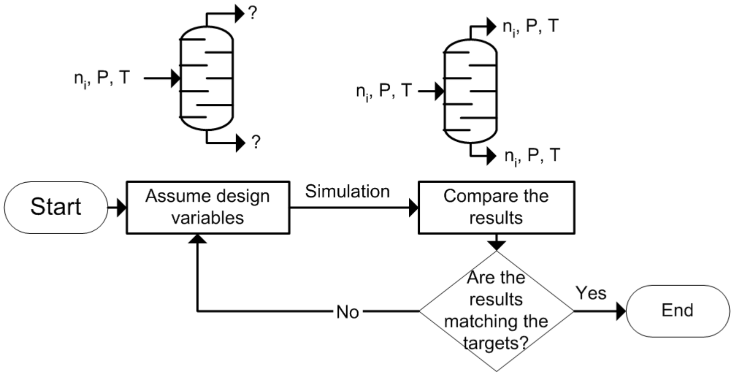


Figure 4.21: Conventional trials and error method for design

for the distillation process group is based on the driving force concept by Gani and Bek-Pedersen[14] (see section 2.1.3 page 7). For the reactor process group it is based on the work by Horn[25] on the attainable region concept (see section 2.1.3 page 8). For the single stage separation process group, the driving force concept has been generalized.

In the following sections, the resolution of the mass balance at the flowsheet level is presented. The resolution of this mass balance is providing the definition of the inlets and the outlets of all the process groups within the flowsheet structure. Then the three reverse simulation methods are presented with the generalized driving force concept for the reverse simulation of single stage separation process groups, the driving force concept for reverse simulation of distillation column process groups and the attainable region for chemical reactions for the reverse simulation of reaction process groups.

4.3.1 The resolution of the mass balance

When a feasible flowsheet structure has been selected, all the process groups and their connections are known. The first step of the reverse simulation method is the resolution of the mass balance between the process groups. As the mass balance on a process group is completely defined this is straightforward with the the exception of the material recycles (where even though the compositions are known, the flowrates are not known). If the flowsheet structure contains some recycles, build up of recycled components must be taken into account by the addition of the necessary purges. In the following sections, first the mass balance at the process group level is presented and then the addition and definition of purges within the recycles.

Mass balance of a process group

Each process group has outlet specifications which are guaranteed to be respected if the connectivity rules of the process group are satisfied. The outlet specifications are based on internal mass balance rules. Note that as each process group is based on the driving force, the mass balance models are simple linear models where the constitutive variable (driving force) is either known (for known input or output) or calculated (for known input and output).

For example, consider the case of the separation of a mixture of 3 components A, B and C into 3 pure streams with only simple distillation columns. This leads to two possible configurations, the direct sequence and the indirect sequence as shown on figure 4.22.

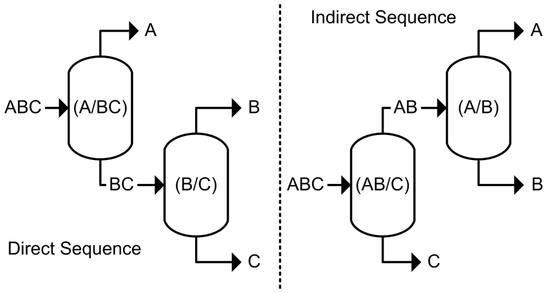


Figure 4.22: Direct and indirect sequence in the separation of a 3 component mixture into 3 pure streams

As the process groups guarantee the recovery of the components in their outlets, it means that during the generation of the alternatives the process group `(A/B)` can be connected to the output of process group `(AB/C)` independently of the composition of the mixture of A, B and C coming in the `(AB/C)` process group as the outputs of `(AB/C)` are ensured to be a mixture of A and B, and, a stream with a high purity and recovery of component C. To summarize, a process group is flexible in its inputs and strict in its outputs.

Definition of the recycle purges

On each recycle loop including a reaction process group, the recycled components are identified. These components pass through the reaction process group, they can be reactants, products or inerts. If products or inerts are recycled, they are either lighter or heavier than the recycled products. If no inerts or products are recycled a purge on the recycle loop is not needed.

In the case of lighter inerts or products, the purge rate is calculated on the light key. Reciprocally, in the case of heavier products or inerts being recycled, the purge rate is calculated on the heavy key as shown on equation (4.14).

$$n_{purge_{key}} = \sum_{inlets} n_{key} - \sum_{outlets} n_{key} \tag{4.14}$$

With $n_{purge_{key}}$ the purge rate of the key component and n_{key} the individual flowrates of the key component in the inlets or outlets of the recycle loops, the inlet of the reaction process group being considered as an outlet and its outlet as an inlet.

As all the inlets are known, the mass balance on each process group is available and the purge rates are available, it then possible to solve the complete mass balance.

4.3.2 The generalized driving force model

The generalized driving force model is the extension of the driving force model for multistage separations like distillation columns to single stage separations such as P,T-flash, evaporation, liquid-liquid extraction, crystallization, etc. In this section, the generalized driving force model is presented followed by its application to the reverse simulation for process groups.

Presentation of the model

In this section the model is presented for vapor/liquid separations. The liquid-liquid, solid-liquid and evaporation (solid-liquid-vapor) separations will follow the same principles, where instead of the energy separation agent, the separation agent may be solvents or mass separating agents.

Considering a 2-phase, vapor-liquid, single stage separation process as shown on figure 4.23, the following equations can be obtained:

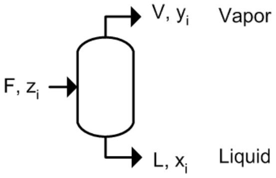


Figure 4.23: Single stage vapor liquid separation process

Global Mass Balance of the separation unit:

$$F = V + L \tag{4.15}$$

Ratio of heavy product rate over light product rate:

$$R = \frac{L}{V} \tag{4.16}$$

Developing equation (4.15) for component i:

$$Fz_i = Vy_i + Lx_i \tag{4.17}$$

Combining the obtained expression with equations (4.15) and (4.16):

$$\frac{L+V}{V} Z_i = \frac{V}{V} y_i + \frac{L}{V} x_i \tag{4.18}$$

$$(R+1)\, z_i = y_i + Rx_i \tag{4.19}$$

It is obtained an expression to determine y_i from the values of R, z_i and x_i:

$$y_i = (R+1)\, z_i - Rx_i \tag{4.20}$$

Combining the driving force equation (4.3) and the last expression:

$$D_{ij} = y_i - x_i = (R+1)\, z_i - (R+1)\, x_i \tag{4.21}$$

The interesting points that can be obtained from equation (4.21) are:

- For $D_{ij} = 0$, $x_i = z_i$. It means that if no driving force is available, there is no possible separation.
- For $x_i = 0$, $D_{ij} = (R+1)z_i$. Thus when $R \to 0$, $D_{ij} = z_i$ at $x_i = 0$. As $R = \frac{L}{V}$ by definition, the intersection of the line going through $D_{ij} = z_i$ on the y-axis and $x_i = z_i$ on the x-axis and equation (4.21) gives the lower limit of the liquid concentration that can be achieved with a feed at z_i.

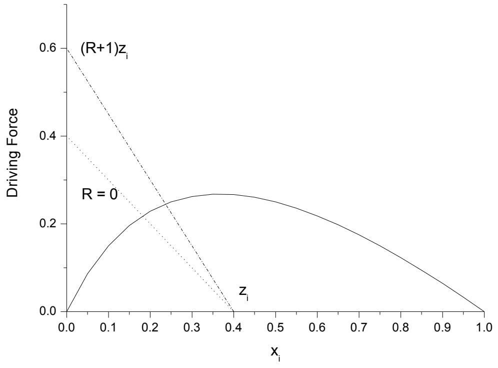


Figure 4.24: Vapor – liquid driving force diagram

An example of driving force diagram for vapor-liquid separation process is shown on figure 4.24. The curve is representing the driving force of the two key components i and j, together with two lines:

Line $R = 0$ It indicates the lower limit of the liquid concentration that can be achieved with a feed at z_i from the intersection of that line and the driving force curve.

Line R The intersection of that line and the driving force curve gives the liquid concentration that is reached for a given R and feed z_i.

It means that given two parameters out of the reflux ratio R, the feed composition z_i and the target driving force D_{ij}, the third is directly available.

Application of the model to the process groups

Considering the flash separation of components i-Butane (index i) and n-Butane with the flash process group `(fA/B)`, having the driving force shown on figure 4.25.

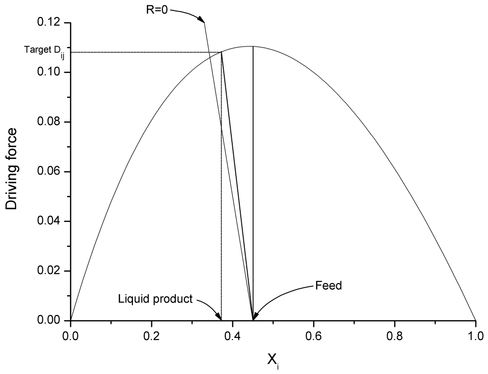


Figure 4.25: Driving force associated with the (fA/B) flash process group

According to the model, given the driving force target $D_{ij} = 0.108$, as this target is within the attainable driving force for the feed at $z_i = 0.45$, the liquid composition is known, with $x_i = 0.37$. From the driving force, the temperature (T) and pressure (P) of operation is also known with $T = 267\ K$ and $P = 1\ atm$. This define the resulting PT-flash separation process completely. An interesting benefit of this method is that the solution presented would be valid for all other binary mixtures having similar driving force diagrams. In some cases, P and T could be adjusted to match the diagram, thereby providing operational alternatives. The problem could be formulated through CAMD as find all binary mixtures which at $P = 1\ atm$, gives $D_{ij} = 0.108$. In this way, the same design (feed, product quality and pressure) will be obtained for different binary mixtures to operate at different temperatures.

4.3.3 Reverse simulation with the driving force for distillation

In this section, the procedure to apply the reverse simulation to a simple distillation process group and a simple example of application are presented.

The reverse simulation for the simple distillation process group (see section D.1 page 179 for more details on this process group) is completely based on the driving force theory presented in section 2.1.3 page 7.

Reverse simulation procedure for the distillation process group

The procedure to get the design parameter of the distillation column in the simple distillation process group is as follows:

1. Given a NC component process group.
2. Order the components by relative volatility and take the key components.
3. Retrieve the maximum driving force between the key components $F_{D_i|Max}$, and the composition of the light key at its maximum D_x, either from experimental data or VLE calculations.
4. Select the target purities or recovery for the key components, by default 99.5%.
5. If the inlet composition is between the requested purities for the bottom and top products. Get the ideal number of stages N_{ideal} for the column from the table of pre-calculated values (see table C.1 page 177).
6. Set the feed plate location of the column to be $N_F = (1 - D_x)N_{ideal}$ (Plate one is the top plate of the column).

Example of reverse simulation for the distillation process group

In this section an example of reverse simulation for the distillation process group with a five component simple distillation process group `(AB/CDE)` is presented. The components and the inlet composition specification are given in table 4.8.

Component	flow rate (kmole/hr)	normal boiling point (K)	relative volatility
A – Propane	45.4	231	7.98
B – i-Butane	136.1	261	3.99
C – n-Butane	226.8	273	3.00
D – i-Pentane	181.4	301	1.25
E – n-Pentane	317.5	309	1.00

Table 4.8: Definition of the five components in the (AB/CDE) process group

The key components are B and C, the i-Butane and the n-Butane. From the driving force plot available on figure 4.26, $F_{D_i|Max} = 0.11$ at $D_x = 0.44$. By doing an interpolation with the table of precalculated values $N_{ideal} = 41$. Given D_x and N_{ideal}, the feed plate location is $N_F = (1 - D_x)N_{ideal} = 23$.

At this stage, the distillation column is fully defined for a possible rigorous simulation.

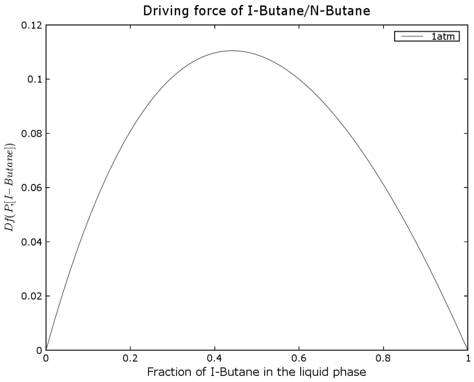


Figure 4.26: Driving force between the i-Butane and the n-Butane

4.3.4 Reverse simulation with the attainable region

In this section, the procedure to apply the reverse simulation to a kinetic model based reactor process group and a simple example of application are presented.

The reverse simulation for the kinetic model based reactor process group (see section D.4 page 188 for more details on this process group) is based on the attainable region theory presented in section 2.1.3 page 8. In this section only the attainable region analysis applied to the kinetic model based reactor process group is presented. More information on attainable regions and their application can be found in Glasser, Hildebrandt and Crowe[19].

Reverse simulation procedure for the kinetic based reactor process group

The kinetic model based reactor process group provides the following information which are necessary to proceed with the attainable region analysis. The generation of the attainable analysis is only done once. If not available for the set of reactions of the reactor process group, the analysis must be performed before the resolution of the mass balance.

- Nature of the components in the reactions.

- Complete definition of the reactions and the associated kinetics.

The attainable region analysis is a five step procedure, composed of the problem definition, the determination of the necessary conditions, the construction of the attainable region (AR), the interpretation of the boundary and finally, the optimization of the objective function. The figure 4.27 shows schematically this procedure.

Problem Definition

I. *Choose the Fundamental Processes*

A fundamental process is any physical phenomenon that can be controlled in the sense that the fundamental process can be used as, and when, required. Typical fundamental processes in chemical engineering include, for example, reaction, heat transfer, mass transfer and bulk mixing. The choice of fundamental processes permitted to occur limits the types of processing units that will appear in the process layout. Ideally, all fundamental processes should be considered in order to obtain the best process.

II. *Select the State Variables*

A set of variables must be chosen that is sufficient to described completely the rates of all the permitted fundamental processes and also the objective function that is to be optimized. Typical variables that describe the state of the system include concentrations, mass fractions, reaction conversion and temperature. In many problems it is also an advantage to choose the variables such that convex mixing laws are obeyed (that is the lever arm rule).

If a system may be fully described by N variables then a point in $\mathbb{R}^N$ space may completely represent the state of the system. Consequently, the outcome is described by a vector $\mathbf{C}$ in $\mathbb{R}^N$, termed the characteristic vector. It may be convenient to relate all the cost variables by assuming a linear relationship so that only a single cost variable characterizes all the fundamental processes. This linear relationship will maintain the convex mixing properties of the cost variables.

III. *Define the Fundamental Process Vectors*

Each of the fundamental processes is represented in the state space by a vector. These fundamental process vectors must be defined such that they represent the instantaneous change in the characteristic vector or state if that fundamental process occurs on its own. For instance, the reaction vector, $\mathbf{R(C)}$ defined at $\mathbf{C}$ represent the instantaneous change in $\mathbf{C}$ if only reaction occurs.

Consider the system where M fundamental processes may occur. These fundamental processes depend on the state of the system $\mathbf{C}$, the choice of

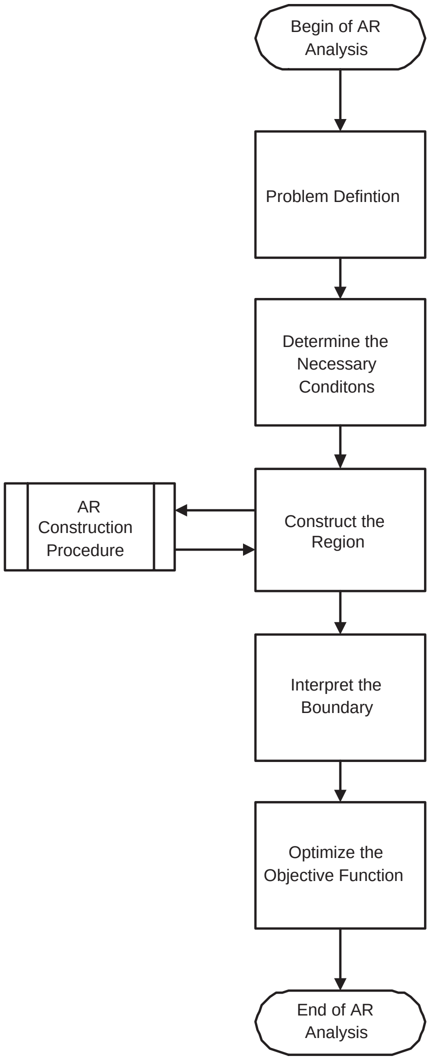


Figure 4.27: Procedure for an Attainable Region Analysis.

the control variables $\mathbf{U}$, and possibly some other achievable state within the system, $\mathbf{C}^*$. The fundamental processes are represented by the vectors, $\mathbf{P_1}(\mathbf{C},\mathbf{U},\mathbf{C}^*)$, $\mathbf{P_2}(\mathbf{C},\mathbf{U},\mathbf{C}^*)$, ..., $\mathbf{P_M}(\mathbf{C},\mathbf{U},\mathbf{C}^*)$. The net effect of all these fundamental processes is expresses by a resultant or net process vector $\mathbf{P}$, such that

$$\mathbf{P}(\mathbf{C},\mathbf{U},\mathbf{C}^*) = \sum_{j=1}^{M} a_j P_j(C,U,C^*) \tag{4.22}$$

where the operating policy for the fundamental process j is denoted a_j indicates if fundamental process j is occurring and the degree to which it is occurring relative to the other fundamental processes operating in the system.

IV. *Define the Systems Constraints*
The degree to which the fundamental processes may occur could be limited because of some physical restriction in the system considered. Hence, the operating policy for process j may be constrained to lie between the bounds $a_j^{min} \leqslant a_j \leqslant a_j^{max}$. The lower bound placed on an operating policy must satisfy $a_j^{min} \geqslant 0$; the upper bound is not limited in value and may not be present in some system (that is $a_j^{max} \to \infty$).

Determine the Necessary Conditions – Hildebrandt and Glasser[24] define the four necessary conditions for the completeness of the attainable region of a given system of reactions. These conditions are the control conditions of the algorithm to construct the region presented in section 4.3.4.

It is necessary that the attainable region $\mathbf{A}$ is such that:

i) it is convex;

ii) no reaction vector on the boundary of $\mathbf{A}$ points outwards;

iii) no reaction vector in the complement of $\mathbf{A}$ can be extrapolated backwards into $\mathbf{A}$;

iv) no two points on a PFR in the complement of $\mathbf{A}$ can be extrapolated back into $\mathbf{A}$.

If one of those conditions is not satisfied for a given attainable region $\mathbf{A}$, it means that $\mathbf{A}$ can be extended in the following ways:

ia) if the region A is not convex, it implies that two points of the region can be mixed in such proportions that the resulting point is outside $\mathbf{A}$.

iia) if a reaction vector on the boundary of $\mathbf{A}$ points outwards, a PFR reactor can be started from this point to extend $\mathbf{A}$.

iiia) if (iii) is not satisfied, a CSTR can be started from **A** to reach the given point in the complement of **A**.

iva) if (iv) is not satisfied, it implies the existence of a recycle reactor that can extend **A**, the fresh feed being in **A** the two points being the mixed point and the rector outlet.

Construct the Attainable Region – Hildebrandt and Glasser[24] describe a general approach for the construction of the attainable region. The boundary is generated by the geometric properties of the fundamental processes. Trial and error, in general are required to solve the problem (see figure 4.28). A brief description of each step is given following.

Step 1 Draw the PFR(s) from the feed point(s).

Step 2 Allow mixing between all the points that can be achieved by the PFR(s). This process is known as finding the *convex hull* of the curve(s).

Step 3 Check whether any reaction vectors point out of the surface of the convex hull. If the reaction vector points over certain regions, then find the CSTR(s) with feed points in the convex hull that extend the attainable region the most. If no reaction vectors point outwards, then check whether necessary conditions (iii) and (iv) are met. If they are not met extend the region using the appropriate reactor.

Step 4 Find the new, enlarged convex hull. If a CSTR lies in the boundary at this stage, the reaction vectors must point out of the region, and the PFR with feed points on the CSTR will extend the region. Extend the region by finding the convex hull with these PFR(s) included.

Step 5 Repeat the last two steps, alternating between PFR's and CSTR's until no reaction vectors point out over the region, and necessary conditions (iii) and (iv) are met.

Interpret the Boundary – Attainable region analysis will not only show that a specific output is achievable, but will also determine the process required to achieve it. The advantage of the geometric approach to finding the attainable region is that the process layout or operating sequence can be determined directly from the geometry of the region.

For a particular output of interest there will be path from the input to the output point. This path can be interpreted in terms of the sequence of fundamental processes required to follow the path. There is usually only one path to a particular point on the boundary and an infinite number of paths to any point in the interior of the attainable region.

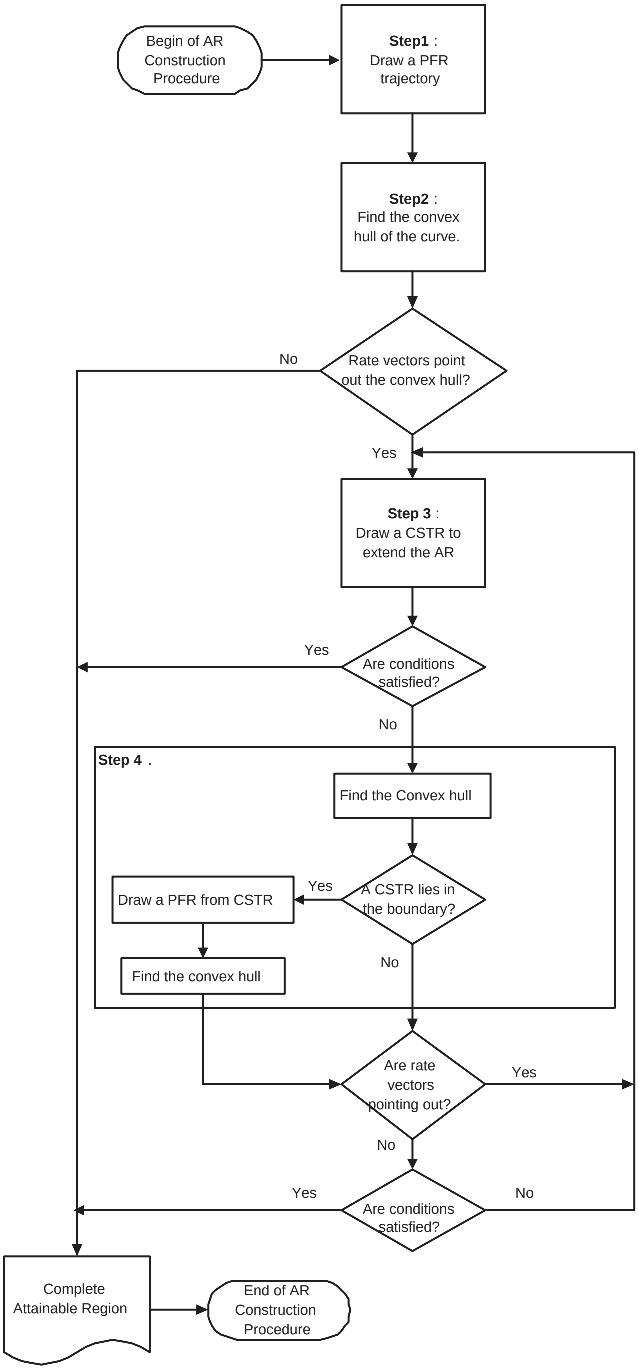


Figure 4.28: Steps for the construction of the Attainable Region

Optimize the Objective Function – The final step is the relatively straightforward procedure of finding the point where the objective function is optimized. In most instances, the objective function will be optimized at a point on the boundary. This is because the boundary points represent extremes in operation.

Since the point that optimizes the objective function has already been interpreted as a process specification and structure, the result is that attainable region analysis determines not only the best value for the objective function, but also the best process design (the best process layout or sequence and the best operating conditions).

Also, the sensitivity of the process design to changes in the objective function can easily be tested. Since the region has already been constructed all that is required is to find the point that optimizes any suitable new objective function. This point is then interpreted as a new process design.

Example of reverse simulation for the kinetic model based reactor process group

To illustrate the attainable region applied to the kinetic model based reactor process group, let us consider the kinetic model based reactor process group `(rA/pABCD)`. Equation (4.23) describes its stoichiometry and the component C is considered as the product of interest.

$$\mathrm{A} \xrightarrow{k_1} \mathrm{B}; \quad \mathrm{A} \xrightarrow{k_2} \mathrm{C}; \quad \mathrm{A} \xrightarrow{k_3} \mathrm{D} \tag{4.23}$$

where the reactions are zero, first and second order, respectively and with the following rate constant values:

$$\begin{array}{ll} k_1 = 0.025 & \mathrm{mol/(L\ min)} \\ k_2 = 0.2 & 1\ /\ \mathrm{min} \\ k_3 = 0.4 & \mathrm{L/(mol\ min)} \end{array}$$

This is the "Trambouze reaction system". The construction of the region and the selection of the design are presented in the next paragraphs.

Construction of the Region **Step 1** – Drawing the PFR trajectory from the feed point.

The equations describing the operation of a PFR for components A and C are, respectively:

$$\frac{dC_A}{d\tau} = r_A = -\left[k_1 + k_2 C_A + k_3 C_A^2\right] \tag{4.24}$$

$$\frac{dC_C}{d\tau} = r_C = k_2 C_A \tag{4.25}$$

where τ is the residence time for a constant density system.

Now, by combining equation (4.25) with equation (4.24) we obtain:

$$\frac{dC_C}{dC_A} = -\frac{k_2 C_A}{k_1 + C_A\left(k_2 + k_3 C_A\right)} \tag{4.26}$$

This reformulation will allow us to visualize, in the concentration space, the PFR trajectory of component C with respect to component A.

If we assume a feed of pure A with a concentration of 1.0 mol/L, then in figure 4.29 is described this trajectory until component A has been exhausted.

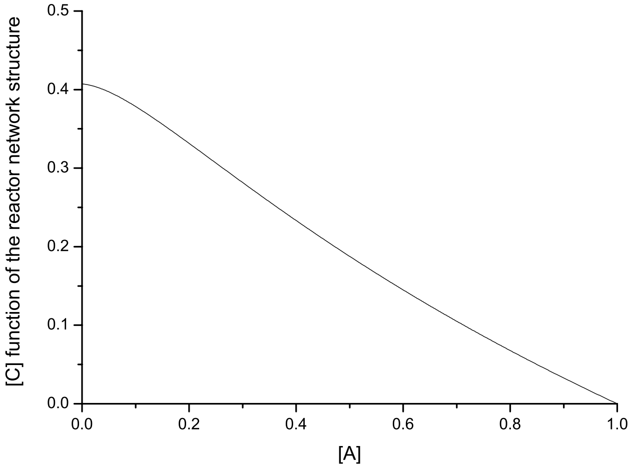


Figure 4.29: Drawing of the first PFR

Step 2 – Finding the convex hull of the curve.

As this profile is not convex, we need to construct the convex hull of this trajectory, as it is seen on figure 4.30 (dashed line).

Step 3 – Drawing the CSTR trajectory to extend the AR.

For component A, the CSTR trajectory has the following expression:

$$F_{A0} - F_A - r_A V_R = 0 \tag{4.27}$$

As in the case of the PFR, it is considered that the density remains constant, therefore, we can express the molar flowrate of A as

$$F_A = Q_R C_A \tag{4.28}$$

thus, equation 4.27 takes the following form

$$Q_R C_{A0} - Q_R C_A - r_A V_R = 0 \tag{4.29}$$

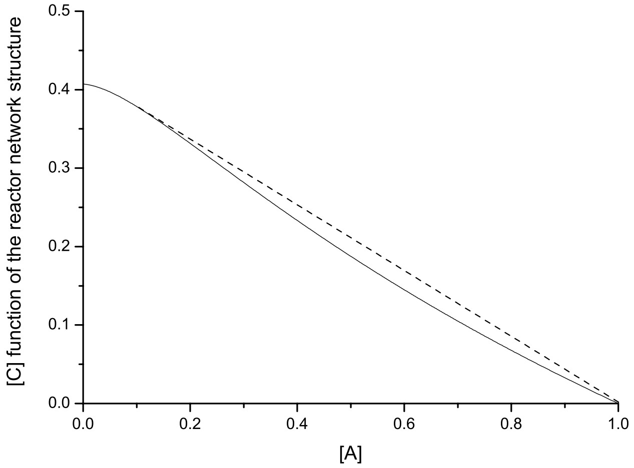


Figure 4.30: First PFR with its convex hull

if we then substitute the rate expression for A, we finally obtain

$$C_{A0} - C_A = \tau \left[k_1 + k_2 C_A + k_3 {C_A}^2\right] \tag{4.30}$$

where $\tau = V_R/Q_R$ is again the residence time.

Similarly for component C, which is the product of interest, we get

$$0 - F_C + r_C V_R = 0 \tag{4.31}$$

and

$$C_c = \tau \left[k_2 C_A\right] \tag{4.32}$$

Now, in order to visualize this trajectory in the concentration space, we find the ratio of equation (4.32) to (4.30), as we did in the case of the PFR, thus

$$C_C = (C_{A0} - C_A)\left[\frac{k_2 C_A}{k_1 + k_2 C_A + k_3 C_A^2}\right] \tag{4.33}$$

Equation 4.33 is the working equation describing the operation of the CSTR. This trajectory will start from the point which extends the region the most, which in this case corresponds to the feed of A to the process ($C_A = 1.0$) as shown on figure 4.31.

Step 4 – Finding the convex hull.

Due to the AR being still not convex (see figure 4.31), it is necessary to find the convex hull of this region, this is done in figure 4.32. This extension of the AR is equivalent to the bypass of the CSTR added in Step 3.

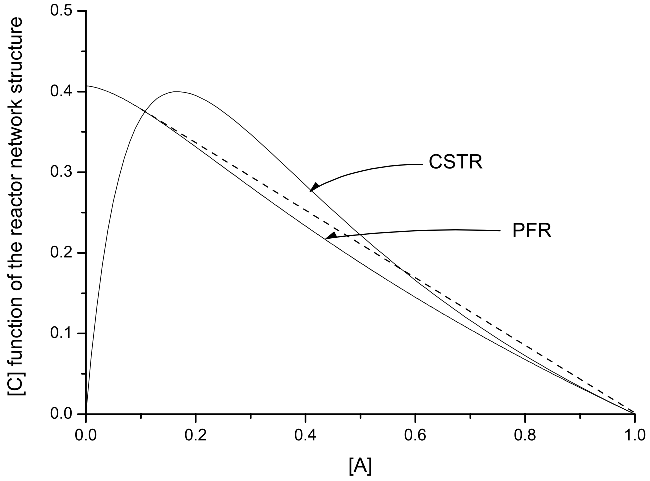


Figure 4.31: Addition of the CSTR in step 3

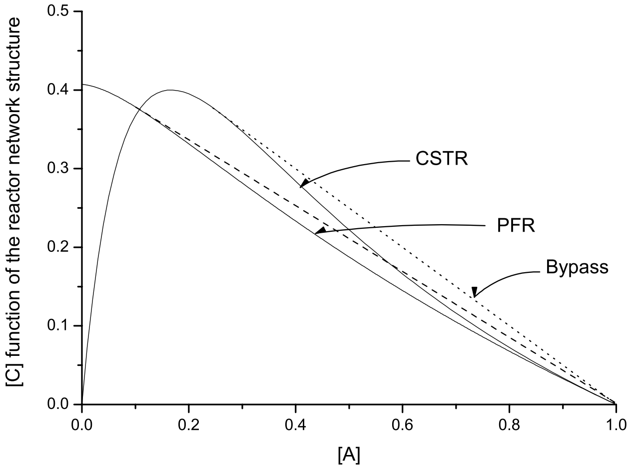


Figure 4.32: Determination of the convex hull

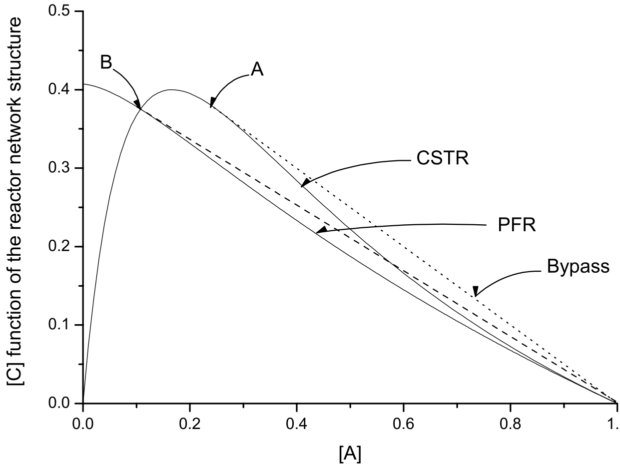


Figure 4.33: Rate vectors pointing out of the AR

Since a CSTR lies in the boundary between points A and B on figure 4.33, then a PFR is added with feed at point A.

Finally, in figure 4.34 is shown the resulting attainable region which satisfies all the necessary conditions.

Final Step – Analysis of the attainable region and optimization.

Let us say we are interested to maximize the concentration ratio of C_{C} over C_{A}: $R_C = C_{\mathrm{C}}/C_{\mathrm{A}}$. Thus, by looking at figure 4.34, the steepest slope is found anywhere along the mixing line OA. Thus, from the properties of the AR, the optimal structure with respect to the concentration ratio is a CSTR with bypass (see figure 4.35).

Selection of a Design From the AR of the Trambouze reaction system, it has been shown that if it is desired to maximize the concentration ratio R_C, then the optimal reactor structure is a CSTR with a bypass. This information permit us to complete the design of the reactor comprising the network, including streams composition and residence time in the CSTR.

Considering the goal of achieving a final outlet concentration $C_{\mathrm{C}_f} = 0.30\,\mathrm{mol/L}$ while maximizing the concentration ratio R_C. This operating condition corresponds to point C in figure 4.36 and it is achieved by mixing a point along the CSTR trajectory with the inlet to the process (point O).

In such a case, point A ensures that the mixing will lie on the boundary of the attainable region, thus maximizing the concentration ratio R_C.

From the lever arm rule, it is possible to calculate the mixing proportions between the bypass ($C_A = 1.0$ at point O) and the outlet of the CSTR (point

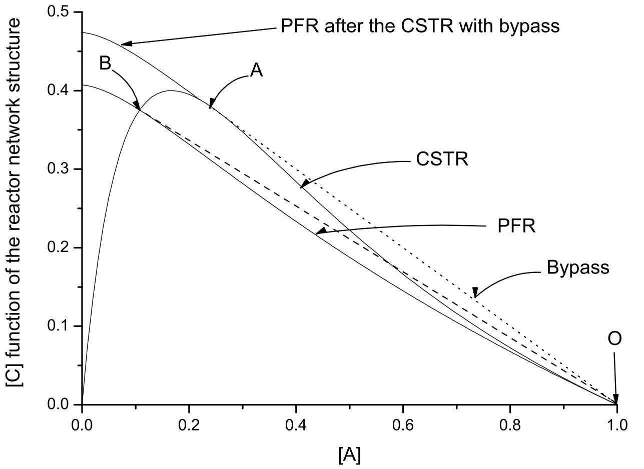


Figure 4.34: Attainable Region for the Trambouze reaction.

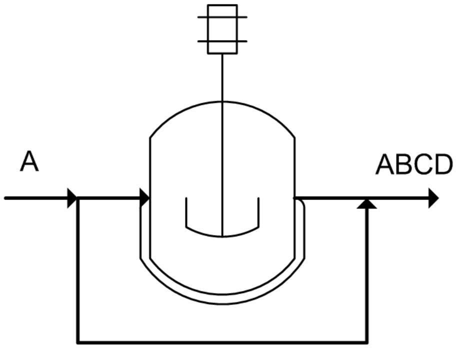


Figure 4.35: Optimal structure maximizing the selectivity to C.

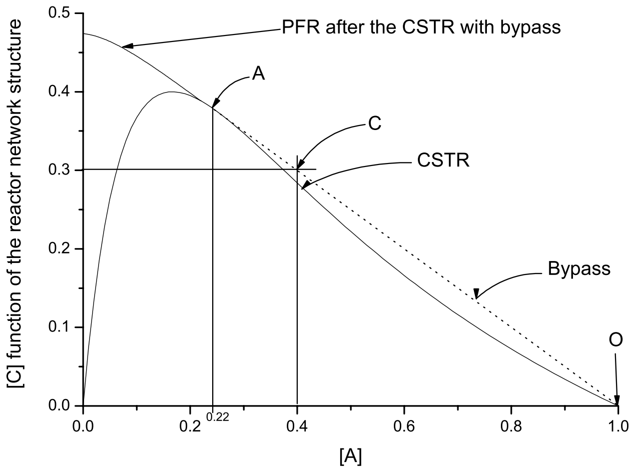


Figure 4.36: Selection of the final design.

A).

We have considered a constant density system, and on the other hand, there is no change in the volumetric flowrate due to reaction, thus, assuming a $Q_O = 1.0\,\mathrm{L/min}$, a mass balance around the mixing point C (and reading from figure 4.36) gives:

$$Q_{cstr} \times (0.4 - 0.22) = Q_{bypass} \times (1.0 - 0.4) \tag{4.34}$$

where Q_{cstr} and Q_{bypass} are the volumetric flowrates at the outlet of the CSTR and the bypass stream, respectively.

Since

$$Q_{cstr} + Q_{bypass} = 1.0\,\mathrm{L/min} \tag{4.35}$$

therefore, substituting equation (4.35) in equation (4.34) and solving for the bypass stream, gives: $Q_{bypass} = 0.23\,\mathrm{L/min}$.

The residence time τ in the CSTR can also be calculated from equation (4.32), as the concentrations of both C and A are known:

$$\tau = \frac{C_{\mathrm{C}}}{k_2 C_{\mathrm{A}}} = 8.52\,\mathrm{min} \tag{4.36}$$

Through this case study, the complete solution of the synthesis and design problem has been shown, from the kinetic expression to the final design of the reactor network.

Note that the reverse approach is a short-cut method, it does not guarantee that the design will be the final optimal design, but it does guarantee that the design will match the constraints and will be a feasible design. It is also a very quick way to get estimates of the important design parameters. Finally, this particular method has been selected for the CAFD approach because of its similarity with the driving force approach for separation processes. In principle, the driving force concept and the attainable region concept are the same. In the same way the driving force diagram provides a connection between the inlet stream and the outlet streams of the separation process groups, the attainable region and the interpretation of its boundary provide a connection between the inlet stream and the outlet stream of the reactor process group. Furthermore, the attainable region provides the reactor network configuration and the corresponding design parameters when an operating point, optimizing the defined objective function, is selected.

5

CAFD Framework and Tools

In this chapter, first the description of ICAS[16], the Integrated Computer Aided System of CAPEC is presented. This software package is integrating the methods and tools developed within CAPEC. Then ProCAFD, the computer implementation of the CAFD framework, is presented. This presentation includes the work flow and the data flow of a process synthesis problem resolution, the ProCAFD software architecture and some details about the integration of ProCAFD within ICAS. Finally, to conclude this chapter some perspectives – results of the experience in managing ICAS and developing ProCAFD during the last three years – about the future of an integrated system are presented.

5.1 ICAS – Integrated Computer Aided System

This work has been carried out within CAPEC, the Computer Aided Process Engineering Center, a research center of the Chemical Engineering Department at the Technical University of Denmark. The main purpose of CAPEC is to develop systematic methods and tools to assist in the resolution of process engineering problems. As a result of this research and development work, a significant body of programs and tools are available within CAPEC for the researcher to capitalize upon.

ICAS[16] is a package of programs and tools communicating with each other to assist an engineer in the resolution of product-process engineering problems. The current version of ICAS is ICAS 8.0, corresponding to more than eight years of development by researchers and students.

The 19 main programs available from the ICAS package are the following:

BRIC – BRIC is the batch operation software. It is used to setup and simulate batch operations like the successive steps of a batch distillation. It is using the DynSim simulator.

CAPSS – CAPSS tool for synthesis of separation trains. Based on the physical insights of the mixture to separate, the appropriate separation tasks are proposed to generate the separation flowsheet.

Database manager – The database manager is the program managing the various databases. The CAPEC databases contain properties and model parameters for more than 12000 components and 400 salts.

Control toolbox – The Control toolbox is used to test control systems of chemical processes through closed-loop dynamic simulations.

DynSim – DynSim is the rigorous dynamic simulator of ICAS.

Heat Integration toolbox – The heat integration toolbox is providing tools to do the pinch analysis of a flowsheet and to determine/analyse the appropriate heat exchanger networks.

ICASSim – ICASSIM is the rigorous steady state simulator of ICAS.

ModDev – ModDev is the Model Development toolbox. This toolbox is automatically generating the model equations from model building blocks. The resulting model can then be solved with MoT.

MoT – MoT, the Modeling Toolbox, was designed with the objective of minimizing the amount of effort needed to specify, solve, and visualize the solution of a system of equations, which could be of algebraic equations (AEs), ordinary and partial differential equations (ODEs or PDEs).

Optimization – The optimization toolbox is used for steady state process optimization and/or solution of non-linear equations.

PDS – Process Design Studio is a design toolbox specialized in design and synthesis distillation columns (simple or complex).

ProCAFD – The implementation of the CAFD method presented in this thesis for the synthesis and design of chemical processes including both reaction and separation.

ProCAMD – ProCAMD is the implementation of the CAMD method. ProCAMD is used to synthesis and design pure components and mixtures from molecular groups and pure component or mixture property targets.

ProPred – ProPred is the Property Prediction tool of ICAS. Given a molecular structure a set of 34 pure component properties can be predicted, from the melting point temperature to the acute toxicity of the component.

Reaction toolbox – The reaction toolbox is a tool to access and update the reaction database. One of its components, the kinetic parameters estimation toolbox, is providing the framework to regress the kinetic parameters of reactions from experimental data. The current database contains 80 reactions.

SoluCalc – SoluCalc is a Solubility Calculation tool for complex chemicals. It is also providing an integrated way to generate from experimental data a custom UNIFAC model for a given set of components.

TML – TML is the Thermodynamic Model Library. It contains a large collection of the thermodynamic models and is linked to the various databases to get access to the needed parameters.

TMS – TMS is the Thermodynamic Model Selection tool. This tool is used to find the right thermodynamic model for a given system. It is knowledge based tool.

WAR toolbox – The WAR toolbox is providing an automated way to calculate the environmental impact of a flowsheet using the WAR algorithm[50] developed by the EPA.

The ICAS package, by the diversity of the available tools it provides, is used to solve a large range of problems within nearly all the stages of the development of a product-process. Tools are available to assist the engineer from the early stages of product-process design with component property analysis through ProPred, chemical design/selection through ProCAMD, to the stages related to control/operation of a process using the Control toolbox and the DynSim simulator. Process synthesis and analysis toolboxes are represented by CAPSS, PDS, ProCAFD, the WAR toolbox and the Heat Integration toolbox. All these tools are relying on the CAPEC database, a comprehensive database of pure component and mixture properties.

5.2 ProCAFD – Computer Aided Flowsheet Design

As presented in the overview of ICAS, ProCAFD, the computer implementation of the CAFD framework, is part of the process synthesis and design tools, including CAPSS and PDS. In this section, first the work flow of the resolution of a synthesis problem is presented. This work flow also provides insights on the integration of ProCAFD within the ICAS package. Then the ProCAFD software architecture is presented, with the translation of the CAFD method into program data structures. Next the computational challenges related to the implementation of the different methods and algorithms is presented, together with a presentation of the methods applied to ensure the reliability of the ProCAFD results.

5.2.1 The ProCAFD work flow

In this section, the CAFD method work flow is presented with the tools used at each step in the resolution of the synthesis problem. Emphasis is given on how the new tool is integrated with the tools available in ICAS as presented in section 5.1.

The ProCAFD work flow is composed of eight steps from which three are optional steps depending on the synthesis problem being solved. Figure 5.1 (page 83) illustrates this work flow together with the methods and external tools being used at each resolution step. At each step, information is needed from the user or created using the in-house methods and tools. The data flow resulting from this creation and manipulation of knowledge is presented in figure 5.2 (page 84). This figure shows, for each resolution step, the information coming from the user and the data generated from the user input and the different tools used during the step.

In the next paragraphs, each step of the process synthesis resolution procedure shown on the work flow and data flow figures are described.

Step 1. Problem definition. The user is providing the general problem definition which is refined during the problem analysis. The necessary user input are the components (chemical) identities, the inlet and outlet stream data (for example, pressure, temperature and individual flowrates) of the desired flowsheet, together with specifications for target properties. In the case of a retrofit at the process level, the fixed part of the process to be kept are also necessary.

Step 2. Problem analysis. As not enough information is available from the problem definition, the problem analysis is performed to generate all the needed information for the process alternatives synthesis and design steps. Several ICAS tools are used for that purpose. The reaction analysis is based on the reaction database of the reaction toolbox. If reaction kinetics data are not available but experimental data are, the reaction toolbox is providing the necessary tools to regress kinetic parameters based on the experimental data.

From the nature of the components, the use of TMS, the thermodynamic model selection toolbox, assists in selecting an appropriate thermodynamic model from those available through TML and to calculate the needed pure component and mixture properties.

If solvents are needed, they can be identified through a database search and/or through ProCAMD.

Step 3. Process group selection and initialization. With the usable knowledge available from the synthesis problem analysis, ProCAFD provides a list of process groups that could be used in the specified problem. The user can then refine the choice from his own experience or with the help of a possible knowledge base. Note that the development of a knowledge base was not part of the objectives of this project.

Step 4. Flowsheet structure alternatives synthesis and test. The synthesis of alternatives is performed by ProCAFD through the algorithm

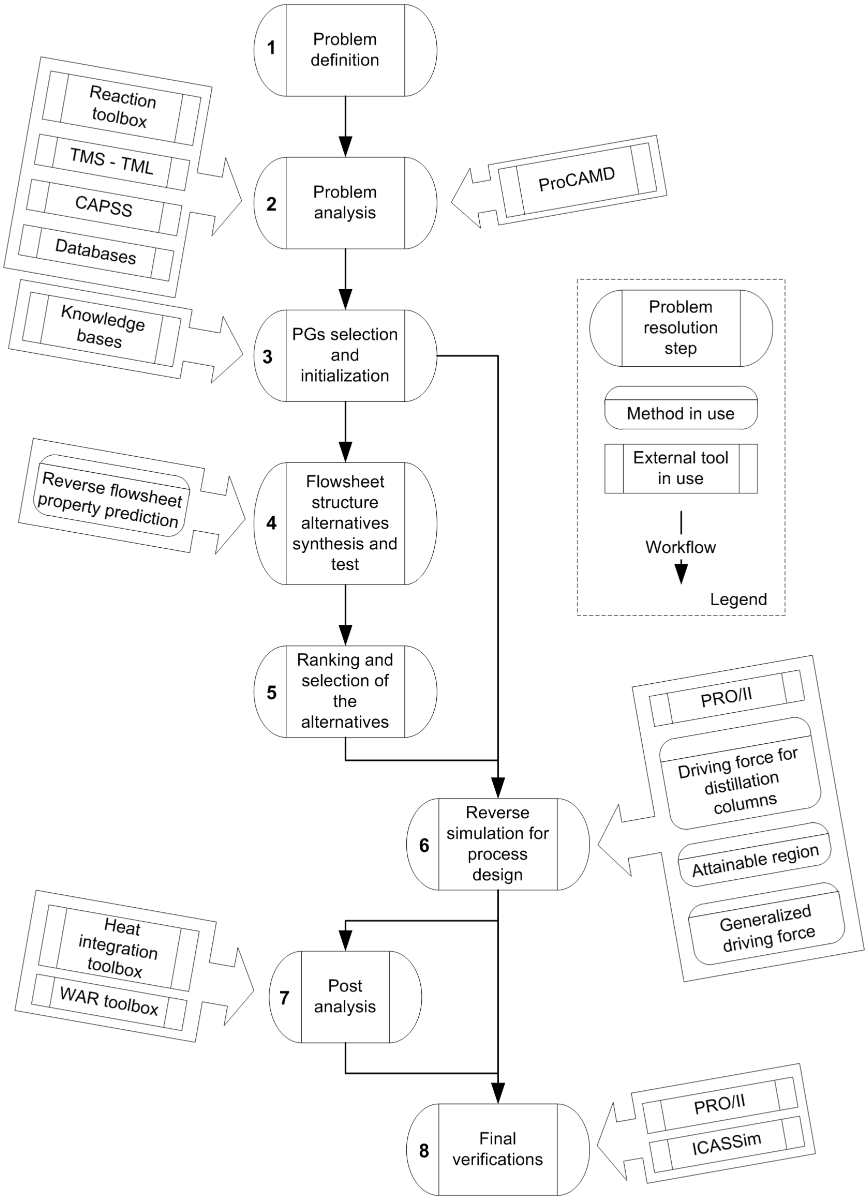


Figure 5.1: The problem resolution work flow of the CAFD framework.

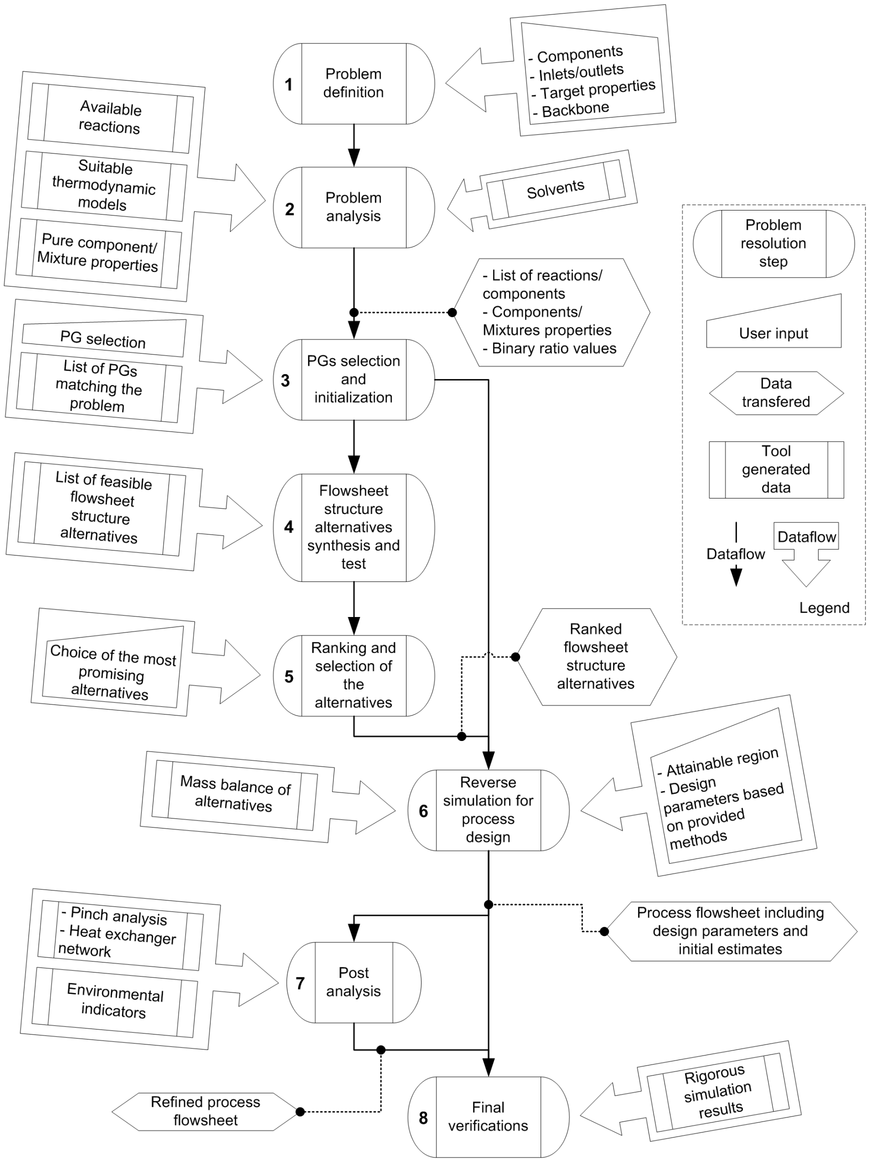


Figure 5.2: The problem resolution data flow of the CAFD framework.

presented on figure 4.17 (page 53) and detailed in section 4.2.3 page 52. During the generation of the flowsheet structure alternatives, additional user input is not needed.

Step 5. Ranking and selection of the alternatives. The ranking is based on the methods provided in section 4.2.4 page 54 and the property targets selected during the synthesis problem definition. The user has then the freedom to select from within the list of alternatives the one he considers as being the most promising for further study. This decision is left to the experience of the user.

Step 6. Reverse simulation for process design. The reverse simulation for process design is composed of two steps for each promising flowsheet structure alternative: the resolution of the mass balance between the process groups and the definition of the design parameters of the unit operations. The resolution of the mass balance can be performed through any simulator (in the first ProCAFD version, the PRO/II simulator has been used). For the definition of the design parameters of the process flowsheet, the reverse simulation algorithms presented in section 4.3 page 56 have been used.

Step 7. Post analysis. The post analysis, which is not a mandatory step, could be followed to improve a selected alternative and to discriminate between several alternatives. The improvements are made in terms of heat integration using the ICAS heat integration toolbox. The heat integration toolbox is providing pinch point analysis based determination of operationally feasible heat exchanger networks. Further discrimination between the promising alternatives can be performed using the WAR algorithm. With the use of the ICAS WAR toolbox, it is possible to automatically calculate the environmental impact of a process flowsheet knowing the conditions and the component flowrates of the process inlet and outlet streams.

Step 8. Final verifications. The final verifications are performed by the user, based on his experience. The use of a rigorous simulator like PRO/II or ICASSim is also possible to further refine the alternatives and to perform optimization of the design parameters.

The process synthesis problem resolution steps presented here have two important characteristics, on one hand the algorithms are both complex and managing complex data structures and on the other hand an important exchange of data between different tools is required. This complexity is addressed by the development of the appropriate software architecture presented in the next section.

5.2.2 Software architecture

As seen in the presentation of the ProCAFD work flow (see figure 5.1), the integration of tools and data is needed to successfully solve a process synthesis problem. In this section, the ProCAFD software architecture is presented. The software architecture of ProCAFD has been developed with a particular attention to the current and future integration needs. This has been possible by the use of the object oriented programming (OOP) paradigm.

First a short introduction to the object oriented programming paradigm is given, and then, the different objects used in ProCAFD and how they are needed to solve the process synthesis problem efficiently are presented.

The object oriented programming paradigm

The idea behind object-oriented programming is that a computer program is composed of a collection of individual units, or objects, as opposed to a traditional view in which a program is a list of instructions to the computer. Each object is capable of receiving messages, processing data, and sending messages to other objects.

Object-oriented programming is claimed to give more flexibility, ease of changes to programs, and is widely popular in large scale software engineering. Furthermore, proponents of OOP claim that OOP is easier to learn for those new to computer programming than previous approaches. Also, they claim that the OOP approach is often simpler to develop and to maintain, thereby lending itself to more direct analysis, coding, and understanding of complex situations and procedures than other programming methods[32].

The objects can have three types of functionalities: the storage of the data, the presentation of the data and the manipulation and interaction with the data. A software architecture presenting a clear distinction between these three functionalities are said to follow the *model*, *view* and *controller* (MVC) design pattern.

A process flowsheet is a good example to illustrate the MVC architecture. In practice, a process flowsheet is presented in two forms, the flowsheet drawing within a process simulator and a text format report providing the details of all the elements in the flowsheet. Within the simulator interface, the user can add new unit operations or new details to already available unit operations. Considering this common interaction with a flowsheet, four objects – one *model*, two *views* and one *controller* – can be used to represent those interactions and the data models. These objects and their interactions are presented on figure 5.3.

The flowsheet *data model object* The flowsheet data model object is the storage all the data related to the flowsheet, like the number of unit operations and their details, the components in the flowsheet or the thermodynamic models in use. When it is modified it is updating the views.

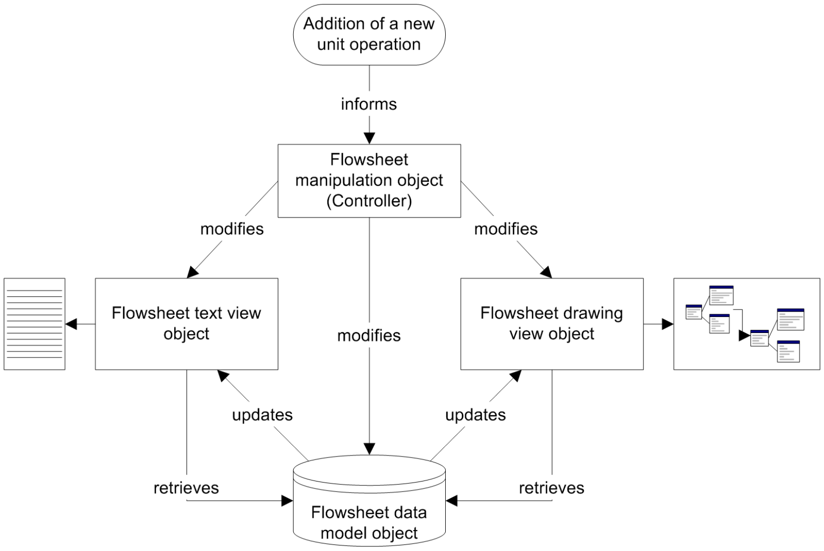


Figure 5.3: Illustration of the MVC architecture for the process flowsheet.

The text *view object* The text view object is an object that will access the flowsheet data model and from the data available in the flowsheet object, will create a text report ready to be printed.

The flowsheet *view object* The flowsheet view object is an object that will access the flowsheet data model and from the data available will create the flowsheet drawing in the process simulator.

The flowsheet manipulation object The flowsheet manipulation object is a *controller* which is taking the input of the user, for example, the need to add a new unit operation, and is updating the flowsheet *data model* accordingly. The *controller* can also modifies the *views*, for example it can modifies the text view object to change the printing of the measurement units from the SI system to the US system.

The ideal situation is to have a clear distinction between the models, the views and the controllers, but sometimes it is more convenient to have one object carrying more than one functionality. This is for example the case when the view is very simple.

The ProCAFD objects

In this section the main objects of the ProCAFD software and their use during the process resolution steps are presented. Emphasis is given on how the steps in the work flow (see figure 5.1 page 83) of the resolution of the process synthesis problem can be achieved efficiently by developing the appropriate software architecture for the implementation of the CAFD method. Note that only the core objects of ProCAFD are presented and not the objects related to the graphical user interface.

To solve the algorithms of the work flow resolution steps, objects are created, duplicated, modified or deleted. All these operations are coordinated through data exchange between the different objects and influenced by the requests made by the user through some of the controller objects. In a way, the object architecture of ProCAFD to solve a given algorithm can be compared to the operation of a biological cell. To achieve a specific macro level functionality, an example of a biological cell, composed of a lot of distinct elements interacting with each others, is used. The cell nucleus is the information center of the cell. The ribosomes are in charge of processing information from the cell nucleus to create proteins and the Golgi apparatus is a controller, ensuring that specific molecules produced by the ribosomes are transfered to the right places.

The process synthesis problem object. The process synthesis problem object is a data container. The purpose of this object is to store all the information needed during the resolution of the synthesis problem. It is the object that is saved and restored when saving and loading a process synthesis problem. At any time, only one instance of the process synthesis problem object exists within ProCAFD, as a critical information center, it could be compared to the cell nucleus.

As illustrated on figure 5.4 with a diagram of some of the data exchange between the process synthesis problem object and some of the ProCAFD objects, the process synthesis problem object is accessed by the objects in need of information about the synthesis problem. The other objects are also able to store results within process synthesis problem object, like for example, the list of generated structure alternatives.

The flowsheet structure object. The flowsheet structure object, as shown on figure 5.5, represents a flowsheet structure as defined in the CAFD method. The flowsheet structure object is composed of process group objects together with their connections. This object is providing a set of methods to update the information of the flowsheet structure and to query some details.

The methods to update the information are to add a process group to the flowsheet structure, to close a connection of a process group with a terminal process group or to connect two process groups within the flowsheet structure. The methods to query the flowsheet structure for information are to get a given property of the flowsheet structure, get the list of all the free connections in the structure or find the connection matching some given criteria. Methods are available to get the SFILES notation of the flowsheet structure and compare the flowsheet structure with another for equality.

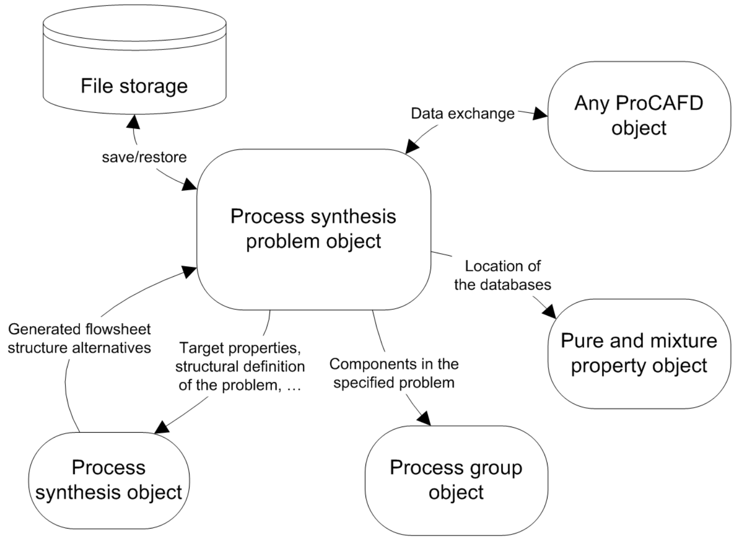


Figure 5.4: Data exchange between the synthesis problem object and other objects.

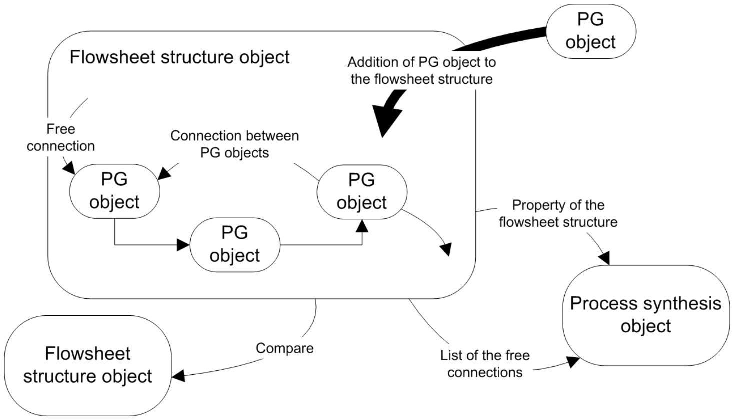


Figure 5.5: Diagram of a flowsheet structure object.

The process group object. The process group object represents a process group as defined in the CAFD method. Different types of process groups exist, one for each process group available in the CAFD method. As for the flowsheet structure object, the process group object is providing methods to update the information and query the information of the process group, but it is also providing methods to store and load itself in and from the databases, and initialization methods.

The method to update the information is to connect the process group to another within a flowsheet structure. The method to query the process group is to retrieve the free connections, the connections matching a given connection from another process group or the contribution of the process group for a given flowsheet property model. As illustrated in the diagram on figure 5.6, the process group object provides the list of mixtures matching with its property dependencies for the components in the specified synthesis problem. It is providing the initialization method from a matching mixture and the synthesis problem. This method, together with the method to get the list of matching mixtures, is used in step 3 of the CAFD framework. A method is available to get the SFILES notation of the process group.

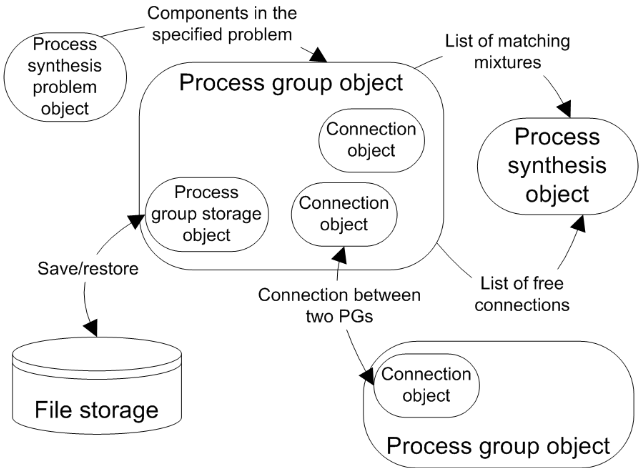


Figure 5.6: Diagram of a process group object with some of its interaction with other objects.

The process group storage object. The process group storage object, as illustrated on figure 5.6, is used to save and load a given process group data. When a process group object needs to be loaded or saved, the corresponding process group storage object is created and populated with the process group object data. The use of a process group storage object independent from the process group object is to be able to easily switch from a text file storage to a relational database storage and this independently of the nature of the process group.

The process group connection object. The process group connection object represents the connection between one process group and another. A connection is either an inlet or an outlet. As shown on figure 5.6, a *complete* connection between two process groups is composed of one inlet connection and one outlet connection. As a connection is simple, the only method to update the state of a connection is to connect a connection to another connection or to no connections.

The pure component and mixture property object. The pure component and mixture property object represents the properties of a component or a binary mixture of components. From this object it is possible to access the 22 pure properties of a component (see appendix A page 173) but also, for binary mixtures, the driving force between the components and the presence of azeotropes. This object also provides methods to compare components for property similarities. As shown on figure 5.7, this object provides an uniform way to retrieve pure component and mixture properties even if the sources of the data are heterogeneous.

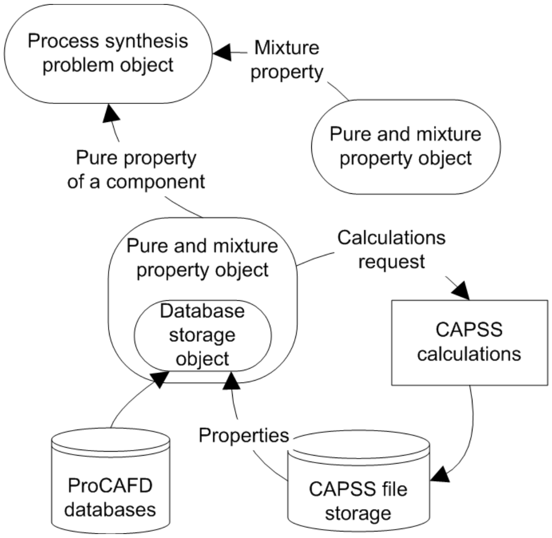


Figure 5.7: Diagram of the pure and mixture property object

The database manager object. The database manager object is a simple object to retrieve and update information in the pure component and mixture property databases. It is used by the objects needing data from the databases, like for example the pure component and mixture property object as shown on figure 5.7.

The SFILES notation object. The SFILES notation of a flowsheet structure is created by the sfiles object, a view object. This object is interesting as it is setup in a domain agnostic way. For any graph representation, the corresponding SFILES notation can be generated. The only constraint is to provide the graph representation in a defined format.

When the SFILES notation of a flowsheet structure is requested, first the flowsheet structure is translated into the defined graph representation format. This representation is then given to the sfiles object. The sfiles object is generating the SFILES notation that is sent back to the flowsheet structure object.

The process synthesis object. The process synthesis object, a controller object, is the object that is accessing all the data models to generate new flowsheet structure alternatives. As the data models are clearly defined, the complexity of the process synthesis object is very low with only a few main methods to load the problem definition from the process synthesis problem object, to perform the preparation of the flowsheet structure starting points, to check if a flowsheet structure has already been analysed or not and sort the resulting flowsheet structures by their properties, together with the main resolution method. It could be compared to the Golgi apparatus in a biological cell, as it is acting as the coordinator of a range of objects.

For example, figure 5.8 illustrates the management by the process synthesis object of the addition of a matching process group object to a flowsheet structure from the working list. First a flowsheet structure is taken from the list of working structures, then its free connections are retrieved (see details on figure 5.5). For each free connection, if a process group object matching the connection is available, the process group object is added and connected to a copy of the flowsheet structure object. Finally, this new flowsheet structure object is added to the list of working structures. When comparing these operations with the theoretical algorithm on figure 4.17 (page 53), it can be seen that this choice of an appropriate software architecture is providing a way to solve the synthesis problem without introducing unnecessary complexity.

The export objects. The export objects are objects used to convert a flowsheet structure into a mass balance flowsheet that can be imported into the PRO/II simulator. They are converting the data from one format to another.

The software architecture of ProCAFD, with objects having a meaning both at the theoretical level as well as at the computational level provides solid foundation to future extensions of the software. The object oriented architecture, with a clear and intuitive correspondence between the theory and the practical implementation, is reducing the complexity of the software to

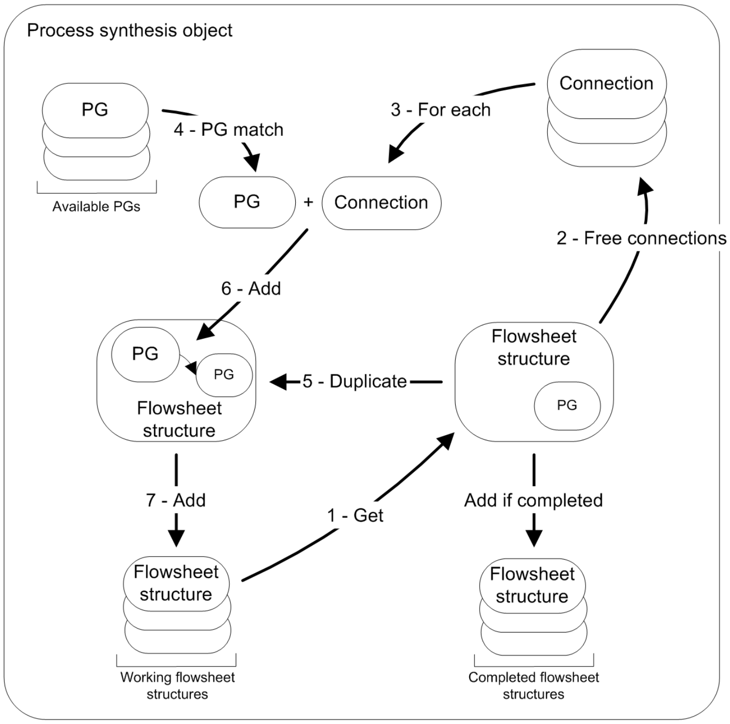


Figure 5.8: Addition of a new process group to a working structure.

the complexity of the algorithms and concepts being manipulated without any extra programming complexity.

5.2.3 Computational challenges

In the alternative generation algorithm of the CAFD method (see figure 4.17 page 53), three lists of flowsheet structure alternatives are needed, the already analysed structures, the completed structures and the working structures. All the flowsheet structures in these lists are unique. This means that before adding a flowsheet structure to a list, the program verifies if an identical structure is in the list or not, if not the flowsheet structure is added to the list.

The software implementation of those three lists of flowsheet structures is composed of five lists: three lists with the SFILES notations of respectively the analysed flowsheet structures, the completed flowsheet structures and the working flowsheet structures, and, two lists with respectively the completed flowsheet structures and the working flowsheet structures.

When adding a new flowsheet structure to a list, it is possible to quickly verify if the structure is already in the list by simply looking in the list of corresponding SFILES notations. This lookup is only composed of fast character string comparisons.

If the SFILES notation would have not been available, the lists would have stored a flowsheet structure object for each flowsheet alternative. So when verifying that a flowsheet structure is not already in a list before adding it, the flowsheet structure to be added is compared with all the existing flowsheet structures in the list. As each comparison is a graph comparison – a graph comparison corresponds to a non-deterministic polynomial (NP) problem to solve – the resulting computational cost would have been very high.

A comparison between the algorithm with or without the use of the SFILES notation is presented in table 5.1. The problem solved in this comparison corresponds to the separation of a five component mixture into five high purity streams using 20 distillation process groups involving two-five component separations. The comparison shows that even on a small problem the use of the SFILES notation is significantly more efficient. It should be pointed out that the use of the SFILES notation also reduces the computer memory usage as the list of already analysed flowsheet structures is storing only the SFILES notations and not complex objects.

Number of completed structures	14
Number of analysed structures	651
Computer time without SFILES	33.2s
Computer time with SFILES	11.6s

Table 5.1: Speed results in the implementation of the SFILES notation within ProCAFD.

5.2.4 Software integrity and documentation

As ProCAFD is intended to be used by engineers from industry, special attention has to be given to the reliability of the results. This can be achieved only by taking care that during all the steps of the software development, the integrity of the code is ensured. This has been achieved mainly by developing automated tests with rigorous code documentation, and at a higher level, the development of a code revision architecture and automatic installer creation.

Unit testing and code documentation

As discussed in section 5.2.2 (page 86), the ProCAFD software architecture is composed of multiples objects, each one interacting with other objects. The exchange of information (data) between two objects is performed through an interface. The direct implication is that the modification of one object can affect another. This situation is often known as *fixing a bug by introducing another one*. To prevent this, automated tests and strict documentation guidelines have been enforced.

The automated tests in ProCAFD are known as *unit testing* in computer science. A *unit test* is the test of a single functionality at the code level. For example, when testing the flowsheet structure object some of the tests are:

- The test of the addition of a new process group to the current flowsheet structure.
- The test of the creation of a connection between two process groups within the flowsheet structure.
- The test of the number of available free connections in a flowsheet structure.
- The test of equality between two flowsheet structures.

The constant updating and addition of tests has proved to be very useful during the development cycle. For example, when the equality check of two flowsheet structures has been changed from a graph comparison to a SFILES notation based comparison, the unit tests have ensured that no functionalities have been broken during this change. Currently 114 unit tests are performed automatically to test the ProCAFD code.

As all the unit tests are documented, they also provide an indirect manual on how to use the objects within the code. Documentation of the code is critical, especially considering that the ProCAFD software development will be taken over by someone without any overlap with the previous developer. Within the ProCAFD code, each method or function is documented together with the non obvious parts of the algorithms.

Code revision architecture and build automation

The code revision architecture is an history manager. It is currently tracking all the changes in the CAPEC software during the last three years and the exact configuration of each version of the CAPEC software delivered to the end users. Thanks to this system, it is possible to see exactly what has been changed between a working version and a version with a bug. This drastically reduces the time to find the origin of a bug.

Considering a bug to be solved, ProCAFD together with the complete ICAS package may need to be rebuild, repackaged with all the documentation, tutorials and examples into an installer ready to be delivered. This represents approximately 1500 files. The number of possible errors is minimized by reducing the number of manual steps to create a new version of the complete ICAS package to only one: a click on a web page. A simple click is launching the retrieval on a dedicated computer of the very last version of the complete ICAS package code, the compilation of all the code, the generation of the documentation, the creation of the ICAS package installer and the delivery of the installer on the member company website for download. This is taking approximately two hours, without any human interventions.

5.3 ProCAFD Manual

In this section, ProCAFD is presented through a case study involving the separation of a mixture composed of 8 alcanes Propane, Isobutane, n-Butane, Isopentane and n-Pentane. The case study highlights the ease of use of ProCAFD for the generation of flowsheet alternatives together with the reverse simulation of a flowsheet alternative. The numbering of the steps corresponds to the steps in the work flow of the framework for CAFD presented in figure 5.1 (page 83).

5.3.1 Step 1 – Definition of the problem

The synthesis problem is to separate the mixture defined in table 5.2 at a minimum energy cost, into 5 pure products.

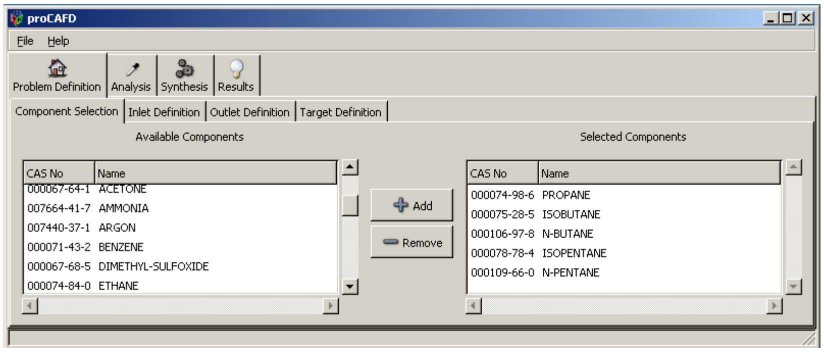
Components	**Feed flowrates** ($kmole/hr$)
A – Propane	45.4
B – i-Butane	136.1
C – n-Butane	226.8
D – i-Pentane	181.4
E – n-Pentane	317.5

Table 5.2: Feed mixture for the hydrocarbon separation.

The structural definition of the synthesis problem is as follows: 1 inlet and

5 outlets of pure products. The flowsheet property target is the minimization of the energy consumption index E_x.

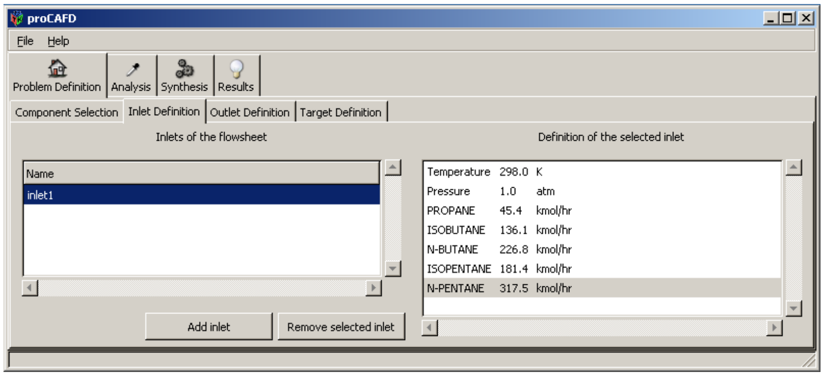
Within ProCAFD, the definition of the problem is performed by first selecting the components involved in the problem (see figure 5.9), then defining the inlet stream (see figure 5.10) and finally the outlet streams (see figure 5.11). At the current stage, ProCAFD implements only the minimization of the energy consumption index.



CAS No	Name
000067-64-1	ACETONE
007664-41-7	AMMONIA
007440-37-1	ARGON
000071-43-2	BENZENE
000067-68-5	DIMETHYL-SULFOXIDE
000074-84-0	ETHANE

CAS No	Name
000074-98-6	PROPANE
000075-28-5	ISOBUTANE
000106-97-8	N-BUTANE
000078-78-4	ISOPENTANE
000109-66-0	N-PENTANE

Figure 5.9: ProCAFD: definition of the components in the problem.



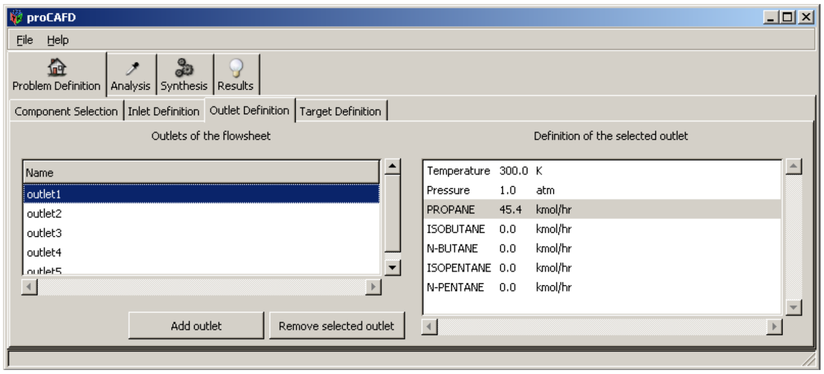
Name
inlet1

Temperature	298.0	K
Pressure	1.0	atm
PROPANE	45.4	kmol/hr
ISOBUTANE	136.1	kmol/hr
N-BUTANE	226.8	kmol/hr
ISOPENTANE	181.4	kmol/hr
N-PENTANE	317.5	kmol/hr

Figure 5.10: ProCAFD: Definition of the inlet stream.

5.3.2 Steps 2/3 – Problem analysis, and, PGs selection and initialization

As the synthesis problem is defined, ProCAFD performs automatically the analysis of the pure component (see table 5.4) and mixture properties. As shown on figure 5.12, ProCAFD automatically identifies the process groups having property dependence matching with the mixtures available from the

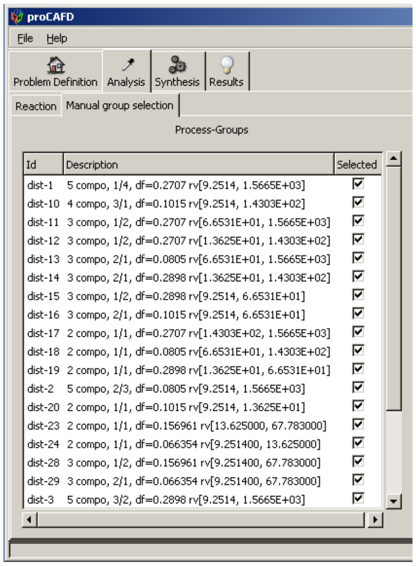


Name
outlet1
outlet2
outlet3
outlet4

Temperature	300.0	K
Pressure	1.0	atm
PROPANE	45.4	kmol/hr
ISOBUTANE	0.0	kmol/hr
N-BUTANE	0.0	kmol/hr
ISOPENTANE	0.0	kmol/hr
N-PENTANE	0.0	kmol/hr

Figure 5.11: ProCAFD: definition of the outlet streams.

synthesis problem. The user can manually reduce the number of selected process groups. Details about the selected process groups, such as, type of operation task performed, mixture initialization or reason for the pre-selection, are available to the user.



Id	Description	Selected
dist-1	5 compo, 1/4, df=0.2707 rv[9.2514, 1.5665E+03]	☑
dist-10	4 compo, 3/1, df=0.1015 rv[9.2514, 1.4303E+02]	☑
dist-11	3 compo, 1/2, df=0.2707 rv[6.6531E+01, 1.5665E+03]	☑
dist-12	3 compo, 1/2, df=0.2707 rv[1.3625E+01, 1.4303E+02]	☑
dist-13	3 compo, 2/1, df=0.0805 rv[6.6531E+01, 1.5665E+03]	☑
dist-14	3 compo, 2/1, df=0.2898 rv[1.3625E+01, 1.4303E+02]	☑
dist-15	3 compo, 1/2, df=0.2898 rv[9.2514, 6.6531E+01]	☑
dist-16	3 compo, 2/1, df=0.1015 rv[9.2514, 6.6531E+01]	☑
dist-17	2 compo, 1/1, df=0.2707 rv[1.4303E+02, 1.5665E+03]	☑
dist-18	2 compo, 1/1, df=0.0805 rv[6.6531E+01, 1.4303E+02]	☑
dist-19	2 compo, 1/1, df=0.2898 rv[1.3625E+01, 6.6531E+01]	☑
dist-2	5 compo, 2/3, df=0.0805 rv[9.2514, 1.5665E+03]	☑
dist-20	2 compo, 1/1, df=0.1015 rv[9.2514, 1.3625E+01]	☑
dist-23	2 compo, 1/1, df=0.156961 rv[13.625000, 67.783000]	☑
dist-24	2 compo, 1/1, df=0.066354 rv[9.251400, 13.625000]	☑
dist-28	3 compo, 1/2, df=0.156961 rv[9.251400, 67.783000]	☑
dist-29	3 compo, 2/1, df=0.066354 rv[9.251400, 67.783000]	☑
dist-3	5 compo, 3/2, df=0.2898 rv[9.2514, 1.5665E+03]	☑

Figure 5.12: ProCAFD: Selection of the process groups.

From the analysis of the structural definition of the synthesis problem, no reactions have been found necessary. From the analysis of the mixture, distillation has been found to be the appropriate separation technique for all the separation tasks that needs to be performed. 20 initialized simple distillation process groups are selected to perform the synthesis of the flowsheet alternatives. These initialized process groups are given in table 5.3.

5 component groups	**4 component groups**	**3 component groups**	**2 component groups**
(A/BCDE)	(A/BCD)	(A/BC)	(A/B)
(AB/CDE)	(B/CDE)	(B/CD)	(B/C)
(ABC/DE)	(AB/CD)	(AB/C)	(C/D)
(ABCD/E)	(ABC/D)	(BC/D)	(D/E)
	(BC/DE)	(C/DE)	
	(BCD/E)	(CD/E)	

Table 5.3: Process-groups for a 5 component simple distillation separation

5.3.3 Step 4 – Generation of the flowsheet structure alternatives

Given the mixture of 5 components defined in table 5.2, to be separated into 5 highly pure products (purity above 99%), using distillation columns represented by 20 process groups (see table 5.3), at the minimum consumption of energy, the flowsheet generation algorithm of the framework for CAFD generates 14 feasible alternatives. Within ProCAFD, after the definition of the synthesis problem and the selection of the initialized process group, as no additional input is needed from the user, the user requests the generation of the alternatives (see figure 5.13) and waits for the generation to be completed (see figure 5.14). The 14 generated feasible flowsheet alternatives are given in table 5.5. Details of the generation, including the number of analysed structures and speed of generation, have already been given in table 5.1 (See section 5.2.3 page 94 on the *Computational challenges* of ProCAFD).

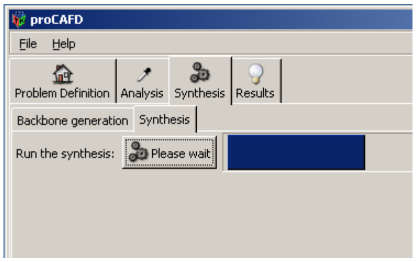


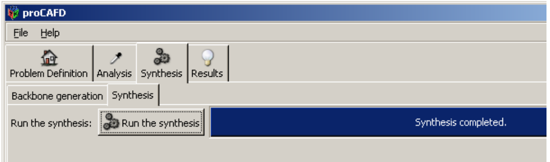
Figure 5.13: ProCAFD: Generation of the flowsheet alternatives.

Property	Propane	i-Butane	n-Butane	i-Pentane	n-Pentane
M_w	44.096	58.123	58.123	72.15	72.15
ω	0.1523	0.1808	0.2002	0.2275	0.2515
T_c	369.83	408.14	425.12	460.43	469.7
P_c	41.924	36.003	37.464	33.368	33.259
Z_c	0.276	0.282	0.274	0.27	0.27
V_c	0.2	0.3	0.3	0.3	0.3
T_b	231.11	261.43	272.65	300.994	309.22
$d_m \times 1 \cdot 10^{-30}$ $(C \cdot m)$	0	0.1319	0	0.1301	0
r_g	2.431	2.948	2.886	3.324	3.337
T_m	85.47	113.54	134.86	113.25	143.42
T_{tp} (K)	85.47	113.54	134.86	113.25	143.42
P_{tp} (Pa)	$1.66E^{-09}$	$1.39E^{-07}$	$6.65E^{-06}$	$1.14E^{-09}$	$6.77E^{-07}$
M_v $(m^3/kmol)$	$7.57E^{-02}$	$1.05E^{-01}$	$9.65E^{-02}$	$1.17E^{-01}$	$1.16E^{-01}$
H_f $(kJ/kmol)$	$-1.05E^{+05}$	$-1.34E^{+05}$	$-1.26E^{+05}$	$-1.54E^{+05}$	$-1.47E^{+05}$
G_f $(kJ/kmol)$	$-2.44E^{+04}$	$-2.08E^{+04}$	$-1.67E^{+04}$	$-1.41E^{+04}$	$-8.81E^{+03}$
S_{IG} $(kJ/(kmol \cdot K))$	$2.70E^{+02}$	$2.95E^{+02}$	$3.10E^{+02}$	$3.44E^{+02}$	$3.49E^{+02}$
H_{fus} $(kJ/kmol)$	$3.52E^{+03}$	$4.54E^{+03}$	$4.66E^{+03}$	$5.15E^{+03}$	$8.40E^{+03}$
H_{comb} $(MJ/kmol)$	$-2.04E^{+06}$	$-2.65E^{+06}$	$-2.66E^{+06}$	$-3.24E^{+06}$	$-3.24E^{+06}$
δ $(\sqrt{kJ/m^3})$	13.1	12.5682	13.7	13.8578	14.4
V_{vw} $(m^3/kmol)$	0.0376	0.0478	0.0478	0.058	0.058
A_{vw} $(m^2/kmol)$	$5.59E^{+08}$	$6.93E^{+08}$	$6.94E^{+08}$	$8.28E^{+08}$	$8.29E^{+08}$
P_{nvap} (Pa)	$9.53E^{+05}$	$3.52E^{+05}$	$2.44E^{+05}$	$9.17E^{+04}$	$6.83E^{+04}$

Table 5.4: Pure component properties in the 5 hydrocarbon separation

(iABCDE)(AB/CDE)[(A/B)[(oA)](oB)](C/DE)[(oC)](D/E)[(oD)](oE)
(iABCDE)(AB/CDE)[(A/B)[(oA)](oB)](CD/E)[(C/D)[(oC)](oD)](oE)
(iABCDE)(ABC/DE)[(A/BC)[(oA)](B/C)[(oB)](oC)](D/E)[(oD)](oE)
(iABCDE)(ABC/DE)[(AB/C)[(A/B)[(oA)](oB)](oC)](D/E)[(oD)](oE)
(iABCDE)(ABCD/E)[(A/BCD)[(oA)](B/CD)[(oB)](C/D)[(oC)](oD)](oE)
(iABCDE)(ABCD/E)[(A/BCD)[(oA)](BC/D)[(B/C)[(oB)](oC)](oD)](oE)
(iABCDE)(ABCD/E)[(ABC/D)[(A/BC)[(oA)](B/C)[(oB)](oC)](oD)](oE)
(iABCDE)(ABCD/E)[(ABC/D)[(AB/C)[(A/B)[(oA)](oB)](oC)](oD)](oE)
(iABCDE)(ABCD/E)[(AB/CD)[(A/B)[(oA)](oB)](C/D)[(oC)](oD)](oE)
(iABCDE)(A/BCDE)[(oA)](B/CDE)[(oB)](C/DE)[(oC)](D/E)[(oD)](oE)
(iABCDE)(A/BCDE)[(oA)](B/CDE)[(oB)](CD/E)[(C/D)[(oC)](oD)](oE)
(iABCDE)(A/BCDE)[(oA)](BCD/E)[(B/CD)[(oB)](C/D)[(oC)](oD)](oE)
(iABCDE)(A/BCDE)[(oA)](BCD/E)[(BC/D)[(B/C)[(oB)](oC)](oD)](oE)
(iABCDE)(A/BCDE)[(oA)](BC/DE)[(B/C)[(oB)](oC)](D/E)[(oD)](oE)

Table 5.5: 14 flowsheet alternatives for the separation of a 5 component mixture into 5 pure product stream using distillation columns.



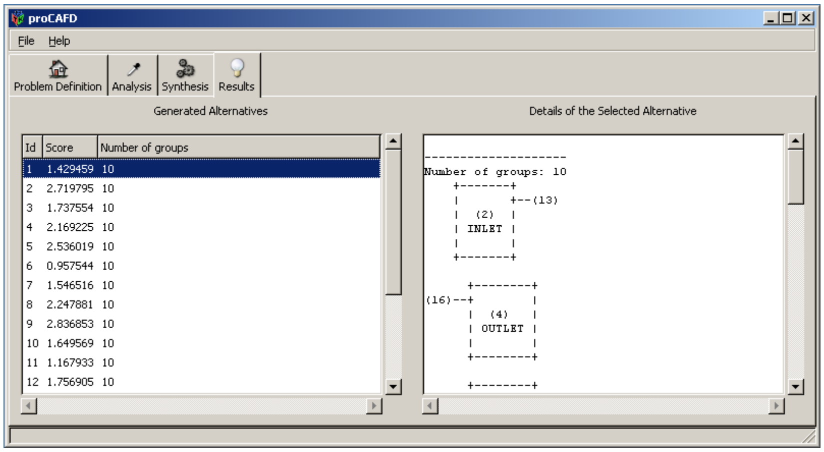
Synthesis completed.

Figure 5.14: ProCAFD: End of the generation of the alternatives.

5.3.4 Step 5 – Ranking and selection of the alternatives

For each of the 14 alternatives, the user can access the details of the alternative, such as, the calculated energy index E_x, the number of process groups and their details (see figure 5.15). Out of the 14 alternatives, the alternative found to be the optimal according to the driving force theory has been generated. This alternative is selected for the reverse simulation:

(iABCDE)(A/BCDE)[(oA)](BC/DE)[(B/C)[(oB)](oC)](D/E)[(oD)](oE)



Id	Score	Number of groups
1	1.429459	10
2	2.719795	10
3	1.737554	10
4	2.169225	10
5	2.536019	10
6	0.957544	10
7	1.546516	10
8	2.247881	10
9	2.836853	10
10	1.649569	10
11	1.167933	10
12	1.756905	10

```
--------------------
Number of groups: 10
    +-------+
    |       +--(13)
    |  (2)  |
    | INLET |
    |       |
    +-------+


      +--------+
(16)--+        |
      |  (4)   |
      | OUTLET |
      |        |
      +--------+

      +--------+
```

Figure 5.15: ProCAFD: Selection of an alternative.

5.3.5 Step 6 – Reverse simulation of an alternative

The reverse simulation of an alternative is composed of two steps: the mass balance through each of the process groups, and, the design of the unit operations underlining the individual process groups.

Mass balance

The mass balance is performed through each process group present in the generated flowsheet alternative. Operating conditions, such as, pressure and temperature of the outlet streams, are given by the process group definition. In the case of the distillation column process groups, the recovery of the components lighter than the light key is equal to 1.0 in the overhead product and the recovery of the components heavier than the heavy key is equal to 1.0 in the bottom product. The recovery of the key components is above 99%. Table 5.7 page 105 is giving the details of the streams in the mass balance flowsheet of the selected alternative, given on figure 5.16.

The mass balance provides the definition (individual molar flowrates, pressure and temperature) of all the inlet and outlet streams of the process groups.

Thus, it is possible to perform the reverse simulation of each process group from the known inlet and outlet streams. The next section is presenting the reverse simulation of the distillation columns.

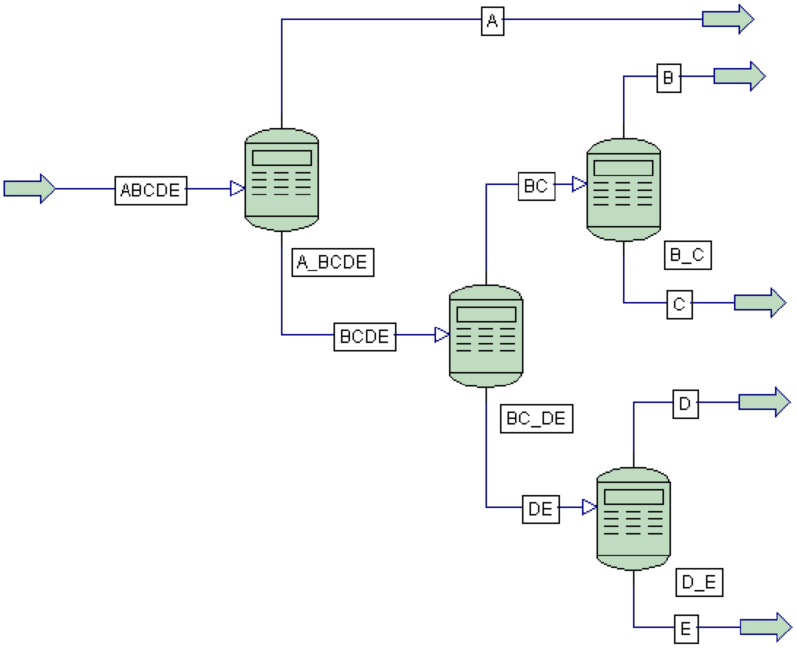


Figure 5.16: Mass balance flowsheet of the separation of 5 alcanes.

Reverse simulation of the distillation column process groups

The reverse simulation of the distillation column is performed using the driving force method presented in section 4.3.3 (page 63). Based on the outlet specifications and on the driving forces as shown on figure 5.17, the reverse simulation of the columns is given the column design parameters in table 5.6.

Column	`(A/BCDE)`	`(BC/DE)`	`(B/C)`	`(D/E)`
Number of stages	18	24	34	59
Feed plate location	7	12	16	33
Purity light key	0.985		0.965	0.98
Recovery light key		0.995		
Purity heavy key			0.985	0.98
Recovery heavy key	0.995	0.995		

Table 5.6: Design parameters of the reverse simulated columns

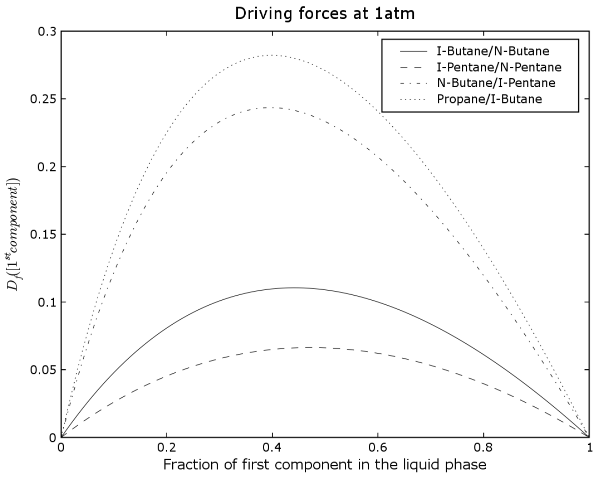


Figure 5.17: Driving forces corresponding to the four consecutive key separations

5.3.6 Step 8 – Final verifications

For final verifications, the mass balance simulation and rigorous simulation have been performed using the PRO/II simulator. Table 5.7 (page 105) show both the details of the mass balance simulation and the rigorous simulation. It can be seen that the mass balance results are very close to the rigorous simulation results. However, the nature of the phase and the presence or not of some of the components in the streams do not always correspond. For example, stream BC is a vapor where a liquid phase is calculated from the mass balance; traces of Pentane are also found in stream BC where the mass balance calculations provide a stream without Pentane. The presence or not of some of the components between the mass balance and the rigorous simulation is explained by the specifications of the mass balance calculations. When performing the mass balance, the components are all fully recovered at the top or the bottom of the column at the exception of the key components which is not the case in the thermodynamically rigorous simulation. The difference of phase is explained by the specification of the process groups, the outlet streams of the distillation process groups are set to be at bubble point temperature. A bubble point temperature corresponds to a total condenser for a distillation column, whereas the rigorous simulations of the columns have been performed with partial condensers.

Rigorous simulation results

Stream Name	ABCDE	A	BCDE	BC	DE	B	C	D	E
Phase	Liquid	Vapor	Liquid	Vapor	Liquid	Vapor	Liquid	Vapor	Liquid
Temperature (K)	276.647	232.448	283.383	269.062	306.159	261.475	272.762	301.464	309.144
Pressure (atm)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Flowrate ($kmole/hr$)	907.200	45.400	861.800	362.898	498.902	137.102	225.796	178.375	320.526
Composition									
PROPANE	0.0500	0.9850	0.0008	0.0019		0.0050			
IBUTANE	0.1500	0.0140	0.1572	0.3733		0.9650	0.0140		
BUTANE	0.2500	0.0010	0.2631	0.6242	0.0005	0.0300	0.9850	0.0013	
IPENTANE	0.2000		0.2105	0.0005	0.3632		0.0008	0.9800	0.0200
PENTANE	0.3500		0.3684	0.0001	0.6363		0.0002	0.0187	0.9800

Mass balance results

Stream	ABCDE	A	BCDE	BC	DE	B	C	D	E
Phase	Liquid	Liquid	Liquid	Liquid	Liquid	Liquid	Liquid	Liquid	Liquid
Temperature (K)	276.647	231.570	283.379	267.726	306.043	261.167	272.861	301.120	309.144
Pressure (atm)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Flowrate ($kmole/hr$)	907.200	45.367	861.833	362.706	499.127	137.775	224.932	177.615	321.512
Composition									
PROPANE	0.0500	0.9850	0.0008	0.0020		0.0052			
IBUTANE	0.1500	0.0150	0.1571	0.3734		0.9650	0.0110		
BUTANE	0.2500		0.2632	0.6222	0.0023	0.0298	0.9850	0.0064	
IPENTANE	0.2000		0.2105	0.0025	0.3616		0.0040	0.9800	0.0200
PENTANE	0.3500		0.3684		0.6361			0.0136	0.9800

Table 5.7: Comparison between the mass balance and the rigorous simulation in the 5 hydrocarbon separation

5.3.7 Conclusion

The results show that the problem formulation and resolution method of the framework for CAFD can be converted into an easy to use computer tool. As the CAFD concepts follow an intuitive representation of the problem and its solution, the software implementation follows the theory. The ability to back calculate the design parameters of the unit operations using the reverse simulation approach also provides a fast and efficient way to obtain the design. It should however be noted, that the mass balance simulation cannot fully describe the dilute components in the mixtures.

5.4 The Future of an Integrated System

In this section, a close look at the current status of ICAS and ProCAFD is presented. From the analysis of the current status, ideas for future developments are proposed.

As shown in table 5.8, the ICAS package is a large application with more than 1 million lines of code, 50 compiled programs, 150 MB of models and data in databases.

Lines of code	1158646 (C/C++: 639563, Fortran: 519083)
Compiled programs	57 (including 16 DLLs)
Databases	156 MB (in 516 files)

Table 5.8: ICAS statistics

The complexity of the ICAS package is for some part necessary, the programs are used to solve complex product-process engineering problems. This engineering complexity cannot be simplified and is naturally translated into complex data structures at the code level. The focus here is not on the engineering complexity but on the software complexity. As many different technologies are used within various ICAS tools, together with a sparse code documentation, the ICAS package has reached a level where the cost of extension is very high. Some of the ideas to improve the development of ProCAFD and its integration within ICAS are exposed below.

- The different core functionalities of the ICAS package should be clearly defined and available in libraries with stable interfaces. The core functionalities are including the thermodynamic calculations, the database management or the flowsheet manipulation. By performing this clear encapsulation of those core functionalities, at least two elements of ProCAFD could be simplified. The component property object could directly access the full properties of the chemical components specified in the synthesis problem this including the properties calculated through correlations and thermodynamic models. It could also be possible to have a more efficient transfer of the generated process flowsheet from ProCAFD to ICASSim.

- The barrier to the contribution of the students to the extension of ProCAFD should be lowered as much as possible. Today, it is difficult to recruit students who have both the process engineering and the computer science knowledge to find their way into complex C/C++ code. On the other hand, complex and efficient models are developed on a regular basis using software environment like Matlab$^{\mathrm{TM}}$. By using dynamic programming languages like Python[46] to write extensions of ProCAFD and ICAS, the ease of development of Matlab-like environments and the power of the already available in-house methods and tools could be combined. This could be for example possible to combine some of the strengths of CAPSS with ProCAFD.
- Before the delivery of the ProCAFD package to the users, it is difficult to assess if critical functionalities are missing or are unavailable, especially with respect to data exchange with external tools. The extension of the automatic tests from ProCAFD to the rest of the ICAS package would improve this situation.
- Finally, the code documentation effort should be maintained to limit the loss of knowledge when ProCAFD and other tools are taken over by other developers.

The other road ahead – By looking at the evolution of the applications used every day by the engineers, it is easy to see a slow but irreversible shift. More and more, the applications are now *web* applications. The best tool to find information is the preferred search engine, emails are available on the network or the accounting of a company is made through a web portal. The increased speed of the network connections together with the increased power of the web applications are opening a world of possibilities. The benefits of transferring ProCAFD from the desktop to the web are numerous. Some of those benefits are presented below together with the main drawback.

Easy maintenance. With only one code base to maintain and the central storage of the data, the scenario leading to bugs could be easy to reproduce and thus improving the speed to fix the bugs. It would also mean that a user had no operation to perform on his computer system to have access to a new version of the software, this would be done transparently.

Collaborative work. By keeping the data in a place accessible for all the users, collaborative work becomes possible and easy to setup. It could be easy for to process engineers to collaborate on the synthesis and design of a new process. Decision choices during the resolution steps of the ProCAFD work flow could easily be made in collaboration.

Energy efficiency. When solving a process synthesis problem, the steps are not all computer intensive tasks. For example the problem definition in step 1 (see figure 5.1 on page 83) and the ranking and selection of the alternatives in step 5 are not computer intensive, but the generation of the flowsheet structure alternatives in step 4 is a very computer intensive task. By sharing the computer load of a large number of users on a GRID computing system – a service for sharing computer power and data storage capacity over the Internet, the overall computer capacities could be used more efficiently, thus reducing the overall energy cost and need of investment in computer hardware.

The main drawback is that the work depends on the network connectivity. But nowadays, the reliability can be very high, as for example common search engines, used by millions of people are virtually available 100% of the time.

6

Case Studies

6.1 Introduction

In this chapter, the application of the framework for CAFD is illustrated through several case studies. Some of the case studies focus on a specific part of the framework (e.g. reverse simulation) in order to highlight the corresponding features of that part. Other cases aim at illustrating the systematic methodology for synthesis and design developed with the framework for CAFD. The main focus for each case study is summarized below:

- In section 6.2, a case study[3, 39] involving the separation of an hydrocarbon mixture composed of 8 components Neopentane, Isopentane, n-Pentane, 2,2-Dimethylbutane, 2,3-Dimethylbutane, 2-Methylpentane, n-Hexane and Benzene is presented. The case study highlights the application of the framework for CAFD with respect to the generation of a large number of alternatives considering that one of the binary pairs (Benzene-Hexane) forms an azeotrope the handling of an azeotropic mixture between the Benzene and the Hexane and the ordering of the alternatives.

- In section 6.3, a case study involving the production of Methyl Acetate from Acetic Acid and Methanol[27] is presented. The case study highlights the application of the framework for CAFD with respect to the representation of a process flowsheet with process groups, the reverse simulation of a reactor process group and the reverse simulation of the pressure swing distillation separation.

- In section 6.4, a case study involving the production of Benzene by Toluene hydrodealkylation[9] is presented. The case study highlights the application of the framework for CAFD with respect to the generation of flowsheet alternatives with retrofit at the process level and the handling of the recycles in the mass balance calculations with the definition of the purges.

- In section 6.5, a case study of the application of the SFILES notation to another field with the Ammonia reaction[11] is presented. The case study highlights the application of the SFILES notation to the representation of reaction pathways.

- In section 6.6, a case study involving flash separation of a Methanol, Ethanol and Propanol mixture is presented. The case study highlights the reverse simulation approach for a single stage separation unit.

6.2 Separation of Aromatic-paraffin Mixture

In this section, the case study involving the separation of an hydrocarbon mixture composed of 8 components Neopentane, Isopentane, n-Pentane, 2,2-Dimethylbutane, 2,3-Dimethylbutane, 2-Methylpentane, n-Hexane and Benzene[3, 39] is presented. The case study highlights the application of the framework for CAFD with respect to the generation of a large number of alternatives considering that one of the binary pairs (Benzene-Hexane) forms an azeotrope, the handling of an azeotropic mixture between Benzene and Hexane and the ordering of the alternatives. Comparison has been made between the flowsheet obtained using the framework for CAFD against two other designs reported in the literature[3, 39]. The numbering of the steps corresponds to the steps in the work flow of the framework for CAFD presented in figure 5.1 (page 83).

6.2.1 Step 1 – Definition of the problem

The synthesis problem is to separate the mixture defined in table 6.1 at the minimum energy cost, into 8 pure products.

Components	**Feed flowrates** ($kmole/hr$)
A – 2,2-Dimethyl-Propane	32.51
B – i-Pentane	59.97
C – n-Pentane	62.71
D – 2,2-Dimethyl-Butane	6.64
E – 2,3-Dimethyl-Butane	10.80
F – 2-Methyl-Pentane	62.71
G – n-Hexane	39.62
H – Benzene	80.33

Table 6.1: Feed mixture for the aromatic-paraffin separation

The structural definition of the synthesis problem is as follows: 1 inlet and 8 outlets of pure products. The flowsheet property target is the minimization of the energy consumption index E_x.

6.2.2 Step 2 – Problem analysis

From the analysis of the mixture, distillation has been found to be the appropriate separation technique for all separation tasks that needs to be performed

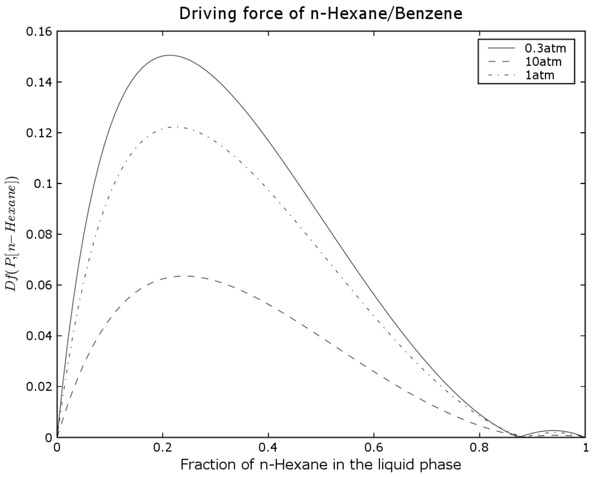


Figure 6.1: Non pressure dependency of the n-Hexane/Benzene azeotrope

except for the Benzene-Hexane binary pair, which forms an azeotrope. The azeotrope is not pressure dependent as shown on figure 6.1. Further analysis of the binary azeotrope indicates that the mixture can be separated either using extractive distillation with an appropriate solvent or through membranes. Dimethylformamide DMF has been found to be a suitable solvent for the extractive distillation[37] and the membrane could be a zeolite membrane[37].

As the separation of the Benzene-Hexane is to be performed either by a membrane or a solvent based separation, the problem definition is refined to be composed of the inlet and 7 outlets, one being a mixture of the binary pair of Benzene-Hexane. The separation of the azeotropic mixture will be considered afterward as a sub-problem.

6.2.3 Step 3 – Process groups selection and initialization

From the analysis of the problem, distillation has been found to be the appropriate separation technique to be performed. From the database of process groups, the distillation process groups are retrieved and matched with the possible mixtures of the 6 components and the binary pair. A distillation process group is initialized with a given mixture if the relative volatilities of the components are in the range supported by the process group, the driving force between the two key components is matching with the process group driving force and if the number of components in the mixture corresponds to the number of components of the process group. 56 initialized process groups are found to be usable in the problem, they are listed in table 6.4.

Property	NeoPentane	i-Pentane	n-Pentane	2,2-DimethylButane
M_w	72.15	72.15	72.15	86.177
ω	0.1964	0.2275	0.2515	0.235
T_c	433.78	460.43	469.7	488.78
P_c	31.572	33.368	33.259	30.407
Z_c	0.269	0.27	0.27	0.272
V_c	0.3	0.3	0.3	0.4
T_b	282.65	300.994	309.22	322.88
$d_m \times 1 \cdot 10^{-30}\ (C \cdot m)$	0	0.1301	0	0
r_g	3.161	3.324	3.337	3.476
T_m	256.6	113.25	143.42	174.28
$T_{tp}\ (K)$	256.6	113.25	143.42	174.28
$P_{tp}\ (Pa)$	$3.53E^{-01}$	$1.14E^{-09}$	$6.77E^{-07}$	$3.87E^{-05}$
$M_v\ (m^3/kmol)$	$1.20E^{-01}$	$1.17E^{-01}$	$1.16E^{-01}$	$1.34E^{-01}$
$H_f\ (kJ/kmol)$	$-1.68E^{+05}$	$-1.54E^{+05}$	$-1.47E^{+05}$	$-1.85E^{+05}$
$G_f\ (kJ/kmol)$	$-1.71E^{+04}$	$-1.41E^{+04}$	$-8.81E^{+03}$	$-8.75E^{+03}$
$S_{IG}\ (kJ/(kmol \cdot K))$	$3.06E^{+02}$	$3.44E^{+02}$	$3.49E^{+02}$	$3.58E^{+02}$
$H_{fus}\ (kJ/kmol)$	$3.15E^{+03}$	$5.15E^{+03}$	$8.40E^{+03}$	$5.79E^{+02}$
$H_{comb}\ (MJ/kmol)$	$-3.25E^{+06}$	$-3.24E^{+06}$	$-3.24E^{+06}$	$-3.84E^{+06}$
$\delta\ (\sqrt{kJ/m^3})$	13.0526	13.8578	14.4	13.7662
$V_{vw}\ (m^3/kmol)$	0.058	0.058	0.058	0.0682
$A_{vw}\ (m^2/kmol)$	$8.48E^{+08}$	$8.28E^{+08}$	$8.29E^{+08}$	$9.83E^{+08}$
$P_{nvap}\ (Pa)$	$1.71E^{+05}$	$9.17E^{+04}$	$6.83E^{+04}$	$4.25E^{+04}$

Table 6.2: Pure component properties in the aromatic paraffin mixture

Property	2,3-DimethylButane	2-MethylPentane	n-Hexane	Benzene
M_w	86.177	86.177	86.177	78.114
ω	0.2461	0.2774	0.3013	0.21
T_c	499.98	497.5	507.6	562.16
P_c	30.861	29.706	29.854	48.339
Z_c	0.269	0.267	0.266	0.271
V_c	0.4	0.4	0.4	0.3
T_b	331.13	333.41	341.88	353.24
$d_m \times 1 \cdot 10^{-30}$ $(C \cdot m)$	0	0	0	0
r_g	3.636	3.784	3.769	3.004
T_m	145.19	119.55	177.83	278.68
T_{tp} (K)	145.19	119.55	177.83	278.68
P_{tp} (Pa)	$1.49E^{-07}$	$9.10E^{-11}$	$8.90E^{-06}$	$4.70E^{-02}$
M_v $(m^3/kmol)$	$1.31E^{-01}$	$1.33E^{-01}$	$1.31E^{-01}$	$8.95E^{-02}$
H_f $(kJ/kmol)$	$-1.77E^{+05}$	$-1.75E^{+05}$	$-1.67E^{+05}$	$8.29E^{+04}$
G_f $(kJ/kmol)$	$-3.13E^{+03}$	$-5.34E^{+03}$	$-6.63E^{+01}$	$1.30E^{+05}$
S_{IG} $(kJ/(kmol \cdot K))$	$3.66E^{+02}$	$3.81E^{+02}$	$3.89E^{+02}$	$2.69E^{+02}$
H_{fus} $(kJ/kmol)$	$7.99E^{+02}$	$6.27E^{+03}$	$1.31E^{+04}$	$9.87E^{+03}$
H_{comb} $(MJ/kmol)$	$-3.85E^{+06}$	$-3.85E^{+06}$	$-3.86E^{+06}$	$-3.14E^{+06}$
δ $(\sqrt{kJ/m^3})$	14.2998	14.4176	14.9	18.7296
V_{vw} $(m^3/kmol)$	0.0682	0.0683	0.0683	0.0484
A_{vw} $(m^2/kmol)$	$9.62E^{+08}$	$9.63E^{+08}$	$9.64E^{+08}$	$6.00E^{+08}$
P_{nvap} (Pa)	$3.13E^{+04}$	$2.82E^{+04}$	$2.02E^{+04}$	$1.26E^{+04}$

Table 6.3: Pure component properties in the aromatic paraffin mixture (Continued)

(A/B)	(A/BC)	(A/BCD)
(A/BCDE)	(A/BCDEF)	(A/BCDEFGH)
(AB/C)	(AB/CD)	(AB/CDE)
(AB/CDEF)	(AB/CDEFGH)	(ABC/D)
(ABC/DE)	(ABC/DEF)	(ABC/DEFGH)
(ABCD/E)	(ABCD/EF)	(ABCD/EFGH)
(ABCDE/F)	(ABCDE/FGH)	(ABCDEF/GH)
(B/C)	(B/CD)	(B/CDE)
(B/CDEF)	(B/CDEFGH)	(BC/D)
(BC/DE)	(BC/DEF)	(BC/DEFGH)
(BCD/E)	(BCD/EF)	(BCD/EFGH)
(BCDE/F)	(BCDE/FGH)	(BCDEF/GH)
(C/D)	(C/DE)	(C/DEF)
(C/DEFGH)	(CD/E)	(CD/EF)
(CD/EFGH)	(CDE/F)	(CDE/FGH)
(CDEF/GH)	(D/E)	(D/EF)
(D/EFGH)	(DE/F)	(DE/FGH)
(DEF/GH)	(E/F)	(E/FGH)
(EF/GH)	(F/GH)	

Table 6.4: Initialized distillation process groups in the aromatic-paraffin mixture separation

6.2.4 Step 4 – Generation of the flowsheet structure alternatives

Given the mixture of 8 components defined in table 6.1, to be separated into 6 highly pure products (purity above 99%) and one binary mixture of Hexane-Benzene, using distillation columns represented by 56 process groups (see table 6.4), at the minimum consumption of energy, the flowsheet generation algorithm of the framework for CAFD generates 132 feasible alternatives. From the table 6.5, it is interesing to see that even when the algorithm is analysing more than 27000 structures when generating the alternatives, it is not memory bounded with the stack of working structures never exceeding 36 structures and the maximum memory usage of the program is approximately 16MB. The complete resolution time on a desktop computer was around 20 minutes, which is reasonable, considering that the algorithm has not been fully optimized and is implemented in the Python programing language without machine level compilation.

6.2.5 Step 5 – Ranking and selection of the alternatives

Out of the 132 alternatives, three are of interest and presented in table 6.6. These are the first alternative and the ones found to be optimal for this separation by Shah and Kokossis[39] and by Bek-Pedersen[2]. It should be noted that

Number of completed alternatives	132
Maximum length of the working list	36
Number of analysed structures	27715
Total time	$\approx 1150s$
Maximum memory usage	$\approx$ 16MB

Table 6.5: Aromatic-paraffin mixture computational statistics

in the original publications by Shah and Kokossis, and, Bek-Pedersen, they are both adding another distillation column to separate the Benzene-Hexane mixture into pure Hexane and the azeotrope Benzene-Hexane. As the difference is only for one distillation column at the end of the train of distillation columns, it is still possible to compare the three alternatives as including the same additional distillation column to the three alternatives would not change the rank in terms of energy usage.

Rank	**SFILES notation**	E_x
1 – A	`(iABCDEFGH)(ABCDEF/GH)[(ABCD/EF)[(ABC/D)[(A/BC)[(oA)](B/C)[(oB)](oC)](oD)](E/F)[(oE)](oF)](oGH)`	379.327889
89 – B	`(iABCDEFGH)(ABC/DEFGH)[(A/BC)[(oA)](B/C)[(oB)](oC)](DEF/GH)[(D/EF)[(oD)](E/F)[(oE)](oF)](oGH)`	392.507506
90 – C	`(iABCDEFGH)(A/BCDEFGH)[(oA)](BC/DEFGH)[(B/C)[(oB)](oC)](DEF/GH)[(D/EF)[(oD)](E/F)[(oE)](oF)](oGH)`	392.507641

Table 6.6: Alternatives of interest in the aromatic-paraffin separation

The first remark when looking at the three alternatives is that the alternatives from Shah and Kokossis[39] ranked 89 and by Bek-Pedersen[2] ranked 90 out of 132 are far from being the optimal ones. They are also found to be close to each other as being consecutively ranked, with the alternative from Shah and Kokossis being slightly better. These results are in contradiction to Bek-Pedersen. Bek-Pedersen is claiming that the alternative C is better than alternative B with an 8% lower energy consumption, when including the distillation column to separate the Benzene and Hexane. Additional comparisons of these results are performed in section 6.2.7 page 120.

6.2.6 Step 6 – Reverse simulation of the azeotropic separation

The most promising alternative being selected, the next steps are the resolution of the mass balance between the process groups and the reverse simulation, i.e. the determination of the design parameters, of the unit operations. As the reverse simulation of distillation columns has already been presented in the case study involving the separation of hydrocarbon mixture in section 5.3, only the details related to the further processing of the azeotropic mixture of Benzene and Hexane will be presented here.

Mass balance

The mass balance is performed through each process group present in the generated flowsheet alternative. Operating conditions, such as, pressure and temperature of the outlet streams are given by the process group definition. In the case of the distillation column process groups, the recovery of the components lighter than the light key is assumed equal to 1.0 in the overhead product and the recovery of the components heavier than the heavy key is assumed equal to 1.0 in the bottom product. The recovery of the key components is above 99%. Table 6.7 on page 118 provides the details of the streams in the mass balance flowsheet of alternative A given on figure 6.2.

The mass balance provides the definition (individual molar flowrates, pressure and temperature) of all the inlet and outlet streams of the process groups. Thus, it is possible to perform the reverse simulation of each process group from the known inlet and outlet streams. The next section is presenting the reverse simulation of the solvent based separation.

Reverse simulation of the columns

Considering the separation of Hexane from Benzene, if the Hexane is not needed in a very high purity, a simple distillation column can recover a high purity Benzene and the Benzene/Hexane azeotrope as shown on figure 6.1 page 111. We are here considering that both Benzene and Hexane should be recovered with a high purity. From the problem analysis, a solvent based separation has been found to be a suitable choice.

Through a search of the CAPEC database[16], two possible solvents are found, DMSO (figure 6.3) and DMF (figure 6.4). They can both break the azeotrope as shown on figure 6.5 and figure 6.6, but the DMF is chosen as used by one of the processes of the Institut Francais du Petrole (IFP)[37].

From the solvent free driving force, it is possible to select a solvent fraction of 0.625 to obtain a solvent free maximum driving force above 0.4. From the mass balance calculations and the solvent fraction it is possible to determine that an inlet of 200 kmole/hr of DMF is needed in the first column. The DMF is mixed with the Hexane and Benzene mixture which contains also trace of 2-MethylPentane. The first column is recovering a near pure stream of Hexane at the top. The bottom product of the first column is sent to the second column to perform the Benzene/DMF separation. From the solvent free driving force and the Benzene/DMF driving force in figure 6.7, the design parameters for the columns are back calculated and given in table 6.8.

It is interesting to note that the thermodynamic model used in the non

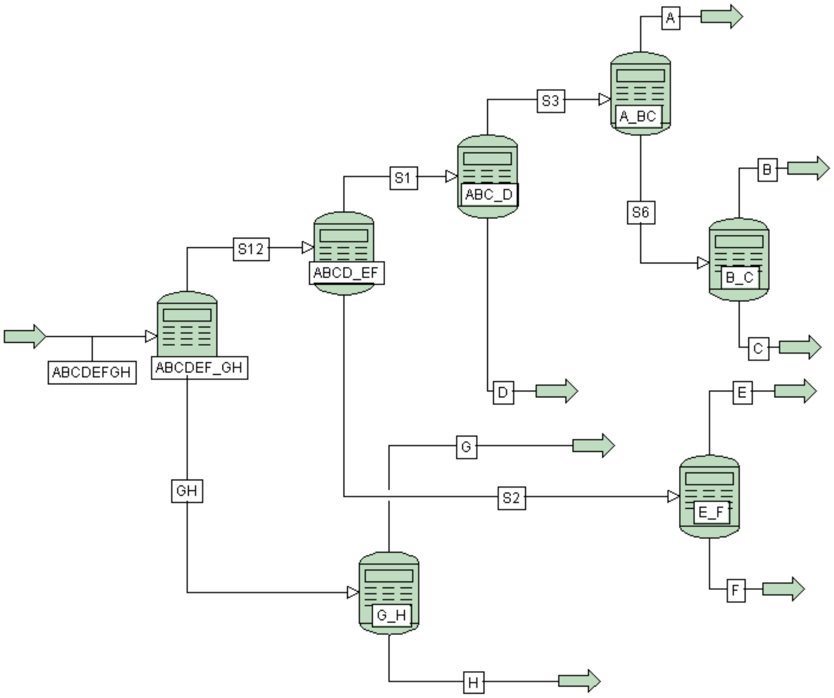


Figure 6.2: Mass balance flowsheet of alternative A in the aromatic-paraffin mixture separation.

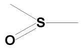


Figure 6.3: DMSO molecular structure

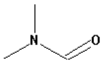


Figure 6.4: DMF molecular structure

Stream Name	ABCDEFGH	A	B	C	D	E	F	GH	G	H
Phase	Liquid	Liquid	Liquid	Liquid	Liquid	Liquid	Liquid	Liquid	Liquid	Liquid
Temperature (K)	298	282.74	301.26	309.07	322.25	331.29	333.51	344.96	341.85	352.73
Pressure (atm)	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Flowrate ($kmol/hr$)	355.29	32.81	59.63	62.47	6.97	10.53	62.81	120.07	39.33	80.73
Composition										
22PR	0.0915	0.9909	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
IPENTANE	0.1688	0.0091	0.9855	0.0145	0.0	0.0	0.0	0.0	0.0	0.0
NC5	0.1765	0.0	0.0145	0.9850	0.0450	0.0	0.0	0.0	0.0	0.0
22MB	0.0187	0.0	0.0	0.0005	0.9473	0.0	0.0	0.0	0.0	0.0
23MB	0.0304	0.0	0.0	0.0	0.0077	0.9800	0.0068	0.0	0.0	0.0
2MP	0.1765	0.0	0.0	0.0	0.0	0.0200	0.9900	0.0026	0.0080	0.0
NC6	0.1115	0.0	0.0	0.0	0.0	0.0	0.0032	0.3283	0.9920	0.0050
BENZENE	0.2261	0.0	0.0	0.0	0.0	0.0	0.0	0.6690	0.0	0.9950

Table 6.7: Mass balance results for the aromatic paraffin separation

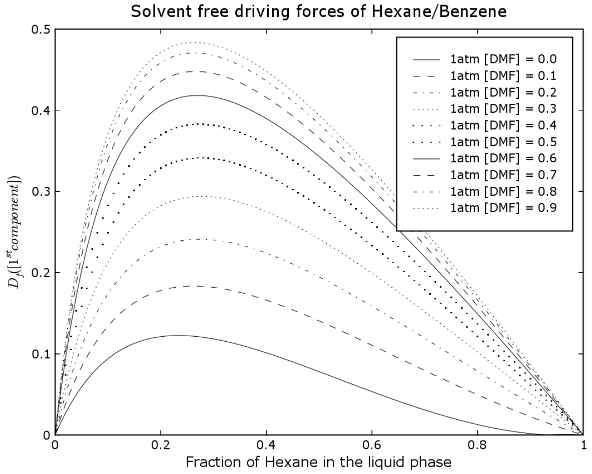


Figure 6.5: Solvent free driving force of the Hexane/Benzene binary pair with the N,N-dimethylformamide (DMF)

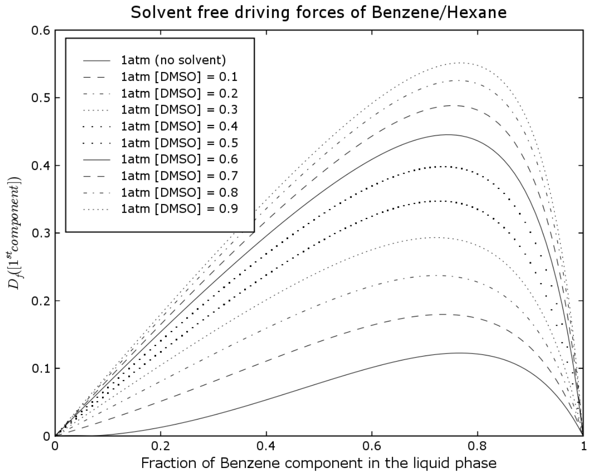


Figure 6.6: Solvent free driving force of the Benzene/Hexane binary pair with the dimethylsulfoxide (DMSO)

azeotropic separations was the SRK[43] equation of state whereas for the azeotropic distillations, the NRTL model[36] has been used for the non-ideal liquid phase and SRK for the vapor phase.

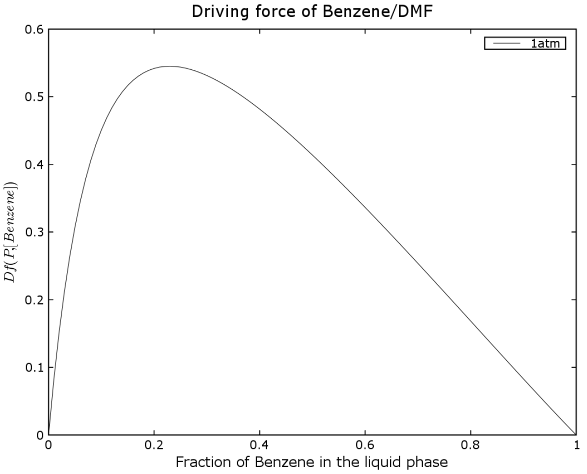


Figure 6.7: Benzene/DMF driving force at 1 atm.

Column	Hexane extraction	Benzene/DMF separation
Number of plates	50	30
Feed plate location	40	22
Purity of light key	0.984	0.986

Table 6.8: Design parameters of the columns in the Hexane Benzene separation.

It was not possible to obtain the 99.5% purity of Hexane as Hexane and DMF are forming an azeotrope. This implies also that a make up of DMF is necessary to compensate for this loss. A possible solution, is to inject steam in the top section of the column to reduce this solvent loss. This solution, which is used in the IFP process has not be explored. The final flowhseet is presented in figure 6.8 page 121.

6.2.7 Step 8 – Comparison of the alternatives by rigorous simulation

As the results found using the CAFD method seem to be in contradiction with those reported in the literature, three rigorous simulations have been performed to compare these alternatives. To be able to really compare the three alternatives, all the distillation columns are set to have 50 stages, except

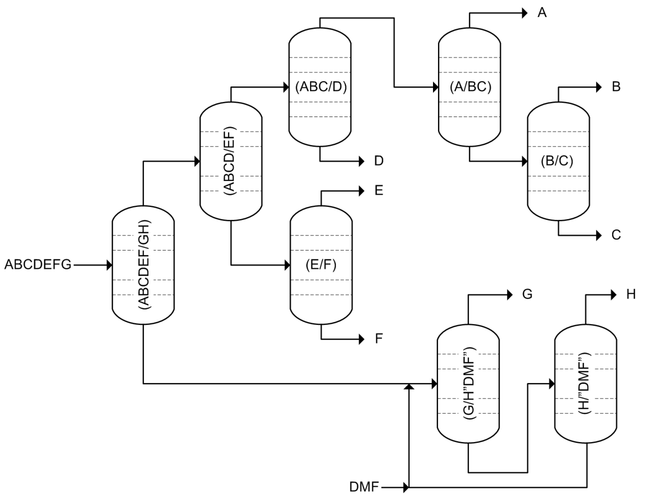


Figure 6.8: Final flowsheet of the separation of the aromatic-paraffin mixture.

for the E/F split where the desired recovery cannot be achieved with only 50 stages, thus a 200 stages column is used. Each column in the flowsheet had the feed plate location optimized to reduce the overall flowsheet energy consumption from a driving force based initialization.

The resulting energy consumptions are presented in table 6.9. The rigorous simulations are both confirming that the best alternative is found through the CAFD method and that the alternative by Shah and Kokossis is slightly better than the one found by Bek-Pedersen.

Alternative	**Energy** (GJ/hr)
A – CAFD method	499.46
B – Shah and Kokossis	523.62
C – Bek-Pedersen	526.16

Table 6.9: Energy consumption from the rigorous simulations

The resulting stream compositions and flowrates for the inlet and the outlet streams are presented in tables 6.10, 6.11 and 6.12.

Bek-Pedersen has found the alternative C to be 8% better than B in terms of energy consumption, whereas this work found that B is better than C by less than 1%. This could be explained by the fact that Bek-Pedersen may not have used a fixed number of stages in the columns for each separation task when comparing the alternatives B and C, or, did not consider driving force based feed locations for the columns in alternative B. As the design paramaters for the columns were only provided for the alternative C, it is not possible draw other conclusions. It should be noted that Shah and Kokossis have found

Stream Name	ABCDEFGH	A	B	C	D	E	F	GH
Phase	Liquid	Vapor	Vapor	Liquid	Vapor	Vapor	Liquid	Liquid
Temperature (K)	298	282.66	301.19	309.16	322.82	331.30	333.53	344.97
Pressure (atm)	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Flowrate ($kmol/hr$)	355.29	32.54	59.95	62.59	6.77	11.04	62.78	119.62
Composition								
22PR	0.0915	0.9982	0.0005	0.0	0.0	0.0	0.0	0.0
IPENTANE	0.1688	0.0018	0.9953	0.0038	0.0001	0.0	0.0	0.0
PENTANE	0.1765	0.0	0.0042	0.9960	0.0185	0.0	0.0	0.0
22MB	0.0187	0.0	0.0	0.0002	0.9752	0.0018	0.0	0.0
23MB	0.0304	0.0	0.0	0.0	0.0048	0.9698	0.0009	0.0001
2MP	0.1765	0.0	0.0	0.0	0.0014	0.0284	0.9917	0.0010
HEXANE	0.1115	0.0	0.0	0.0	0.0	0.0	0.0013	0.3306
BENZENE	0.2261	0.0	0.0	0.0	0.0	0.0	0.0062	0.6683

Table 6.10: Driving force based simulation results of the aromatic-paraffin separation

Stream Name	ABCDEFGH	A	B	C	D	E	F	GH
Phase	Liquid	Vapor	Vapor	Liquid	Vapor	Vapor	Liquid	Liquid
Temperature (K)	298	282.66	301.19	309.16	322.82	331.30	333.53	344.97
Pressure (atm)	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Flowrate ($kmol/hr$)	355.29	32.54	59.95	62.59	6.78	11.04	62.78	119.62
Composition								
22PR	0.0915	0.9982	0.0005	0.0	0.0	0.0	0.0	0.0
IPENTANE	0.1688	0.0018	0.9953	0.0038	0.0001	0.0	0.0	0.0
PENTANE	0.1765	0.0	0.0042	0.9960	0.0185	0.0	0.0	0.0
22MB	0.0187	0.0	0.0	0.0002	0.9751	0.0018	0.0	0.0
23MB	0.0304	0.0	0.0	0.0	0.0048	0.9698	0.0009	0.0001
2MP	0.1765	0.0	0.0	0.0	0.0014	0.0284	0.9917	0.0010
HEXANE	0.1115	0.0	0.0	0.0	0.0	0.0	0.0013	0.3306
BENZENE	0.2261	0.0	0.0	0.0	0.0	0.0	0.0062	0.6683

Table 6.11: Simulation results of the aromatic-paraffin separation using the Shah and Kokossis configuration

Stream Name	ABCDEFGH	A	B	C	D	E	F	GH
Phase	Liquid	Vapor	Vapor	Liquid	Vapor	Vapor	Liquid	Liquid
Temperature (K)	298	282.66	301.19	309.16	322.69	331.30	333.53	344.97
Pressure (atm)	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Flowrate ($kmol/hr$)	355.29	32.54	59.95	62.59	6.77	11.04	62.79	119.61
Composition								
22PR	0.0915	0.9982	0.0005	0.0	0.0	0.0	0.0	0.0
IPENTANE	0.1688	0.0018	0.9953	0.0038	0.0	0.0	0.0	0.0
PENTANE	0.1765	0.0	0.0042	0.9960	0.0185	0.0	0.0	0.0
22MB	0.0187	0.0	0.0	0.0002	0.9753	0.0018	0.0	0.0
23MB	0.0304	0.0	0.0	0.0	0.0048	0.9698	0.0009	0.0001
2MP	0.1765	0.0	0.0	0.0	0.0014	0.0284	0.9915	0.0010
HEXANE	0.1115	0.0	0.0	0.0	0.0	0.0	0.0013	0.3306
BENZENE	0.2261	0.0	0.0	0.0	0.0	0.0	0.0063	0.6683

Table 6.12: Simulation results of the aromatic-paraffin separation using the CAFD method configuration

the alternative C to be the optimal alternative with respected to an objective function that includes the cost related to the number of stages.

6.2.8 Conclusion

The verification (simulation) results confirm, that the rank obtained by using the energy index consumption model is reliable. The results also show that the algorithms of the framework for CAFD are able to generate and test a large number of alternatives. The flowsheet property model for energy index consumption has been shown to be valid and to provide results that match those obtained through rigorous simulations. The ability to back calculate the design parameters of the unit operations using the reverse simulation approach also provides a fast and efficient way to obtain the design. It should however be noted that the comparison of a flowsheet using different types of process groups are limited by the availability of the energy index consumption model. However, this limitation can be overcome by performing the appropriate analysis of the problem and performing a comparison on a subset of the final flowsheet structure, as shown in the case study.

6.3 Methyl Acetate Production

In this section, a case study involving the production of Methyl Acetate from Acetic Acid and Methanol[27] is presented. The case study highlights the application of the framework for CAFD with respect to the representation of a process flowsheet with process groups, the reverse simulation of a reactor process group and the reverse simulation of the pressure swing distillation separation. The numbering of the steps corresponds to the steps in the work flow of the framework for CAFD presented in figure 5.1 (page 83).

6.3.1 Step 1 – Definition of the problem

The problem definition in this case study involves the application of the framework for CAFD for the design of the flowsheet for the production of Methyl Acetate. Methyl Acetate is produced by esterification of Methanol and Acetic Acid, producing Water as a side product. The reaction is known to take place in the liquid phase[20] at high temperature and pressure. The process flowsheet to be designed is shown in figure 6.9. This physically feasible process flowsheet, identified by Jaksland[27] is composed of five main parts: the reactor with Methyl Acetate, Acetic Acid, Methanol and Water as effluents; then a distillation column which separates the reactor effluents into an overhead product with Methyl Acetate, Methanol and trace of Water and a bottom product with Water and Acetic Acid. The overhead product is then first desiccated using a molecular sieve based separation and the Methanol/Methyl Acetate is separated through a pressure swing distillation. The bottom product of the column separating the effluents of the reactor containing a mixture of Water and

Acetic Acid is separated using a pressure swing distillation. Both the Methanol and the Acetic Acid are recycled back to the reactor.

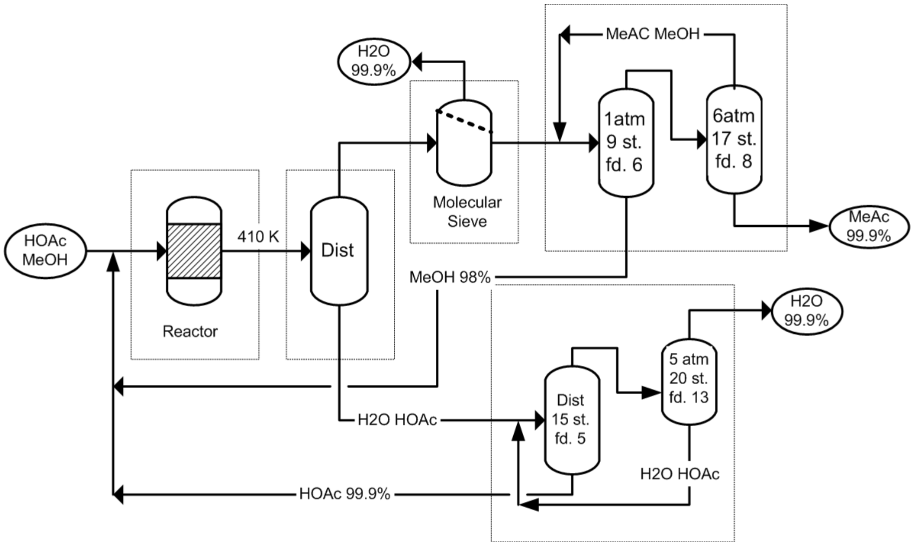


Figure 6.9: Flowsheet for the production of Methyl Acetate

6.3.2 Step 2 – Problem Analysis

To perform the representation of the process flowsheet in figure 6.9 using process groups, the systematic method presented in section 4.1.1 (page 27) is applied. The required information for the method is the process flowsheet and a list of process groups to be matched against the process flowsheet.

Pure component and mixture properties. The match of the process groups against the process flowsheet requires the pure component and mixture properties from the components in the specified problem. The components are Methyl Acetate, Acetic Acid, Methanol and Water. The 22 pure component properties retrieved from the CAPEC databases[16] or calculated using CAPSS[27] are presented in table 6.13.

The analysis of the mixtures in the specified problem shows that three binary pairs are forming azeotropes. The compositions of the three azeotropes at 1 atm are given in table 6.14. However it should be noted that the azeotropes are found to be pressure dependent. The Methanol/Methyl Acetate azeotrope (see figure 6.10) and the Methyl Acetate/Water azeotrope have a considerable pressure dependent composition variation. This is also the case – but to a smaller extent – for the Acetic Acid/Water azeotrope (see figure 6.11).

Property	Methyl-Acetate	Methanol	Water	Acetic-Acid
M_w	74.079	32.042	18.015	60.053
ω	0.3313	0.564	0.3449	0.4665
T_c	506.55	512.64	647.13	591.95
P_c	46.879	79.911	217.666	57.103
Z_c	0.257	0.224	0.229	0.211
V_c	0.2	0.1	0.1	0.2
T_b	330.09	337.85	373.15	391.05
$d_m \times 1 \cdot 10^{-30}$ $(C \cdot m)$	1.6787	1.6996	1.8495	1.7386
r_g	2.996	1.552	0.615	2.61
T_m	175.15	175.47	273.15	289.81
T_{tp} (K)	175.15	175.47	273.16	289.81
P_{tp} (Pa)	$1.00E^{-05}$	$1.10E^{-06}$	$6.04E^{-03}$	$1.26E^{-02}$
M_v $(m^3/kmol)$	$7.98E^{-02}$	$4.06E^{-02}$	$1.81E^{-02}$	$5.76E^{-02}$
H_f $(kJ/kmol)$	$-4.12E^{+05}$	$-2.01E^{+05}$	$-2.42E^{+05}$	$-4.33E^{+05}$
G_f $(kJ/kmol)$	$-3.24E^{+05}$	$-1.62E^{+05}$	$-2.29E^{+05}$	$-3.75E^{+05}$
S_{IG} $(kJ/(kmol \cdot K))$	$3.20E^{+02}$	$2.40E^{+02}$	$1.89E^{+02}$	$2.83E^{+02}$
H_{fus} $(kJ/kmol)$	$7.97E^{+03}$	$3.22E^{+03}$	$6.00E^{+03}$	$1.17E^{+04}$
H_{comb} $(MJ/kmol)$	$-1.46E^{+06}$	$-6.38E^{+05}$	0.00	$-8.15E^{+05}$
δ $(\sqrt{kJ/m^3})$	19.3549	29.5906	47.8127	19.0078
V_{vw} $(m^3/kmol)$	0.0425	0.0217	0.0124	0.0333
A_{vw} $(m^2/kmol)$	$6.44E^{+08}$	$3.58E^{+08}$	$2.26E^{+08}$	$5.18E^{+08}$
P_{nvap} (Pa)	$2.86E^{+04}$	$1.68E^{+04}$	$3.17E^{+03}$	$2.08E^{+03}$

Table 6.13: Pure component properties in the Methyl Acetate production

Binary pair	**Azeotropic composition**, x_i at P = 1atm
Water/Methyl Acetate	0.0776
Water/Acetic Acid	0.8384
Methyl Acetate/Methanol	0.6685

Table 6.14: Composition of the azeotropes identified between the components in the problem.

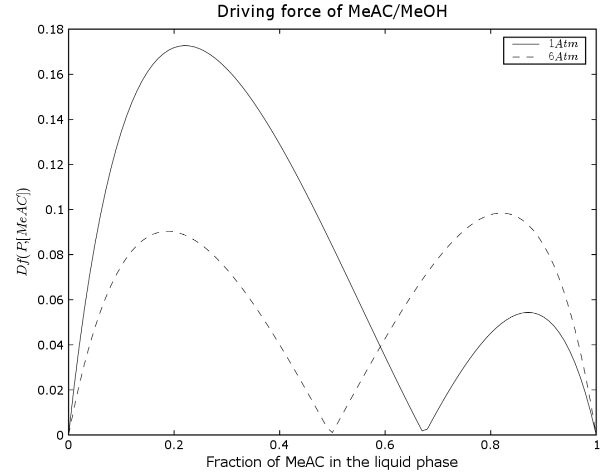


Figure 6.10: Pressure dependency of the MeOH/MeAC azeotrope.

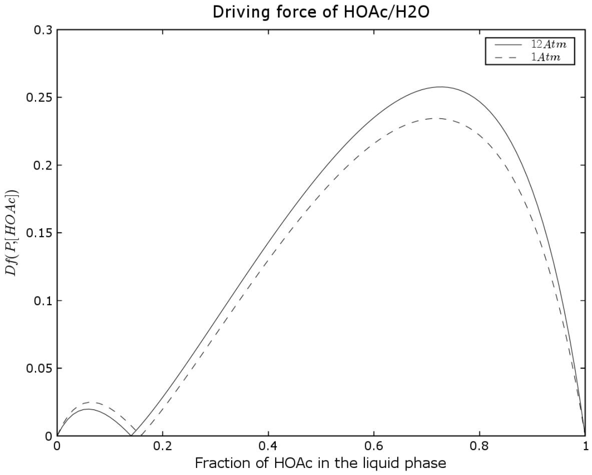


Figure 6.11: Pressure dependency of the Acetic Acid/Water azeotrope.

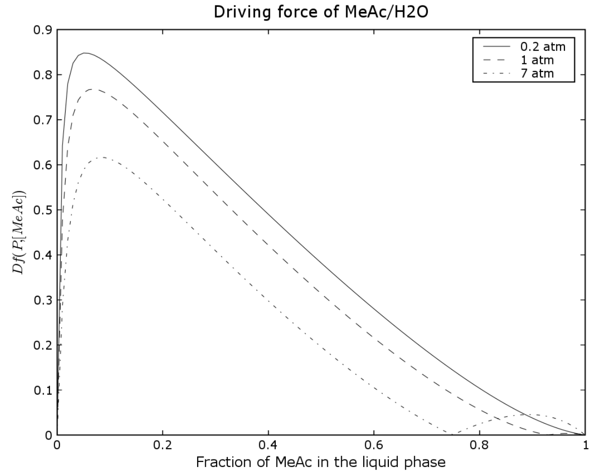


Figure 6.12: Pressure dependency of the Methyl Acetate/Water azeotrope.

Reaction problem properties. The analysis of the reaction properties in the process design problem is performed after retrieving the data from the CAPEC reaction database[16]. The search of the reactions to produce Methyl Acetate returned the reaction shown by equation (6.1). The reaction is an esterification of Methanol and Acetic Acid to produce Methyl Acetate. Water is a side product of the reaction. The reaction database contains the kinetic parameters. These kinetic parameters are given in table 6.17 (page 134). From the CAPEC reaction database the reaction is assumed to occur in the liquid phase with no reaction in the gas phase. The pressure of operation is at 40 atm. As the reaction is assumed to be performed in the liquid phase, and as the MethylAcetate and Water mixture is a low boiling point azeotrope mixture one could see that the vaporization of some of the Water during the reaction will have an entrainer effect on the MethylAcetate. This removal of products from the liquid phase could thus improve the overall conversion.

$$\mathrm{HOAc} + \mathrm{MeOH} \rightleftarrows \mathrm{MeOAc} + \mathrm{Water} \tag{6.1}$$

6.3.3 Step 3 – Process groups selection and initialisation

From the design problem analysis, all the needed information is available to apply the method to represent the process flowsheet with process groups (see section 4.1.1 page 27). First, the available process groups are ordered by decreasing number of underlining unit operations (see table 4.1 page 26) to prioritize the match of the process groups with the unit operations in the process flowsheet. The pressure swing distillation scheme contains two distillation columns, a mixer, a valve and a compressor. It has thus the highest priority for the match. It is followed by the solvent based azeotropic distillation scheme with two distillation columns and a mixer. The process groups selection and initialization is as follows:

MeOH/MeAC pressure swing distillation. The separation task of the azeotropic binary mixture composed of Methanol and Methyl Acetate is identified. The separation is performed with two distillation columns. The first column is at atmospheric pressure and the second column is at a higher pressure of 6 atm. From the problem analysis (Step 2), the Methanol/Methyl Acetate azeotropic mixture has been found to have a pressure dependent azeotrope. Thus, this separation can be represented with a pressure swing distillation process group. The inlet stream of the process group is a mixture of Methanol and Methyl Acetate, and the two outlet streams are a high purity Methanol stream and a high purity Methyl Acetate stream.

H2O/HOAc pressure swing distillation. The separation task of the azeotropic binary mixture composed of Water and Acetic Acid is identified. The separation is performed with two distillation columns. The first column is at atmospheric pressure and the second column is at a higher pressure of 5 atm. From the problem analysis, the Water/Acetic Acid azeotropic mixture has been found to have a pressure dependent azeotrope. Thus, this separation can be represented with a pressure swing distillation process group. The inlet stream of the process group is a mixture of Water and Acetic Acid, and the two outlet streams are a high purity Water stream and a high purity Acetic Acid stream. It should be noted however that the mixture analysis has shown that the azeotrope has only a slight variation with the pressure. We could thus expect that this pressure swing distillation will be difficult to perform.

Reactor. The esterification reaction to convert Methanol and Acetic Acid into Methyl Acetate and Water is identified. The reaction is performed with a single reactor. From the problem analysis, the reaction, together with the kinetic parameters, is known. The reactor can be represented with a kinetic-model based reactor process group. The inlet stream of the process group is a mixture of Methanol and Acetic Acid, and, the outlet stream is a mixture of non reacted Methanol and Acetic Acid together with the Methyl Acetate and Water products.

Desiccation of the Methyl Acetate/Methanol mixture. The thorough removal of the Water (desiccation) from the Methyl Acetate/Methanol mixture is identified. The separation is performed with a molecular sieve. From the analysis of the ratio of the pure component properties, the dipole moment ratios in the table 6.15 are found.

Binary pair	Dipole moment ratio
Water/Methyl Acetate	1.10
Water/Methanol	1.09
Methanol/Methyl Acetate	1.01

Table 6.15: Dipole moment ratios in the MeAC/MeOH/Water mixture.

The dipole moment ratios shown in table 6.15 indicate that a polar molecular sieve process group can be used to perform this separation. The inlet stream of the process group is a Water, Methyl Acetate and Methanol mixture with the Water in the dilute range. The outlets streams are a pure stream of Water, and, a stream with a mixture of Methyl Acetate and Methanol.

Reactor effluents separation. The separation task of the reactor effluents is identified. This separation is performed with a distillation column. The one-feed two-products distillation column has the overhead product composed of a mixture of Methyl Acetate, Methanol and traces of Water and the bottom product composed of Water and Acetic Acid. The two key components in the separation are Methanol (light key) and Water (heavy key). From the problem analysis, it is found that the Water is forming an azeotrope with the Methyl Acetate. As the Water is the heavy key and the Methyl Acetate is present in the overhead product, the property dependence of a simple distillation column process group cannot be satisfied. A simple distillation column process group cannot separate an azeotropic mixture. Following strictly the property dependency rules, this separation cannot be represented with a simple distillation column process group. However the calculations of the Methyl Acetate/Water azeotrope has not been performed taking into account the other components in the mixture. Keeping in mind that the effect of the other components has not been taken into account in the azeotrope calculations, the separation is represented with a simple distillation process group. The distillation process group has one inlet stream composed of the reactor effluents, and, two outlet streams: in the overhead product a mixture of Methyl Acetate and Methanol with traces of Water, and, in the bottom product a mixture of Acetic Acid and Water.

Process inlet and outlet streams. Finally, the process inlet and the three process outlet streams are represented with inlet and outlet process groups.

Keeping in mind the particular case of the reactor effluents separation with a distillation column process group due to the Methyl Acetate/Water azeotrope, the process flowsheet of the Methyl Acetate production can be represented with process groups as shown on figure 6.13.

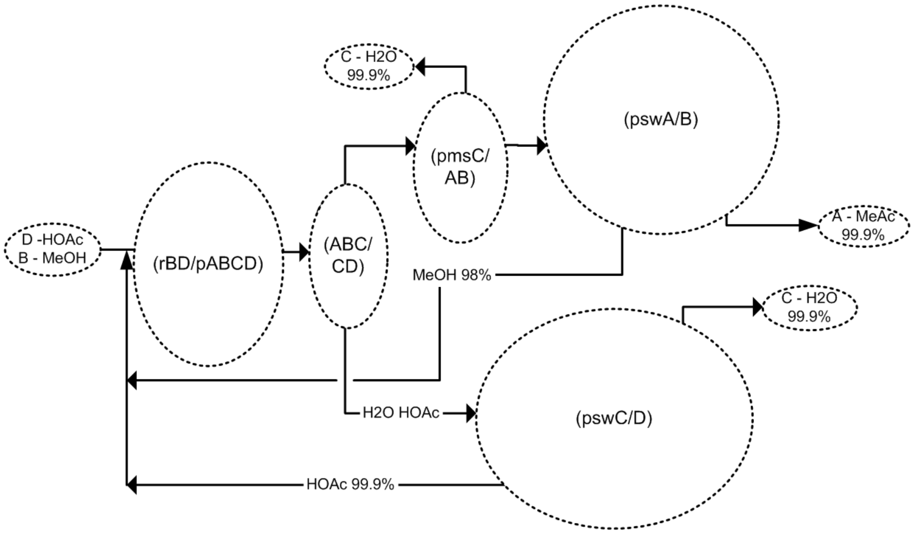


Figure 6.13: Process group representation of the Methyl Acetate production process flowsheet.

6.3.4 Step 6 – Reverse simulation

The flowsheet structure of the Methyl Acetate production being available, the next steps are the resolution of the mass balance between the process groups and the reverse simulation, i.e. the determination of the design parameters, of the unit operations. In this case study, emphasis is given on the reverse simulation of the reactor process group and the pressure swing distillation process group. The mass balance is performed through each process group present in the flowsheet structure. Operating conditions, such as, pressure and temperature of the outlet streams, are given by the process group definition. In the case of the kinetic model based reactor process group, the mass balance calculations rely on the availability of the attainable region diagram. In the next sections, the analysis of the attainable region for the esterification reaction (6.2) is presented. The mass balance for the flowsheet structure (see Fig. 6.13 and Fig. 6.19) is then presented, followed by the reverse simulation of the reactor process group and a pressure swing distillation process group.

Attainable region analysis

The Problem The reaction of Acetic Acid (HOAc) and Methanol (MeOH) to produce Methyl Acetate (MeOAc) and Water is an exothermic, reversible reaction of first order with respect of each reactant which follows the form of Eq. (6.2); it is assumed also that the reaction is taking place in the liquid phase.

$$\text{HOAc} + \text{MeOH} \rightleftarrows \text{MeOAc} + \text{Water} \tag{6.2}$$

Table 6.16 presents the table of moles for the esterification reaction system under study.

Table 6.16: Table of Moles

	Initial	Final	Composition
	$C_{\text{HOAc},0}$	$C_{\text{HOAc},0} - \xi'$	$(C_{\text{HOAc},0} - \xi')/C_{T,0}$
	$C_{\text{MeOH},0}$	$C_{\text{MeOH},0} - \xi'$	$(C_{\text{MeOH},0} - \xi')/C_{T,0}$
	$C_{\text{MeOAc},0}$	$C_{\text{MeOAc},0} + \xi'$	$(C_{\text{MeOAc},0} + \xi')/C_{T,0}$
	$C_{\text{Water},0}$	$C_{\text{Water},0} + \xi'$	$(C_{\text{Water},0} + \xi')/C_{T,0}$
Total	$C_{T,0}$	$C_{T,0}$	1

where ξ' is the extent of reaction.

The kinetic expression for the esterification reaction system is

$$-r_{\xi} = k_f C_{\text{HOAc}} C_{\text{MeOH}} - k_b C_{\text{MeOAc}} C_{\text{water}} \tag{6.3}$$

where the kinetic constants k_f and k_b for the forward and backward reactions, respectively, follow an Arrhenius expression

$$k = A_0 e^{-E_a/RgT} \tag{6.4}$$

Table 6.17 shows the kinetic constants available in the CAPEC reaction database[16] for the forward and backward reactions.

Table 6.17: Kinetic constants for MeAc production

	A_0, $\mathrm{m^3kmol^{-1}min^{-1}}$	E_a, $\mathrm{Jmol^{-1}}$
Forward	9.616×10^5	53670
Backward	3.684×10^6	69170

Problem Definition I. *Fundamental Processes* – We are interested in finding for a given feed material all final conditions that can be achieved in a reactor system of a given volume using all possible combinations of reaction and mixing.

A constant density and constant heat capacity with ideal mixing is also assumed. The characteristic vector and thus the space in which the construction of the Attainable Region is performed is then modified under these assumptions to become the true concentrations and space time of the system, respectively. II. *State Variables* – The residence time (τ) is defined as the ratio of the total volume of the reactor system to the volumetric flow rate of the feed. If the ratio of the heat of reaction to the specific heat capacity of the mixture ($\hat{C}p_m$) is constant, then from the energy balance for an adiabatic reaction the relation between the extent of reaction and the operating temperature can be expressed as

$$T = T_b^\circ + T_{ad}x \tag{6.5}$$

T_b° is referred to as the *basis temperature* and T_{ad} is the increase of temperature of the mixture if it was adiabatically reacted to form pure HOAc; x, is the conversion of the limiting reactant (HOAc). The relationship between the conversion and extent of reaction is given by

$$x = \frac{C_{\mathrm{HOAc},0} - C_{\mathrm{HOAc}}}{C_{\mathrm{HOAc},0}} = \frac{\xi'}{C_{\mathrm{HOAc},0}} \tag{6.6}$$

therefore, substituting Eq. (6.6) in Eq. (6.5)

$$T = T_b^\circ + T_{ad}\xi \tag{6.7}$$

where $\xi = \xi'/C_{\mathrm{HOAc},0}$.

III. *Fundamental Process Vector* – For the above assumptions it is possible to construct the attainable region in the space defined by the characteristic

vector $\mathbf{C} = (\xi, T, \tau)$. The adiabatic reaction vector $\mathbf{R(C)}$, the local change in the characteristic vector due to reaction, is given by $\mathbf{R} = (r_\xi, T_{ad} r_\xi, 1)$, where r_ξ is the rate of depletion of HOAc in terms of the extent of reaction and temperature of the mixture only. However, for an adiabatic reactor the basis temperature T_b° is fixed and the temperature and conversion are thus related via Eq. (6.7); there are consequently only two independent variables, namely, space time and extent of reaction. The space in which the attainable region is constructed in and thus the characteristic vector $\mathbf{C}$ can be reduced to $\mathbf{C} = (\xi, \tau)$; consequently, as the the reaction vector is only function of the temperature and the extent of reaction, and as a relationship exists between the temperature and the extent of reaction, the reaction vector will be reduced to $\mathbf{R} = (r_\xi, 1)$.
IV. *Systems Constraints* – We will not consider restrictions beyond any physical limitations such as materials of constructions and therefore, the maximum allowable temperature in the reactor.

Necessary Conditions These conditions have already been listed in section 4.3.4. It should be noted that when constructing the attainable region in two dimensions, a reactor locus or trajectory divides the space into regions. This is not true for three and higher dimensions.

Construct the Region. The 2D-space example. If it is assumed a stoichiometric feed of 1 kmol/m^3 of HOAc and 1 kmol/m^3 of MeOH and that no MeOAc nor Water is present at the beginning of the reaction, then Eq. (6.3) can be reduced to

$$-r_\xi = k_f (1-\xi)^2 - k_b \xi^2 \tag{6.8}$$

Figure 6.14 is a diagram that shows the extent of reaction ξ as a function of the reactor temperature T for the exothermic reversible esterification reaction (6.2). It is helpful for understanding the reactor layout derived once the attainable region has been generated.

In the exothermic reversible reaction, the rate of the forward reaction increases with increasing temperature but at the same time the equilibrium conversion decreases with increasing temperature. The equilibrium conversion is the locus of points ξ_{eq}(T,ξ) for which the rate $-r$ is zero. This curve, indicated by Ψ_{eq} on figure 6.14, is determined as follow:

$$-r_\xi = k_f (1-\xi)^2 - k_b \xi^2 = 0 \tag{6.9}$$

thus

$$\frac{\xi^2}{1-\xi^2} = \frac{k_f}{k_b} = K_{eq} \tag{6.10}$$

With the kinetic constants of table 6.17 and equation (6.4), it thus possible to determine the curve Ψ_{eq}:

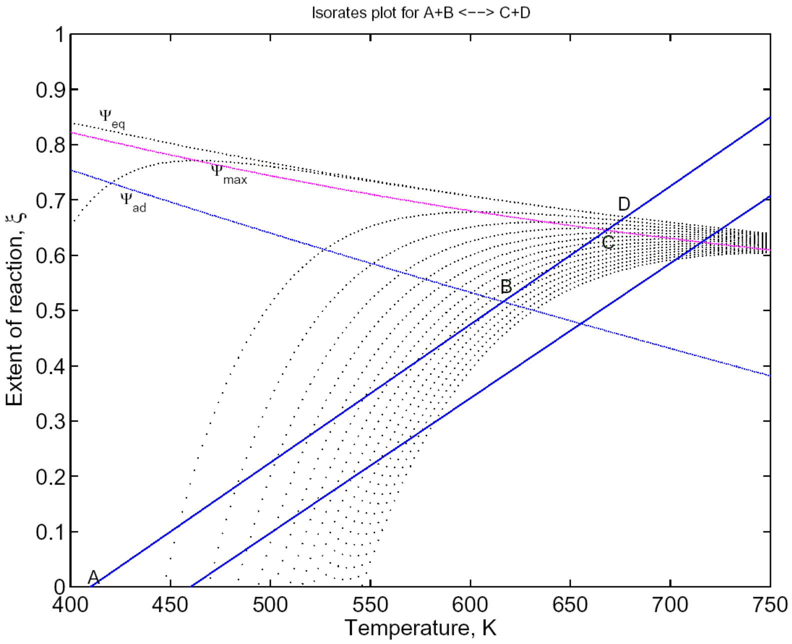


Figure 6.14: Isorates plot for the Esterification reaction.

$$\xi_{eq} = \frac{K_{eq}^{1/2}}{1 + K_{eq}^{1/2}} \tag{6.11}$$

with $K_{eq} = 0.2610e^{\frac{15500}{8.3144T}}$. Thus for $T = 400\ K$, $\xi_{eq} = 0.8401$.

On the other hand, one might be interested to determine the location where the maximum conversion Ψ_{max} can be achieved in a CSTR at constant temperature, $\xi_{max}(T)$. Points on this curve are found by differentiation of r with respect to T:

$$\frac{\partial r}{\partial T} = 0 \tag{6.12}$$

Given $E_f = E_{a_f}/R_g$ and $E_b = E_{a_b}/R_g$:

$$\frac{\partial r}{\partial T} = \frac{E_f}{T^2} k_f (1-\xi)^2 - \frac{E_b}{T^2} k_b \xi^2 = 0 \tag{6.13}$$

$$\frac{E_f \cdot k_f}{E_b \cdot k_b} = K_{eq} \frac{E_f}{E_b} = \frac{\xi^2}{(1-\xi)^2} \tag{6.14}$$

Thus ξ_{max} can be expressed as a function of T:

$$\xi_{max} = \frac{\sqrt{K_{eq} \frac{E_f}{E_b}}}{1 + \sqrt{K_{eq} \frac{E_f}{E_b}}} \tag{6.15}$$

For $T = 400\ K$, $\xi_{max} = 0.8223$, and for $T = 670\ K$, $\xi_{max} = 0.644$ representing point C on figure 6.14.

The meaning of the curve Ψ_{ad} will now be explained. The equation (6.5) is shown as line AD on figure 6.14. This corresponds to all the possible operating points of an adiabatic reactor system, with $T_b^\circ = 410\ K$ and $T_{ad} = 400\ K$. The maximum rate along line AD occures at point B, where the operating line is tangent to an isorate curve. For a smaller T_{ad} this rate cannot be achieved, whereas for a higher T_{ad}, two solutions exist: a high ξ_1 at a high temperature T, and, a low ξ_2 at a lower temperature T.

Thus for a fixed value of T_{ad}, it is possible to draw the curve Ψ_{ad} that corresponds to the tangency points for adiabatic operation at different T_b°. This curve will be referred to as the adiabatic maximum rate curve[18].

In order to determine this curve, it is necessary to remember that we are constrained to adiabatic operation and that there is thus a relationship between T and ξ, which is T(ξ) (see equation (6.5)), as described by the energy balance. The point where the rate is a maximum along the adiabatic operating line is described by

$$\frac{\mathrm{d}r(\xi, T(\xi))}{\mathrm{d}\xi} = 0 = \left.\frac{\partial r(\xi, T)}{\partial \xi}\right|_T + \left.\frac{\partial r(\xi, T)}{\partial T}\right|_\xi \frac{\mathrm{d}T(\xi)}{\mathrm{d}\xi} \tag{6.16}$$

thus for the case where Eq. (6.7) is valid, the curve Ψ_{ad} will follow the form

$$\frac{\mathrm{d}r(\xi, T(\xi))}{\mathrm{d}\xi} = 0 = \left.\frac{\partial r(\xi, T)}{\partial \xi}\right|_T + T_{ad} \left.\frac{\partial r(\xi, T)}{\partial T}\right|_\xi \tag{6.17}$$

where

$$\begin{aligned} \left.\frac{\partial r(\xi, T)}{\partial \xi}\right|_T &= 2k_f(1-\xi) + 2k_b\xi && (6.18) \\ \left.\frac{\partial r(\xi, T)}{\partial T}\right|_\xi &= -k_f(E_f/T^2)(1-\xi)^2 + k_b(E_b/T^2)\xi^2 && (6.19) \end{aligned}$$

substituting Eqs. (6.18) and (6.19) in Eq. (6.17) gives

$$\Psi_{ad} \equiv c + 2b\xi + a\xi^2 = 0 \tag{6.20}$$

where

$$\begin{aligned} a &= \frac{T_{ad}}{T^2}(k_b E_b - k_f E_f) \\ b &= k_f T_{ad} \frac{E_f}{T^2} + k_b - k_f \\ c &= k_f(2 - T_{ad}(E_f/T^2)) \end{aligned}$$

For $T = 400\ K$, $\xi_{ad} = 0.7542$ and for $T = 617.5\ K$, $\xi_{ad} = 0.515$ representing point B on figure 6.14.

Step 1 – Drawing of the PFR trajectory from the feed point.

The equation for a PFR under the above-mentioned assumptions takes the following form:

$$\frac{d\xi}{d\tau} = -r_\xi \tag{6.21}$$

Figure 6.15 describes the PFR trajectory (equation (6.21)) from the feed point, this is for a $\xi = 0$ until the equilibrium has been reached for a basis temperature $T_b^\circ = 410\,K$. Point A corresponds to the inlet of the PFR, with the extent of reaction ξ equals to zero.

Step 2 – Convex hull finding.

We now allow the mixing between all the achievable points by the PFR by making convex the region. The convex hull region for this trajectory can be seen in Fig. 6.16.

Step 3 – Checking of reaction vector direction.

By evaluating the rate vector along this region and projecting it on the orthogonal to the tangent at this region, we see that there are some rate vectors pointing out of the surface of the convex hull, which it will require then to consider additional CSTR trajectories with feed points in the convex hull that extend the region the most, this is at the feed point A in Fig. 6.16.

Step 4 – Finding the new region.

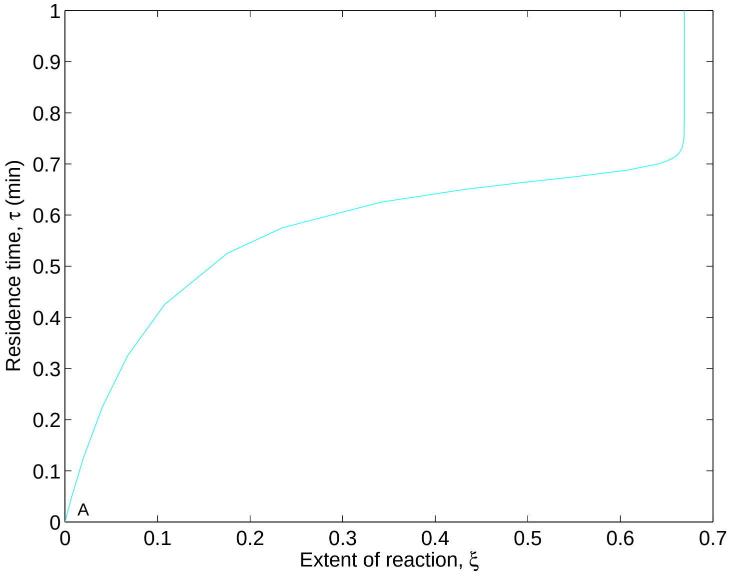


Figure 6.15: PFR profile for Step 1.

Drawing the CSTR trajectory (see Eq. (6.22)) from the feed point A leads to Fig. 6.17.

$$\frac{\xi}{\tau} - r_{\xi} = 0 \tag{6.22}$$

We note, however that this new trajectory is also non-convex and that there are some portions overlapping the PFR trajectory, therefore we fill in the concavities by allowing again mixing between all achievable points. At the minimum of the (τ vs. ξ) curve (point B), $\tau = 1$ (min) at $\xi = 0.52$. This gives $-r = 5.2 moles\ L^{-1} min^{-1}$. From figure 6.14 and r, one obtains $T = 532.26K$. When this isorate curve is followed until $\xi = 0.52$, we see on figure 6.14 that the situation corresponds to an adiabatic reaction with $T_b^{\circ} = 410\ K$ and $T_{ad} = 400\ K$.

Step 5 – Extension of the region and meeting of necessary conditions.

From point B, one alternative (in order to use the full attainable region) is to continue with a PFR until equilibrium and check if there are any rate vectors pointing out of this new extension, which is not the case. Therefore in Figure 6.18 is shown the attainable region for this example, which is explained below under ”interpret the boundary”.

Interpret the Boundary The boundary of the attainable region for feed material A is made up of a CSTR operating at point B, mixing with the feed

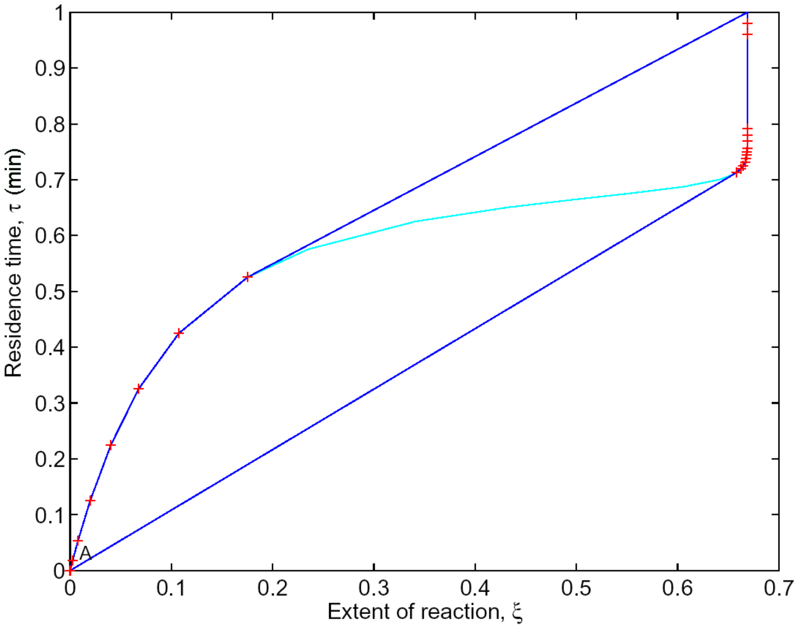


Figure 6.16: PFR convex hull.

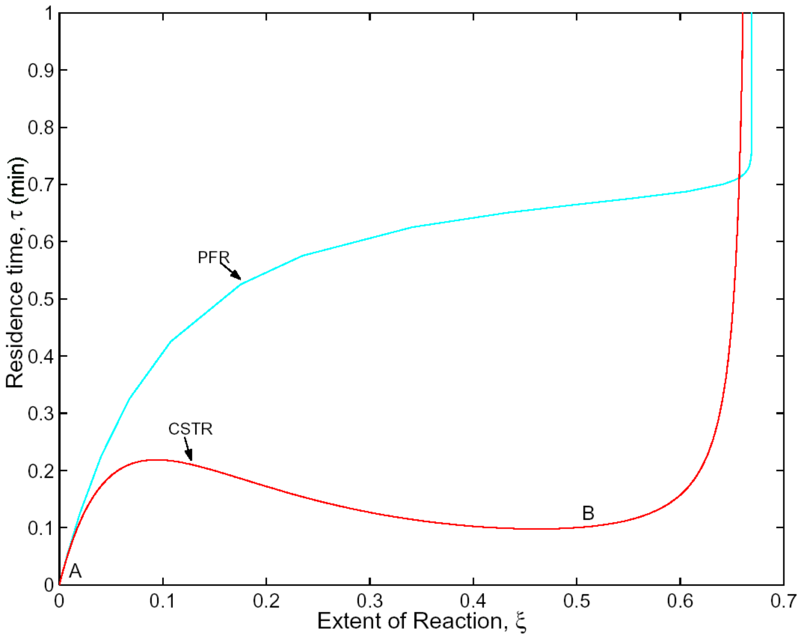


Figure 6.17: CSTR trajectory.

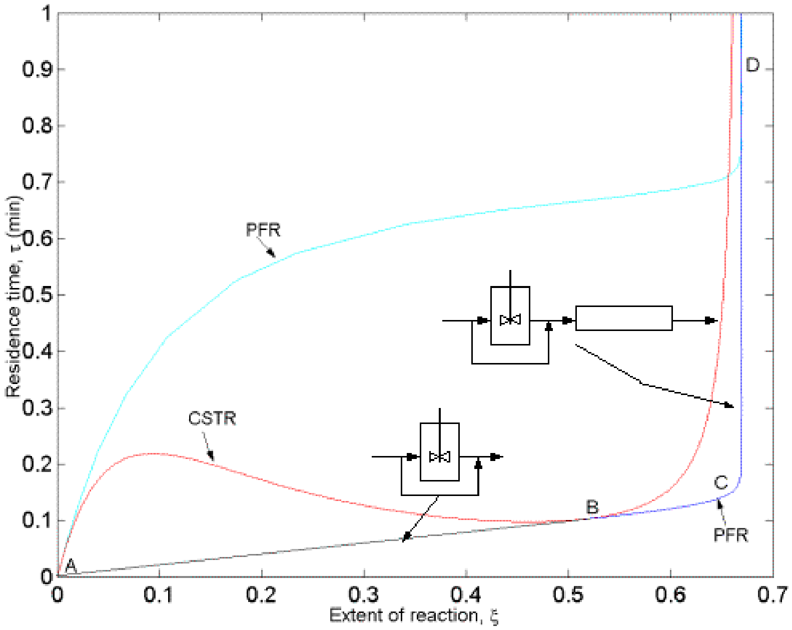


Figure 6.18: Attainable Region for $T_b^\circ = 410\,\mathrm{K}$.

material to give line AB, and a plug flow reactor with feed at point B operating along curve BD. All the points above ABD and on ABD can be reached (attained) if the reaction kinetics and associated assumptions match those given in table 6.17.

The point B corresponds to the maximum reaction rate that can occur in an adiabatic reactor. This can be seen from the properties of a CSTR; the line AB is collinear with the rate vector $(r_\xi, 1)$. Line AB fills in the concavity in the CSTR locus and must thus have the minimum slope of all possible lines between the feed point A and points on the locus; consequently point B must thus correspond to the maximum value of r_ξ along the reactor locus. Let us consider now this layout on Fig. 6.14. The adiabatic energy balance as given by Eq. (6.7) and shown as line AD, correspond to all the possible operating points of an adiabatic reactor system with basis temperature of 410 K. The maximum rate along line AD (for adiabatic operation) occurs at point B, where the operating line is *tangent* to an isorate curve and does not correspond to point C on the maximum rate curve Ψ_{max}. Note that even if the operating line were not straight but a curve, the maximum rate in an adiabatic reactor would still correspond to the point where the operating (energy balance) curve was tangent to an isorate line. Note, that for the flowsheet described in Fig. 6.19, a single CSTR operating at point B is considered.

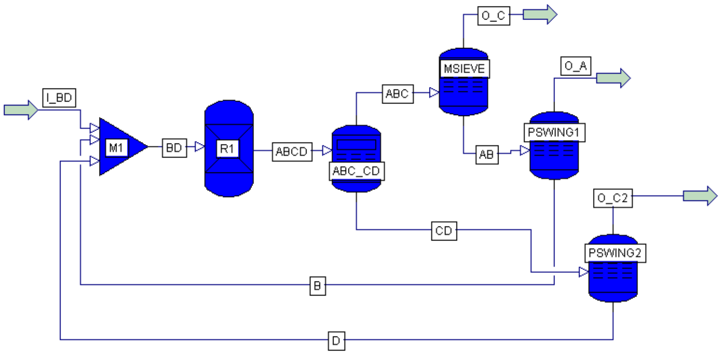


Figure 6.19: Mass balance simulation of the Methyl-Acetate production.

Mass balance

The mass balance is now performed through all the process groups in the flowsheet structure. With the exception of the reactor process group, all the other process groups have their mass balance specifications defined. The mass balance specifications of the reactor process group is available from the corresponding attainable region. The attainable region of the esterification reaction is shown on figure 6.18 (page 142). From the attainable region, it is possible to observe that to maximize the extent of reaction while minimizing the residence time, the reactor network should be either: a CSTR with a bypass (line AB on figure 6.18); a single CSTR at point B; or a CSTR with a bypass, followed with a plug flow reactor. To operate with the simplest reactor network, a CSTR operating at point B is chosen. This implies an extent of reaction $\xi = 0.52$.

With the definition of the extent of reaction from the attainable region, all the process groups mass balance specifications are defined. The mass balance is performed and the results are presented in table 6.18. They correspond to the mass balance simulation given in figure 6.19. Note, that it is not possible yet, to know the pressure at the outlets of the pressure swing distillation columns.

Reverse simulation of the reaction process group

The extent of the esterification reaction was defined to be $\xi = 0.52$ for the flowsheet structure mass balance. Given the mass balance results in table 6.18, the attainable region on figure 6.18 and the extent of reaction, the design parameters of the CSTR reactor such as inlet temperature and residence time can be determined.

The inlet temperature of the CSTR reactor is 410K, temperature at which the attainable region analysis has been performed. The residence time in the reactor is directly available from the attainable region. At $\xi = 0.52$, the residence time $\tau = 6$ *sec*. From the reaction stoichiometry the molar flowrate is constant in the reactor. According to the PRO/II simulator, used to perform

Stream	BD	ABCD	ABC	CD	O_C	AB
Phase	Liquid	Mixed	Mixed	Mixed	Water	Liquid
Temperature (K)	326.089	410.00	326.513	362.202	326.513	326.513
Pressure (atm)	1.000	40.000	1.000	1.000	1.000	1.000
Total (kmole/hr)	3.817	3.817	1.918	1.898	0.010	1.908
Composition						
MEAC	0.0013	0.2606	0.5185			0.5212
MEOH	0.4987	0.2394	0.4763			0.4788
WATER	0.0013	0.2606	0.0052	0.5187	1.0000	
HOAC	0.4987	0.2394		0.4813		

Stream	O_A	B	O_C2	D	I_BD
Phase	Liquid	Liquid	Mixed	Liquid	Liquid
Temperature (K)	330.373	338.276	362.202	375.707	298.000
Pressure (atm)	1.000	1.000	1.000	1.000	1.000
Total (kmole/hr)	0.994	0.914	0.984	0.914	2.000
Composition					
MEAC	0.9954	0.0054			
MEOH	0.0046	0.9946			0.5000
WATER			0.9954	0.0054	
HOAC			0.0046	0.9946	0.5000

Table 6.18: Mass balance results for the Methyl-Acetate production

the mass balance, the actual flowrate in the outlet of the reactor is 1.557 $m3/hr$ in a liquid state at $T = 532.26\,K$ and $P = 40\,atm$. This implies to have a small reactor with an active volume of 2.6 L (Height of 0.3 m and diameter of 0.11 m considering a 10% increase in size for the internals). The CSTR must be equipped with a cooling jacket for the temperature control.

Reverse simulation of the pressure swing distillation

From the MeOH/MeAC azeotrope pressure dependency shown on figure 6.10, it can be seen that the azeotropic concentration varies from 0.68 of MeAc at 1atm, to 0.5 of MeAc at 6atm. From the mass balance, the inlet of the `(pswA/B)` process group is at 0.52 of MeAc, so the first column will be operated at atmospheric pressure and the second column at 6atm. The first column will recover pure MeOH in the bottom product and a mixture at azeotropic composition in the overhead product. The second column will recover pure MeAc as the bottom product and a mixture at azeotropic composition in the overhead product. The overhead product of the second column will be recycled back to the first column.

The maximum driving force for the first column is 0.17 and for the second 0.1. This leads to a feed stage position at 0.674% and 0.27% in columns one and two respectively, as the feed stage position is identified on the driving force between the azeotrope composition and the pure component. From the tables of precalculated data in appendix C, the reflux ratios are set to 3.53 and 6.74 in columns one and two respectively. The configuration is given in figure 6.20. In the outlet stream of MeAc and the outlet stream of MeOH, the MeAc composition is 0.999 and the MeOH composition is 0.998.

6.3.5 Step 8 – Final verifications

The final verifications have been performed using the PRO/II rigorous simulator. It has been possible to confirm the design of all the unit operations with the exception of the separation of Water and Acetic Acid using pressure swing distillation. The thermodynamic model, NRTL, describes the azeotrope, but the rigorous simulation of the first column using the Chemdist resolution model in PRO/II performs directly the Water/Acetic Acid separation, crossing the azeotrope.

6.3.6 Conclusion

The results show that the framework for CAFD is providing a fast, efficient and systematic approach for process design. By first solving the mass balance based on the process group specifications and then calculating the design parameters of the unit operations through the reverse simulation methods, the process flowsheet design can be performed independently for the unit operations of each process group. As long as the design of the unit operations satisfy the

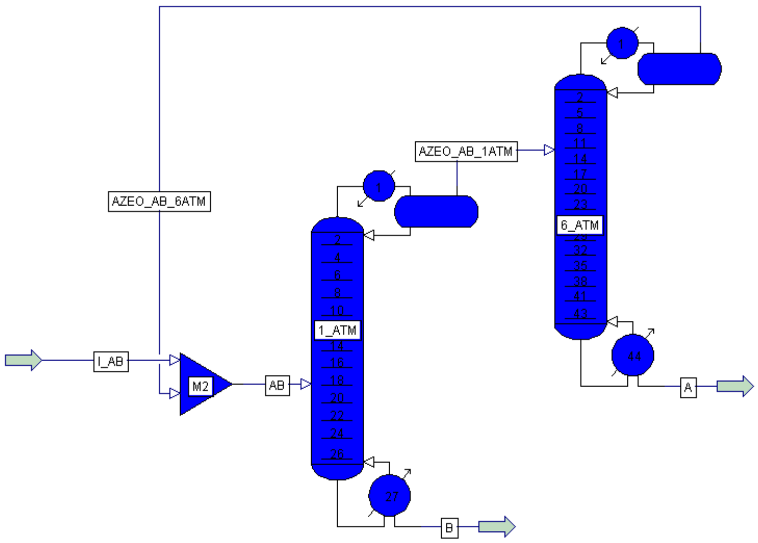


Figure 6.20: Pressure swing distillation for the separation of MeOH and MeAC.

specifications of the process groups, the mass balance is satisfied and does not need to be performed again with the design of each unit operation. In particular, the results show that the reverse simulation of reaction process group by applying the attainable region analysis method, provides an efficient way to obtain the optimal reactor network together with all the necessary design parameters, such as, temperature of operation, residence time and sizing of the reactor. The results also show, that the framework for CAFD, by providing a systematic method for the analysis of the process design problem, gives the engineer all the necessary information to perform the appropriate review of the rigorous simulation results. It should however be noted, that, in some cases, the process groups cannot represent a feasible process flowsheet. This limitation can be progressively overcome by the addition in the framework for CAFD of new process groups to describe new type of separations or particular cases of already supported separation types.

6.4 Benzene Production

In this section, the case study involving the production of Benzene by Toluene hydrodealkylation is presented. The case study highlights the application of the framework for CAFD with respect to the generation of flowsheet alternatives with retrofit at the process level and the handling of the recycles in the mass balance calculations with the definition of the purges. The numbering of the steps corresponds to the steps in the work flow of the framework for CAFD presented in figure 5.1 (page 83).

6.4.1 Step 1 – Definition of the problem

The problem definition is to apply the framework for CAFD to the retrofit at the process level and design the Benzene production flowsheet. The synthesis problem is to produce Benzene with the highest conversion of the available raw materials at the minimum energy cost subject to the conservation of an already existing part of a process. For technical and financial reasons, the part of the process to be kept fixed is composed of the reactor and the immediate downstream separation through a flash separation (see figure 6.21). Within the catalytic reactor, Toluene hydrodealkylation is performed at high pressure (35 atm) and high temperature (900 K). The effluents of the reactor are then cooled with a heat exchanger (not shown) and sent to the flash for a first separation of the light ends. The available raw materials to produce the Benzene are Toluene and Hydrogen. Hydrogen is only available with Methane impurities.

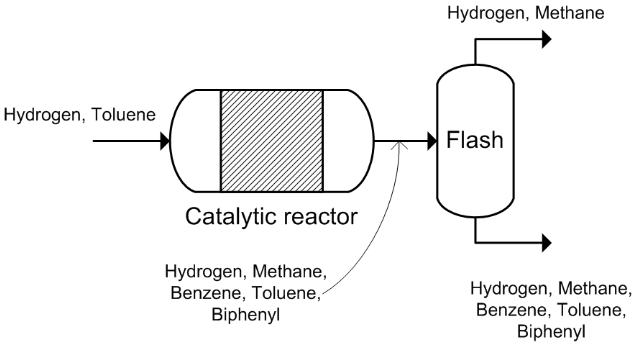


Figure 6.21: Fixed part in the Benzene production retrofit.

The structural definition of the synthesis problem is as follows: 2 inlets of Toluene and Hydrogen, 1 outlet of Benzene and a fixed process flowsheet part or backbone. The backbone has itself 1 inlet of a Toluene/Hydrogen mixture and 2 outlets of an Hydrogen/Methane mixture and a mixture composed of Hydrogen, Methane, Benzene, Toluene and Biphenyl.

6.4.2 Step 2 – Analysis of the problem

From the analysis of the mixture, distillation has been found to be the appropriate separation technique for the separations between Methane, Benzene, Toluene and Biphenyl. From the ratio of pure component properties in table 6.19 (page 149), flash separation has been found to be an appropriate separation technique for the separation of Hydrogen and Methane from the other components in the problem. The liquid membrane, the gas membrane and the molecular sieve separation techniques have been found to be able to selectively separate Benzene and/or Toluene from a mixture of Hydrogen, Methane, Benzene and Toluene. No azeotropes have been found.

The reaction analysis has confirmed the possibility to produce Benzene by Toluene hydrodealkylation (see equation (6.23)). The search in the reaction database also confirmed the secondary reaction of producing Biphenyl from Benzene (see equation (6.24)). The reaction database provides only conversion rates for the reactions. As the Biphenyl is found to have financial value, the structural problem definition of the synthesis problem is refined to include one outlet of Biphenyl.

$$\text{Hydrogen} + \text{Toluene} \rightleftarrows \text{Methane} + \text{Benzene} \tag{6.23}$$

$$2 \cdot \text{Benzene} \rightleftarrows \text{Hydrogen} + \text{Biphenyl} \tag{6.24}$$

6.4.3 Step 3 – Process group selection and initialization

In the particular case of a retrofit at the process level problem, both, process groups must be selected and initialized, and, the backbone (see figure 6.21) must be represented with process groups. The procedure to represent the backbone with process groups is similar to the procedure for the representation of a process flowsheet with process groups (see section 4.1.1 page 27). The only difference is that the inlet and outlet streams of the backbone are not connected to inlet and outlet process groups.

From the analysis of the problem, distillation, liquid membrane based separation, gas membrane based separation, flash separation and molecular sieve based separation have been found to be the appropriate separation techniques to be performed. From the database of process groups, the corresponding process groups are retrieved and matched with the possible mixtures of the 5 components. For the gas membrane, liquid membrane, flash and molecular sieve process groups, the matching conditions are given in table 6.20. A distillation process group is initialized with a given mixture if the relative volatilities of the components are in the range supported by the process group, the driving force between the two key components is matching with the process group driving force and if the number of components in the mixture corresponds to the number of components of the process group. A reactor process group is initialized with a given mixture based on the components in the reactions. The corresponding list of 31 initialized process groups is given in table 6.21.

The fixed part of the process (backbone) is represented with the process groups available in the list of initialized process groups. Two process groups are found to be needed to represent the backbone, a fixed conversion reactor process group and a flash process group: `(rAD/pABCDE)(fAB/ABCDE)`.

Property	A – Hydrogen	B – Methane	C – Benzene	D – Toluene	E – Biphenyl
M_w	2.016	16.043	78.114	92.141	154.211
ω	−0.216	0.0115	0.21	0.2621	0.3654
T_c	33.19	190.564	562.16	591.8	789.26
P_c	12.958	45.389	48.339	40.523	37.997
Z_c	0.305	0.286	0.271	0.264	0.295
V_c	0.1	0.1	0.3	0.3	0.5
T_b	20.39	111.66	353.24	383.78	528.15
$d_m \times 1 \cdot 10^{-30}$ $(C \cdot m)$	0	0	0	0.3597	0
r_g	0.371	1.118	3.004	3.472	4.834
T_m	13.95	90.694	278.68	178.18	342.37
T_{tp} (K)	13.95	90.694	278.68	178.18	342.2
P_{tp} (Pa)	$7.13E-02$	$1.15E-01$	$4.70E-02$	$4.18E-07$	$9.25E-04$
M_v $(m^3/kmol)$	$2.86E-02$	$3.80E-02$	$8.95E-02$	$1.07E-01$	$1.55E-01$
H_f $(kJ/kmol)$	$0.00E+00$	$-7.45E+04$	$8.29E+04$	$5.02E+04$	$1.82E+05$
G_f $(kJ/kmol)$	$0.00E+00$	$-5.05E+04$	$1.30E+05$	$1.22E+05$	$2.80E+05$
S_{IG} $(kJ/(kmol \cdot K))$	$1.31E+02$	$1.86E+02$	$2.69E+02$	$3.21E+02$	$3.94E+02$
H_{fus} $(kJ/kmol)$	$1.17E+02$	$9.41E+02$	$9.87E+03$	$6.64E+03$	$1.86E+04$
H_{comb} $(MJ/kmol)$	$-2.42E+05$	$-8.03E+05$	$-3.14E+06$	$-3.73E+06$	$-6.03E+06$
δ $(\sqrt{kJ/m^3})$	6.6478	11.6	18.7296	18.3242	19.2549
V_{vw} $(m^3/kmol)$	0.0063	0.017	0.0484	0.0595	0.0917
A_{vw} $(m^2/kmol)$	$1.43E+08$	$2.88E+08$	$6.00E+08$	$7.42E+08$	$1.07E+09$
P_{nvap} (Pa)	$3.94E+07$	$1.87E+07$	$1.26E+04$	$3.78E+03$	$-2.04E+02$

Table 6.19: Pure component properties in the Benzene production synthesis problem.

Process group	**Minimum ratio of the specific properties for the key components**
Gas membrane	ratio $(V_{vw}) > 1.07$
	ratio $(T_c) > 1.10$
Liquid membrane	ratio $(r_g) > 1.01$
	ratio $(M_v) > 1.02$
	ratio $(\delta) > 1.2$
Flash	ratio $(T_b) > 1.23$
Molecular sieve	ratio $(V_{vw}) > 1.07$

Table 6.20: Mixture matching conditions for process group initialization.

`(ABC/D)`	`(ABC/DE)`	`(ABCD/E)`	`(ABD/E)`
`(AC/D)`	`(AC/DE)`	`(ACD/E)`	`(AD/E)`
`(BC/D)`	`(BC/DE)`	`(BCD/E)`	`(BD/E)`
`(C/D)`	`(C/DE)`	`(CD/E)`	`(D/E)`
`(AB/CDE)`	`(AB/CD)`	`(AB/C)`	`(msC/BA)`
`(gmemDC/BA)`	`(lmemDC/BA)`	`(gmemD/CBA)`	`(gmemC/BA)`
`(msDC/BA)`	`(lmemD/C)`	`(msD/CBA)`	`(lmemD/CBA)`
`(gmemD/C)`	`(fAB/ABCDE)`	`(rAD/pABCDE)`	

Table 6.21: Initialized process groups for the Benzene production synthesis problem.

6.4.4 Step 4 – Generation of the flowsheet structure alternatives

Given the Hydrogen (Methane impurities not considered) and Toluene raw materials to be converted into Benzene and Biphenyl, using 5 separation techniques and a reactor represented by 31 process groups (see table 6.21) and one backbone, at the minimum consumption of energy, the flowsheet generation algorithm of the framework for CAFD generates 31 feasible alternatives. From table 6.22, it is interesting to see that the algorithm is analysing more than 1000 structures when generating the alternatives. The complete resolution time on a desktop computer was around only 20s.

Number of completed alternatives	31
Number of analysed structures	1117
Total time	$\approx 20s$
Maximum memory usage	$\approx$ 9MB

Table 6.22: Benzene production computational statistics

6.4.5 Step 5 – Ranking and selection of an alternative

The table 6.23 is providing the list of the 31 generated flowsheet alternatives. For the purpose of simplification of the SFILES strings, the two process groups of the backbone have been replaced with `(backbone)` and the inlet streams of Hydrogen and Toluene have been merged into one inlet process group: `(iAD)`. Out of the 31 generated feasible flowsheet alternatives, it should be noted that the first 14 have the same energy index E_x. This is explained by the fact that the energy index E_x is only calculated for the distillation process groups, the other process groups do not contribute to the energy index (contribution equals to 0.0). The following alternative is chosen for the reverse simulation:

`(iAD)(backbone)1<2<1(AB/CDE)(oAB)](C/DE)[(oC)](D/E)2(oE)`

The SFILES string fragment `(iAD)(backbone)1<2<1` corresponds to:

`(iAD)(rAD/pABCDE)<1<2(fAB/ABCDE)1`.

This shows that the overhead product of the flash is recycled back to the reactor.

Table 6.23: Generated flowsheet structure alternatives.

Alternative SFILES string	**Energy index E_x**
`(iAD)(backbone)1<2<1(ABC/DE)(msC/BA)[(oC)](oAB)](D/E)2(oE)`	0.031717
`(iAD)(backbone)1<2<1(ABC/DE)(gmemC/BA)[(oC)](oAB)](D/E)2(oE)`	0.031717
`(iAD)(backbone)1<2<1(ABCD/E)(lmemD/CBA)2(msC/BA)[(oC)](oAB)](oE)`	0.031717
`(iAD)(backbone)1<2<1(ABCD/E)(lmemD/CBA)2(gmemC/BA)[(oC)](oAB)](oE)`	0.031717
`(iAD)(backbone)1<2<1(ABCD/E)(msDC/BA)[(lmemD/C)2(oC)](oAB)](oE)`	0.031717
`(iAD)(backbone)1<2<1(ABCD/E)(msDC/BA)[(gmemD/C)2(oC)](oAB)](oE)`	0.031717
`(iAD)(backbone)1<2<1(ABCD/E)(msD/CBA)2(msC/BA)[(oC)](oAB)](oE)`	0.031717
`(iAD)(backbone)1<2<1(ABCD/E)(msD/CBA)2(gmemC/BA)[(oC)](oAB)](oE)`	0.031717
`(iAD)(backbone)1<2<1(ABCD/E)(gmemD/CBA)2(msC/BA)[(oC)](oAB)](oE)`	0.031717
`(iAD)(backbone)1<2<1(ABCD/E)(gmemD/CBA)2(gmemC/BA)[(oC)](oAB)](oE)`	0.031717
`(iAD)(backbone)1<2<1(ABCD/E)(lmemDC/BA)[(lmemD/C)2(oC)](oAB)](oE)`	0.031717
`(iAD)(backbone)1<2<1(ABCD/E)(lmemDC/BA)[(gmemD/C)2(oC)](oAB)](oE)`	0.031717

Continued on next page

Alternative SFILES string	**Energy index E_x**
(iAD)(backbone)1<2<1(ABCD/E)(gmemDC/BA)[(lmemD/C)2(oC)](oAB)](oE)	0.031717
(iAD)(backbone)1<2<1(ABCD/E)(gmemDC/BA)[(gmemD/C)2(oC)](oAB)](oE)	0.031717
(iAD)(backbone)1<2<1(ABCD/E)(ABC/D)2(msC/BA)[(oC)](oAB)](oE)	0.064096
(iAD)(backbone)1<2<1(ABCD/E)(ABC/D)2(gmemC/BA)[(oC)](oAB)](oE)	0.064096
(iAD)(backbone)1<2<1(ABCD/E)(msDC/BA)[(C/D)2(oC)](oAB)](oE)	0.098749
(iAD)(backbone)1<2<1(ABCD/E)(lmemDC/BA)[(C/D)2(oC)](oAB)](oE)	0.098749
(iAD)(backbone)1<2<1(ABCD/E)(gmemDC/BA)[(C/D)2(oC)](oAB)](oE)	0.098749
(iAD)(backbone)1<2<1(ABC/DE)(AB/C)[(oAB)](oC)](D/E)2(oE)	0.245415
(iAD)(backbone)1<2<1(ABCD/E)(lmemD/CBA)2(AB/C)[(oAB)](oC)](oE)	0.245415
(iAD)(backbone)1<2<1(ABCD/E)(msD/CBA)2(AB/C)[(oAB)](oC)](oE)	0.245415
(iAD)(backbone)1<2<1(ABCD/E)(gmemD/CBA)2(AB/C)[(oAB)](oC)](oE)	0.245415
(iAD)(backbone)1<2<1(ABCD/E)(ABC/D)2(AB/C)[(oAB)](oC)](oE)	0.277794
(iAD)(backbone)1<2<1(ABCD/E)(AB/CD)[(oAB)](lmemD/C)2(oC)](oE)	0.292408
(iAD)(backbone)1<2<1(ABCD/E)(AB/CD)[(oAB)](gmemD/C)2(oC)](oE)	0.292408
(iAD)(backbone)1<2<1(ABCD/E)(AB/CD)[(oAB)](C/D)2(oC)](oE)	0.35944
(iAD)(backbone)1<2<1(AB/CDE)(oAB)](CD/E)[(lmemD/C)2[(oC)](oE)	0.411659
(iAD)(backbone)1<2<1(AB/CDE)(oAB)](CD/E)[(gmemD/C)2(oC)](oE)	0.411659
(iAD)(backbone)1<2<1(AB/CDE)(oAB)](C/DE)[(oC)](D/E)2(oE)	0.465658
(iAD)(backbone)1<2<1(AB/CDE)(oAB)](CD/E)[(C/D)2(oC)](oE)	0.478691

6.4.6 Step 6 – Reverse simulation of a flowsheet structure alternative

A flowsheet alternative being selected, the next steps are the resolution of the mass balance between the process groups and the reverse simulation, i.e. the determination of the design parameters, of the unit operations. From the SFILES string of the alternative, it can be observed that two recycles exist from separation process groups to the reactor process group:

`(iAD)(rAD/pABCDE)<1<2(fAB/ABCDE)1(AB/CDE)(oAB)]`
`(C/DE)[(oC)](D/E)2(oE)`

This case study gives emphasis on the handling of these recycles when performing the mass balance. The two recycle loops are given on figure 6.22.

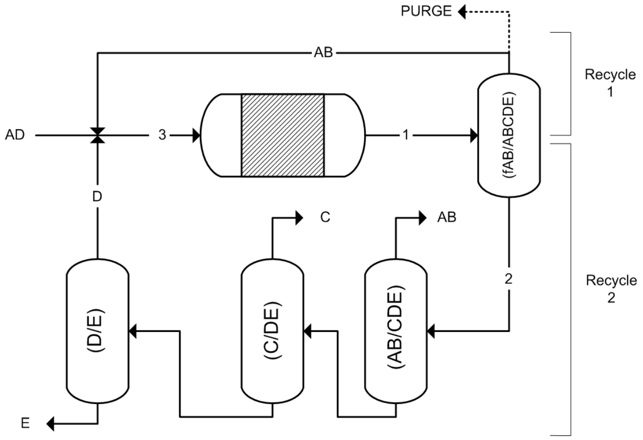


Figure 6.22: Recycles in the selected Benzene production alternative.

Analysis of the recycle 2. The recycle 2 is going through 5 process groups: the reactor and flash process groups and then 3 distillation process groups. By applying the recycle analysis method presented in section 4.3.1 (page 59), only one component, D – Toluene, is found to be present in all the loop. As Toluene is a reactant, no purges are needed to be defined on this recycle loop.

Analysis of the recycle 1. The recycle 1 is going through 2 process groups: the reactor and the flash process groups. Two components are found to be present in all the loop, A – Hydrogen and B – Methane. As Hydrogen is a reactant, no purge is needed for this component, however Methane is a by-product of the reaction. To avoid a build up of Methane in the recycle 1, a purge needs to be defined.

The purge is positioned on the stream from the flash to the reactor. From the equation (4.14) (page 60), the following purge rate of Methane is found:

$$n_{purge} = (n_{AD} + n_D + n_1) - (n_2 + n_3) \tag{6.25}$$

with n, the individual flowrate of Methane in the given streams (see figure 6.22).

From the specifications of the process groups and the definition of the purge, the mass balance calculations can be performed. The results of the mass balance calculations are presented in table 6.24. It should be noted in the results of the mass balance calculations, that the specifications of overhead product of the `(AB/CDE)` distillation process group has been changed from bubble point temperature to dew point temperature as achieving a bubble point specification for an Hydrogen/Methane mixture is not appropriate with a calculated temperature of 4 K.

6.4.7 Step 7 – Post analysis

From the mass balance calculations, a first post analysis of the flowsheet can be performed. The mass balance calculations provide the flowrate, pressure and temperature of all the streams within the flowsheet. As presented in the definition of the problem, the inlet of the reactor need to be pre-heated and reactor effluents are then cooled before the flash separation. Douglas[9] proposes the complete heat integration of this process flowsheet. To perform the complete heat integration of this process flowsheet, the duties of the distillation columns are necessary. The heat integration of the distillation columns is not possible from the mass balance calculations, as the condenser and reboiler duties are not available. However, the mass balance calculations provide the necessary data to generate basic heat integration as presented on figure 6.23. This simple heat integration is performed by using the hot effluents of the reactor to pre-heat the reactants. To further reduce the cold utilities before the flash separation, the liquid outlet of the flash is divided to quench the effluents of the reactor. On figure 6.23, HX1 represents the cold side and HX2 the hot side of the same heat exchanger.

6.4.8 Step 8 – Final verifications

The final verifications have been performed using the PRO/II rigorous simulator, with the results being presented in table 6.25. The rigorous simulation confirms the mass balance calculations. This rigorous simulation can be used as a basis for further heat integration analysis.

6.4.9 Conclusion

The results show that the framework for CAFD is providing a fast and efficient approach for retrofit at the process level. The retrofit at the process level is simply performed by building the flowsheet alternatives around an existing backbone composed of process groups. The results show that the synthesis algorithm is able to handle different types of process groups, such as, molecular

Stream	AD	C	E	D	AB	PURGE
Phase	Mixed	Liquid	Liquid	Liquid	Vapor	Vapor
Temperature (K)	297.43	349.47	480.24	384.07	338.90	310.93
Pressure (atm)	37.43	1.02	1.02	1.02	10.20	31.64
Flowrate ($kmol/hr$)	772.00	265.11	5.08	89.85	19.07	482.73
Composition						
Hydrogen	0.6054				0.0664	0.3978
Methane	0.0319	0.0003			0.8635	0.5940
Benzene		0.9980		0.0148	0.0701	0.0073
Toluene	0.3627	0.0019	0.0876	0.9849		0.0008
Biphenyl			0.9124	0.0003		

Table 6.24: Mass balance results for the Benzene production

Stream	AD	C	E	D	AB	PURGE	QUENCH
Phase	Mixed	Vapor	Liquid	Liquid	Vapor	Vapor	Liquid
Temperature (K)	302.75	353.15	516.30	384.44	303.15	307.69	310.93
Pressure (atm)	37.43	1.02	1.02	1.02	9.87	31.64	31.64
Flowrate ($kmol/hr$)	772.00	273.25	4.66	90.17	12.11	481.98	119.30
Composition							
Hydrogen	0.6054	0.0001			0.1002	0.3969	0.0032
Methane	0.0319	0.0224			0.8835	0.5960	0.0443
Benzene		0.9772		0.0030	0.0151	0.0065	0.7035
Toluene	0.3627	0.0003	0.0193	0.9970	0.0011	0.0007	0.2370
Biphenyl			0.9807	0.0001			0.0120

Table 6.25: Rigorous simulation results for the Benzene production

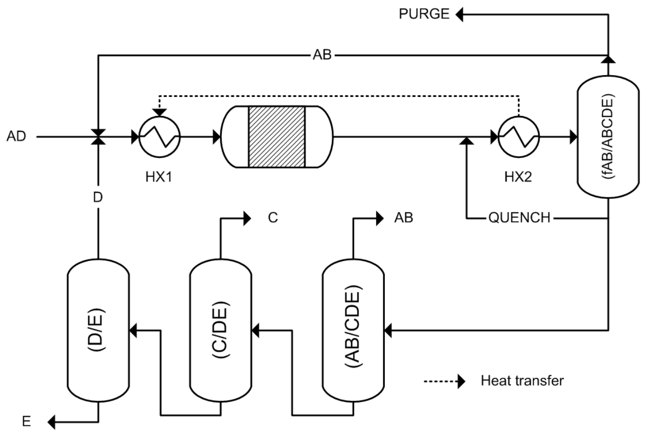


Figure 6.23: Heat integration of the Benzene production from the mass balance calculations.

sieve PG, reactor PG, flash PG, liquid/gas membrane PG, without adding complexity to the problem. However, it should be noted that when a process group has a zero contribution to the energy index, E_x, ranking of the different flowsheet alternatives is difficult. This has also been shown in the case study on the Aromatic-paraffin mixture separation (see section 6.2 page 110).

The results also show, that the systematic design approach provides the guidance to setup the appropriate purges on the recycles to satisfy the mass balance. It should however be noted that the mass balance calculations do not provide all the necessary data for a heat integration including the distillation columns. It is still possible to determine a first simple heat integration structure based on the available hot and cold streams.

6.5 Ammonia Reaction

In this section, the case study of the application of the SFILES notation to another field such as the Ammonia reaction system is presented. The case study highlights the application of the SFILES notation to the representation of reaction pathways. The mechanisms or pathways of the catalytic synthesis of ammonia from Nitrogen and Hydrogen, $N_2 + 3H_2 \rightleftharpoons 2NH_3$, have been investigated because of its long history and enormous economic implication. From the literature Fan et al.[11] have identified 6 mechanisms or pathways from a list of 11 plausible elementary reactions.

The list of plausible elementary reactions are given in table 6.26. In the elementary reactions l represents an active site on the catalyst, when alone it represents a free active site and when associated with a molecule or atom it indicates that the molecule or atom is associated with the active site. The 6 identified mechanisms or pathways are given in table 6.27.

(1)	$H_2 + l \rightleftharpoons H_2l$
(2)	$H_2l + l \rightleftharpoons Hl + Hl$
(3)	$N_2 + l \rightleftharpoons N_2l$
(4)	$N_2l + l \rightleftharpoons Nl + Nl$
(5)	$N_2l + H_2l \rightleftharpoons N_2H_2l + l$
(6)	$N_2H_2l + l \rightleftharpoons NHl + NHl$
(7)	$Nl + Hl \rightleftharpoons NH_2l + l$
(8)	$NHl + Hl \rightleftharpoons H_2l$
(9)	$NHl + H2l \rightleftharpoons NH_3l + l$
(10)	$NH_2l + Hl \rightleftharpoons NH_3l + l$
(11)	$NH_3l \rightleftharpoons NH_3 + l$

Table 6.26: List of plausible elementary reactions

The SFILES notation is able to represent directed graphs with the nodes and the bonds. To use the SFILES notation for the representation of the reaction pathways the following elements are needed:

- The graph representation of a pathway with the definition of the nodes, the bonds and the direction of the bonds;
- the invariant or a method to calculate the invariant of the pathway graph.

6.5.1 Graph representation of a network of reactions

To represent the network of reactions with a graph, a process graph[13] (P-Graph) representation has been chosen. A process graph represents both the reactions and the chemical entities as nodes. The bonds connect chemical entities to reactions to represent the process of chemical entities by reactions. As all the reactions are reversible, only the forward step of the reactions are represented. For example on figure 6.24, the Hydrogen molecule is fixed on an active site by reaction (1). The reaction (2) converts an Hydrogen molecule on an active site with another active site into two Hydrogen atoms on two active sites. A molecule or atom and the same molecule or atom fixed on an active site are considered as two distinct chemical entities. A free active site is also considered as a distinct chemical entities.

By applying this graph representation, the corresponding P-graph representation of the first reaction pathway in table 6.27 is shown on figure 6.25.

6.5.2 Invariant of the pathway graph

The 11 plausible elementary reactions involve the following 12 chemical entities: H_2, l, H_2l, Hl, N_2, N_2l, Nl, NHl, N_2H_2l, NH_2l, NH_3l and NH_3. The result is a maximum of 23 distinct nodes in a pathway that would involve the 12

	Pathway 1		**Pathway 2**
(1)	$H_2 + l \rightleftharpoons H_2l$	(1)	$H_2 + l \rightleftharpoons H_2l$
		(2)	$H_2l + l \rightleftharpoons Hl + Hl$
(3)	$N_2 + l \rightleftharpoons N_2l$	(3)	$N_2 + l \rightleftharpoons N_2l$
		(4)	$N_2l + l \rightleftharpoons Nl + Nl$
(5)	$N_2l + H_2l \rightleftharpoons N_2H_2l + l$		
(6)	$N_2H_2l + l \rightleftharpoons NHl + NHl$		
		(7)	$Nl + Hl \rightleftharpoons NH_2l + l$
		(8)	$NHl + Hl \rightleftharpoons H_2l$
(9)	$NHl + H2l \rightleftharpoons NH_3l + l$		
		(10)	$NH_2l + Hl \rightleftharpoons NH_3l + l$
(11)	$NH_3l \rightleftharpoons NH_3 + l$	(11)	$NH_3l \rightleftharpoons NH_3 + l$
	Pathway 3		**Pathway 4**
(1)	$H_2 + l \rightleftharpoons H_2l$	(1)	$H_2 + l \rightleftharpoons H_2l$
(2)	$H_2l + l \rightleftharpoons Hl + Hl$		
(3)	$N_2 + l \rightleftharpoons N_2l$	(3)	$N_2 + l \rightleftharpoons N_2l$
(4)	$N_2l + l \rightleftharpoons Nl + Nl$	(4)	$N_2l + l \rightleftharpoons Nl + Nl$
(5)	$N_2l + H_2l \rightleftharpoons N_2H_2l + l$		
(7)	$Nl + Hl \rightleftharpoons NH_2l + l$	(7)	$Nl + Hl \rightleftharpoons NH_2l + l$
(8)	$NHl + Hl \rightleftharpoons H_2l$	(8)	$NHl + Hl \rightleftharpoons H_2l$
		(9)	$NHl + H2l \rightleftharpoons NH_3l + l$
(10)	$NH_2l + Hl \rightleftharpoons NH_3l + l$	(10)	$NH_2l + Hl \rightleftharpoons NH_3l + l$
(11)	$NH_3l \rightleftharpoons NH_3 + l$		
	Pathway 5		**Pathway 6**
(1)	$H_2 + l \rightleftharpoons H_2l$	(1)	$H_2 + l \rightleftharpoons H_2l$
		(2)	$H_2l + l \rightleftharpoons Hl + Hl$
(3)	$N_2 + l \rightleftharpoons N_2l$	(3)	$N_2 + l \rightleftharpoons N_2l$
(4)	$N_2l + l \rightleftharpoons Nl + Nl$		
(5)	$N_2l + H_2l \rightleftharpoons N_2H_2l + l$	(5)	$N_2l + H_2l \rightleftharpoons N_2H_2l + l$
(6)	$N_2H_2l + l \rightleftharpoons NHl + NHl$	(6)	$N_2H_2l + l \rightleftharpoons NHl + NHl$
(7)	$Nl + Hl \rightleftharpoons NH_2l + l$		
(8)	$NHl + Hl \rightleftharpoons H_2l$	(8)	$NHl + Hl \rightleftharpoons H_2l$
(10)	$NH_2l + Hl \rightleftharpoons NH_3l + l$	(10)	$NH_2l + Hl \rightleftharpoons NH_3l + l$
(11)	$NH_3l \rightleftharpoons NH_3 + l$	(11)	$NH_3l \rightleftharpoons NH_3 + l$

Table 6.27: Independent pathways resulting from the set of eleven elementary reactions.

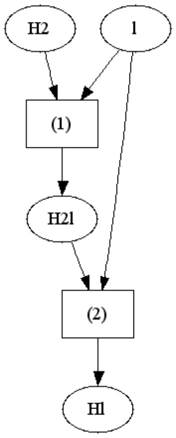


Figure 6.24: P-graph representation of the network comprising the forward steps of elementary reactions (1) and (2)[11].

chemical entities and the 11 elementary reactions. The master rank of the 23 nodes are given in table 6.28, this rank is only valid for this system. As a reaction pathway is not involving all the reactions and the chemical entities, the master rank is used to generate the invariant of the pathway.

6.5.3 SFILES string of the reaction pathways

By applying the algorithm for the generation of the SFILES string using the SFILES notation to the graphs representing the reaction pathways, six SFILES strings are generated (see table 6.29 page 164). The first SFILES string corresponding to the reaction path in figure 6.25 can be explained as follows:

- `(H2)(1)<1(H2l) ... (l)1`: H_2 reacts with l in reaction (1) to form H_2l.
- `(H2l)2(5)4<3(l) ... (N2l)3 ... (N2H2l)<4`: H_2l reacts with N_2l in reaction (5) to form N_2H_2l and l.
- `(l)1<6<5[(3)[(N2l)3]<(N2)]`: l reacts with N_2 in reaction (3) to form N_2l.
- `(l) ... (6)[(NHl) ... <(N2H2l)`: l reacts with N_2H_2l in reaction (6) to form NHl.
- `(H2l)2 ... (l)>1<6<5 ... (NHl)(9)5<2(NH3l)` : H_2l reacts with NHl in reaction (9) to form $NH3l$ and l.
- `(l)>1<6<5 ... (NH3l)(11)6(NH3)`: NH_3l reacts in reaction (11) to form l and NH_3.

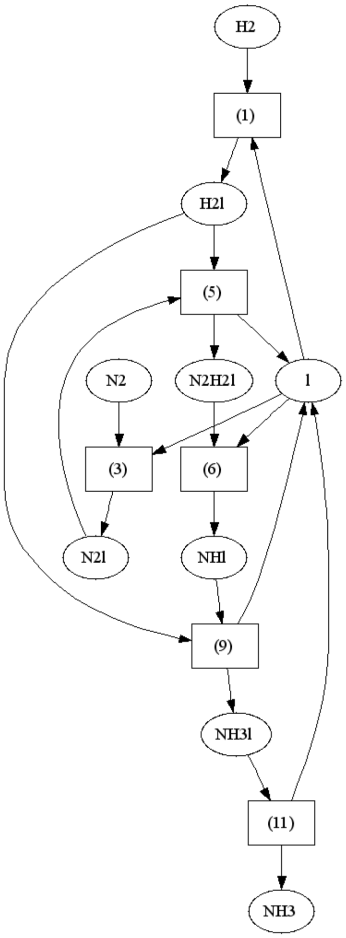


Figure 6.25: P-graph representation of the first reaction pathway.

Node	**Rank**
H_2	1
N_2	2
l	3
H_2l	4
Hl	5
N_2l	6
Nl	7
NHl	8
N_2H_2l	9
NH_2l	10
NH_3l	11
NH_3	12
(1)	13
(2)	14
(3)	15
(4)	16
(5)	17
(6)	18
(7)	19
(8)	20
(9)	21
(10)	22
(11)	23

Table 6.28: Master rank for the elementary reactions and the chemical entities in the ammonia synthesis.

6.5.4 Conclusion

The results show that a network of reactions can be represented in the form of a SFILES string. This has been illustrated with the reaction pathways of the catalytic synthesis of Ammonia from Hydrogen and Nitrogen. The SFILES strings are providing a rapid and efficient way to compare reaction pathways for equalities. As a SFILES string carries the complete description of a network of reactions it is possible from the SFILES string to represent the network of reactions. The results show the flexibility of the SFILES notation with the ability to use the SFILES string to easily compare or exchange graph data with no loss of information. Note that as already pointed out, exactly the same is also possible for comparison of process flowsheets and molecular structures.

6.6 Flash Separation of Alcohols

In this section, the case study involving the separation of a Methanol, Ethanol and Propanol mixture using a flash separation is presented. The case study highlights the reverse simulation approach for a single stage separation unit. The reverse simulation of the flash separation corresponds to step 6 of the work flow of the framework for CAFD presented in figure 5.1 (page 83).

Given the feed definition in table 6.30, the PT-flash separation target is to obtain in the liquid product a fraction of Methanol on a Propanol free basis of $x_{Methanol} = 0.40$ at a pressure of 1 atm.

Given the feed definition and the pressure, the driving force curve is obtained using the ICAS Utility toolbox. The Propanol fraction is fixed to a fraction corresponding to the feed conditions. The resulting driving force curve is shown on figure 6.26.

From the driving force curve and the target fraction of Methanol, the feasibility of the separation is immediately available with the corresponding driving force $D_{ij} = 0.1504$ (see figure 6.27).

From the driving force equation (4.21) (page 61) expressed as a function of z_i, x_i and R, the ratio R is calculated: $R = 0.504$. From the definition of L and V, and the feed of 900 $kmol/hr$, $V = 598.4$ $kmol/hr$ and $L = 301.6$ $kmol/hr$. As $y_i = (R{+}1)z_i - Rx_i$, y_i is now directly available, with $y_i = 0.5504$. From the driving force data it is possible to determine the temperature of the separation unit: $T = 345K$.

Comparison of the results obtained against the simulator results are given in table 6.31, the fractions are Propanol free molar fractions. The comparisons between the reverse simulation results and the rigorous simulation results for a feed with no Propanol, 100kmol/hr and 200kmol/hr of Propanol are presented in deviation diagrams in figures 6.28, 6.29, and 6.30, respectively. It can be seen that the differences between the simulation results and the graphical results (reverse simulation) are minimal with the exception of the ratio $R = L/V$. It is interesting to note that the Propanol free molar fractions are not very sensitive to the difference of R.

Pathway	Corresponding SFILES string
1	(H2)(1)<1(H21)2(5)4<3(1)>1<6<5[(3)[(N21)3]<(N2)](6)[(NH1)(9)5<2(NH31)(11)6(NH3)]<(N2H21)<4
2	(H2)(1)<1(H21)(2)<2(H1)(8)4<3(1)<7>6<5>2>1(3)[(N21)(4)<6(N1)(7)5(NH1)3]<(N2)]<(11)[(NH3)]<(NH31)<(10)7<(NH21)<4
3	(H2)(1)<1(H21)2(2)<3(H1)5>4(7)7<6(1)<11<10<9>8>3>1(3)[(N21)[(4)<8(N1)6](5)<2>9(N2H21)]<(N2)]<(11)[(NH3)]<(NH31)<(10)<4>10<(NH21)<(8)<5>11<(NH1)<7
4	(H2)(1)<1(H21)(9)<3>2(1)8<7<6<5<4>1(3)[(N21)(4)<8(N1)(7)<8>4(NH1)3(8)<9>5(NH21)(10)6(NH31)<2(11)7(NH3)]<(H1)9>8]<(N2)
5	(H2)(1)<1(H21)(5)<3>2(1)9>8<7<6<5<4>1(3)[(N21)3(4)<9(N1)(7)<9>4(NH1)[(8)<10>5(NH21)(10)6(NH31)(11)7(NH3)]<(H1)10>9]<(6)<8<(N2H21)<2]<(N2)
6	(H2)(1)<1(H21)2(2)<3(H1)4(8)6<5(1)<9>8<7>3>1(3)[(N21)[(4)](5)<2>7(N2H21)(6)<8(NH1)5]<(N2)]<(11)[(NH3)]<(NH31)<(10)<4>9<(NH21)<6

Table 6.29: SFILES strings of six Ammonia reaction pathways.

Component	Flow ($kmol/h$)
Methanol	450
Ethanol	450
Propanol	100

Table 6.30: Feed definition

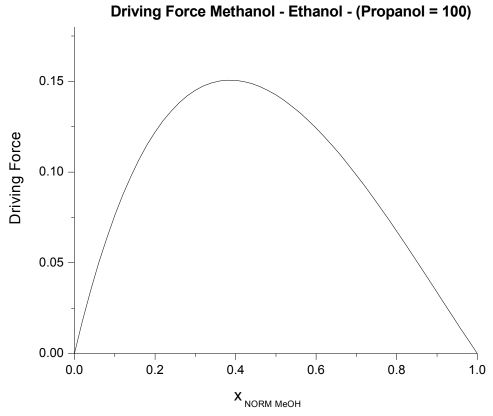


Figure 6.26: Driving force between Methanol and Ethanol for a fixed amount of Propanol (100kmol/hr)

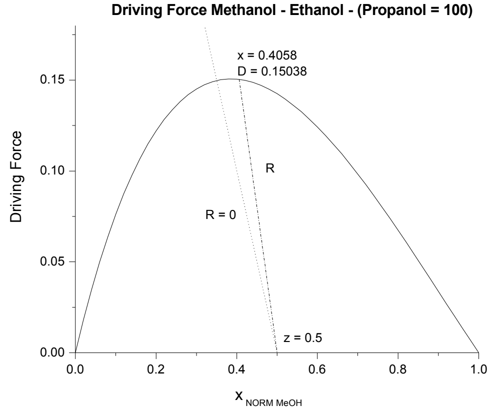


Figure 6.27: Calculations of the driving force for the desired product specifications

	Simulator	Graphical
x_{MeOH}	0.4058	0.40
x_{EtOH}	0.5942	0.60
y_{MeOH}	0.5572	0.5504
y_{EtOH}	0.4428	0.4496
$T(K)$	346.50	345
R	0.5964	0.504

Table 6.31: Comparison of the results between the driving force approach and the simulation

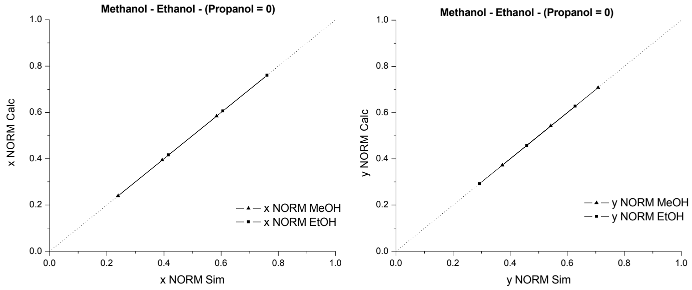


Figure 6.28: Comparison between the calculated reverse simulation results and the simulation results for the Methanol, Ethanol separation.

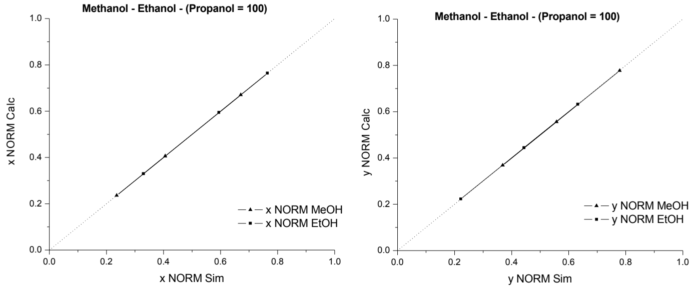


Figure 6.29: Comparison between the calculated reverse simulation results and the simulation results for the Methanol, Ethanol, Propanol separation (100 kmol/hr Propanol).

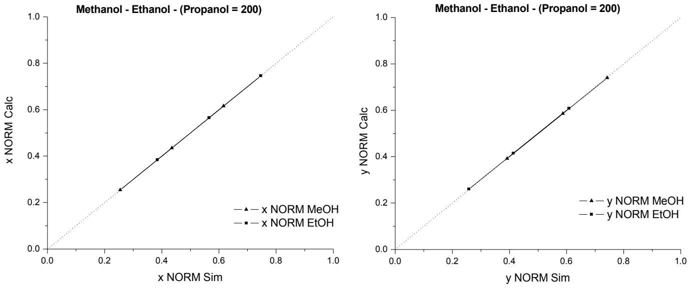


Figure 6.30: Comparison between the calculated reverse simulation results and the simulation results for the Methanol, Ethanol, Propanol separation (200 kmol/hr Propanol).

6.6.1 Conclusion

The results show that the reverse simulation approach for PT-flash separation provides the necessary data to fully describe the separation. The results also show that the it is possible to immediately know if the target property, such as, driving force, component free composition, is feasible or not. If the target property is attainable, the complete definition of the PT-flash is immediately available. However, it should be noted, that the ratio R is very sensible to the calculated driving force where the normalized molar fractions of the key components are not very sensitive to the variation of R.

7

Conclusions

7.1 Achievements

In this thesis the main achievements are the development of the concepts, methods and tools at the foundations of Computer Aided Flowsheet Design (CAFD). The process group concept – representing a unit operation or a set of unit operations, the process group connectivity rules – to combine process groups to generate feasible flowsheet structures – and the process group contribution based flowsheet property model – to estimate the performance of the generated flowsheet structures – have been developed and integrated into a framework for CAFD. This framework is in particular similar to CAMD and represents a successful transfer of algorithms/methods/tools from one area of application to another. That is, from the synthesis of chemicals to the synthesis of chemical process flowsheets.

The framework for Computer Aided Flowsheet Design (CAFD) is composed of eight main steps. *i) The definition of the process synthesis problem* – where the user provides information on the available raw materials and the desired products. *ii) The analysis of the process synthesis problem* – to refine and extract usable knowledge from the problem definition. This is performed by applying physical insights and knowledge based methods to identify and select a set of feasible process operation tasks for the defined synthesis problem. *iii) The selection of the process groups matching with the synthesis problem* – based on the selection of the appropriate process tasks in the analysis, the corresponding process groups are matched, based on property dependence, with the mixtures involved in the problem. *iv) The synthesis and test of the flowsheet structure alternatives* – the framework for CAFD provides a unique way, a reverse flowsheet property prediction method, to efficiently generate process alternatives by combining process groups according to connectivity rules and ensuring the feasibility (similar to chemical stability of a molecule) of generated flowsheet structures. *v) The ranking of the generated alternatives and the selection of the most promising alternatives* – with a process group contribution based flowsheet property model, the performance of the alternatives are predicted. Through defined property target values, it is not only possible to find the optimal process alternative but also to compare the generated alternatives. *vi) The design of the selected flowsheet structure alternatives*, is achieved by applying a reverse simulation approach to determine the design parameters of the unit operations from the specifications inherited from the underlining process groups. *vii) The post analysis of the designed alternatives* – includes issues related to heat integration and/or environmental impact. *viii) The final verifications* – the synthesised design and model of the selected process flowsheet is verified through plant-data and/or rigorous simulation.

Highlights of the framework for CAFD are outlined below:

- *A framework that is applicable to a large range of problems.* The framework for CAFD addresses a large range of problems through the development of process groups representing a large range of process operations: distillation column, solvent based separation, vapor-liquid single-stage separation, fixed conversion reactor, kinetically controlled reactor, pressure-swing distillation, molecular sieve based separation, membrane based separation, crystallization and adsorption.

- *A framework that does not need to employ rigorous models at each decision step.* The combination of the process groups to form flowsheet structure alternatives does not depend on the resolution of the heat and mass balance, as the connectivity rules are *a priori* defined from the built-in mass/energy balance of the process groups. It means that during the generation of the alternatives, no detailed simulation needs to be performed, and only a simple mass balance is needed at the end of the generation of a flowsheet structure before the reverse simulation step.

- *A framework that supports the retrofit of existing processes and the generation of new alternatives.* Retrofit design/analysis can be performed at two different levels. First the generation of new alternatives with some imposed (fixed) parts in the resulting flowsheet structures. Second, given an already existing equipment the reverse simulation approach can help to obtain a better design of the process/operation.

- *A framework that provides the ability to build and reuse knowledge over time.* Addition of new process groups, new feasibility rules and new selection rules to the framework by the user will extend the applicability of the method.

- *A framework that is easy to use and in which the synthesis/design problems are easy to setup.* The resolution of the synthesis and design problem is easy as the concepts employed follow an intuitive representation of the problem and its solution.

- *A framework that is modular.* This is, in particular, achieved through the clear distinction between the synthesis of the alternatives and the design of the alternatives in a process synthesis and design framework. Usually synthesis and design are performed simultaneously. The modularity of the framework provides the ability to integrate new design and post analysis methods.

7.2 Remaining challenges and future work

Even though the framework for CAFD shows a strong foundation for expansion/application, challenging problems remain to be considered.

Using the method based on thermodynamic insights for selection of process groups, the list of available process groups could be extended to support a larger range of process operations, such as, cryogenic separation, liquid-liquid extraction or desublimation.

As the framework for CAFD is modular, extensions could be made at various steps of the CAFD work flow. The analysis of the synthesis problem could be extended to provide more insight of the reaction paths. It should be possible to preselect the reaction path that will minimize the production of components having a negative environmental impact. In the same way in ProCAMD, reference solvents are provided to assist the user in the selection of the right molecule alternative, in step 5 of the CAFD framework, the selection of the alternatives could be supported by the availability of a database of reference flowsheet structure alternatives.

An important part of the processes in the fine chemical and pharmaceutical industry are batch processes. The framework could be extended to generate a set of the most promising batch sequences to be then further discriminated based on scheduling considerations.

Further testing of the framework for CAFD should be performed on more difficult synthesis and design problems both for completely new alternatives and retrofit of existing processes.

The difficulty to compare alternatives having a diversity of types of process groups is opening a challenging problem of the extension of the current flowsheet property model. The model should provide a flowsheet property that would be able to predict the performance of process groups of different types.

When combining the process groups to form flowsheet structures, the generated flowsheet alternatives are implicitly providing the superstructure of all the possible alternatives. The generation of all the alternatives for a 15 component separation using only distillation columns (sharp split) leads 2674440 feasible alternatives. The number of generated alternatives points to the second challenge. Two possible approaches could provide a way to limit the number of alternatives without sacrifying the optimal solution. First it could be possible to evaluate critical flowsheet properties during the generation of the alternatives to discard during the generation step the less promising alternatives. Second, the generation of the alternatives could be divided into simpler sub-problems with the assistance of the engineer. The choice of the sub-problems would be left to the experience of the engineer.

A

Pure Component Properties

M_w	Molecular Weight (g/mol)
ω	Acentric Factor
T_c	Critical Temperature (K)
P_c	Critical Pressure (bar)
Z_c	Critical Compressibility Factor
V_c	Critical Volume $(m^3/kmol)$
T_b	Normal Boiling Point (K)
d_m	Dipole Moment $\times 1 \cdot 10^{-30}$ $(C \cdot m)$
r_g	Radius of Gyration (nm)
T_m	Melting Point (K)
T_{tp}	Triple Point Temperature (K)
P_{tp}	Triple Point Pressure (Pa)
M_v	Molar Volume $(m^3/kmol)$
H_f	Ideal Gas Heat of Formation $(kJ/kmol)$
G_f	Ideal Gas Gibbs Energy of Formation $(kJ/kmol)$
S_{IG}	Ideal Gas Absolute Entropy $(kJ/(kmol \cdot K))$
H_{fus}	Heat of Fusion at T_m $(kJ/kmol)$
H_{comb}	Standard Net Heat of Combustion $(MJ/kmol)$
δ	Solubility parameter $(\sqrt{kJ/m^3})$
V_{vw}	Van der Waals Volume $(m^3/kmol)$
A_{vw}	Van der Waals Area $(m^2/kmol)$
P_{nvap}	Normal Vapour Pressure (Pa)

Table A.1: Properties calculated for each pure component in the synthesis problem

B

List of Common Solvents

Name	**CAS number**	**Type**
1,1,1-Trichloroethane	000071-55-6	Halide
1,2-Dichloroethane	000107-06-2	Halide
1-Butanol	000071-36-3	Alcohol
2-Butanol	000078-92-2	Alcohol
2-Propanol	000067-63-0	Alcohol
Acetone	000067-64-1	Ketone
Acetonitrile	000075-05-8	Nitrile
Benzene	000071-43-2	Aromatic-HC
Butyl Acetate	000123-86-4	Ester
Butyl Cellosolve	000111-76-2	Alcohol-Ether
Carbon Tetrachloride	000056-23-5	Chloride
Chlorobenzene	000108-90-7	Chloride
Chloroform	000067-66-3	Chloride
Cyclohexane	000110-82-7	Cyclic-HC
Cyclohexanol	000108-93-0	Alcohol
cyclohexanone	000108-94-1	Ketone
Dichloromethane	000075-09-2	Chloride
Diisopropyl Ether	000108-20-3	Ether
Dimethyl formamide	000068-12-2	Amide
Dimethylacetamide	000127-19-5	Amide
Dimethylsulphoxide	000067-68-5	S-oxide
Ethanol	000064-17-5	Alcohol
Diethyl Ether	000060-29-7	Ether
Ethyl Acetate	000141-78-6	Ester
Ethyl Benzene	000100-41-4	Aromatic
1,2-Dimethoxyethane	000110-71-4	Ether
Furfural	000098-01-1	Ether-Aldehyde
Heptane	000142-82-5	Alkane
Hexane	000110-54-3	Alkane
1-pentanol	000071-41-0	Alcohol

Table B.1: Common solvents

Name	**CAS number**	**Type**
1-hexanol	000111-27-3	Alcohol
1-heptanol	000111-70-6	Alcohol
1-octanol	000111-87-5	Alcohol
Methanol	000067-56-1	Alcohol
Methyl Acetate	000079-20-9	Ester
Methyl Cellosolve	000109-86-4	Alcohol-Ether
Methyl ethyl ketone	000078-93-3	Ketone
Methyl isobutyl ketone	000108-10-1	Ketone
Dioxane	000505-22-6	Ether
Pentane	000109-66-0	Alkane-HC
1-Propanol	000071-23-8	Alcohol
Pyridine	000110-86-1	Amine
Tetrahydrofuran	000109-99-9	Ester
Toluene	000108-88-3	Aromatic-HC
Water	007732-18-5	Aqueous
Xylenes	000108-38-3	Aromatic-HC
Propionic acid	000079-09-4	Acid
2-Ethylhexanol	000104-76-7	Alcohol
Ethylene Glycol	000107-21-1	Alcohol
Isopropyl acetate	000108-21-4	Ester
Mesitylene	000108-67-8	Aromatic-HC
Methyl cyclohexane	000108-87-2	Cyclic-HC
n-Propyl acetate	000109-60-4	Ester
Diethylene glycol butyl ether	000112-34-5	Ether-Alcohol
Dimethyl ether	000115-10-6	Ether
Triethylamine	000121-44-8	Amine
N,N-Dimethylaniline	000121-69-7	Amine
t-Butyl acetate	000540-88-5	Ester
1-Methyl-2-Pyrrolidone	000872-50-4	Amide
Methyl tertbutyl ether	001634-04-4	Ether

Table B.2: Common solvents (Continued.)

C

Pre-calculated Values for the Driving Approach

The pre-calculated values in table C.1 are by Bek-Pedersen[2].

Table C.1: Pre-calculated values of reflux ratio, minimum reflux ratio, number of ideal stages, product purities and driving force for ideal distillation.

$F_{Di\|max}$	$X_{LK,Dist}$	$X_{LK,Bot}$	RR_{min}	$RR_{min} \cdot C$	$N_{ideal} \cdot C$
0.045	0.995	0.005	9.89	14.83	96
	0.98	0.02	9.56	14.36	71
	0.95	0.05	8.90	13.35	54
	0.90	0.10	8.22	12.33	41
0.065	0.995	0.005	7.33	11.00	67
	0.98	0.02	7.10	10.65	50
	0.95	0.05	6.64	9.96	38
	0.90	0.10	6.64	8.58	29
0.101	0.995	0.005	4.50	6.74	44
	0.98	0.02	4.35	6.52	33
	0.95	0.05	4.05	6.08	25
	0.90	0.10	3.56	5.33	19
0.146	0.995	0.005	2.92	4.41	31
	0.98	0.02	2.84	4.26	23
	0.95	0.05	2.63	3.95	18
	0.90	0.10	2.29	3.44	14
0.172	0.995	0.005	2.35	3.53	27
	0.98	0.02	2.26	3.40	20
	0.95	0.05	2.09	3.13	15
	0.90	0.10	1.80	2.70	12
0.195	0.995	0.005	2.06	3.09	24
	0.98	0.02	1.89	2.97	18
	0.95	0.05	1.82	2.74	14
	0.90	0.10	1.57	2.35	11

Continued on next page

$F_{Di\|max}$	$X_{LK,Dist}$	$X_{LK,Bot}$	RR_{min}	$RR_{min} \cdot C$	$N_{ideal} \cdot C$
0.225	0.995	0.005	1.73	2.60	21
	0.98	0.02	1.67	2.50	16
	0.95	0.05	1.53	2.30	12
	0.90	0.10	1.37	1.97	9
0.268	0.995	0.005	1.37	2.06	18
	0.98	0.02	1.31	1.97	13
	0.95	0.05	1.20	1.80	10
	0.90	0.10	1.02	1.52	8
0.382	0.995	0.005	0.82	1.23	13
	0.98	0.02	0.78	1.17	10
	0.95	0.05	0.70	1.05	8
	0.90	0.10	0.57	0.86	6
0.478	0.995	0.005	0.54	0.81	10
	0.98	0.02	0.51	0.76	8
	0.95	0.05	0.44	0.67	6
	0.90	0.10	0.34	0.51	5

D

Available Process Groups

In this section the process groups available with the CAFD method are presented. Twelve types of process groups are available with the CAFD method, addressing a large range of problems. They are the distillation column, the solvent based azeotropic separation, the flash, the kinetic and conversion rate based reactors, the pressure swing distillation, the polar molecular sieve, the molecular sieve, the liquid membrane, the gas membrane, the crystallization and the absorption process groups.

D.1 Simple Distillation Column Process Group

The simple distillation column process group is representing conventional column with one feed and two products for sharp and non sharp separations. The synopsis of this process group is available in table D.1.

Name	Simple distillation column
Specific properties	Driving force Relative volatility
Unit operations	distillation column
SFILES notation examples	`(A/BC)`, `(p25A/BC)`, `(BC/CDE)`
Reverse simulation	Available

Table D.1: Simple distillation synopsis

D.1.1 Procedure to generate the SFILES notation

To generate the SFILES notation of a simple distillation process group, the following procedure is applied:

1. Start the SFILES with an opening parenthesis `(`.
2. Retrieve the list of components in the top outlet of the simple column process group.
3. Order the components by decreasing relative volatility.

4. For each component, append to the SFILES the letter corresponding to the component.
5. Append to the SFILES a slash character `/` representing the split.
6. Retrieve the list of components in the bottom outlet of the simple column process group.
7. Order the components by decreasing relative volatility.
8. For each component, append to the SFILES the letter corresponding to the component.
9. End the SFILES with a closing parenthesis `)`.

Example of SFILES notation – Considering the simple distillation process group presented on figure D.1, considering that the components C_1 to C_4 have for corresponding letters B to E and are ordered by decreasing relative volatility. By following the procedure to generate the SFILES notation, the corresponding SFILES notation is `(BC/CDE)`.

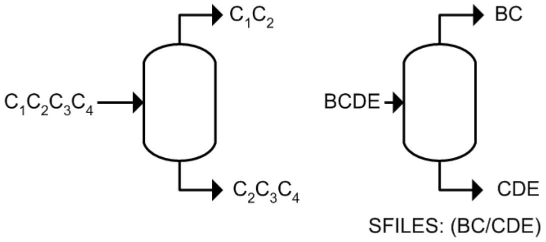


Figure D.1: Simple non sharp distillation

D.1.2 Property dependence

The simple distillation process group can be used with a mixture of NC components if the following property dependence is satisfied:

1. The simple distillation process group is a NC component distillation process group.
2. The maximum driving force between light and heavy key components must be within the driving force range of the simple distillation process group.
3. The relative volatility of each of the components must be within the range of relative volatilities of the process group.

4. No azeotropes must be present between the key components and between pair of components with one having higher relative volatility than the light key and the other a lower relative volatility than the heavy key.

If all the property dependence are satisfied the distillation process group can be initialized with the given mixture.

D.1.3 Connectivity rules and specifications

The simple distillation process group is a one inlet two outlet process group.

The inlet connection must contain the exact same components that the ones set during the initialization. Pressure and temperature of the inlet are not critical.

The two outlet connections are ensuring to have at least 99.5% purity in pure streams and above 99.5% of recovery of key components if mixtures. The outlet pressure is for both outlets set to the process group operationnal pressure and the temperature is set to the bubble point of the mixtures. The assumption is made that there is no pressure drop in the column.

D.1.4 Regression of the energy index model parameters

The regression of the parameters for the energy index model is based on rigorous simulation data. Each column in the rigorous simulation has been simulated using the standard setup presented in table D.2 to ensure the coherence of the data.

Number of stages	50
Minimum recovery of key components	99.5%
Minimum purity of a *pure* key	99.5%
Molar flowrate in the inlet	$1kmol/hr$
Conditions of the inlet	Standard

Table D.2: Standard setup for the rigorous distillation simulation

The feed plate location is initialized using the driving force method and optimized with the PRO/II optimizer to minimize the total duties of the column as shown on the objective function (D.1).

$$Obj = E_{reboiler} - E_{condenser} \tag{D.1}$$

The results of the simulations are presented in table D.3. The regressed parameters of the distillation process groups depending on the number of components, the maximum driving force between the key components, the number of components in the overhead products and the relative volatility range of the components are given in table D.4.

Number of rigorous simulations	105
Number of 2 components separations	15
Number of 3 components separations	24
Number of 4 components separations	24
Number of 5 components separations	22
Number of 6 components separations	9
Number of 7 components separations	5
Number of 8 components separations	6

Table D.3: Overview of the simulations for the model parameters regression

Table D.4: Contributions of the simple distillation process groups.

Group id	NC$_{total}$	NC$_{top}$	$F_{Di\mid max}$	a_k	$\min \alpha_{i_{ref}}$	$\max \alpha_{i_{ref}}$
dist-1	2	1	0.1570	0.01173	13.625	67.783
dist-2	2	1	0.0664	0.04027	9.2514	13.625
dist-3	2	1	0.1092	0.01845	10.313	9.2514
dist-4	2	1	0.0632	0.09744	7.5854	10.313
dist-5	2	1	0.0168	6.33727	7.0563	7.5854
dist-6	3	1	0.1570	0.01016	9.2514	67.783
dist-7	3	2	0.0664	0.06629	9.2514	67.783
dist-8	3	1	0.0664	0.08074	10.313	13.625
dist-9	3	2	0.1092	0.01310	10.313	13.625
dist-10	3	1	0.1092	0.01476	7.5854	9.2514
dist-11	3	2	0.0632	0.52620	7.5854	9.2514
dist-12	3	1	0.0632	23.71149	7.0563	10.313
dist-13	3	2	0.0168	5.98355	7.0563	10.313
dist-14	3	1	0.0636	0.86774	4.0472	7.0563
dist-15	4	1	0.1570	0.00808	10.313	67.783
dist-16	4	2	0.0664	0.05101	10.313	67.783
dist-17	4	3	0.1092	0.00911	10.313	67.783
dist-18	4	1	0.0664	0.06426	7.5854	13.625
dist-19	4	2	0.1092	0.01113	7.5854	13.625
dist-20	4	3	0.0632	0.30529	7.5854	13.625
dist-21	4	1	0.1092	0.01278	7.0563	9.2514
dist-22	4	2	0.0632	1.03656	7.0563	9.2514
dist-23	4	3	0.0168	5.41363	7.0563	9.2514
dist-24	4	1	0.0168	5.97814	4.0472	7.5854
dist-25	4	2	0.0636	0.07139	4.0472	7.5854
dist-26	5	1	0.1570	0.01406	7.5854	67.783
dist-27	5	2	0.0664	0.04336	7.5854	67.783
dist-28	5	3	0.1092	0.00809	7.5854	67.783
dist-29	5	4	0.0632	0.19277	7.5854	67.783

Continued on next page

Group id	NC$_{total}$	NC$_{top}$	$F_{Di\|max}$	a_k	$\min \alpha_{i_{ref}}$	$\max \alpha_{i_{ref}}$
dist-30	5	1	0.0664	0.02434	7.0563	13.625
dist-31	5	2	0.1092	0.00992	7.0563	13.625
dist-32	5	3	0.0632	0.52071	7.0563	13.625
dist-33	5	4	0.0168	5.11638	7.0563	13.625
dist-34	5	1	0.0632	0.12088	4.0472	10.313
dist-35	5	2	0.0168	5.68497	4.0472	10.313
dist-36	5	3	0.0636	0.05739	4.0472	10.313
dist-37	6	1	0.1570	0.01262	7.0563	67.783
dist-38	6	2	0.0664	0.01754	7.0563	67.783
dist-39	6	3	0.1092	0.00830	7.0563	67.783
dist-40	6	4	0.0632	0.02942	7.0563	67.783
dist-41	6	5	0.0168	4.53667	7.0563	67.783
dist-42	6	1	0.1092	0.00896	4.0472	9.2514
dist-43	6	2	0.0632	0.07976	4.0472	9.2514
dist-44	6	3	0.0168	5.16771	4.0472	9.2514
dist-45	6	4	0.0636	0.04244	4.0472	9.2514
dist-46	7	1	0.0664	0.01506	4.0472	13.625
dist-47	7	2	0.1092	0.00823	4.0472	13.625
dist-48	7	3	0.0632	0.06012	4.0472	13.625
dist-49	7	4	0.0168	4.89430	4.0472	13.625
dist-50	7	5	0.0636	0.03375	4.0472	13.625
dist-51	8	1	0.1570	0.01022	4.0472	67.783
dist-52	8	2	0.0664	0.01144	4.0472	67.783
dist-53	8	3	0.1092	0.00714	4.0472	67.783
dist-54	8	4	0.0632	0.01527	4.0472	67.783
dist-55	8	5	0.0168	4.33034	4.0472	67.783
dist-56	8	6	0.0636	0.02561	4.0472	67.783
dist-57	2	1	0.2199	0.01474	1.5479	4.0472
dist-58	3	2	0.2199	0.00965	1.5479	4.993e+015
dist-59	3	1	0.2199	0.00696	0.028517	4.0472
dist-60	4	3	0.2199	0.00712	1.5479	1e+100
dist-61	4	2	0.2199	0.00516	0.028517	4.993e+015
dist-62	5	3	0.2199	0.00409	0.028517	1e+100
dist-63	5	4	0.7065	0.02241	0.028517	1e+100

D.1.5 Reverse simulation

The reverse simulation of the simple distillation column process group is using the driving force approach presented in section 4.3 page 56.

D.2 Solvent Based Azeotropic Separation Process Group

The solvent based azeotropic separation process group is representing the separation of an azeotropic mixture with a solvent. Two typical configurations to separate two components E and O forming an azeotrope using a solvent S, depending on the ability to mix the solvent and the azeotropic mixture in the first column or the need to *wash* the column with the solvent are shown on figure D.2. The synopsis of this process group is available in table D.5.

Name	Solvent based azeotropic separation
Specific properties	Solvent free driving force Relative volatility Azeotrope
Unit operations	distillation column, mixer
Representation example	`(cycA/B)`
Reverse simulation	Partially available

Table D.5: Solvent based separation azeotropic synopsis

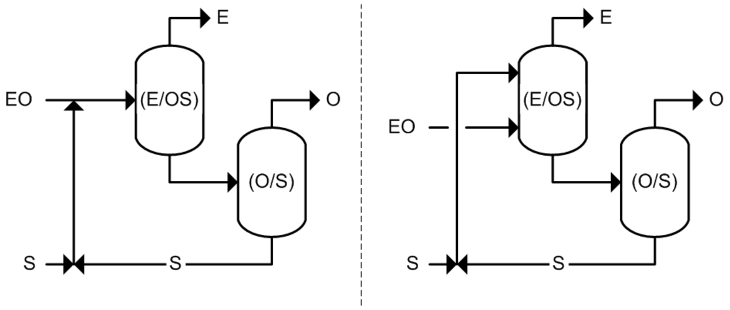


Figure D.2: Solvent based azeotropic distillation

D.2.1 Procedure to generate the SFILES notation

To generate the SFILES notation of a solvent based azeotropic separation process group, the following procedure is applied:

1. Start the SFILES with an opening parenthesis `(`.
2. Append the identification string `cyc` of the process group to the SFILES.
3. Retrieve the two components being separated and order them by increasing boiling point.

4. Append to the SFILES the letter corresponding to the low boiling component.

5. Append to the SFILES a slash character `/` representing the split.

6. Append to the SFILES the letter corresponding to the high boiling component.

7. End the SFILES with a closing parenthesis `)`.

D.2.2 Property dependence

The solvent based azeotropic separation process group can be used with a binary mixture if the following property dependence is satisfied:

1. The mixture is a binary mixture.

2. The mixture is an azeotropic mixture.

3. The solvent free maximum driving force between the two components is within the driving force range of the solvent based azeotropic separation. This implies the availability of an adequate solvent for the separation of the binary pair.

4. The relative volatility of each of the components must be within the range of relative volatilities of the process group.

D.2.3 Initialization procedure

When initializing a solvent based separation process group with an azeotropic mixture, two cases are possible. Either the binary mixture and the corresponding solvent are known and matching the property dependence, or the binary mixture is matching the property dependence but no corresponding solvent is known.

For the first case, the solvent based separation process group is initialized with the binary mixture and the solvent.

For the second case, as no solvent is known, the following procedure is applied to find a matching solvent. First a database search in the most common solvent database (see tables B.1 and B.2) is performed to look for a potential solvent. If no solvent is found, a CAMD problem formulation[23] is setup with the targets being to match the solvent free driving force. If more than one potential solvent are found, the WAR algorithm[50] can be applied to keep the solvent with the least environmental impact.

D.2.4 Connectivity rules and specifications

The solvent based azeotropic separation process group is a one inlet two outlets process group. The uptake of solvent is not an inlet of the process group.

The composition of the two outlets is the one being the highest purity between 99.5% purity or above the azeotrope.

D.2.5 Regression of the energy index model parameters

The regression of the parameters for the solvent energy index model is based on rigorous simulation data. As in the case of the simple distillation, a standard setup for all the columns has been defined. The setup and the recovery in the columns is using the same standard setup of the simple distillations as shown in table D.2 page 181, with the added constraints that the purity of the pure key components and the recovery of the key components must be above the azeotrope.

The simulation is then performed for a $1kmol/hr$ feed of the azeotropic mixture at standard conditions. It means that the total feed in the first column will be $2kmol/hr$ in the case of a solvent fraction of 0.5. The feed plate location is initialized using the driving force method for the second column and if the mixture and the solvent feed can be merged, for the first column, else it is initialized to get a feasible simulation. The feed plate locations are then optimized with the PRO/II optimizer to minimize the duties as shown on the objective function (D.2). A controller ensures the correct solvent fraction in the first column.

$$Obj = (E_{reboiler} - E_{condenser})_1 + (E_{reboiler} - E_{condenser})_2 \tag{D.2}$$

D.2.6 Reverse simulation

The reverse simulation of the solvent based azeotropic separation process group is performed using the driving force approach presented in section 4.3 page 56 on a solvent free basis.

The target solvent free driving force in the first column is fixing the solvent fraction and thus fixing the solvent recycle from the second column. The specifications of the columns is also fixing the needs in uptake of solvent.

The reverse simulation is not fully available when the solvent needs to be separated from the azeotropic mixture in the first column. In that case, only the second column can be fully designed.

D.3 Flash Separation Process Group

The flash process group is representing a liquid vapor flash separation. The flash is a non sharp single stage separation. The synopsis of this process group is available in table D.6.

Name	Flash
Specific properties	Driving force Relative volatility Boiling point
Unit operations	flash
Representation example	`(fABC/BCD)`
Reverse simulation	Available

Table D.6: Flash synopsis

D.3.1 Procedure to generate the SFILES notation

To generate the SFILES notation of a flash process group, the following procedure is applied:

1. Start the SFILES with an opening parenthesis `(`.
2. Append the identification string `f` of the process group to the SFILES.
3. Retrieve the list of components in the top outlet of the flash process group.
4. Order the components by decreasing relative volatility.
5. For each component, append to the SFILES the letter corresponding to the component.
6. Append to the SFILES a slash character `/` representing the split.
7. Retrieve the list of components in the bottom outlet of the flash process group.
8. Order the components by decreasing relative volatility.
9. For each component, append to the SFILES the letter corresponding to the component.
10. End the SFILES with a closing parenthesis `)`.

D.3.2 Property dependence

The flash process group can be used with a mixture of NC components if the following property dependence is satisfied:

1. No azeotropes must be present between the key components and between pair of components with one having higher relative volatility than the light key and the other a lower relative volatility than the heavy key.

2. The binary ratio for the boiling point between the key components must be greater than 1.23.

If the NC component mixture is satisfying the property dependence, the flash process group can be initialized with the mixture.

D.3.3 Connectivity rules and specifications

The flash process group is a one inlet two outlets process group.

The top outlet of the flash process group is a vapor and the bottom outlet is a liquid. The pressure and temperature are within the range of the liquid vapor equilibria between the two key components.

The components having a relative volatility higher than the heavy key and lower than the light key, together with the key components are available in both the top outlet and the bottom outlet.

The components having a relative volatility lower than the heavy key and higher than the light key are available respectively only in the bottom outlet and the top outlet.

D.3.4 Reverse simulation

The reverse simulation of the flash process group is based on the generalized driving force concept presented in section 4.3.2 page 60.

D.4 Kinetic Model Based Reactor Process Group

A kinetic reactor process group represents a reactor or a network of reactors. The synopsis of this process group is presented in table D.7.

Name	Kinetic Rector
Property dependence	Kinetic reactions
Unit operations	CSTR, PFR, mixer, divider
Representation example	`(rABC/nE/pABCD)`
Reverse simulation	Available

Table D.7: Kinetic based reactor synopsis

D.4.1 Procedure to generate the SFILES notation

To generate the SFILES notation of a kinetic reactor process group, the following procedure is applied:

1. Start the SFILES with an opening parenthesis `(`.
2. Append the identification string `r` to the SFILES.

3. Retrieve the list of components being reactants.
4. Order the reactants by boiling points.
5. For each reactant component, append to the SFILES the letter corresponding to the component.
6. Retrieve the list of components being inerts.
7. If no inert components, go to step 12.
8. Append to the SFILES a slash character `/`.
9. Append the identification string `n` to the SFILES.
10. Order the inerts by boiling points.
11. For each inert component, append to the SFILES the letter corresponding to the component.
12. Append to the SFILES a slash character `/`.
13. Append the identification string `p` to the SFILES.
14. Retrieve the list of reactant and product components.
15. Order the reactant and product by boiling points.
16. For each reactant and product component, append to the SFILES the letter corresponding to the component.
17. End the SFILES with a closing parenthesis `)`.

D.4.2 Property dependence

The kinetic reactor process group can be used with a mixture of NC components if the following property dependence is satisfied:

1. The components in the mixture have the same reacting properties as defined in the the process group.
2. All the reactants needed for the process group are available in the mixture.

If the mixture is satisfying the property dependence the process group can be initialized with the mixture.

Example of a satisfied reacting property dependence – Considering the kinetic reactor process group for the reaction given in equation (D.3). The SFILES notation would be `(rAB/pABC)`.

$$A + B \xrightarrow{k_1} C \tag{D.3}$$

The mixture composed of D, E and F can use this kinetic reactor process group if the reacting properties of D, E and F are the same as A, B and C. This means that the equation (D.4) is satisfied with k_1 and k_2 being of the same form.

$$D + E \xrightarrow{k_2} F \tag{D.4}$$

D.4.3 Connectivity rules and specifications

The kinetic based reactor process group is a one inlet one outlet process group. The inlet and the outlet being connected, kinetic based reactor process group supports also the addition of other inlets coming from recycled streams.

The inlet rules are:

- The inlet of the kinetic reactor process group must contain all the reactants.
- The inlet of the kinetic reactor process group may contain the product.
- The inlet of the kinetic reactor process group may contain the inerts.

The outlet rules are:

- The outlet of the kinetic reactor process group must contain the reactants and the products.
- The outlet of the kinetic reactor process group may contain the inerts if and only if available in the inlet.

The inlets coming from recycled streams must follow the following rules:

- The recycled inlet must contain at least one of the reactants.
- The recycled inlet must no contain a product.
- The components in recycled inlet must be composed of reactants or inerts.

The pressure and the temperature of the outlet is set to the operational conditions of the reactor. The outlet contains a mixture of all the reactants, inerts and products.

D.4.4 Reverse simulation

The reverse simulation of the kinetic based reactor process group is based on the attainable region concept presented in section 4.3.4 page 64.

D.5 Fixed Conversion Reactor Process Group

The fixed conversion reactor process group represents a reactor or a network of reactors. The difference with the kinetic based reactor is that the kinetics of the reactions involved are not known but global conversion rates between one component to another are known. This process group is always superseded by the kinetic reactor if kinetics data are available. The synopsis of this process group is presented in table D.8.

The goal of the fixed conversion reactor is to provide an easy way to include the reaction section of a process in the synthesis algorithm, thus providing the ability to include the recycles in the synthesis.

Name	fixed conversion reactor
Property dependence	Conversion rate
Unit operations	CSTR, PFR, mixer, divider
Representation example	`(rABC/nE/pABCD)`
Reverse simulation	Not available

Table D.8: Conversion reactor synopsis

The SFILES notation of the fixed conversion reactor process group is the same as the kinetic based reactor process group (see section D.4.1 page 188).

D.5.1 Property dependence

The fixed conversion reactor process group can be used with a mixture of NC components if the following property dependence is satisfied:

1. All the reactants needed for the process group are available in the mixture.
2. The conversion rates are known.

If the mixture is satisfying the property dependence the process group can be initialized with the mixture.

D.5.2 Connectivity rules and specifications

The fixed conversion reactor process group is a one inlet one outlet process group. The connectivity rules and specifications are the same as the kinetic based reactor process group (see section D.4.3)

D.6 Pressure Swing Distillation Process Group

The pressure swing distillation process group is representing the separation of an azeotropic mixture with two columns operating at different pressures. The synopsis of this process group is presented in table D.9.

Name	Pressure swing
Property dependence	Azeotrope pressure dependent Driving force Relative volatility
Unit operations	mixer, compressor, valve distillation column
Representation example	`(swA/B)`
Reverse simulation	Available

Table D.9: Pressure swing distillation synopsis

Depending on the chemical system, the pressure can effect more or less the vapor-liquid equilibrium. In the case of system forming an azeotrope the pressure can affect the composition of the azeotrope and in some case it can even suppress it.

As shown on figure D.3, in the system Ethanol-Benzene, the azeotrope is moved from a fraction of 0.45 of Ethanol to 0.35 of Ethanol with a pressure drop from 1 atm to 0.3 atm.

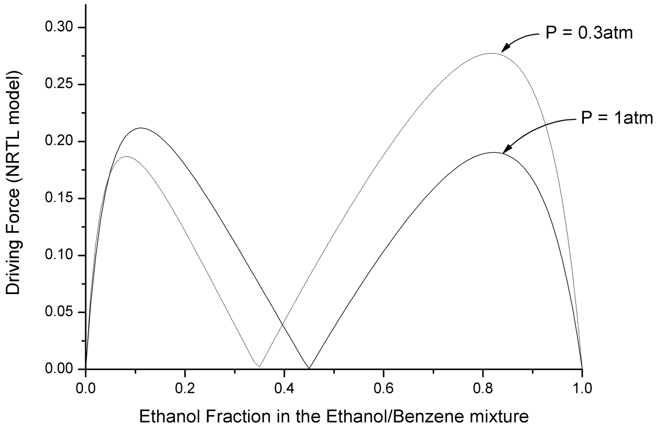


Figure D.3: Ethanol Benzene driving force function of the pressure

By operating one column at 0.3 atm it is possible to obtain pure Ethanol and mixture with a fraction of 0.35 of Ethanol. Taking this azeotropic mixture as the inlet of a column operating at 1 atm, the azeotrope limit is crossed as the at 1 atm the azeotrope fraction is of 0.45 of Ethanol. Thus, it is possible to obtain a pure stream of Benzene and mixture with a fraction of 0.45 of Ethanol. This mixture is recycled back to the inlet of the first column.

D.6.1 Procedure to generate the SFILES notation

To generate the SFILES notation of a solvent based azeotropic separation process group, the following procedure is applied:

1. Start the SFILES with an opening parenthesis `(`.
2. Append the identification string `sw` of the process group to the SFILES.
3. Retrieve the two components being separated and order them by increasing boiling point.
4. Append to the SFILES the letter corresponding to the low boiling component.
5. Append to the SFILES a slash character `/` representing the split.
6. Append to the SFILES the letter corresponding to the high boiling component.
7. End the SFILES with a closing parenthesis `)`.

D.6.2 Property dependence

The pressure swing distillation process group can be used with a binary mixture if the following property dependence is satisfied:

1. The mixture is a binary azeotropic mixture.
2. The azeotrope is pressure dependent.
3. The pressure variation to get a usable variation of the azeotrope composition and the maximum pressure are within the process group range.
4. The relative volatility of the components must be within the range of relative volatilities of the process group.

D.6.3 Connectivity rules and specifications

The pressure swing distillation process group is a one inlet two outlet process group. The two outlets are two near pure stream of the components and above azeotropic conditions. One outlet is at the low pressure condition and the other is at the high pressure condition. Both are bubble point temperature.

D.6.4 Reverse simulation

The reverse simulation of the pressure swing distillation process group is performed using the driving force approach presented in section 4.3 page 56 and configuration rules presented below.

The sequence high pressure, low pressure or low pressure high pressure of the distillation columns depend on the conditions of the inlet mixture. The following rules, based on the principles of pressure conservation and the driving force are applied to make the selection.

- If the inlet mixture pressure is already near the high pressure conditions, the first column is set to be at high pressure.
- If the inlet mixture is already near the low pressure conditions, the first column is set to be at low pressure.
- If the pressure conditions of the inlet mixture are not significant, the first column is set to maximize the available driving force.

Example of choice of the first column pressure – Considering the Ethanol - Benzene mixture presented in figure D.3. If the inlet mixture is at near 0.3 atm, the first column will be at 0.3 atm, if near 1 atm, the first column will be at 1 atm.

If the pressure is between 0.3 and 1 atm and not significantly near 0.3 or 1 atm, three cases are possible depending of the composition of the mixture.

$C_{Ethanol} < 0.35$ – the first column will be at 1 atm as the available driving force for this composition is higher at this pressure.

$0.35 < C_{Ethanol} < 0.45$ – the first column will be at 0.3 atm as the available driving force for this composition is higher at this pressure.

$0.35 < C_{Ethanol}$ – the first column will be at 0.3 atm as the available driving force for this composition is higher at this pressure.

D.7 Polar Molecular Sieve Based Separation Process Group

The polar molecular sieve process group is representing the adsorption of components on a molecular sieve due to the polarisation of the components. The synopsis of this process group is presented in table D.10.

D.7.1 Procedure to generate the SFILES notation

To generate the SFILES notation of a polar molecular sieve process group, the following procedure is applied:

Name	Polar molecular sieve
Specific properties	Dipole moment
Unit operations	Molecular sieve Regeneration process
Representation example	`(pmsABC/D)`
Reverse simulation	Not available

Table D.10: Polar molecular sieve synopsis

1. Start the SFILES with an opening parenthesis `(`.
2. Append the identification string `pms` of the process group to the SFILES.
3. Retrieve the list of components not being adsorbed.
4. Order the components by decreasing dipole moment.
5. For each component, append to the SFILES the letter corresponding to the component.
6. Append to the SFILES a slash character `/` representing the split.
7. Retrieve the list of components being adsorbed.
8. Order the components by decreasing dipole moment.
9. For each component, append to the SFILES the letter corresponding to the component.
10. End the SFILES with a closing parenthesis `)`.

D.7.2 Property dependence

The polar molecular sieve process group can be used with a mixture of NC components if the following property dependence is satisfied:

1. The binary ratio for the dipole moment between the key components must be greater than 1.08.

D.7.3 Connectivity rules and specifications

The polar molecular sieve process group is a one inlet two outlets process group.

The adsorbed components are all the components having higher dipole moment than the key component having the higher dipole moment and this key component.

The adsorbed components are considered as fully removed from the inlet stream.

D.8 Molecular Sieve Based Separation Process Group

Name	Molecular sieve
Specific properties	Kinetic diameter Van der Waals volume
Unit operations	Molecular sieve regeneration process
Representation example	`(msABC/D)`
Reverse simulation	Not available

Table D.11: Molecular sieve synopsis

D.8.1 Procedure to generate the SFILES notation

To generate the SFILES notation of a polar molecular sieve process group, the following procedure is applied:

1. Start the SFILES with an opening parenthesis `(`.
2. Append the identification string `ms` of the process group to the SFILES.
3. Retrieve the list of components not being adsorbed.
4. Order the components by decreasing van der Waals volume.
5. For each component, append to the SFILES the letter corresponding to the component.
6. Append to the SFILES a slash character `/` representing the split.
7. Retrieve the list of components being adsorbed.
8. Order the components by decreasing van der Waals volume.
9. For each component, append to the SFILES the letter corresponding to the component.
10. End the SFILES with a closing parenthesis `)`.

D.8.2 Property dependence

The molecular sieve process group can be used with a mixture of NC components if the following property dependence is satisfied:

1. The binary ratio for the kinetic diameter between the key components must be greater than 1.05.

2. The binary ratio for the van der Waals volume between the key components must be greater than 1.07.

D.8.3 Connectivity rules and specifications

The molecular sieve process group is a one inlet two outlets process group.

The adsorbed components are all the components having lower kinetic diameter than the key component having the lower kinetic diameter and this key component, or, all the components having lower van der Waals volume than the key component having the lower van der Waals volume and this key component.

The adsorbed components are considered as fully removed from the inlet stream.

D.9 Liquid Membrane Based Separation Process Group

The liquid membrane process group is representing a membrane separation in liquid conditions. The synopsis of this process group is available in table D.12.

Name	Liquid membrane
Specific properties	Radius of gyration Molar volume Solubility parameter
Unit operations	Membrane
Representation example	`(lmemABC/D)`
Reverse simulation	Not available

Table D.12: Liquid membrane synopsis

D.9.1 Procedure to generate the SFILES notation

To generate the SFILES notation of a liquid membrane process group, the following procedure is applied:

1. Start the SFILES with an opening parenthesis `(`.
2. Append the identification string `lmem` of the process group to the SFILES.
3. Retrieve the list of components in the retentate.
4. Order the components by decreasing molar volume.
5. For each component, append to the SFILES the letter corresponding to the component.

6. Append to the SFILES a slash character `/` representing the split.
7. Retrieve the list of components in the permeate.
8. Order the components by decreasing molar volume.
9. For each component, append to the SFILES the letter corresponding to the component.
10. End the SFILES with a closing parenthesis `)`.

D.9.2 Property dependence

The liquid membrane process group can be used with a mixture of NC components if the following property dependence is satisfied:

1. The binary ratio for the radius of gyration between the key components must be greater than 1.03.
2. The binary ratio for the molar volume between the key components must be greater than 1.08.
3. The binary ratio for the solubility parameter between the key components must be greater than 1.28.

D.9.3 Connectivity rules and specifications

The liquid membrane process group is a one inlet two outlets process group.

The components recovery of the components in the permeate is 99.5%.

D.10 Gas Membrane Based Separation Process Group

The gas membrane process group is representing a membrane separation in vapor conditions. The synopsis of this process group is available in table D.13.

Name	Gas membrane
Specific properties	Van der Waals volume Critical temperature
Unit operations	Membrane
Representation example	`(gmemABC/D)`
Reverse simulation	Not available

Table D.13: Gas membrane synopsis

D.10.1 Procedure to generate the SFILES notation

To generate the SFILES notation of a gas membrane process group, the following procedure is applied:

1. Start the SFILES with an opening parenthesis `(`.
2. Append the identification string `gmem` of the process group to the SFILES.
3. Retrieve the list of components in the retentate.
4. Order the components by decreasing molar volume.
5. For each component, append to the SFILES the letter corresponding to the component.
6. Append to the SFILES a slash character `/` representing the split.
7. Retrieve the list of components in the permeate.
8. Order the components by decreasing molar volume.
9. For each component, append to the SFILES the letter corresponding to the component.
10. End the SFILES with a closing parenthesis `)`.

D.10.2 Property dependence

The gas membrane process group can be used with a mixture of NC components if the following property dependence is satisfied:

1. The binary ratio for the van der Waals volume between the key components must be greater than 1.07.
2. The binary ratio for the critical temperature between the key components must be greater than 1.10.

D.10.3 Connectivity rules and specifications

The gas membrane process group is a one inlet two outlets process group.

The components recovery of the components in the permeate is 99.5%.

D.11 Crystallization Separation Process Group

The crystallization process group is representing a crystallizer. The synopsis of this process group is available in table D.14.

Name	Crystallization
Specific properties	Melting point
Unit operations	Crystallizer
Representation example	`(crsABC/D)`
Reverse simulation	Available

Table D.14: Crystallization synopsis

D.11.1 Property dependence

The crystallization process group can be used with a mixture of NC components if the following property dependence is satisfied:

1. The binary ratio for the melting point between the key components must be greater than 1.20.

D.11.2 Connectivity rules and specifications

The crystallization process group is a one inlet two outlets process group.

The crystallized components are all the components having a higher melting point than the key component having the highest melting point, this key component being included.

The recovery of the crystallized components is 99.5%.

D.11.3 Reverse simulation

The reverse simulation of the crystallization process group is based on the generalized driving force concept presented in section 4.3.2 page 60.

D.12 Absorption Separation Process Group

The absorption process group is representing an absorption column. The synopsis of this process group is available in table D.15.

Name	Absorption
Specific properties	Solubility parameter
Unit operations	Column
Representation example	`(abEAB/eF/EABF/EF)`
Reverse simulation	Not available

Table D.15: Absorption synopsis

D.12.1 Procedure to generate the SFILES notation

To generate the SFILES notation of an absorption process group, the following procedure is applied:

1. Start the SFILES with an opening parenthesis `(`.
2. Append the identification string `ab` of the process group to the SFILES.
3. Retrieve the list of the components in the mixture to separate.
4. Order the components by decreasing solubility parameter.
5. For each component, append to the SFILES the letter corresponding to the component.
6. Append to the SFILES a slash character `/`.
7. Append `e` to the SFILES.
8. Append to the SFILES the letter corresponding to the entrainer.
9. Append to the SFILES a slash character `/`.
10. Retrieve the list of the components in the outlet low in entrainer.
11. Order the components by decreasing solubility parameter.
12. For each component, append to the SFILES the letter corresponding to the component.
13. Append to the SFILES a slash character `/`.
14. Retrieve the list of the components in the other outlet.
15. Order the components by decreasing solubility parameter.
16. For each component, append to the SFILES the letter corresponding to the component.
17. End the SFILES with a closing parenthesis `)`.

D.12.2 Property dependence

The absorption process group can be used with a mixture of NC components and the entrainer E, if the following property dependence is satisfied:

1. It exists to set of components within the NC components of the mixture such that the binary ratio for the solubility parameter between the components in set one and the entrainer E is greater than 1.11, and, the binary ratio for the solubility parameter between the components in the set two and the entrainer E is lower than 1.11.

In that case the components in set one will be available in the outlet low in entrainer and the components in set two will be in the outlet high in entrainer.

D.12.3 Initialization procedure

When initializing an absorption process group can with a given mixture, two cases are possible. Either the possible entrainers are known or not.

For the first case, the absorption process group is initialized with the mixture and the entrainer.

For the second case, as no entrainers are known, the following procedure is applied to find a matching entrainer. First a database search in the most common solvent database (see tables B.1 and B.2) is performed to look for a potential entrainer. If no entrainers are found, a CAMD problem formulation[23] is setup with the targets being to match the needed solubility parameter providing the definition of the two sets of components, the adsorbed and the non adsorbed. If more than one potential entrainer are found, the WAR algorithm[50] can be applied to keep the entrainer with the least environmental impact.

D.12.4 Connectivity rules and specifications

The absorption process group is a two inlets two outlets process group.

The recovery of the absorbed components is 99.5%.

Nomenclature

α_{ij}	Relative separability of component i with respect to component j
$\alpha_{i_{ref}}$	Relative separability of component i with respect to n-Octane
$\mathbf{A}$	Attainable region
A, B, C, ...	Component identities
a_j	Operating policy for process j
a_k	Regressed contribution of molecular group or process group k
$\mathbf{C}$	Characteristic vector or state of the system
$\mathbf{C}^*$	Achievable state within the system
C_A	Concentration of component A
$\overline{D_i}$	Maximum driving force of a separation task
D_{ij}	Driving force between component i and j
D_x	Composition of the light key at $F_{D_i\|Max}$
E	Exchange of energy
E_x	Energy consumption index
F_1	Flowsheet structure 1
F_A	Molar flowrate of component A
$F_{D_i\|Max}$	Maximum driving force between the key components
$f(P)$	Flowsheet property function
h^L	Liquid stream enthalpy
h^V	Vapor stream enthalpy
I	Exchange of information
k_i	Rate constant of reaction i
L	Liquid product flowrate
M	Exchange of materials
n	Number of different molecular groups
n	Number of process groups in a flowsheet structure
n	Number of flowsheet structure alternatives
N	Number of state variables to describe a state point of the attainable region
N_F	Feed plate location
N_{ideal}	Number of ideal stages in a distillation column
n_k	Number of each type of molecular group k
n_{key}	Individual flowrates of the key component
$n_{purge_{key}}$	Purge rate of the key component

NC	Number of component in a process group
NG	Number of process groups
nt	Number of separation tasks to perform before task k in the ideal case
P	Chemical process
p_1	First property to optimize
$\mathbf{P_1}(\mathbf{C}, \mathbf{U}, \mathbf{C}^*)$	Fundamental process vector
p_k	Topology factor
pos_k	Topology factor
Q	Volumetric flowrate
$R = L/V$	Ratio of the flowrates of product 1 (liquid) and product 2 (vapor)
R_A	Reaction set A
S_C	Selectivity to component C
τ	Residence time
U	The universe
$\mathbf{U}$	Control variables
u_i	Unit operation i
V	Vapor product flowrate
ξ	Extend of reaction
x_i	Fraction of component i in the first phase
y_i	Fraction of component i in the second phase
Z	Objective function
z_i	Feed composition of component i

References

[1] Y. Avramenko and A. Kraslawski. Decision supporting system for preselection of column internals in reactive distillation. *Chemical Engineering & Processing*, 44(6):609–616, 2005.

[2] E. Bek-Pedersen. *Synthesis and Design of Distillation based Separation Schemes*. PhD thesis, Technical University of Denmark, 2003.

[3] E. Bek-Pedersen and R. Gani. Design and synthesis of distillation systems using a driving-force-based approach. *Chemical Engineering and Processing*, 43:251–262, 2004.

[4] S. Bruggemann and W. Marquardt. Rapid screening of design alternatives for nonideal multiproduct distillation processes. *Computers and Chemical Engineering*, 29(1):165–179, 2004.

[5] L. Constantinou and R. Gani. New group contribution method for estimating properties of pure compounds. *AIChE Journal*, 40(10):1697–1709, 1994.

[6] L. d'Anterroches and R. Gani. Group contribution based process flowsheet synthesis, design and modelling. *Fluid Phase Equilibria*, 228-229:141–146, 2005.

[7] L. d'Anterroches, R. Gani, P.M. Harper, and M. Hostrup. Design of molecules, mixtures and processes through a novel group contribution method. In *Proceedings of the 7th World Congress of Chemical Engineering*, Glasgow, July 2005.

[8] I. Das and J. Dennis. A closer look at drawbacks of minimizing weighted sums of objectives for pareto set generation in multicriteria optimization problems. *Structural Optimization*, 14(1):63–69, 1997.

[9] J.M. Douglas. *Conceptual Design of Chemical Processes*, chapter 8. McGraw-Hill, 1988.

[10] A.P. Duvedi and L.E. Achenie. Designing environmentally safe refrigerants using mathematical programming. *Chemical Engineering Science*, 51(15):3727–3739, 1996.

[11] L.T. Fan, B. Bertok, F. Friedler, and S. Shafie. Mechanisms of ammonia-synthesis reaction revisited with the aid of a novel graph-theoretic method for determining candidate mechanisms in deriving the rate law of a catalytic reaction. *Hungarian Journal of Industrial Chemistry*, 29(1):71–80, 2001.

[12] J.L. Franklin. Prediction of heat and free energies of organic compounds. *Industrial & Engineering Chemistry*, 41(5):1070–1076, 1949.

[13] F. Friedler, K. Tarjan, Y.W. Huang, and L.T. Fan. Graph-theoretic approach to process synthesis: Polynomial algorithm for maximal structure generation. *Computers & Chemical Engineering*, 17(9):929–942, 1993.

[14] R. Gani and E. Bek-Pedersen. Simple new algorithm for distillation column design. *AIChE Journal*, 46(6):1271–1273, 2000.

[15] R. Gani, P.M. Harper, and M. Hostrup. Automatic creation of missing groups through connectivity index for pure-component property prediction. *Industrial and Engineering Chemistry Research*, 44(18):7262–7269, 2005.

[16] R. Gani, G. Hytoft, C. Jaksland, and A.K. Jensen. Integrated computer aided system for integrated design of chemical processes. *Computers & Chemical Engineering*, 21(10):1135–1146, 1997.

[17] M.R. Garey. *Computers and intractability: A guide to the theory of NP-completeness*. W. H. Freeman, 1979.

[18] B. Glasser, D. Hildebrandt, and D. Glasser. Optimal mixing for exothermic reversible reactions. *Industrial Engineering and Chemistry Research*, 31:1541–1549, 1992.

[19] D. Glasser, D. Hildebrandt, and C. Crowe. A geometric approach to steady flow reactors: The attainable region and optimization in concentration space. *Industrial Engineering and Chemistry Research*, 26:1803–1810, 1987.

[20] P.H. Groggins, editor. *Unit Processes in Organic Synthesis*. McGraw-Hill, 1952.

[21] I.E. Grossmann, P.A. Aguirre, and M. Barttfeld. Optimal synthesis of complex distillation columns using rigorous models. *Computers and Chemical Engineering*, 29(6):1203–1215, 2005.

[22] P.M. Harper. *A Multi-Phase Multi-Level Framework for Computer Aided Molecular Design*. PhD thesis, Technical University of Denmark, 2000.

[23] P.M. Harper and R. Gani. A multi-step and multi-level approach for computer aided molecular design. *Computers and Chemical Engineering*, 24(2-7):677 – 683, 2000.

[24] D. Hildebrandt and D. Glasser. The attaianble region and optimal reactor structures. *Chemical Engineering Science*, 45(8):2161–2168, 1990.

[25] F. Horn. Attainable and non-attainable regions in chemical reaction technique. In *Proc. 3rd Eur. Symp. on Chemical Reaction Engineering*, pages 293–303, 1964.

[26] M. Hostrup. *Integrated Approach to Computer Aided Process Synthesis*. PhD thesis, Technical University of Denmark, 2002.

[27] C. Jaksland. *Separation Process Design and Synthesis Based on Thermodynamic Insights*. PhD thesis, Technical University of Denmark, 1996.

[28] C. Jaksland and R. Gani. An integreated approach to process/product design and synthesis based on properties-process relationship. *Computers and Chemical Engineering*, 20(Supplement 1; Supplement to Computers & Chemical Engineering Part A):S151–S156, 1996.

[29] K.G. Joback and R.C. Reid. Estimation of pure-component properties from group contributions. *Chemical Engineering Communications*, 57:233–243, 1987.

[30] C.J. King. *Separation Processes*. McGraw-Hill, Inc., New York, 1980.

[31] B. Linnhoff and E. Hindmarsh. The pinch design method for heat exchanger networks. *Chemical Engineering Science*, 38:745–763, 1983.

[32] S. Lipczynska. Power to the people: the case for wikipedia. *Reference Reviews incorporating ASLIB Book Guide*, 19(2):6–7, 2005.

[33] J. Marrero and R. Gani. Group-contribution based estimation of pure component properties. *Fluid Phase Equilibria*, 183-184:183–208, 2001.

[34] K.P. Papalexandri and E.N. Pistikopoulos. A process synthesis modelling framework based on mass/heat transfer module hyperstructure. *Computers and Chemical Engineering*, 19:71–76, 1995.

[35] I.C. Parmee. Human-centric intelligent systems for exploration and knowledge discovery. *The Analyst*, 130(1):29–34, 2005.

[36] H. Renon and J.M. Prausnitz. Local compositions in thermodynamic excess functions for liquid mixtures. *AIChE Journal*, 14(1), 1968.

[37] C. Ruiz, J. Coca, A. Vega, and F.V. Diez. Extractive distillation of hydrocarbons with dimethylformamide: Experimental and simulation data. *Industrial & Engineering Chemistry Research*, 36(11):4934–4939, 1997.

[38] T. Seuranen, M. Hurme, and E. Pajula. Synthesis of separation processes by case-based reasoning. *Computers and Chemical Engineering*, 29(6):1473–1482, 2005.

[39] P.B. Shah and A. Kokossis. Design targets of separator and reactor-separator systems using conceptual programming. *Computers & Chemical Engineering*, 21:S1013–S1018, 1997.

[40] P.B. Shah and A.C. Kokossis. New synthesis framework for the optimization of complex distillation systems. *AIChE Journal*, 48(3):527–550, 2002.

[41] M.R. Shelton and I.E. Grossmann. Optimal synthesis of integrated refrigeration systems. i. mixed-integer programming model. *Computers & Chemical Engineering*, 10(5):445–59, 1986.

[42] J.J. Siirola. Strategic process synthesis: advances in the hierarchical approach. *Computers and Chemical Engineering*, 20(Supplement 2; Supplement to Computers & Chemical Engineering Part B):S1637–S1643, 1996.

[43] G. Soave. Equilibrium constants from a modified redlich-kwong equation of state. *Chemical Engineering Science*, 27(6):1197–203, 1972.

[44] P.J. Trambouze and E.L. Piret. Continuous stirred tank reactors: Designs for maximum conversions of raw material to desired product. *AIChE Journal*, 5:384, 1959.

[45] J.C. Trinder and Y. Wang. Automatic road extraction from aerial images. *Digital Signal Processing*, 8(4):215–224, 1998.

[46] G. van Rossum. Python tutorial. Technical Report CS-R9526, Centrum voor Wiskunde en Informatica, Amsterdam, May 1995.

[47] V. Venkatasubramanian, K. Chan, and J.M. Caruthers. Computer-aided molecular design using genetic algorithms. *Computers & Chemical Engineering*, 18(9):833–844, 1994.

[48] D. Weininger. Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences*, 28(1):31–6, 1988.

[49] D. Weininger, A. Weininger, and J.L. Weininger. Smiles: 2. algorithm for generation of unique smiles notation. *Journal of Chemical Information and Computer Sciences*, 29(2):97–101, 1989.

[50] D. Young, R. Scharp, and H. Cabezas. The waste reduction (war) algorithm: environmental impacts, energy consumption, and engineering economics. *Waste Management*, 20(8):605–615, 2000.