A PROPERTY BASED APPROACH TO INTEGRATED

PROCESS AND MOLECULAR DESIGN

Except where reference is made to the work of others, the work described in this thesis is my own or was done in collaboration with my advisory committee. This thesis does not include proprietary or classified information.

_____
Fadwa Tahra Eljack

Certificate of Approval:

_____
W. Robert Ashurst
Assistant Professor
Chemical Engineering

_____
Mario R. Eden, Chair
Assistant Professor
Chemical Engineering

_____
Ram B. Gupta
Professor
Chemical Engineering

_____
Christopher B. Roberts
Professor and Chair
Chemical Engineering

_____
Mahmoud M. El-Halwagi
Professor
Chemical Engineering
Texas A&M University
College Station, TX

_____
George T. Flowers
Interim Dean
Graduate School

A PROPERTY BASED APPROACH TO INTEGRATED

PROCESS AND MOLECULAR DESIGN

Fadwa Tahra Eljack

A Dissertation

Submitted to

the Graduate Faculty of

Auburn University

in Partial Fulfillment of the

Requirements for the

Degree of

Doctor of Philosophy

Auburn, Alabama
May 10, 2007

A PROPERTY BASED APPROACH TO INTEGRATED

PROCESS AND MOLECULAR DESIGN

Fadwa Tahra Eljack

_____
Signature of Author

_____
Date of Graduation

VITA

       Fadwa Tahra Eljack, daughter of Rashida Eljack and Abdalla Eljack, wife of Nazar Suliman, and mother of Nour Suliman, was born on June 6, 1977, in Saskatoon, Canada. She graduated from Auburn High School as National Merit Finalist. She pursued a Bachelor's degree in Chemical Engineering at Auburn University and graduated in June 1999. She later entered the graduate doctoral program at Auburn University in 2003. Fadwa is currently a new faculty member at Qatar University, Qatar.

DISSERTATION ABSTRACT

A PROPERTY BASED APPROACH TO INTEGRATED

PROCESS AND MOLECULAR DESIGN

Fadwa Tahra Eljack

Doctor of Philosophy, May 10, 2007
(B.Sc., Auburn University, 1999)

191 Typed Pages

Directed by Mario Richard Eden

In this work, a new simple yet effective, systematic method to synthesize and design molecules is presented. Visualization of the problem is achieved by employing an annex to the recently developed property clustering techniques, which allows a high-dimensional problem to be visualized in two or three dimensions by employing the concepts of reverse problem formulation. Group contribution methods are used to predict the properties of the formulated molecule. For the molecular design problem the target properties as well as the molecular groups that make up the formulations are identified on a ternary diagram. The target properties are represented as individual points if given as discrete values or as a region if given as intervals. The formulation of the desired

molecule is achieved via linear "mixing" of molecular groups in order to match the desired performance.

A significant advantage of the developed methodology is that for problems that can be satisfactorily described by just three properties, the process and molecular design problems can be simultaneously solved visually on ternary diagrams, irrespective of how many molecular fragments are included in the search space. The process design problem is solved for the desired target properties using property clusters. This is the solution of a reverse simulation problem, where the process design problem is solved in terms of constitutive variables and without having to commit to any component a priori. The target properties as well as a selection of molecular building blocks (groups) are used as input into the molecular design algorithm. The problem is now visualized on a molecular ternary cluster diagram. The structure and identity of candidate components is then identified by combining or "mixing" molecular fragments until the resulting properties match the targets. The designed candidate formulations are screened using the developed necessary and sufficient conditions for the synthesis of molecules. Finally, the feasible molecular formulations are mapped back to the process domain for verification.

Although, the molecular property clustering framework provides a property interface for the simultaneous consideration of process and molecular design problems, it should be emphasized that the developed tools can also be used to solve just molecular synthesis problems (e.g. solvent design). As a CAMD tool, this algorithm has the added feature of visual synthesis for those problems that can be described using three clusters or properties; and for those requiring more than three, an algebraic approach for the formulation and solution of molecular design problems is outlined.

ACKNOWLEDGEMENT

Style manual or journal style: Computers and Chemical Engineering Journal

Computer software used: Microsoft Word, Excel, and Visio

TABLE OF CONTENTS

`

LIST OF FIGURES

LIST OF TABLES

1. Introduction

The terms chemical (product) synthesis and design designate problems involving identification and selection of formulations (compounds) or mixtures that are capable of performing certain tasks or possess certain qualities (properties). Since the properties of the compound or mixture dictate whether or not the design is useful, the basis for solution approaches in this area should be based on the properties themselves. In fact, for molecular design techniques e.g. Computer Aided Molecular Design (CAMD), the desired target properties are required inputs to the algorithm. The performance requirements for the formulations are usually determined by process needs. Thus, the identification of the desired formulation properties should be driven by the desired process performance.

Although this integrated relationship between product and process design problems is recognized; traditionally they have been treated as two separate problems, with little or no feedback between the two. Generally the objective in the design or optimization of processes is to find a balance between satisfying process unit requirements and the use of appropriate raw materials in order to maximize profit and minimize cost. The raw materials used, are selected from a list of pre-defined candidate components, therefore limiting performance to the listed components. The problem here is that these decisions are made ahead of design and are usually based on qualitative process knowledge and/or experience and thus possibly yield a sub-optimal design,

re-emphasizing that the main setback in finding an optimal solution is that process and molecular design problems have been decoupled. Each problem has been conveniently isolated.

Why does the simultaneous consideration of process and molecular design problems

present such a challenge?

When considering interfacing product and process design using conventional methods such as mathematical programming, most algorithms face a bottleneck when it comes to using property models; suitable models for product design may not be available for process design and vice versa (Gani, 2001). In addition, once a property model is selected for inclusion into the process model, its application range is restricted by the availability of model parameters for those molecules. Mathematical programming techniques suffer from discontinuities in the solution trajectory in response to changes in the model equations. Inclusion of multiple models for the same property into the algorithm may make it more difficult to achieve convergence. Hence, understanding how property models fit into design is key in resolving some of these issues.

Recent contributions to understanding the roles of property models in the solution of Computer Aided Process Engineering (CAPE) problems brought about the development of the reverse problem formulation (RPF) framework (Eden, 2003). In principle, process models are made up of balance equations, constraint equations, and constitutive equations. The constitutive equations are used to represent property models

in terms of intensive variables (e.g. temperature, pressure, and composition). Often, the complexity of property models is indicative of non-linear behavior of process model equations; leading to intense computations and consequently difficulty in reaching convergence. The RPF methodology decouples the constitutive equations from the balance and constraint equations, so the traditional forward problem can be solved as two reverse problems. It is a technique analogous to how a molecular design problem is formulated as a reverse property prediction problem. Here, the first problem (reverse-simulation) solves the balance and constraint equations in terms of the constitutive variables (properties), providing the design targets. The second reverse problem (reverse property prediction) solves the constitutive equations to identify unit operations, operating conditions and/or products that possess the targeted property values, set forth by the first reverse problem. The key advantage to this targeting approach is the exclusion of the constitutive equations, which allows for easy solution of the balance and constraint equations, which are generally linear. Now, the algorithm gains the freedom to use different property models at any point during the solution step as long as the property targets are matched; making for a robust design algorithm. Once the design targets are identified, only then are the constitutive equations solved to identify the intensive variables. Thus RPF lowers the complexity of the design problem without sacrificing accuracy of the design.

The use of RPF in design is the first step towards developing of a simultaneous approach for solving process and molecular design problems. However, there is the still the question of how to link the two problems. There is a need for a method to facilitate the flow of information from the process level to the molecular level, and vice versa.

Recall, that process unit performance is gauged by properties such as boiling temperature, heat of vaporization etc. Furthermore, properties are required inputs for solution algorithm in molecular design. Therefore it would make sense to use a property based platform as the link between process and molecular design approaches.

Introduction of the property clustering framework allow for representation of process streams and units from a properties perspective (Shelley and El-Halwagi, 2000). Recognizing that properties are not conserved and thus can not be tracked, the property clustering concept fashioned the idea of property operators which are functions of physical properties. The property clusters are conserved and posses the unique feature of linear mixing rules; even though the operators themselves might not be linear (e.g. the inverse of the mixture density of two components is the result of the sum of the inverse density of the individual components).

The property integration framework is based on reverse problem formulations and utilizes property clusters to provide a representation of the constitutive variables of a system. Within this framework, only the process design algorithm has been developed; where process needs are targeted ahead of design and used as input. For cases where the system can be described by just three properties, the process design problem can be visualized on a ternary cluster diagram. Discrete points are used to represent property values while feasibility regions are used for a range of accepted property values. Visualization of the problem allows for easy identification of optimum recycle strategies, while the unique feature of linear mixing rules allows for the use of simple lever arm analysis to solve the problem. Thus, this framework allows for the representation and solution of process design problems that are driven by properties.

4

In chemical process design, there is a general need for reliable and accurate property estimation methods. It is critical to the solution of most simulation problems, where convergence is often dependent on the reliability of predicted physical and thermodynamic properties. More so for molecular design algorithms, where predictive property models are at the heart of all solution strategies. Almost all property estimation methods used in CAMD techniques are based on Group Contribution Methods (GCM), where the properties of a compound are expressed in terms of a function of the number of occurrences of predefined groups in the molecule. The novel techniques developed in this work, merge the concepts of group contribution methods with those of the property integration framework, to provide a property based platform capable of simultaneous handling of process and molecular synthesis/design needs. In chemical processes the utilization of such an approach enables identification of the desired system properties by targeting the optimum process performance without committing to any components during the solution step. The identified property targets can then be used as inputs for solving the molecular design problem, which returns the corresponding components.

The purpose of the work presented here is to develop a property based molecular design algorithm within the general property clustering framework. The mixing rules will invariably be functionally different for molecular groups and process streams; however since they represent the same property, they can still be visualized on the same diagram. Once visualized it is possible to solve the process design problem by identifying the system/product properties corresponding to the desired process performance. On the ternary diagram the target product properties will be represented as either a single point or a region depending on whether the target properties are discrete or given as intervals.

5

The structure and identity of candidate molecules are then identified by combining or "mixing" molecular fragments until the resulting properties match the targets.

A key advantage of the developed technique is that for problems, which can be satisfactorily described by just three properties, the process and molecular design problems are solved visually and simultaneously on ternary diagrams; irrespective of how many molecularly fragments are included in the search space. Furthermore, the molecular property cluster framework can be used as a visual CAMD tool for solvent design. By visualizing the property constraints, as a feasibility region on the ternary cluster diagram along with a wide range of molecular groups, the search space is minimized by excluding those fragments that do not aid in reaching the property targets. In addition, an algebraic approach has been developed in recognition of the fact that not all design problems can be described in terms of just three properties. To take advantage of the developed molecular property operators in lowering the dimensionality of the design problem, this technique provides a simple method for formulating the molecular design problem as a set of linear algebraic equality and inequality equations. The benefits gained through utilization of this technique, is that the molecular design problem traditionally formulated as mixed integer non-linear program (MINLP) can now be presented as a simple linear program (LP)

The conventional process and molecular design methods, tools and techniques upon which the molecular property technique was developed, are discussed in Chapter 2. The property clustering technique that provided the building blocks for a property driven approach to solving process design is presented in Chapter 3. Section 3.2 introduces the foundations of the molecular property clustering methodology developed in this work,

6

including the incorporation of group contribution methods and detailed development of the molecular property operators. In Sections 3.3 and 3.4 the methodology behind the visual synthesis of molecular formulations and the algebraic approach are outlined. Chapters 4 and 5 provide application examples of the developed methodology. In chapter 4, aniline extraction and blanket wash solvent design problems illustrate the advantages of using this new methodology for molecular design problems. It highlights how the algorithm handles cases where group contribution property models do not exist, demonstrating the algorithm's ability to handle multiple models. The blanket wash solvent design problem has been previously solved as mixed-integer non-linear problem (MINLP), which allowed for comparing the formulation, designed using the molecular cluster technique versus other established methods. In chapter 5, the simultaneous consideration of process and molecular design using the clustering framework, is presented via two application examples. The metal degreasing example supplied property targets as well as the constraints placed on the process design problem. The problem is solved as a maximization problem of the available in house resource. Using lever-arm analysis the process design problem was solved for the cluster values that optimized the process needs. Those cluster values are then converted property targets used as input into the molecular framework, where a set of candidate formulations are generated and screened using a set of conditions established as part of the framework. The resulting three valid formulations were mapped back to the process design level, where the validity of the solution was established. The gas purification example aims at identifying a solvent that will replace methyl diethanol amine, one of three streams currently fed to the gas treatment process unit (sink). The process design objectives and

7

requirements dictate that the mixed stream properties must match that of the sink. Lever arm analysis is used to determine the mixture property requirements; and the calculations steps are highlighted in Section 5.2.2. Finally chapter 6 summarizes the conclusions of the thesis and provides future directions for further development of the framework.

2. Theoretical Background

2.1. Process and Product Design

In the chemical processing industry product design arises from a "need" and involves finding a product that exhibits certain desirable behavior or involves finding an additive (chemical) that when added to another chemical or product enhances its desirable functional properties (Achenie et al., 2003).

In product design, the identity of the final product is unknown, however the general behavior or characteristics of the product (goal) is known. The objective is to find the most appropriate chemical or a mixture of chemicals that will satisfy this goal. Once possible solutions to the problem are generated, the next step is processing of the raw materials. Cussler and Moggridge (2001) suggested the following steps for product design:

1. Define needs (formulate the problem)

2. Generate ideas to meet the needs (generate molecular or flowsheet structures matching the problem)

3. Select among ideas (rank the generated alternatives to get the best alternative)

4. Manufacture product

The first step is to understand the design objectives. Regardless, if it is a molecular design problem or a mixture/blend design problem the solution strategy is the same (see Figure 2.1). Product designers/chemists depend on their understanding and knowledge of the matter and suggest a list of raw materials they believe will lead to the desired product(s); in other words they generate possible solutions and select among those alternatives.
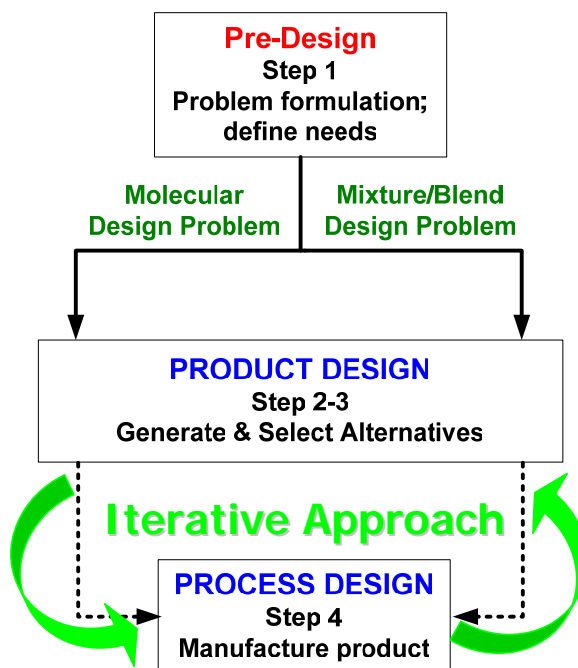
Figure 2.1: Product design approach according to Cussler and Moggridge

Next product designers transfer this information to process designers in order to manufacture the final desired product, i.e. satisfy initial consumer "needs". This final step is labeled process design; where process designers obtain the list of suggested alternatives provided by the product designers and they look at performing feasibility and profitability analysis. Engineers here gain a detailed understanding of how the process flowsheet will function; and a common objective is to determine feasible and preferably

10

optimal configurations in terms of selecting equipment and conditions of operation for the parts of the process being considered (Hostrup, 2002). Process designers also consider environmental impact as well as health and safety issues. After addressing all these tasks, they determine whether the generated alternatives (step 2 and 3) are practical. If they conclude that the designs are infeasible, then all of their findings are passed back to the product designers to make further alterations to their design. Such changes may include altering the chemistry make up or the starting raw materials, etc. Once new alternative designs are generated then they are again passed to process engineers for further study, thus making the approach iterative (see Figure 2.1). Iteration can lead to inefficiency and is a result of an apparent gap that exists between products/molecular and process design approaches.

The work presented in this thesis aims at bridging the gap between process and molecular design approaches, through the development of design tools that address both design objectives. First the scope of the problem will be clearly identified and the overall objectives will be stated. The following sections intend to provide an overview of some of the current approaches and techniques used in process (Section 2.4) and molecular synthesis/design (Section 2.6). Furthermore, this chapter discusses the methods and tools that provided the foundation for the design techniques developed in this thesis.

2.2. Scope of the Problem

Traditionally, process and molecular design problems have been treated as two separate entities. In molecular design, the general approach is often based on trial and error experimentation. Although not efficient, this method is the only available option in

cases where property models are not available to predict the properties of the desired components. In cases where models do exist, the algorithm (see Figure 2.2) is given the overall objectives and a set of molecular building blocks (e.g. $-CH_3$, $-OH$, etc.). and the goal is to identify a set of candidate components that meet a given set of criteria (e.g. physical or chemical properties). The importance of property models to design will be further discussed in Section 2.7.

In process design, generally the objective is to find a balance between satisfying process unit requirements/constraints and the use of appropriate raw materials and processing chemicals e.g. solvents, in order to maximize profit and minimize cost (see Figure 2.2). The chemicals used as input into process algorithms, are selected from a list of pre-defined candidate components, therefore limiting performance to the listed components. The problem here is that (1) molecular/product designers make these decisions ahead of design (process) and (2) often these decisions are based on qualitative process knowledge and/or experience and thus possibly yield a sub-optimal design.

Hence a major obstacle in finding an optimal design is that the process and molecular design problems have been decoupled from each other. Each problem has been conveniently isolated.
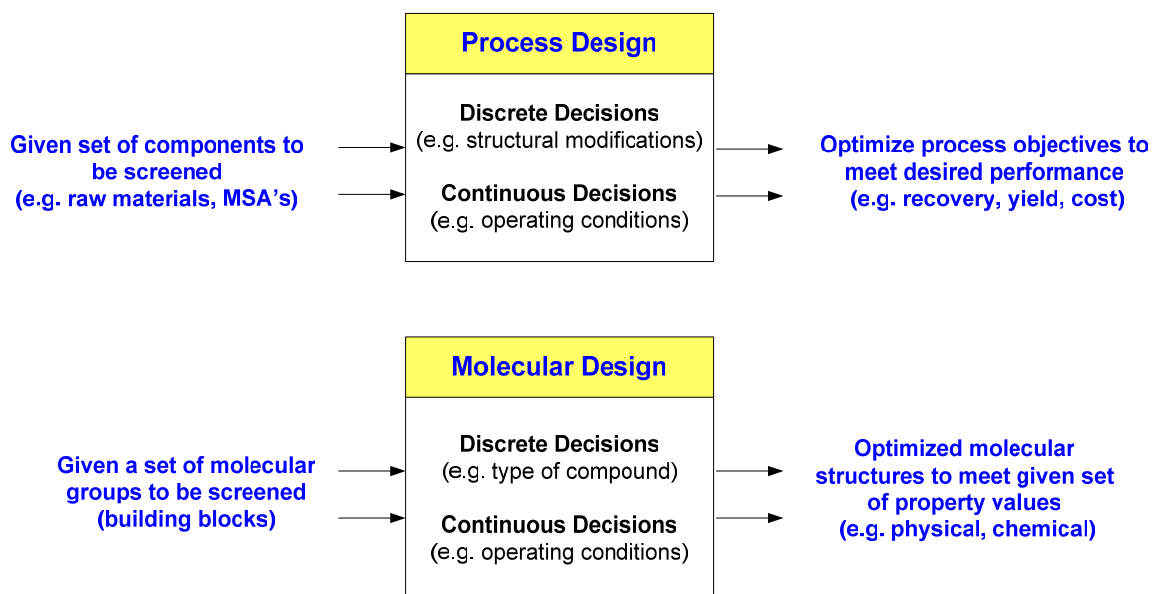
Figure 2.2: Traditional approach to molecular and process design
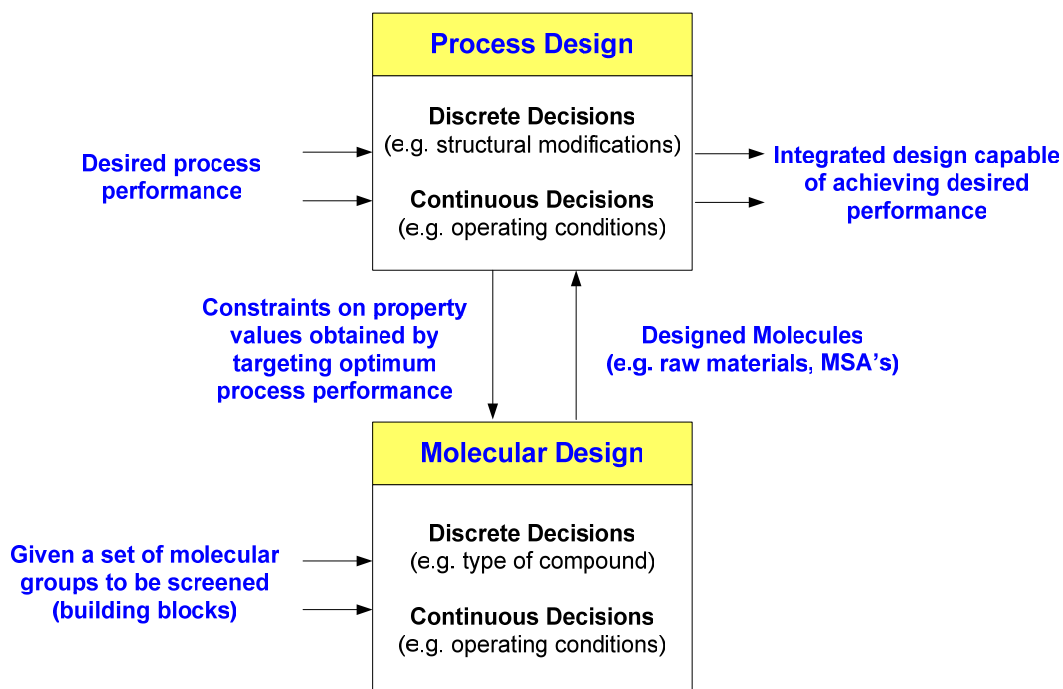
Figure 2.3: Integrated approach to process and molecular design

13

The work of researchers like Cussler and Moggride (2001) highlighted at the beginning of this chapter recognize the potential benefits to be gained, by allowing the flow/exchange of information between process and molecular design algorithms. It is the objective of this thesis to introduce a unified design approach to overcome the limitations of decoupling the two problems. Figure 2.3 outlines the proposed approach for handling design problems. Where the process design algorithm takes in only the desired process performance requirements and solves the design problem for optimal process operating conditions and desired functionalities. These requirements along with a set of molecular building blocks are now the input to the molecular design algorithm; where the objective is to formulate candidate components that posses these properties. The generated list of candidate feasible components are guaranteed to posses all of the required criteria, and other screening criteria (e.g. environmental impact, cost etc.) can be used to rank the candidate components.

2.3. General Problem Definition

The general formulation of process/product synthesis/design problems can be described by the following set of equations with x and y as the optimization real and integer variables, respectively:

$$F_{obj} = \min \; \{A^T y + f(x)\} \qquad \text{Objective function} \qquad (2.1)$$

s.t.

$$h_1(\frac{\partial x}{\partial z}, x, y) = 0 \qquad \text{Process/product model} \qquad (2.2)$$

$$h_2\,(x,y) = 0 \qquad\qquad \text{Equality constraints} \qquad\qquad (2.3)$$

$$g_1\,(x) > 0 \qquad\qquad \text{Process inequality constraints} \qquad\qquad (2.4)$$

$$g_2\,(x) > 0 \qquad\qquad \text{Product inequality constraints} \qquad\qquad (2.5)$$

$$B \cdot y + C \cdot x > d \qquad\qquad \text{Structural constraints} \qquad\qquad (2.6)$$

Many variations of the above mathematical formulation can be derived to represent different synthesis and/or design problems. Some of the equations or terms may be excluded, depending on the type of problem solved. If the objective is to simply generate a feasible solution to the process/molecular design problem, then only the equality, inequality and structural constraints are considered. However, some approaches utilize mathematical optimization tools that aim at identifying optimal solutions, which require solving equations 2.1-2.6. The various approaches for process synthesis/design are reviewed in the following section.

2.4. Process Synthesis and Design Approaches

Process synthesis deals with the activities where the various process elements are integrated and the flowsheet of the system is generated to meet certain objectives. To gain a detailed understanding of how the process behaves and whether the process objectives are met, process analysis tools such as ASPEN Plus, PRO/II, and HYSYS are often utilized (El-Halwagi, 2006). The common objective is to determine feasible and preferably optimal configurations in terms of the selection of equipment and conditions of operation for each part of the process (Hostrup, 2002). Once a feasible flowsheet has been identified, it is analyzed/tested to make sure the process objectives are met.

Iterations between process synthesis and analysis are continued until the desired goals are met (El-Halwagi, 1997). Biegler et al. (1997) lists the basic steps in flowsheet/process synthesis as: (1) gathering information, (2) representation of alternatives, (3) assessment of preliminary design, and (4) the generation and search among alternatives.

Several approaches exist that aim at developing and improving ideas for the design of process including:

- *Brainstorming different scenarios* by a select group of experts in the scientific and engineering fields dealing with the specific process. The "optimal" generated design using this approach is determined by the ability to generate alternatives and absence of bias towards a specific solution. The problem here is that these decisions are made ahead of design which might lead to a sub-optimal solution.

- Another method to solving a process design problem is *adapting an old solution* to a similar design challenge and then improving on it (El-Halwagi, 2006). The limitation here is that this solution can not guarantee optimality.

- *Heuristic based approaches*: Process engineers have classified most processes into groups or categories, and each is assigned a group of possible solutions. This approach uses rules to analyze the problem and to fix many of the discrete variables a priori to reduce the size of the search space. The rules come about as direct observation of recurring behavior in a given type of problem. Heuristics are used as a tool to aid in choosing how decisions should be made and which decisions we should make. Without heuristics, design problems are often too difficult to converge and/or too large to search, however here again the optimality

of the generated solution is not guaranteed (Westerberg, 2004). The approach is useful only in cases where the problem at hand is closely related to the class of problems for which the solution has been derived (El-Halwagi, 1997).

- *Mathematical optimization approaches* for process design require (Westerberg, 2004): a problem formulation that can express goals and describe them as an actionable task; the ability to enumerate all alternatives; and the capability of narrowing down the search space by eliminating alternatives. Mathematical optimization approaches are excellent because they can guarantee optimality solutions; however, they can not always guarantee convergence, i.e. you can not depend on the approach to always generate a solution. Usually, representation of such large optimization problems is in the form of Mixed Integer Non-Linear Programs (MINLPs). The algorithm identifies the integer variables (e.g. determine the existence or absence of a certain piece of equipment) and continuous variables, that determine design and operating parameters such as temperature, pressure, flowrate and the size of the equipment.

Process synthesis reviews are readily available in the open literature. The various approaches are organized into two categories, those that are structure-independent (also known as targeting approaches), and those that are structure based. All the approaches follow the basic steps of process synthesis and design summarized by Biegler and co-workers (1997). The approaches vary in the generation of alternatives and the manner in which optimal solutions are identified from amongst all the alternatives. The first category looks at solving the synthesis problem by breaking it down into multiple stages

to reduce the dimensionality of the problem. Within each stage, the design targets are identified and used in the following stage. The structure-dependent approaches, like superstructures, include the structure of the process design (i.e., equipment identity and connectivity) as well as all the design and operating parameters for each piece of equipment as part of its formulation, therefore the superstructure encompasses many redundant paths and equipment alternatives for achieving the design objectives. Superstructure optimization is the process of stripping away the unessential pathways and equipment alternatives to find the "best" solution. Two separate and distinct problems still limit the use of superstructure optimization techniques: (1) how to generate the initial superstructure to guarantee it contains the "best" solution (2) how to solve the large optimization problems inherent in practical synthesis problems (Barnicki and Siirola 2004).

Recognizing, that structure-dependent approaches are generally more robust than their independent counter part, El-Halwagi (1997) and Westerberg (2004) identified several issues that process design algorithms should address: first, the methodology has to be able to enumerate all alternatives and represent them in a common space. Failure to include some possible configuration can lead to sub-optimal solutions. This is related to the ability to systematically narrow down the search space. The second issue is that mathematical optimization problems of such magnitude often fail to converge due to the complexity of the non-linear properties included in the formulation. Finally, to avoid obtaining a biased configuration, due to the influence of personal or engineering evaluation, all insights should already be part of the problem formulation.

The novel concept of property clusters (Shelly and El-Halwagi, 2000; Eden, 2003) provides design tools that can address the issues raised by El-Halwagi (1997) and Westerberg (2004). Property clustering methods lower the complexity of the design problem, by mapping properties to a lower domain and by perceptively using given information as guidelines for placing bounds on the search space. This methodology provides the platform on which the tools presented by thesis are built. Recognizing that clustering methods have only been recently developed, a detailed review of the concept will be covered in Section 3.1.

2.5. Process Integration

First attempts at optimization of processes came in the form of process integration. *Process integration* is defined as "a holistic approach to process design and optimization, which encompasses design, retrofitting and operations of the process" (El-Halwagi, 1997). The aim is to allow us to see "the big picture first, and the details later". Integration requires the ability to state the objective in "actionable tasks" or in terms of quantified engineering parameters (e.g. maximizing profit can be translated into minimizing raw material usage or waste material generation etc.) The designer needs to identify global performance targets ahead of any development activity and identify the optimal strategy to reach it (Sirinivas, 1997). It is important to find and evaluate the maximum performance benchmarks ahead of synthesizing the design to obtain insights about potential opportunities. In that sense, an efficient methodology must include the ability to identify the search space, generate solutions knowledgably; and finally the capability to select amongst the alternatives.

Processes are generally characterized by the flow of materials/mass and energy. Mass flow includes flow of raw materials including solvents, feed material etc. utilized within the process to make the products. Energy flow in the form of water, heating and cooling power, coal or gases etc. is needed to process the mass flow to desired products (see Figure 2.4), (Srinivas, 1993).

Energy and mass integration are systematic methods for identifying energy and mass performance targets respectively. Energy integration aims at heat recovery within a process. It can also identify the optimal system configuration for the minimal energy consumption. Mass integration techniques/tools provide means of identifying optimum performance targets by generating and selecting among alternatives for allocating the flow of material (species) in the process.
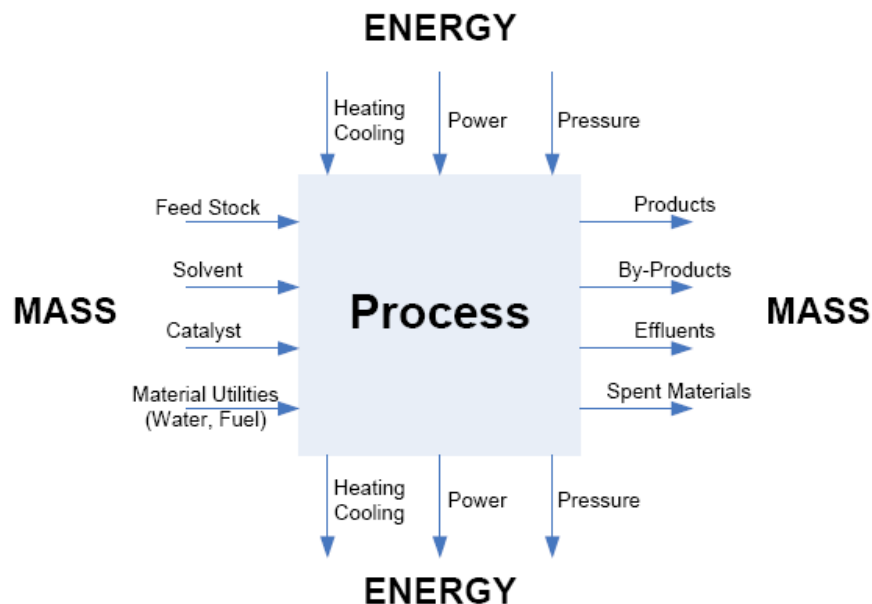
Figure 2.4: Mass-energy matrix of a process (Garrison et al., 1996)

Numerous successful applications worldwide in a range of industries testify to the value of process integration technology in reducing energy costs and increasing capacity through debottlenecking (Gundersen and Ness, 1988). Some of the principal tools used by the two integration fields are highlighted below.

2.5.1. Heat Integration

Process integration efforts began in the late 1970s; and were initially rooted in energy conservation and, in particular, the design of heat exchanger networks (HEN) (Linnhoff and Hindmarsh, 1983; Papoulias and Grossmann, 1983; Gundersen and Ness, 1988; and Cerda et al., 1983). Tools were developed to find ways to increase energy conservation/utilization, in response to the rise in energy cost. Efforts led to the development of a variety of tools, the best-known of which are composite curves and pinch design methods or analysis used to identify minimum utility targets ahead of designing of the HEN.

In a chemical process there are generally several streams that require heating or cooling before they satisfy the process unit requirements. The use of external utilities; (e.g. steam and cooling water) for each stream requiring cooling or heating is not economically efficient, consequently it is desirable to lower the use of external utilities by maximizing the transfer of available internal energy from hot to cold streams prior to implementation.

The synthesis of HENs involves optimum allocation of energy within a process via maximizing the exchange of internal energy between process hot streams and cold

streams; which are process streams that require cooling and heating, respectively. In synthesizing such networks the following process information is given:

- Number of hot and cold streams

- Heat capacity flowrate of hot (HH) and cold (HC) streams = flowrate (F) x specific heat (C_p)

- Hot streams supply (inlet) temperature (T^s) and target (outlet) temperature (T^t) Cold streams inlet or supply temperature (t^s) and outlet or target temperature (t^t)

- T^s and T^t of available external heating and cooling utilities

The design tasks are as follows (El-Halwagi, 2006):

- Which heat and/or cooling utilities should be employed?

- What is the optimal heat load to be removed/added by each utility?

- How should the hot and cold streams be matched, i.e. stream pairings?

- What is the optimal system configuration, e.g. how should the heat exchangers be arranged? Should any streams be mixed or split?

Hohmann (1971) introduced the "thermal pinch diagram", the first graphical approach aimed at identifying the minimum utility requirements. Linnhoff and collaborators led the efforts to advance the development of this technique (Linnhoff et al., 1982; Linnhoff and Hindmarsh, 1983). The method is based on the ability to thermodynamically transfer heat from any hot streams with temperature T to any cold stream with temperature t, with a minimal driving force of $\Delta T_{min.}$. The minimum hot stream temperature where heat transfer is feasible is given by equation 2.7:

$$T = t + \Delta T_{min} \tag{2.7}$$

First, a hot composite stream representing all hot process streams must be constructed. The diagram represents the amount of enthalpy exchanged by each hot stream vs. temperature, assuming ideal thermodynamics and constant heat capacities. The composite stream is a global representation of all the hot process streams as a function of the heat they exchanged vs. temperature. An example of a hot composite stream for two hot streams is shown on Figure 2.5, where the tail and head of each arrow represents the supply (T^s) and target (T^t) temperatures, respectively. The amount of energy or heat lost by a hot stream (HH) and analogously gained by a cold stream (HC) is calculated according to equations 2.8 – 2.9. The hot composite stream is created by superposition, see Figure 2.5. In a similar manner a cold composite stream can be constructed. Next both streams are plotted on the same diagram; this is possible by having two temperature scales, where the cool composite temperature scale shifts by ΔT_{min}. The position of the hot composite stream is going to always be on the right of the cold composite stream because the temperature of the hot stream is always higher than or equal to the temperature of the cold with a minimum temperature gradient of ΔT_{min}, as seen in equation 2.7.
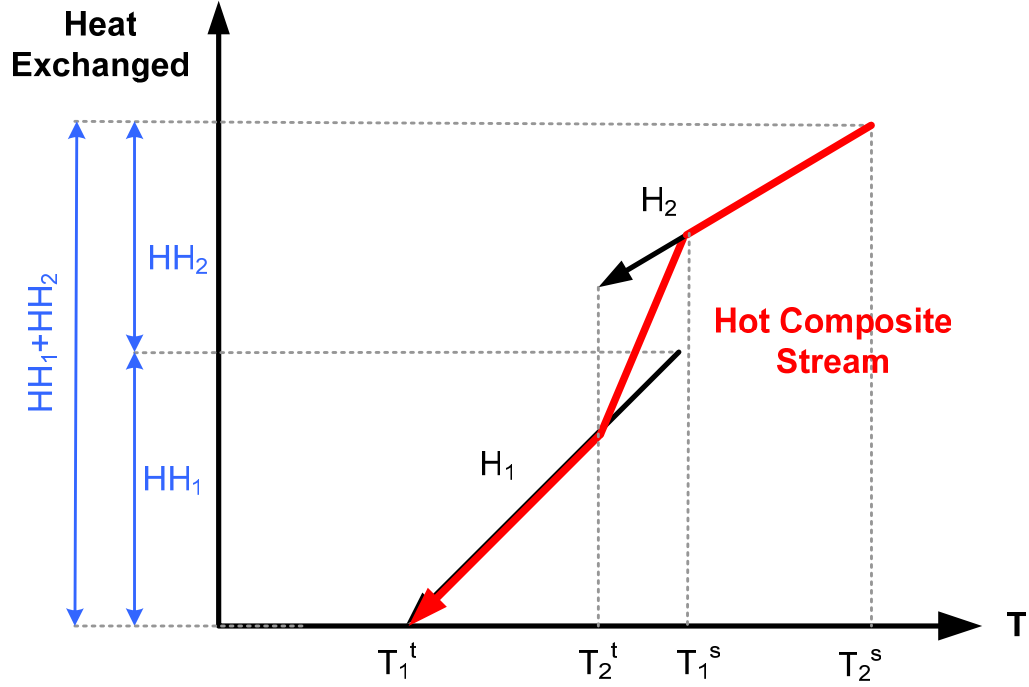
Figure 2.5: Representation of hot composite stream

$$HH = FC_p(T^s - T^t) \tag{2.8}$$

$$HC = FC_p(t^s - t^t) \tag{2.9}$$

The overlap between the composite curves provides a target for the heat recovery opportunities, labeled integrated heat exchange, see Figure 2.6A. Hence, the overlap in enthalpy that occurs between the two curves on this diagram guarantees the ability to exchange heat from hot to cold streams without the use of external utilities. Those duties that cannot be satisfied by internal energy recovery must be serviced by external heating and cooling utilities.

The cold composite stream can be moved up/down on the diagram, where final location of the stream determines the amount of heat being exchanged, see Figure 2.6A.

Here, the construction of the diagram provides a tool for the determination of maximum heat recovery targets. The minimum external utility requirements can be identified by sliding the cooling curve all the way down until the two curves touch, Figure 2.6B. This point is named "the thermal pinch point". If the cooling curve is moved up and away from the pinch this signifies a penalty in terms of the amount of energy being exchanged between the streams, and thus additional external utility are required. If the cooling curve is moved down passed the pinch point, then heat integration potential is lost; yet again resulting in a need for additional external utilities. To avoid loss of potential integration, Linnhoff et al. (1982) developed rules to identify minimum external heating and cooling utilities once the pinch point has been identified:

- No heat should be passed though the pinch

- No external cooling utilities used above the pinch

- No external heating utilities used below the pinch
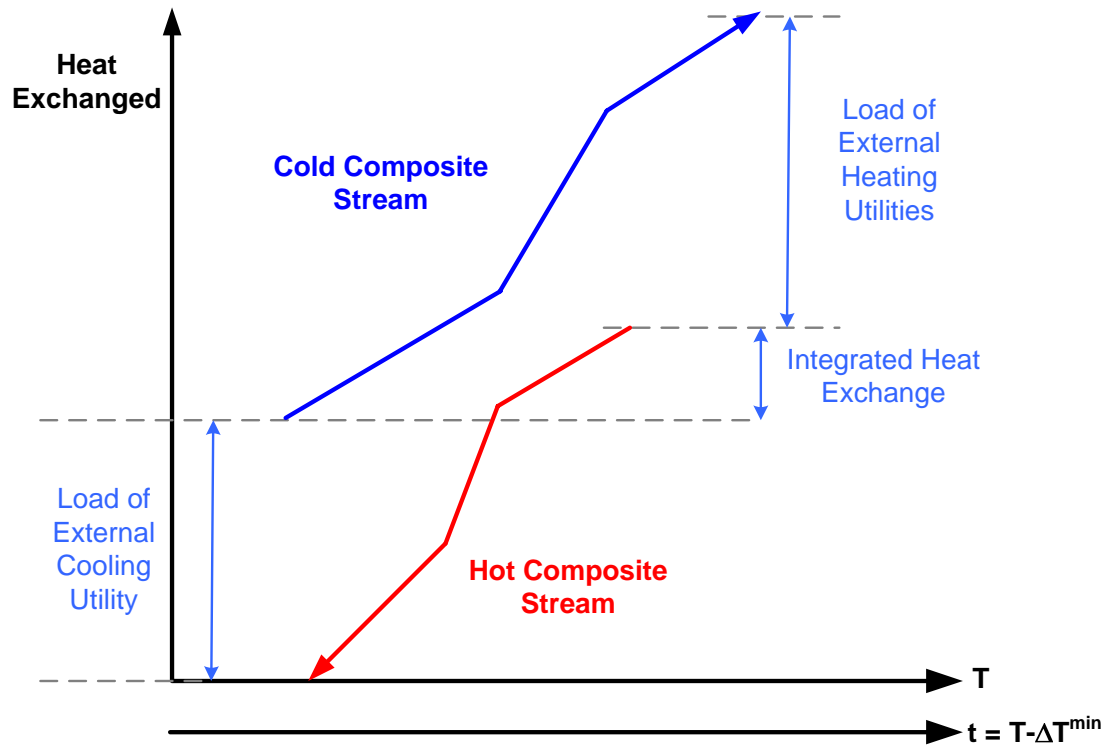
Figure 2.6A: Composite heat diagram with partial integration
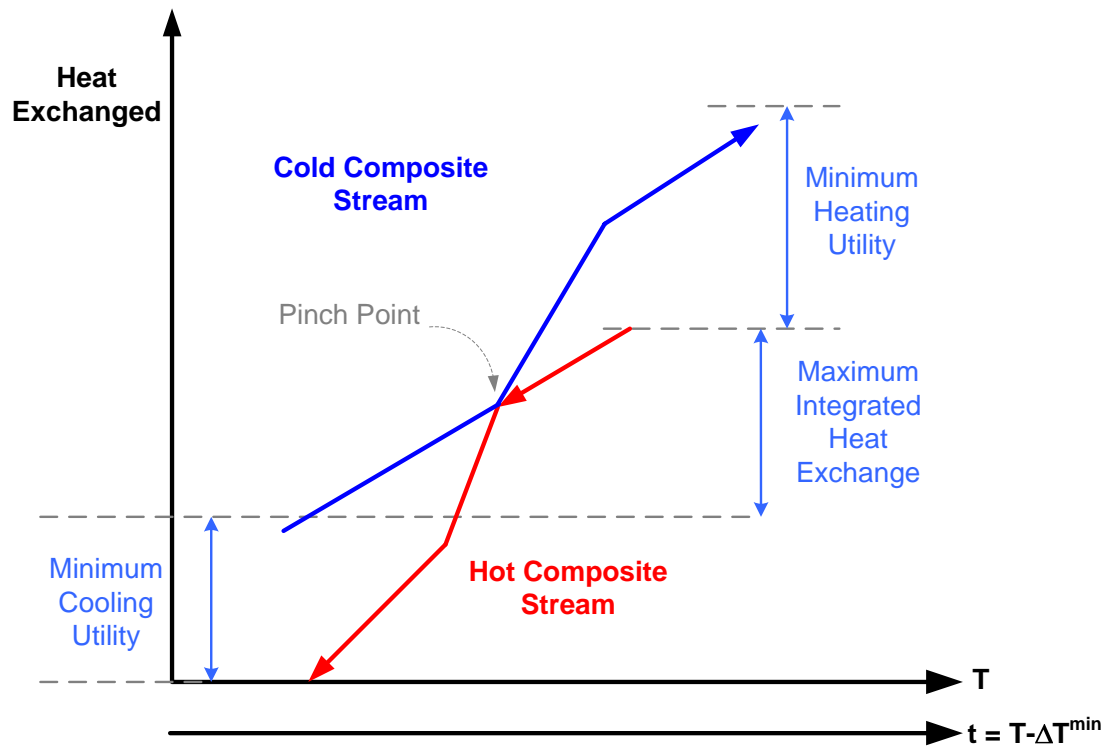
Figure 2.6B: Thermal pinch diagram – maximum heat integration

It should be recognized that the identification of minimum external utilities does not necessarily translate to minimum total cost. The required amount of external utilities can be lowered by decreasing ΔT_{min}; however, the decrease in driving force translates to larger heat exchanger area and in turn higher exchanger unit cost. Thus there is a trade-off between minimum utility requirements and the number/size of the heat exchangers that will need to be implemented. The optimal solution to heat integration problems has also been successfully identified using mathematical optimization methods (Floudas, 1995). Non-Linear Program (NLP) and Linear Program (LP) transshipment models representing superstructures for each possible heat exchanger network were solved to give the optimal number of heat exchangers and minimal external duties (Papoulias and Grossman, 1983); however, these problems suffer from nonconvexities that can result in suboptimal solutions. Reformulation/Linearization techniques (RLTs) have also been used to solve such nonconvex problems (Sherali and Adams, 1999); however, such methods are harder to implement than the pinch analysis methods. Pinch analysis is a powerful tool, which illustrates the cumulative cooling and heating requirements of the process in a single diagram. It is however based on an assumption of ideal thermodynamics and constant heat capacities. To address this limitation, simulated annealing methods have been employed to include detailed thermodynamics in addition to property correlations (Nielsen et al., 1996).

2.5.2. Mass Integration

By the end 1980s, the development of process integration tools was extended beyond just heat integration. El-Halwagi and Manousiouthakis (1989) created new tools

for designing mass exchange networks using the same philosophy as utilized in the thermodynamic analysis of heat exchanger networks. Due to more stringent environmental regulations on the chemical industry, later work focused on the particular sub-problem of water networks (Takama et al., 1980; Wang et al., 1994; and Doyle et al., 1997). The design objective in water-reusing networks is to minimize water consumption by maximizing water reuse. This led to the development of new general design methodologies, specifically the formation of mass pinch analysis and source sink mapping diagrams. This new paradigm is collectively referred to as mass integration (El-Halwagi and Manousiouthakis, 1989; El-Halwagi and Spriggs, 1996; and El-Halwagi, 1997).

Mass integration enables identification of the optimal path for the recovery and allocation of process species or resources by the use of systematic design and analysis tools (El-Halwagi, 2006). Mass integration aims at improving yield, debottlenecking the process, conserving energy and reducing waste in a cost effective manner. In other words it aims at determining achievable performance targets ahead of detailed design by the use of fundamentals such as thermodynamics, transport phenomena and mathematical optimization (Sirinivas, 1993).

Mass integration uses mass-separating agents (MSAs) to remove undesirable materials from waste streams (rich streams). To understand exactly where these MSAs (lean streams) should be used and which streams need to be intercepted is the challenge in synthesizing mass exchange networks (MENs). Mass exchange networks aim at:

- Selecting the mass exchange operation needed
- Choosing the MSA

- Matching MSAs with waste streams

- Deciding the arrangement of mass exchangers and where to split and mix streams

The approach in solving such tasks needs to be very systematic, due to the combinatorial nature of each of the tasks. Several approaches have been used to solve such problems, e.g. enumeration techniques which proved to be complicated (discussed in Section 2.6) and it is not able to guarantee a feasible much less an optimal solution because of all the decisions that are involved (El-Halwagi, 1997). On the other hand, "targeting approaches" simplify the design challenge by identifying performance targets, such as minimizing the cost of MSAs and the number of mass exchange units ahead of design without committing to a MEN configuration. Mass pinch analysis is a graphical approach for analyzing available process MSAs along with mass exchangers from a thermodynamic limitations point of view (El-Halwagi and Manousiouthakis, 1989). By understanding the available MSAs within a process and maximizing their use, the goal of minimizing the cost of external MSAs can be realized. Mass pinch analysis is very similar to the thermal pinch analysis in its approach. The construction of a pinch diagram is as follows:

- Each rich and lean stream is represented by an arrow on a mass exchanged vs. composition diagram.

- The slope of the line corresponds to the flowrate of the stream. The tail and the head of the arrow on the diagram represent the maximum (supply) and minimum (target) compositions of the rich streams, and vice versa for the lean streams,

where the minimum is labeled as supply composition, while the maximum is the target composition.

- Construct the composite rich and lean streams by using the "diagonal rule" of superposition, also known as linear superposition to add up the mass exchanged in regions where overlapping occurs.

The order in which rich streams are stacked is by placing the streams with the lowest target composition first. Lean streams are stacked with the MSA having the lowest supply composition first. On the diagram, the rich composite curve represents the cumulative mass of the pollutant lost by all rich streams. Similarly, the lean composite curve represents the mass of pollutant gained by all process MSAs.

Next, both the rich and lean composite streams are placed on the same graph, and the lean stream is slid down vertically on the graph until the two curves meet. If the lean process stream is to the left of the rich that means that mass can be exchanged between the two streams (see Figure 2.7).
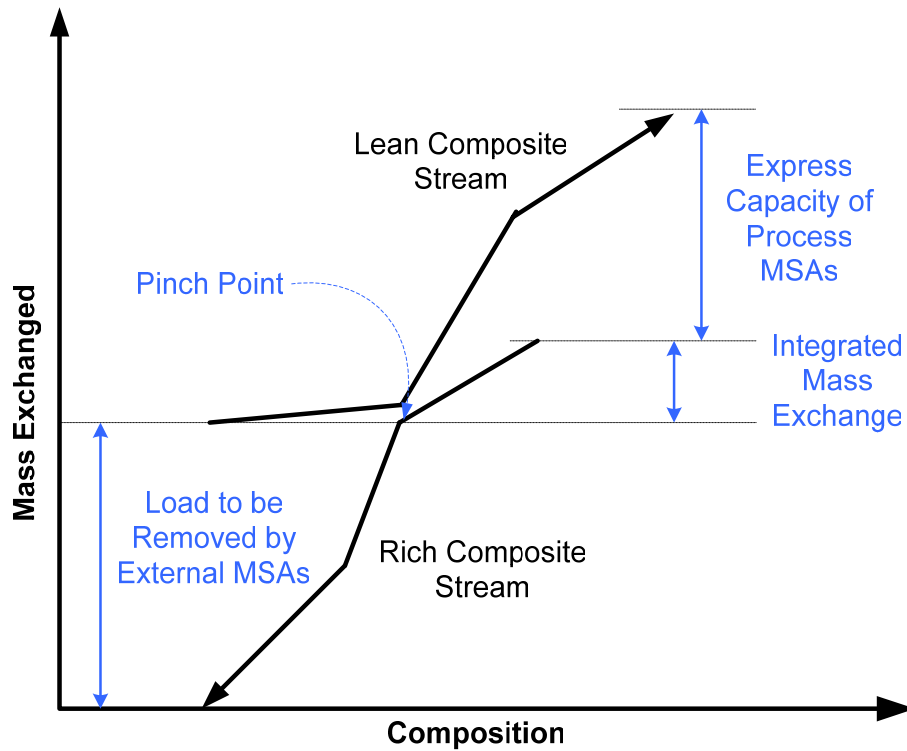
Figure 2.7: Mass pinch diagram.

No external MSA's should be used above the pinch. This point is emphasized because mass exchanged above pinch would require the use of added external MSAs, which translates to higher cost of external MSAs. However, the load below the lean composite stream must be removed using external MSAs; and the vertical overlap between the two composite curves represents the maximum amount of mass that can be exchanged internally, with the use of already available MSAs. This region is labeled integrated mass exchange on Figure 2.7. Anything above the integrated mass exchange is excess capacity of the internal process streams which can be eliminated by lowering the flowrates of those streams or lowering the outlet composition.
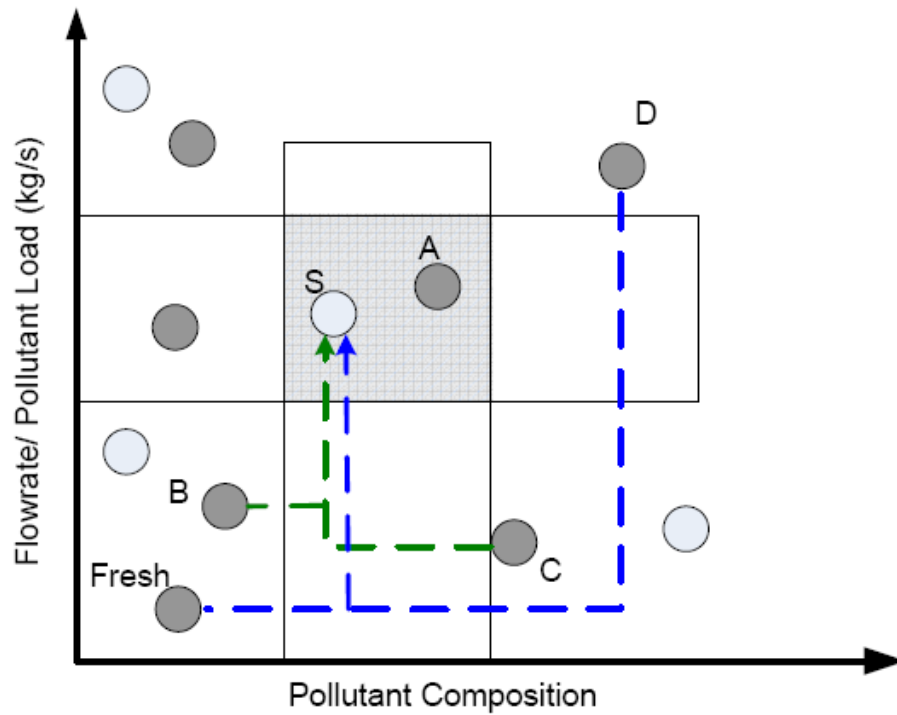
Figure 2.8: Process source-sink mapping diagram

The source-sink mapping diagram is a visualization tool used to determine feasible recycle strategies within a process (El-Halwagi and Spriggs, 1996; El-Halwagi, 2006). The sources correspond to available process waste streams that are available for recycle and whose flowrate and targeted pollutant composition are known; while the sinks are process units that have certain constraints in terms of input flowrate and allowed maximum composition of pollutant species. The diagram is constructed as pollutant load/flowrate vs. composition; with sources and sinks being represented by shaded and hollow circles on the diagram, respectively (see Figure 2.8). The sink flowrate and composition constraints are represented by the horizontal and vertical bands, respectively; with the shared spaces representing area of acceptable loads and

compositions available for recycle. For instance, source A (Figure 2.8) can be directly recycled into sink S (El-Halwagi, 2006). The location of a mixture of sources B and C can be determined using lever arm analysis, and if the mixture is located within the band then it is a feasible recycle stream into sink S as well. Source D, located above the band can be rerouted to sink S by mixing it with a fresh source in order to lower the load of source D and minimize the use of fresh source, again following lever-arm rules and material balance equations. A general rule is sources with the shortest arm to the sink should be recycled first.

There is a rich volume of information available in literature that covers the development and uses of energy and mass integration tools (Cerda et al., 1983; Linhoff and Hindmarsh, 1983; Gundersen and Ness, 1988; Douglas, 1988; Shenoy, 1995; El-Halwagi, 1997; Dunn and El-Halwagi, 2003; Dunn and El-Halwagi, 2003; and Rossiter, 2004).

The previous sections gave an overview of currently utilized process synthesis and design methods. The next sections will concentrate on the methodology behind molecular design algorithms; the importance of property models to design, and how it challenges developed methodologies.

2.6. Computer Aided Molecular Design Methods (CAMD)

By definition, a CAMD problem is (Brignole and Cismondi, 2003): Given a set of building blocks and a specified set of target properties, determine the molecule or molecular structure that matches these properties.

A class of CAMD software for chemical synthesis developed by Molecular Knowledge Systems Inc focuses on three major steps in the formulation of molecules; it illustrates the general methodology behind most CAMD methods (Joback, 2006):

1. *Identifying target physical property constraints.* Translate performance requirements in terms of constraints on properties: e.g. if a certain chemical must be liquid at certain conditions, it should be translated in terms of constraints on melting and boiling temperature.

2. *Automatically generating molecular structures.* CAMD software is used to generate molecular structures based on the groups of molecular building blocks given as the input. Types of molecules being generated can be controlled (e.g. alcohols could be removed by simply excluding –OH group from the pool of building blocks, same can be for amines, amides, chlorines etc.)

3. *Estimating physical properties.* Using structural groups as building blocks enables the use of group contribution estimation techniques to predict the properties of all generated formulations. (More on group contribution in Section 3.2.1).

In property prediction, the component's structural information is used to predict its properties; therefore, the identity of the formulation is required as input to the algorithm. The solutions generated by design algorithms that employ these methods are limited to the list of "pre-selected" components, which can lead to "sub-optimal" designs. CAMD is able to avoid such tribulations by solving the reverse of the property prediction

problem. It uses available property models to formulate the design problem in terms of target values for the identifiable set of properties. The property constraints are used as input into its algorithm, then it determines candidates of molecules (or mixture of molecules) that match the specified property targets values without limiting the search space (Eden, 2003). Hence, with the problem well defined, in terms of properties, CAMD methods are able to design novel formulations that otherwise might not of have been part of the available database.

A rich volume of investigative research regarding CAMD is available in literature and can be grouped into three main categories: mathematical programming, stochastic optimization, and enumeration techniques (Harper, 2000; Harper and Gani, 2000):

Mathematical programming solves the CAMD problem as an optimization problem where the property constraints are used as mathematical bounds and the performance requirements are defined by an objective function. Solutions techniques to such optimization problems include Mixed Integer Non-linear Programming (MINLP) solution methods. Although widely used and proven to be effective, MINLP methods suffer from a large computational load and it lacks the guarantee of finding a globally optimal solution. (Odele and Macchietto, 1993; Vaidyanathan and El-Halwagi, 1994; Duvedi and Acheni, 1996; Pistikopoulos and Stefanis, 1998).

Stochastic optimization, where the solution alternatives are based on the successive pseudo-random generation method. Like the previously mentioned approach, this method aims at finding the optimal value for the objective function, but the technique it uses varies. One important aspect is that stochastic

optimization methods do not require any gradient information, giving it the freedom to specify discontinuous properties as design targets. There are two forms of stochastic optimization: (1) uses the Simulated Annealing (SA) method and (2) uses Genetic Algorithm, which is based on Darwin's evolutionary theory. The Simulated Annealing technique requires the formulation of the problem in form of states and moves. States refer to an instance of design parameters and possible parameter modifications are the moves. The algorithm runs as an iterative process where moves generate new states, according a set of perturbation probabilities (Marcoulaki and Kokossis, 1998). The generated parameters (states) are tested against previous to satisfy a probability criterion. The advantage of using SA is its ability to easily deal with highly non-linear models (e.g. predictive property models) and large numbers of decision variables (e.g. numerous alternative molecular structures). In the second approach, populations of potential solutions are obtained from the previous populations based on "survival of the fittest"; it also takes into account how attributes are passed from "parent" to "offspring", i.e. from one solution population to the next population. Because of the stochastic nature, both approaches are capable of handling non-linear models, although as the problem complexity increases, the genetic algorithm approach reports limitations in terms of computational time. (Holland, 1975; Venkatasubramanian et al., 1994; Marcoulaki and Kokossis, 1998; Ourique and Telles, 1998).

Enumeration techniques aim at satisfying the feasibility and property constraints by first generating molecules using a combinatorial approach and then test against

the specifications, where molecules that fail to satisfy the constraints are eliminated. Thus, the generation and screening of molecules are performed separately. As with the stochastic methods, no gradient information is needed; however, a disadvantage of this approach is that solving a simple enumeration of a CAMD problem can lead to combinatorial explosion. Meaning that even with today's fast computers, excessive computational time is needed (Gani et al., 1991; Pretel et al., 1994; Joback and Stephanopolous, 1995; Constantinou et al., 1996; Friedler et al., 1998).

Another method labeled as "generate and test approach" was introduced by Harper (2000). It is an approach where only feasible formulations are generated from molecular building blocks using a rule based combinatorial approach. The difference between this and the enumeration techniques mentioned previously, is that this method uses a multi-level CAMD approach that controls the generation and testing of molecules. Harper (2000) proved that a solution algorithm of a "generate and test" type can be successful without suffering from "combinatorial explosion", even when considering detailed molecular models. The employed method consists of four levels (see Figure 2.9). Each level has a generation and a screening step. In the generation step the molecular structures are created while the properties of the generated compound are predicted and compared against the design specifications in the screening step. The first two levels operate on molecular descriptions based on groups while the latter two rely on atomic representations. The outline for the individual levels has the following characteristics (Harper, 2000):
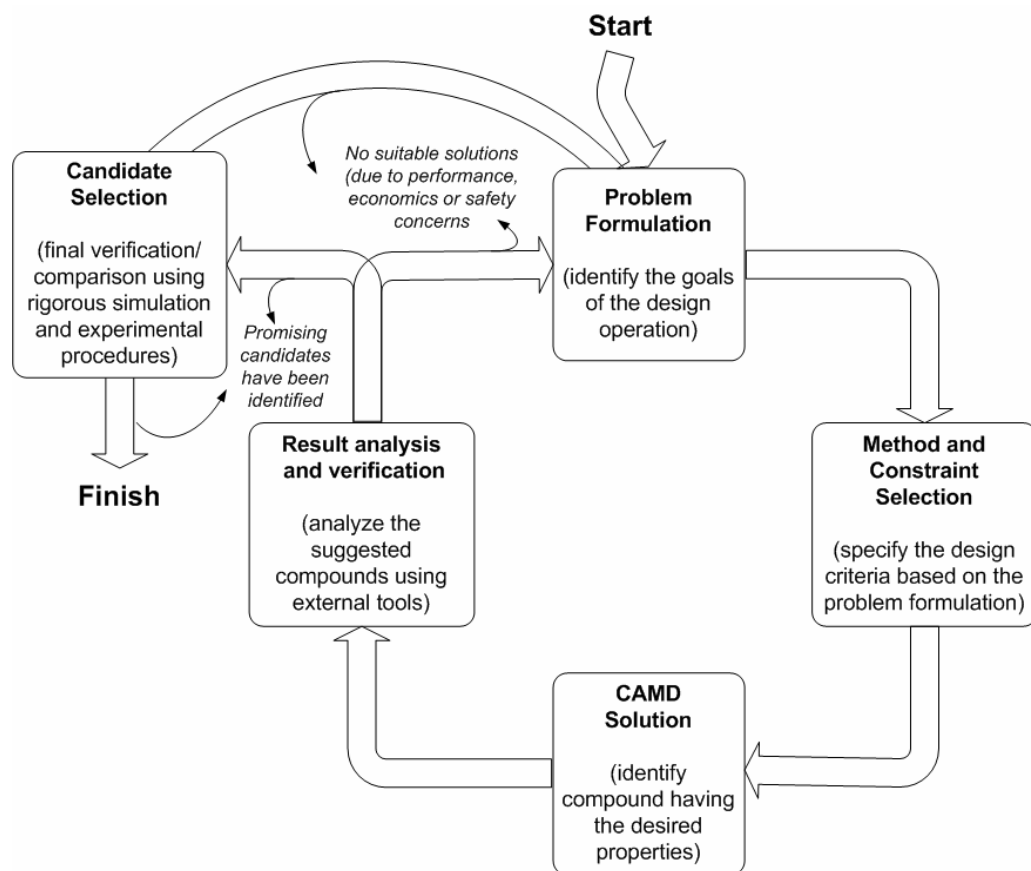
Figure 2.9: Flow diagram of the multi-level CAMD framework (Harper, 2000)

Level 1   In the first level, a group contribution approach (generation of group vectors) is used with group contribution property prediction methods. Group vectors are generated from the set of building blocks identified in the pre-design step. The generation step does not suffer from the so-called "combinatorial explosion" as it is controlled by rules regarding the feasibility of a compound consisting of a given set of groups (Gani et al., 1991). Only the candidate molecules fulfilling all the requirements are allowed to proceed onto the next level.

Level 2   At the second level, corrective terms to the property predictions are introduced. These terms are based on identifying substructures in molecules. At this level molecular structures are generated using the output from the first level as a starting point and the second order groups are identified using a pattern matching algorithm. The generation step at this level is a tree building process where all the possible legal combinations of the groups in each group vector are generated.

Level 3   In the third level the molecular structures are converted into an atomic representation by expanding the group representations. The conversion into an atomic representation enables the use of molecular encoding techniques (Harper & Gani, 1999). The use of molecular encoding techniques makes it possible to re-describe the candidate compounds using other group contribution schemes thereby further broadening the range of properties that can be estimated as well as giving the opportunity to estimate the same properties using different methods for comparison.

Level 4   In the fourth level the atomic representations from level three are further refined to 3-dimensional representations. This conversion can create further isomer variations and enables the use of molecular modeling techniques as well as creating molecules ready for structural database searches in the post-design step.

This methodology has been implemented by the Computer Aided Process Engineering Center (CAPEC) in their ProCAMD software as part of the Integrated Computer Aided System (ICAS) (CAPEC, 2006).

Regardless of the method of choice, stating the objective (Pre-Design Phase) of the problem is a prerequisite for solving any CAMD problem. Here the goals/targets of the design (numerical property constraints) and a selection of molecular building blocks are used as input into the CAMD algorithm (see Figure 2.10). The design phase includes the generation of molecular formulations and testing their ability to satisfy the property constraints placed on the problem. Next, the post-design phase involves using other prediction methods, database sources, engineering insight, and if possible, simulation in order to screen and rank the designed compound(s) based on suitability and capabilities (e.g. environmental impact, health and safety aspects, production cost or availability).
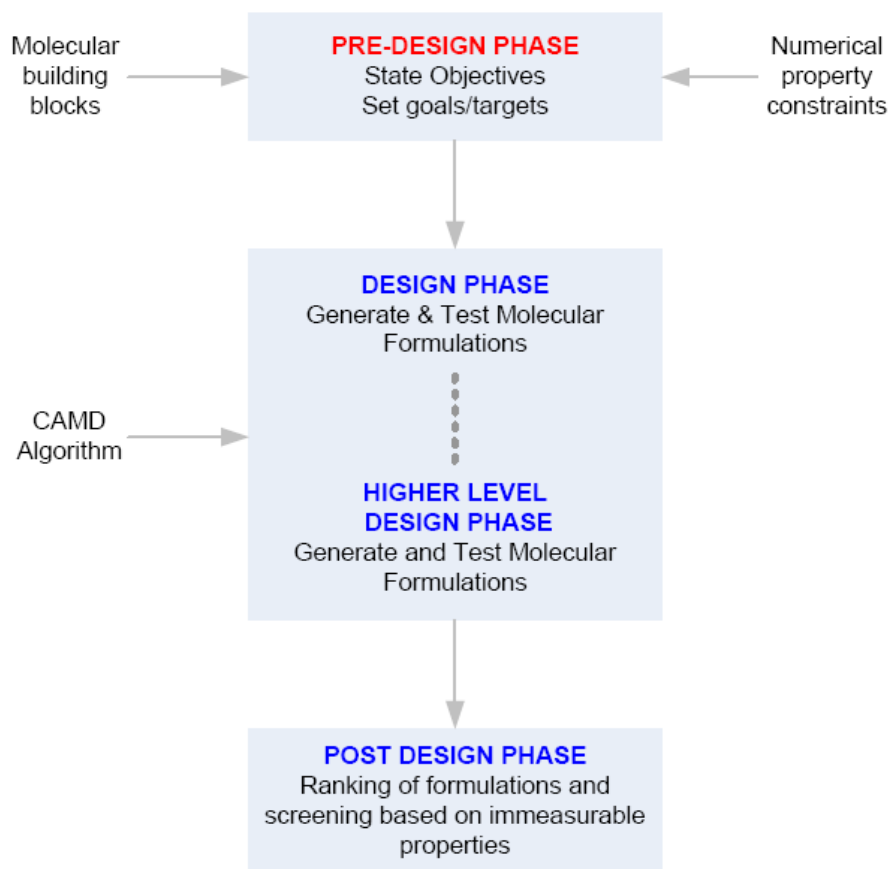
Figure 2.10: Formulation and solution of a CAMD problem

2.7. Roles of Property Models & Reverse Problem Formulation

In design, process or molecular, the formulation of the problem will always include a property model (see equations 2.1-2.6). Property models play an important role in design and it is the non-linear nature of these property models that often lead to complications within design calculations. The following section takes a closer look at property models.

The widespread availability of powerful computers and user-friendly software has made process modeling an integral part of chemical engineering practice. An essential requirement of successful process modeling is the availability of thermo-physical

41

property models that are accurate, reliable, and computationally efficient over a very large range of temperatures, pressures, and compositions. However, property models can be more than just generators of property values. They can provide insight and guidance to the efficient solution of process engineering problems. Gani and O'Connell (2001) describe property models as having three distinctive roles in computer-aided product and process engineering (CAPE).

The first is a *service role* where the property models are used to provide the needed property values when prompted by the process model. The second is a *service plus advice role*. In addition to providing property information, the models advise on feasibility. The third role, considered the most comprehensive in CAPE problems is the *service/advice/solve role*, where in addition to the previous roles the property models take part in the solution of design problems.

In the advice role, the choice of property model will dictate the resulting properties, and usually the property model complexity can lead to observed non-linear behavior of the process model equations, causing difficulties in achieving convergence. The use of an inappropriate property model and/or property parameters can lead to erroneous numerical results that cause further complications by causing bottlenecking, over-sizing and even resulting in wrong process configurations (Whiting and Xin, 1999).

In CAMD, the property model plays first an advice role in the formulation of the design problem, by advising on which properties to target. Once the CAMD algorithm has generated the candidate formulations, it prompts the property model in the service role to verify that the properties of the designed molecules satisfy the targets. Notice that without the service role, the search space for candidate molecules would be too great to

handle, regardless of the particular method of solution for the CAMD problem. Thus, the advice role of property models helps to narrow down the search space for solutions to the design problem.

Eden (2003) described how the various roles of property (constitutive) models fit into the overall solution of the design problem. In Figure 2.11, the process model provides intensive variables such as temperature, pressure and composition to the property model; and requests property values during the solution step. Here the property models are in the service role. The process model can act as the basis for a process simulator, where the effects of changing various parameters can be analyzed. Furthermore, the simulator can be connected to a process synthesis/design algorithm in order to update the process parameters based on the results from each simulation. Thus, the operating conditions that yield the desired process performance are identified.
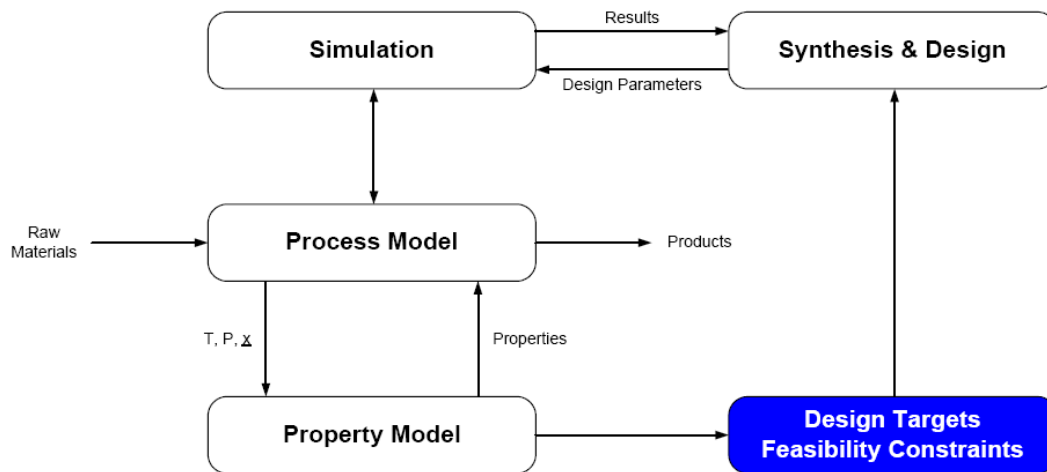
Figure 2.11: Property models presented in various roles (Eden, 2003)

Property models now advise the synthesis/design algorithm on feasible and optimal operation/process conditions, thereby narrowing down the search space for the process design problem.

The ability to provide design targets and feasibility constraints, enables the property model to be included as part of the solution routine. In the service/advice/solve role, the property models are decoupled from the process model and solved separately. Figure 2.12 shows how the information to and from the process model is reversed, i.e. the process model is solved in terms of constitutive variables and the property model is called upon to determine the corresponding intensive properties.
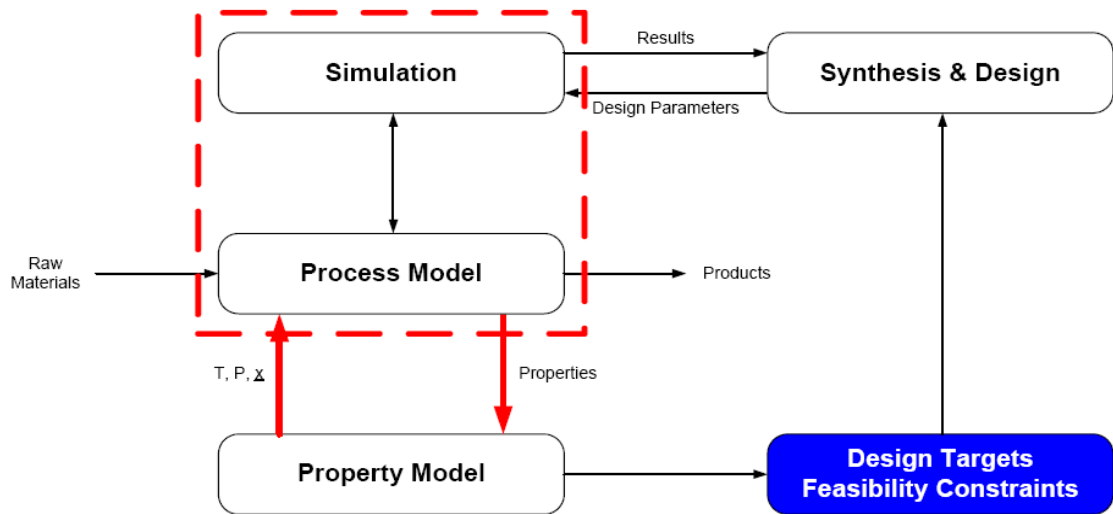
Figure 2.12: Property model in service, advice and solve role.

Decoupling of the property/constitutive model from the process model is key to lowering the dimensionality of design problems because as previously mentioned the

nonlinearity of the process model is generally attributed to the complexity of the constitutive model. Now different constitutive models can be used at different stages of solving the design problem, the selection of the model depending on the data given by the process model. To fully take advantage of property models as a powerful tool in the solve role, requires a methodology capable of handling multiple property models.

Recently, Gani and Pisitikopoulos (2002) and Eden et al. (2002) proposed the solution of process as well as product design problems as a series of reverse problems (as the approach is part of the technique developed in this thesis, more details on this methodology in Section 3.1). Based on understanding the roles of property models discussed above, Gani and Pistikopoulos (2000) and Eden et al. (2002) have shown that applying the reverse problem formulation to process or product design does not require the use of property models in the process model equations since the unknown design targets are functions of the target properties. This means that the target properties can be determined from the solution of the reverse simulation problem by solving a set of linear equations (in most cases) and from these, the design targets are calculated. As long as these targets are matched, any number of property models may be used at the various stages of the solution step. The advantage of using reverse problem formulation is that the problem complexity is significantly reduced (by decoupling the constitutive property models from the design step) without sacrificing solution accuracy.

The design tools developed in this thesis have recognized the importance of reverse problem formulations in design, and to property integration.

2.8. Property Integration – Motivation, Need & Limitations

While integrating process and product design problems presents a logical approach to optimality, there is still the problem of a common platform that allows for such incorporation. As Figure 2.13 shows, current process design algorithms take in overall objectives along with available data (e.g. pre-selected list of possible components or mixture data) and through the various methods discussed in Section 2.4, generate a solution(s) to meet the performance requirements. In molecular design, advances in the area of property prediction, specially using GCM, proved valuable to the success of CAMD tools such as the multi-level generate and test approach (Harper, 2000). In product/process development, the pre-design needs are described in terms of the products' performance or properties. Process unit performance is measured as a function of properties or "functionalities" (e.g. condensers dependence on vapor pressure, reactors on heat capacities etc.). Moreover, in molecular design, the designed formulations' properties are later checked to make sure that they satisfy these needs. Therefore, it makes sense to use a property-based platform to link the two previously decoupled design problems.

The need for systematic methodologies based on properties was recognized by El-Halwagi and collaborators. They introduced the concept of *Property integration;* and it is defined as "a functionality-based holistic approach for the allocation and manipulation of streams and processing units, which is based on functionality tracking, adjustment, and matching of functionalities throughout the process" (Shelley and El-Halwagi, 2000; El-Halwagi et al., 2004 and Eden, 2003). Introduction of the property integration framework enabled for representation of process and products from a properties perspective.

The need for solving process design problems in terms of properties, involves addressing the challenge that while chemical components are conserved, properties are not conserved entities. Another difference between component-based "chemo-centric" approaches and property-based approaches is that the mixing of components is linear, while the mixing of properties is not necessarily linear.
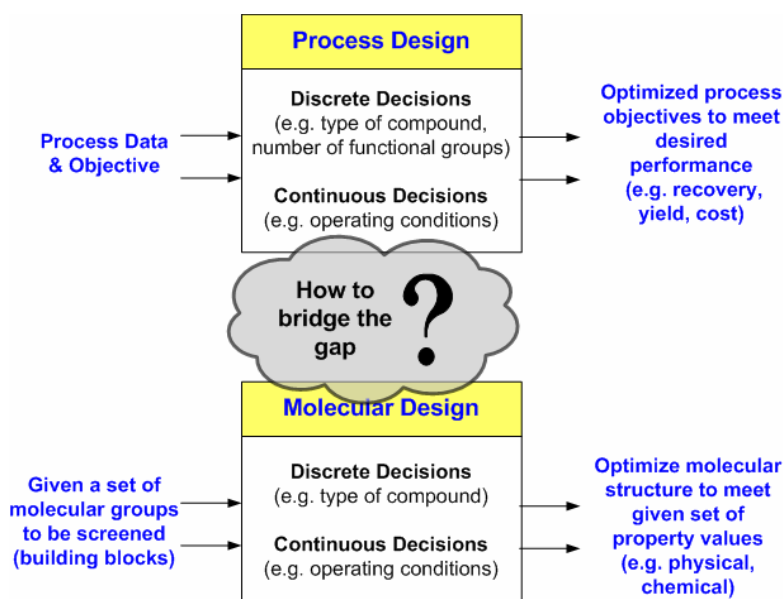
Figure 2.13: Conventional approach for process and molecular design problems.

To overcome this limitation, Shelley and El-Halwagi (2000) introduced conserved quantities called clusters. The clustering concept utilizes property operators, defined as functionalities of the non-conserved raw physical properties. The operators are formulated to posses simple linear mixing rules, even though the properties themselves might be nonlinear (e.g. the reciprocal value of the density of a mixture of two streams is the summation of the reciprocal densities for each stream). Property clusters are the basis for the developed property integration framework that allows for representation of

process and products from a properties perspective. The technique constitutes the basis of the methods developed in this research; hence, more in depth information regarding the property integration framework will be covered in chapter 3.

Utilizing the property clustering concept as the basis, Eden (2003) introduced a systematic property integration framework for the formulation and solution of property-driven process design problems. Although analyzed in chapter 3 brief aspects of this technique are highlighted here:

- Using a reformulation strategy based on decoupling the constitutive equations from the balance and constraint equations, the traditional forward process design problem is converted into two reverse problems. The first solves the balance and constraint equations in terms of the constitutive (property) variables to identify the design targets. The second problem solves the constitutive equations to reveal the unit operating conditions and/or components that match the design targets set by the first reverse problem.

- By solving the constitutive equations separate from the balance and constraint equations, the method provides an important feature - a framework capable of handling multiple property models as needed to describe entire processes.

- For problems, that can be described using three properties, the problem and solution are visualized on a ternary diagram. Thus, the ability to use visualization insight to identify the desired properties needed to satisfy optimum process performance targets without committing to any component during the solution step (Eden, 2003).

To overcome the limitation of using only three properties and to increase the application range, Qin et al. (2004) developed an algebraic approach for property integration. Process constraints and stream characterization are described using bounds on intensive properties and flows. The specific mathematical structure of the set of operator constraints is used to develop a constraint-reduction algorithm, which provides rigorous bounds on the feasibility region. This algebraic approach allows for the expansion of the design problem to include any number of properties, and formulates the problem as a set of equality and inequality equations. By lowering the general dimensionality of the problem, the algebraic approach provides an easier solution to the design problem.

Kazantzi et al. (2005) further extended the application of the clustering concept into a targeting procedure for material reuse in property based applications. They developed new property-based pinch analysis and visualization techniques. For systems characterized by one key property the developed technique determines rigorous targets for minimum fresh usage, maximum recycle, and minimum waste discharge (Kazantzi et al., 2004 a,b). This is a generalization of the conventional material reuse/recycle pinch diagram (Section 2.8) which can be modified to include property operators to track the properties as streams are segregated, mixed, and recycled. The graphical technique provides visualization insights for targeting and network synthesis (Foo et al., 2006).

For solvent design Hostrup et al. (1990) and Linke and Kokossis (2002) have recently reported problems with simultaneous solution approaches for process-product design. Mathematical programming approaches incorporating product and process design, while attractive, needed first to overcome a problem in handling property models.

As pointed out by Gani (2001), the property models for product design may not be suitable for process design and vice versa. In addition, once a property model is selected for inclusion into the process model, the application range in terms of additional new mixtures (generated by the product design steps) is restricted, since either the model parameters may be unavailable or the property model may not be suitable for the generated molecules. In view of the fact that in mathematical programming techniques the changing of model equations (included as equality constraints) will cause discontinuities in the solution trajectory, it may become extremely difficult to achieve convergence if multiple versions of models for the same properties were to be used (Gani et al., 2003).

Eden (2003), proposed the simultaneous consideration of both process and molecular design problems as illustrated in Figure 2.14. The molecular building blocks are used as input into the molecular design algorithm; while the process algorithm takes in the desired performance goals. The result of the simultaneous consideration is the identification of design variables needed to facilitate the process performance targets and the generation of molecular formulations that aim at satisfying the property constraints determined from solution of the process design algorithm. Eden (2003) succeeded in developing a systematic method for the solution of process design problems, where the objectives are driven by properties. The methodology is also capable of identifying property values that correspond to optimum process performance without committing to any components by ahead of design. Thus, Eden (2003) developed the process end of this integrated approach. The efforts of the presented thesis explores the problem of identifying a property based molecular design methodology capable of incorporating the

clustering techniques developed by Eden (2003) in order to bridge the gap between the two design problems.
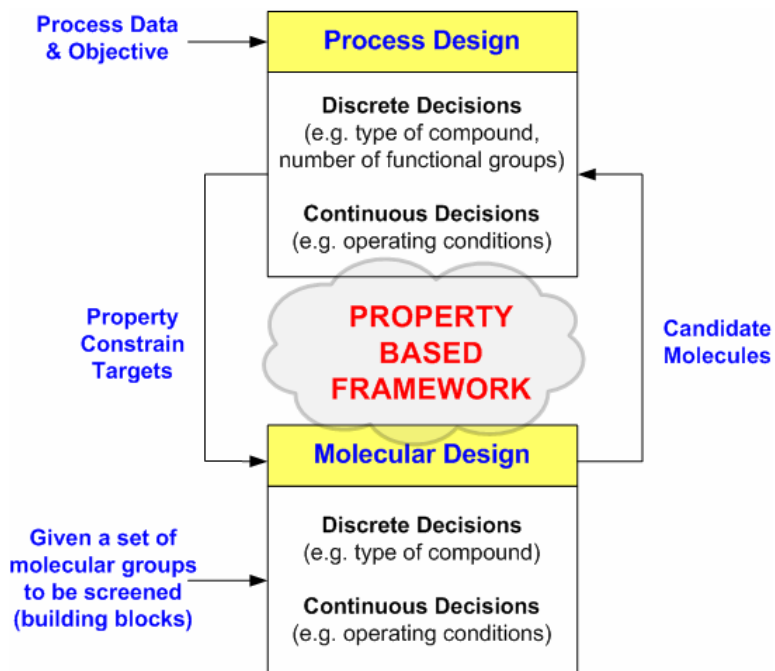
Figure 2.14: New approach to process and molecular design problems

Detailed coverage regarding property clusters and the property integration framework is presented in Section 3.1. These techniques constitute the foundation for the generated molecular design approach presented in Section 3.2.

2.9. Property Prediction and Group Contribution

Almost all CAMD algorithms rely on the ability to predict pure component and mixture properties for the analysis and design of formulations. In addition, the need for reliable and accurate property estimation methods is critical to the solution of various

simulation problems where convergence is often related to failures in the reliability of predicted physical and thermodynamic properties (Constantinou and Gani, 1994). Most property estimation methods used in CAMD techniques are based on Group Contribution Methods (GCM), where the properties of a compound are expressed in terms of a function of the number of occurrences of predefined groups in the molecule (Harper, 2000). The group contribution method is totally predictive, meaning that, as long as the structure can be fully described with the groups, the properties of the structure are immediately available. The method can be used to synthesize new structures easily as the evaluation of the properties of a structure is straightforward, given the models and the group contributions (d'Anterroches and Gani, 2005).

Group contribution based design methodologies are built on the following general premise: the structure is composed of groups and the targets are properties. The formulation of a group contribution based problem is defined as identifying structures that possess target properties (e.g. molecular weight, melting temperature etc.) while matching structural constraints (e.g. no cyclical groups, no alcohols, etc.). The goal is to generate the molecular structures that match the target properties within the structural constraints.

Group Contribution Methods (GCM) allow for the prediction of pure component physical properties from structural information. Property prediction of pure compounds was initially estimated as a summation of the contributions of simple first-order groups that occur in the molecular structure (Lydersen, 1955; Joback and Reid, 1983; Reid et al., 1987; Lyman et al., 1990; and Hovarth, 1992). The advantage of using such methods is it provides quick estimates without requiring substantial computational resources; however,

the molecular structure is often oversimplified, making isomers indistinguishable. To overcome such issues, Jalowka and Daubert (1986) have employed another class of group contribution involving particulate grouping of atoms in the presence of other atoms. Despite the increase in accuracy, this method is complex and in order to produce the desired properties it requires the use of other determined properties (Reid et al., 1987).

Other methods have proposed correlating critical properties and normal boiling point to the number of carbon atoms in the molecule of homologous series (e.g. alkanes and alkanols) (Kreglenski and Zwolinski, 1961; Tsonopoulos, 1987; and Teja et al., 1990). An increase of accuracy was observed, however the application range has been questioned (Tsonopoulos and Tan, 1993).

Constantinou et al. (1993, 1995) proposed an additive method where the molecular structure of a compound is viewed as a hybrid of a number of alternative formal arrangements of valence electrons (conjugate forms), and the property of a compound is a linear combination of the contributions. The systematic inclusion of more molecular information allowed this method to substantially increase the accuracy of the predicted property and capture the difference among isomers; however, a shortcoming of this method is that it requires a symbolic computing environment for the generation and enumeration of conjugate forms.

In 1990, Gani et al. reported that group contribution based computational tools need to accommodate two separate molecular structure descriptions: one for the prediction of properties for pure compounds and another for the mixture property estimation. To overcome this limitation, Constantinou and Gani (1994) proposed the use of first order group contribution; and defined them as the set of groups commonly used

for the estimation of mixture properties, where each group has a single contribution, independent of the type of compound involved (e.g. acyclic or cyclic). The shortcoming here is that this method cannot differentiate between isomers.

In their work, Constantinou and Gani (1994) also included a two-level approach to property estimation. The basic level has contribution from first-order groups for mixture properties, and the next level has a set of second-order groups that have the first-order groups as building blocks. The role of the second order groups is to consider, to some extent, the proximity effects and to distinguish amongst isomers. Despite the advantages of second order GC methods, the application range is limited by the relatively simple compounds that make-up its small data bank (more details regarding available property models and groups for first-order estimation are included in Section 3.2.1)

Marrero and Gani (2001) developed a GCM that performs estimation in three levels. The first level is made-up of simple groups that describe a wide range of organic compounds; however, it still cannot distinguish between isomers. The second level involves groups that permit a better description of proximity effects and differentiation among isomers. The third level has groups that provide more structural information about the molecular fragments of compounds whose first and second level description is not sufficient. This level allows for the estimation of complex heterocyclic and large (C $\leq$ 60) polyfunctional acyclic compounds. The following is the full GC property-estimation model that includes 1st, 2nd and 3rd order groups:

$$\underbrace{f(X)}_{\substack{\text{Property} \\ \text{Function}}} = \underbrace{\overbrace{\sum_i N_i C_i}^{1^{st}\ \text{order}} + \overbrace{\sum_j M_j D_j}^{2^{nd}\ \text{order}} + \overbrace{\sum_k O_k E_k}^{3^{rd}\ \text{order}}}_{\text{Group Contribution Terms}}$$

(2.10)

C_i is the first-order group type i which occurs N_i times, D_j is the contribution of the second-order group type j which occurs M_j times and E_k is the third-order group type k which occurs O_k times.

Any applications of group contribution rely on availability of groups to describe the structure as well as tables giving the contributions of each group (Franklin, 1949). The group-contribution property data has been developed from regression using a large data bank of more than 2000 compounds collected at the Computer Aided Process and Engineering Center at Technical University of Denmark (CAPEC-DTU). Properties that are predicted using GCM (e.g. critical properties, boiling point temperature etc.) are referred to as primary properties, and all other properties (e.g. density, viscosity, vapor pressure, heat of vaporization etc.) are classified by Jaksland (1996) as secondary properties, usually predicted as functions of primary or critical properties (Marrero and Gani, 2001).

2.10. Design of Experiments

Industrial experimenters that deal with formulation of mixtures are often forced to deal with a rising number of variations in data samples, usually arising from external influences (e.g. raw materials, environmental condition or human operating error). To dampen these effects, some try to run large-scale experiments; however, they are too expensive and time-consuming. Others hope to isolate the underlying cause of the variations by going back and forth changing and then re-testing one parameter at a time; however this one-factor at a time approach (OFAT) is not able to provide any insight on

the interaction of different factors (variables). In addition, OFAT is more expensive than large scale experiments.

A two-leveled factorial design (TLFD) method was developed as a statistical strategy that simultaneously adjusts all factors at two levels, i.e. the low and high values for each factor. Using only two levels helps in limiting the number of needed experiments, however more levels can be added if requested, e.g. the mean (midpoint) value of the factor can be included to increase resolution. Box et al. (1978) and Cornell (1990) discussed the basics of TLFD, where as Anderson and Whitcomb (1996) extended the application to chemical engineering problems.

The advancement in technology brought about the inclusion of TLFD strategies into advanced software tools, e.g. Design Expert (Stat-Ease, 1999). Whitcomb (1999) described the general methodology for identifying the optimal mixture Design of Experiments (DOE) as follows:

1. Specify the polynomial order, i.e. first, second, third or beyond, that is needed to model the response.

2. Generate a "candidate set" with more than enough points to fit the specified model

3. Select the minimum number of points, from the candidate set, needed to fit the model.
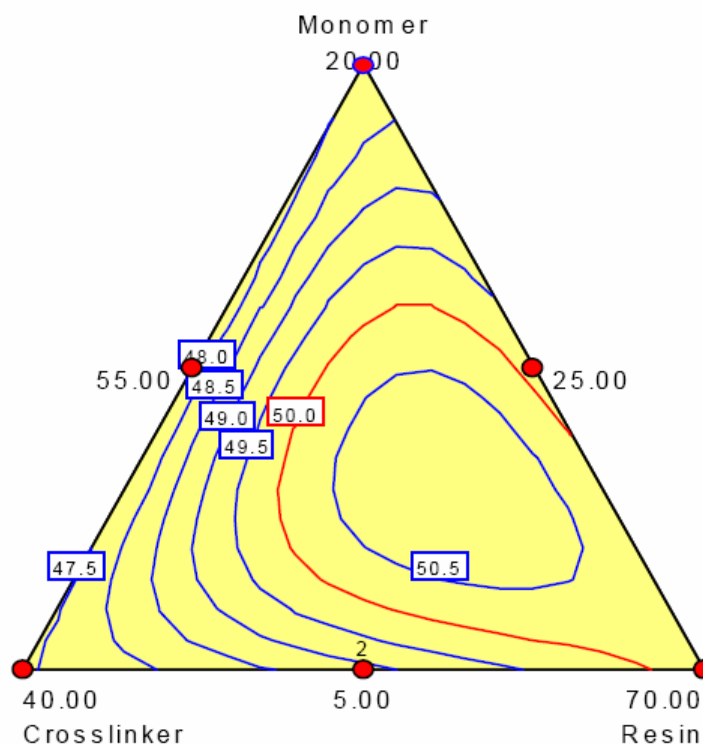
Figure 2.15: Response surface plot

The algorithm is fed the factor constraints along with the specified polynomial order. After statistical analysis, the algorithm yields a sub-set of experiments needed to provide maximum information using the minimum number of experiments. Once the experiments are carried out and the response values measured, a model is generated for each response. Often the result of each design is represented by ternary diagrams that are plotted according to the generated response models. An example of these contours plots are shown in Figure 2.15; and are often called Response Surface Maps. The diagrams are used to identify optimum levels for each factor. The best formulation may be determined without having to prepare or test it. Next the identified solutions should be verified by performing confirmation runs, since the solution was identified based on predictive models.

The advantages of Design of Experiments (DOE) is that factors and/or processes can be changed independently so that main effects can be determined with fewer runs, saving both time and money.

2.11. Ternary Diagrams for Visualization

A ternary diagram uses an equilateral triangle to graphically depict the relationship among three data values which sum to a constant value. It graphically depicts the ratios of three proportions. Geologists use ternary diagrams for a variety of purposes: identification and classification of sedimentary, igneous, and metamorphic rocks. Ternary phase diagrams are used in chemistry to gain insight into the miscibility of a three component system (e.g. ethanol, water and phenol). In this representation the effects of properties like temperature and pressure on component miscibility at various compositions can be visualized. The ternary diagram is read counter clockwise.

As an illustration, four different ternary mixtures are depicted in Figure 2.16. The composition for each of these points is shown below.
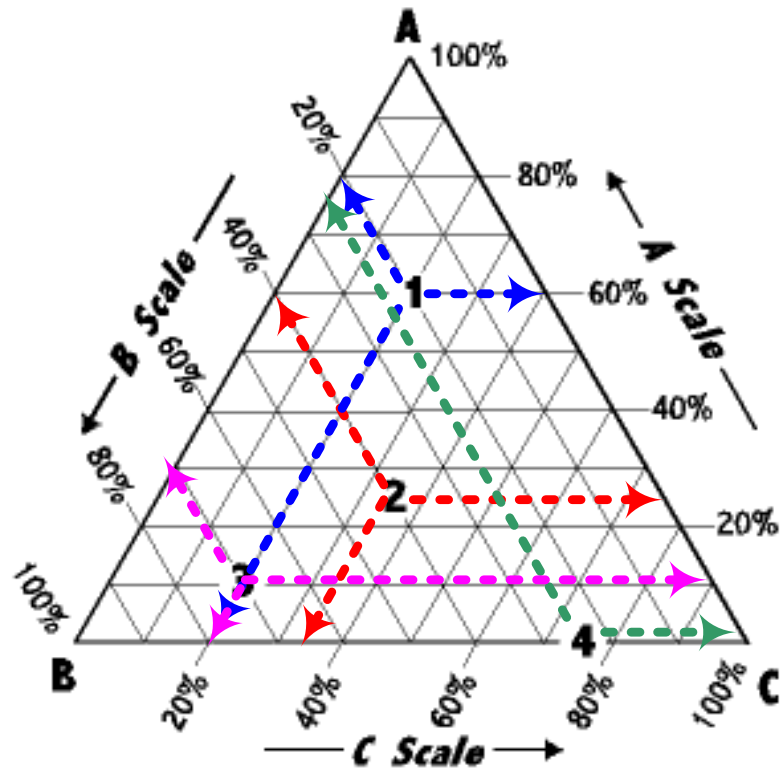
Figure 2.16: Generic ternary diagram.

1. 60% A | 20% B | 20% C = 100%

2. 25% A | 40% B | 35% C = 100%

3. 10% A | 70% B | 20% C = 100%

4. 0.0% A | 25% B | 75% C = 100%

Constructing a ternary diagram can be a tedious and time-consuming process if a plot program e.g. Ternary Plot, is not available. However, a conversion methodology has been developed to plot diagrams using Cartesian coordinates. The methodology is discussed in detail in Section 3.1.4.

2.12. Summary

From the review provided in this chapter it should be evident that process and product design problems are intertwined. Although solving product or molecular design problem separately has its direct benefits (e.g. the design problems are less complex), nevertheless the overall objective in process/product synthesis/design is not just to find any chemical or mixture that satisfies the described objectives. The goal is to achieve an optimal design which addresses cost of raw materials, operation, efficiency and environmental impact.

The targeting approaches used in heat and mass integration have proven very effective. The key feature of the targeting approach is that the design targets are identified without committing to a specific solution. Various targeting tools such as the pinch analysis, the source-sink mapping diagrams, are widely used in industry and have proven very successful in maximizing profit margins while lowering processing cost (e.g. raw material consumption rate, utility cost, and waste generation).

The recently developed property clustering framework is a very powerful targeting tool that provides a platform for solving process design and optimization problems. The novel technique has enabled systematic tracking of properties (in the form of functionalities) throughout a process. The clustering concept has been implemented in various property integration tools, e.g. property-based pinch analysis (Kazantzi et al., 2005), the property integration framework (Eden, 2003), and to further expand the application range of the property integration framework, Qin et al. (2004) developed an algebraic approach. The clustering concept spawned a new generation of design tools

that recognize the importance of property based design. The recognition came about as a direct result of the following observations:

- Many processes are driven by properties NOT components

- Performance objectives are often described by properties

- Often objectives can not be described by composition alone

- Product/molecular design is based on properties

- Insights are often hidden by not integrating properties directly

Current CAMD methods have the ability to design formulations that target property needs; however the property targets are generally set forth by the performance needs of a single process unit (e.g. separator, distillation column etc.). Unlike mass and energy integration tools, CAMD method fall short in taking the requirements of the entire process as part of setting up the design problem. Consider the simple case of designing a solvent for a certain process unit. The impact of this individual solvent is not limited to that specific processing unit; in fact the impact is propagated throughout the entire process (e.g. the remaining streams and other processing units). Hence, a truly effective molecular design approach is one that can handle integrating the needs of the entire process into its molecular design scheme.

A major contribution of the highlighted targeting tools is that they posses a visualization media to help in the formulation and the generation of solutions to the design/optimization problem. Being able to visualize an entire chemical process in terms

of its streams and units provides insights on direct recycle and interception opportunities that might otherwise be hidden.

The discussion presented here provided the motivation that guided the research, and as a result the *objective* of this dissertation is to develop methods and tools that must accomplish the following:

- Integrate process and molecular/product design problems via a systematic methodology within the property integration paradigm.

- The approach needs to be capable of setting up the design performance requirements or "targets" a priori, i.e. a targeting approach.

- Incorporate the concepts of reverse problem formulation and property clusters to aid in the decomposition of the design problem

- In addition, the technique should take advantage of the benefits of using visual tools in the formulation of the problem and as part of its solution algorithm

3. Unified Property Integration Framework

As discussed in the previous chapter, the requirements dictate the need for a design approach that incorporates properties directly as part of the solution algorithm. The property integration framework (Shelley and El-Halwagi, 2000; Eden, 2003) was developed to address this need. It is a novel technique that has proven very useful in the optimization of various industrial processes as well as product blend problems including binary and ternary mixtures. (Shelley and El-Halwagi, 2000; El-Halwagi et al., 2003;

Eden, 2003; Eden et al., 2004; and Eljack et al., 2005). This property based platform was developed as a reverse problem formulation framework with the ability to systematically reformulate design problems and generate solutions visually on a ternary diagram. It differs from conventional techniques because it is non-iterative.

By understanding the roles of process and property models on design and recognizing that the complexity of the design problem is a direct result of the constitutive equations, the framework reformulates the original design problems as two reverse problems and decouples the constitutive equations from the balance and constraint equations (see Figure 3.1). Now the balance and constraint equations are solved in terms of the constitutive variables and the design targets are obtained, this is the reverse of a simulation problem. Next, the second reverse problem solves the constitutive equations to identify the unit operations, operating conditions and components needed to satisfy the design constraints. The key here is that any model can be used to describe the constitutive variables as long as the design targets are matched. This means that more than one solution to the targets can be identified, hence all feasible solutions to the design problem can be determined, and finally the optimal design can be determined based on a performance index (Eden, 2003).
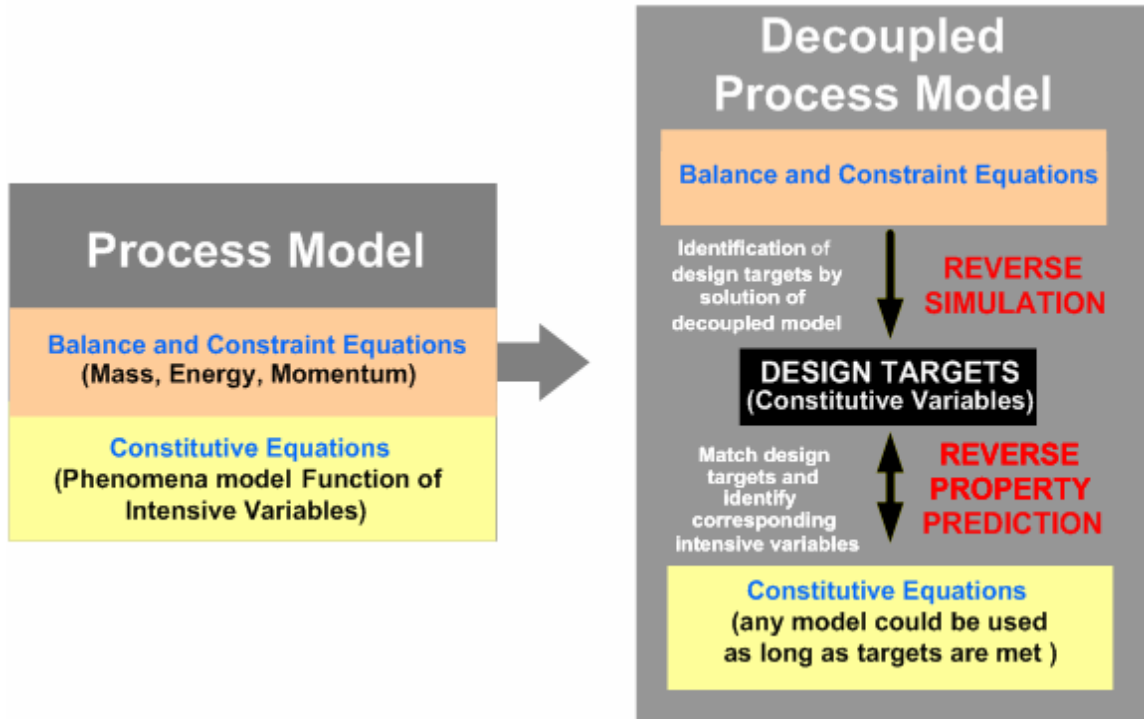
Figure 3.1: Reverse problem formulation methodology

This methodology is a valuable and powerful tool because:

- It allows for the integration of the process and product design problems using properties as a common interface.

- Simplifies the design problem by decoupling the often complex constitutive equations from the process model.

3.1. Property Clustering Fundamentals

3.1.1. Property Operator Description

Property clusters are conserved surrogate properties that are functions of non-conserved properties. The clusters are obtained by mapping property relationships into a low dimensional domain, thus allowing for visualization of the problem (Shelley and El-

Halwagi, 2000). The basis for the property clustering technique is the use of property operators. Although the operators themselves may be highly non-linear, they are tailored to possess linear mixing rules (Eden et al., 2004; El-Halwagi et al., 2004). The operator functions describe a class of properties that can be described by equation 3.1, in which the operator, Ψ_j, of property j is determined for a mixture M. The mixture is made up of N_s streams and can be described using j properties. The operator (Ψ_j) is formulated as the summation of each stream flowrate fraction (x_s) multiplied by the contribution of property j for stream s (P_{js}) (Eden, 2003).

$$\psi_j(P_{jM}) = \sum_{s=1}^{N_g} \frac{F_s}{\sum_{s=1}^{N_s} F_s} \cdot \psi_j(P_{js}) = \sum_{s=1}^{N_g} x_s \cdot \psi_j(P_{js}) \tag{3.1}$$

The operators are always formulated in a manner so that the right hand side (RHS) of equation 3.1 exhibits a linear mixing rule. The operator can be directly defined as a function of the actual property P_{js} (see equation 3.2), where the operator (Ψ_{jM}) of the mixed stream M will be referred to as P_{jM} and that of stream s as P_{js}. The operator can also describe functional relationships as shown for density in equation 3.3. Thus, the property operators can be non-linear functionalities, but the mixing rules have to be linear.

$$P_{jM} = \sum_{s=1}^{N_g} x_s \cdot P_{js} \qquad , \psi_j(P_{jM}) = P_{jM} \qquad \psi_j(P_{js}) = P_{js} \tag{3.2}$$

$$\frac{1}{\rho_M} = \sum_{s=1}^{N_g} x_s \cdot \frac{1}{\rho_s} \qquad , \psi_j(P_{jM}) = \frac{1}{\rho_M} \qquad \psi_j(P_{js}) = \frac{1}{\rho_s} \tag{3.3}$$

In equation 3.4, the property operators are normalized to a dimensionless form by dividing by a reference value. This is a necessary step due to the fact that properties can possess various functional forms and units. The reference value for each operator is chosen, so that various properties used to describe the system are in the same order of magnitude.

$$\Omega_{js} = \frac{\psi_j(P_{js})}{\psi_j(P_j^{ref})} \tag{3.4}$$

An Augmented Property index (AUP) for each stream s is defined as the summation of all the NP normalized property operators:

$$AUP_s = \sum_{j=1}^{NP} \Omega_{js} \tag{3.5}$$

The property cluster C_{js} for property j of stream s is defined as:

$$C_{js} = \frac{\Omega_{js}}{AUP_s} \tag{3.6}$$

3.1.2. Cluster Formulation

Property clusters are formulated to exhibit two fundamental rules:

1.   *Intra-stream conservation* For each stream s, the summation of all NC clusters, which correspond to the NP property operators values add up to unity as shown in equation 3.7. For systems that can be described by only

three properties, a ternary diagram can be used for visualization as seen in Figure 3.2.

$$\sum_{j=1}^{N_C} C_{js} = 1 \qquad (3.7)$$
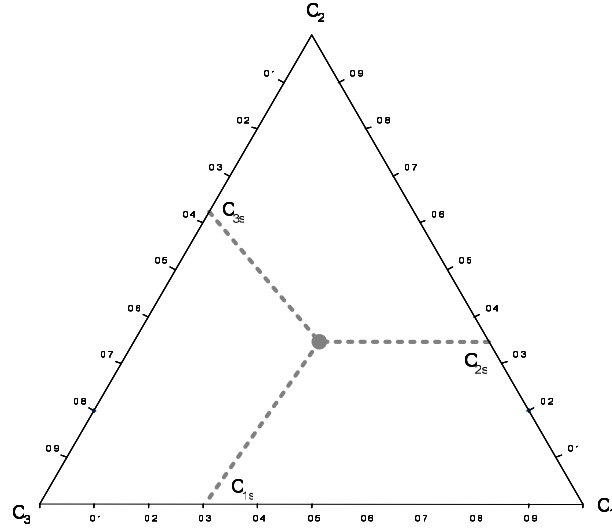
Figure 3.2: Intra-stream conservation of clusters

2.   *Inter-stream conservation* requires that the mixing of two streams should be performed so that the resulting individual clusters are conserved, corresponding to consistent additive rules as seen in equation 3.8.

$$C_{jM} = \sum_{j=1}^{N_C} \beta_s \cdot C_{js} \qquad (3.8)$$

In order to validate the inter-stream conservation rule represented in the equation above, it is first noted that the original definition of a property cluster C_{js} is valid for any

67

cluster. Meaning that it should also apply to the individual cluster values of a mixture, as shown in equation 3.9:

$$C_{jM} = \frac{\Omega_{jM}}{AUP_M} \tag{3.9}$$

Next, the generalized mixing rule given in equation 3.1 is divided by a reference property value. Substituting in the definition of the normalized property operator (equation 3.4) results in equation 3.10. This can be rearranged to show the normalized property operator for a mixture as shown in equation 3.11.

$$\frac{\psi_j(P_{jM})}{\psi_j(P_j^{ref})} = \sum_{s=1}^{N_s} x_s \cdot \frac{\psi_j(P_{js})}{\psi_j(P_j^{ref})} = \sum_{s=1}^{N_s} x_s \cdot \Omega_{js} \tag{3.10}$$

$$\Omega_{jM} = \sum_{s=1}^{N_s} x_s \cdot \Omega_{js} \tag{3.11}$$

Inserting the expression for the normalized operator of a mixture (equation 3.11) into equation 3.9, while rearranging the cluster definition in equation 3.6, yields equation 3.12, indicating inter-stream conservation of the clusters. Simplifying equation 3.12 yields equation 3.13, which defines the relative cluster arm for the individual stream (s).

$$C_{jM} = \frac{\Omega_{jM}}{AUP_M} = \frac{\sum_{s=1}^{N_s} x_s \cdot \Omega_{js}}{AUP_M} = \frac{\sum_{s=1}^{N_s} x_s \cdot AUP_s \cdot C_{js}}{AUP_M} = \sum_{s=1}^{N_s} \beta_s \cdot C_{js} \tag{3.12}$$

$$\beta_s = \frac{x_s \cdot AUP_s}{AUP_M} \tag{3.13}$$

3.1.3. Lever Arm Analysis

Inter-stream conservation, as given in equation 3.12, indicates that the mixture of two streams $S1$ and $S2$ on the ternary diagram can be represented by a straight line, as shown in Figure 3.3. This line corresponds to all possible mixture between the two streams, with the location of the mixture point, C_{jM}, being directly related to the streams fractional flowrate contributions, x_s. The location of the mixture point splits the line into two segments, each represented by β_1 and β_2, corresponding to relative cluster arms for stream $S1$ and $S2$, respectively. The mixture cluster equations developed by Shelley and El-Halwagi (2000) are given in equations 3.14-3.15.

$$\Omega_M = x_1 \cdot \Omega_1 + x_2 \cdot \Omega_2 \tag{3.14}$$

$$AUP_M = x_1 \cdot AUP_1 + x_2 \cdot AUP_2 \quad , \quad AUP_M = \sum_{s=1}^{N_s} x_s \cdot AUP_s \tag{3.15}$$
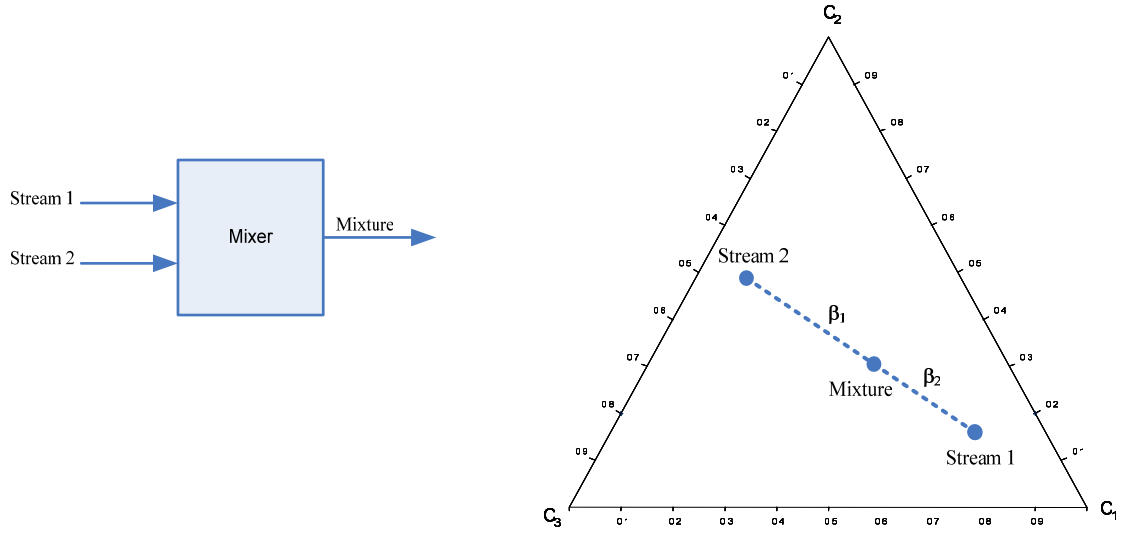
Figure 3.3: Inter-stream conservation of clusters

The relative cluster arm β_s, is a conserved entity and is defined as the *AUP* fractional contribution of each stream s to the mixture stream, (shown previously in equation 3.13). The merging of equations 3.9 and 3.12 generates equation 3.16; the subsequent substitution of equation 3.7 yields the conservation rule for the general relative cluster arm (equation 3.17). The expression for the Augmented Property index of a mixture, AUP_M, as shown in equation 3.15, is a result of combining equation 3.17 with the definition of the cluster arm (equation 3.13).

$$\sum_{j=1}^{N_C}\sum_{s=1}^{N_s}\beta_s \cdot C_{js} = 1 \qquad \Leftrightarrow \qquad \sum_{s=1}^{N_s}\beta_s\sum_{j=1}^{N_c}C_{js} \tag{3.16}$$

$$\sum_{s=1}^{N_s}\beta_s = 1 \tag{3.17}$$

70

These conservation rules are important features of the cluster formulation used in this methodology, as they allow for tracking clusters visually on a ternary diagram. As the conserved clusters are directly related to the raw properties, they enable tracking properties and this provides a unique way of representing processes/products from a properties perspective. The conversion of physical property data to cluster values is outlined in Table 3.1 (Eden, 2003).

Step	Description	Equation
1	Calculate dimensionless stream property value	3.4
2	Calculate stream *AUP* indices	3.5
3	Calculate ternary cluster values for each stream	3.6
4	Plot the points on the ternary cluster diagram	--

Table 3.1: Calculation of cluster values from physical property data

In summary, clusters are obtained by mapping property relationships into a low dimensional domain, thus allowing for visualization of the problem (Shelley and El-Halwagi, 2000). Although the operators themselves may be highly non-linear, they are tailored to possess linear mixing rules, e.g. density does not exhibit a linear mixing rule, however the reciprocal value of density follows a linear mixing rule (Eden et al., 2004; El-Halwagi, Glasgow et al., 2004). The operator expressions will invariably be different for molecular fragments and process streams, however as they represent that same property, it is possible to visualize them in a similar fashion. This is part of the novel work that is presented in this thesis.

3.1.4. Ternary Diagram and Cartesian Coordinate Conversion

The construction of the ternary diagrams in this work is accomplished by converting the cluster points from ternary to Cartesian coordinates. The conversion is used due to the absence of available software that supports ternary plot representations, as is the case here. By converting the ternary coordinates to Cartesian coordinates more common tools like Microsoft Excel can be used.
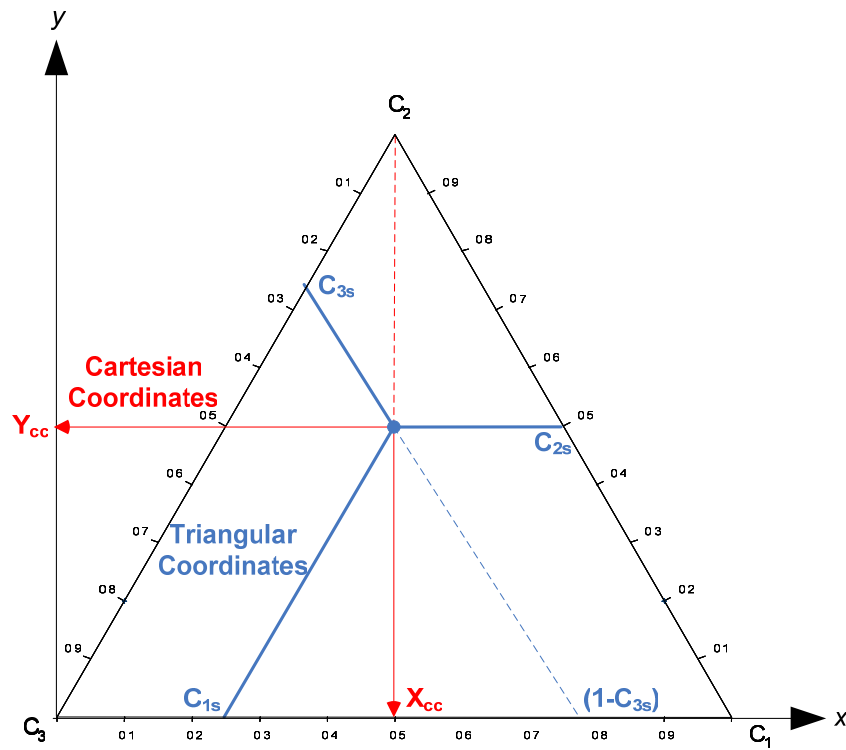
Figure 3.4: Converting ternary to Cartesian coordinates

Figure 3.4 is used to aid in describing the conversion methodology. Points on a ternary plot are represented in three dimensional coordinates (x, y, z) in this case (C_{1s}, C_{2s}, C_{3s}). All axes on a triangular diagram have a length of 1. X_{cc}, the x-value of the Cartesian coordinate set, is determined on the C_3-C_1 axis on the ternary diagram. On this

72

axis C_{1s} and the value of (1-C_{3s}) is known. X_{cc} will always be the arithmetic mean of these two points since the triangular plot is equilateral, as shown in equation 3.18.

$$X_{cc,s} = \frac{C_{1s} + (1 - C_{3s})}{2} = \frac{C_{1s} + (C_{1s} + C_{2s})}{2} = C_{1s} + 0.5 \cdot C_{2s} \qquad (3.18)$$

The y-value of the Cartesian coordinate, Y_{cc}, is directly related to C_{2s} by some scaling factor. From Figure 3.4, it is obvious that the value has to be less than 1. Points C_{3s}, X_{cc} and C_{2s} on the diagram along with the Pythagorean Theorem are used to determine this scaling factor from triangular to Cartesian coordinates. The length of C_3-C_2 axis is 1 and according to equation 3.18 the length of X_{cc}-C_3 is 0.5. Thus, Y_{cc} is calculated to be $0.5 \cdot \sqrt{3}$ using the Pythagorean Theorem (see equation 3.19); this value is constant due the equal length sides of the triangle. This scaling factor is used to convert triangular coordinates to Cartesian y-coordinates (Eden, 2003).

$$\left(\frac{1}{2}\right)^2 + Y_{scaling}^2 = (1)^2 \qquad \Rightarrow \qquad Y_{CC,s} = \frac{\sqrt{3}}{2} \cdot C_{2s} \quad , \quad Y_{scaling} = \frac{\sqrt{3}}{2} \qquad (3.19)$$

3.1.5. Feasibility Region Boundaries

Processes are made up of process units (sinks) and streams (sources). On the ternary diagram, the property values for streams (sources) are represented by discrete points while ranges of property values or property constraints (sinks) are denoted by a feasibility region. For visualization purposes only systems that can be described by three properties are used, with each property bound by a lower (P_j^{min}) and an upper limit (P_j^{max}), see equation 3.20. These values can also be described in terms of dimensionless property operators, as shown in equations 3.21 and 3.22.

$$P_{j,\text{sink}}^{\min} \leq P_j \leq P_{j,\text{sink}}^{\max} \quad , \quad \Omega_{j,\text{sink}}^{\min} \leq \Omega_j \leq \Omega_{j,\text{sink}}^{\max} \tag{3.20}$$

$$\Omega_{j,\text{sink}}^{\min} = \frac{\psi_j(P_{j,\text{sink}}^{\min})}{\psi_j(P_j^{ref})} \tag{3.21}$$

$$\Omega_{j,\text{sink}}^{\max} = \frac{\psi_j(P_{j,\text{sink}}^{\max})}{\psi_j(P_j^{ref})} \tag{3.22}$$

Using the definition of the augmented property index (*AUP*) as given in equation 3.5, the visualization of the sink region is achieved by translating the above dimensionless operators into the following cluster expressions:

$$C_{1,\text{sink}}^{\min} = \frac{\Omega_{1,\text{sink}}^{\min}}{\Omega_{1,\text{sink}}^{\min} + \Omega_{2,\text{sink}}^{\max} + \Omega_{3,\text{sink}}^{\max}}, \qquad C_{1,\text{sink}}^{\max} = \frac{\Omega_{1,\text{sink}}^{\max}}{\Omega_{1,\text{sink}}^{\max} + \Omega_{2,\text{sink}}^{\min} + \Omega_{3,\text{sink}}^{\min}} \tag{3.23}$$

$$C_{2,\text{sink}}^{\min} = \frac{\Omega_{2,\text{sink}}^{\min}}{\Omega_{1,\text{sink}}^{\max} + \Omega_{2,\text{sink}}^{\min} + \Omega_{3,\text{sink}}^{\max}}, \qquad C_{2,\text{sink}}^{\max} = \frac{\Omega_{2,\text{sink}}^{\max}}{\Omega_{1,\text{sink}}^{\min} + \Omega_{2,\text{sink}}^{\max} + \Omega_{3,\text{sink}}^{\min}} \tag{3.24}$$

$$C_{3,\text{sink}}^{\min} = \frac{\Omega_{3,\text{sink}}^{\min}}{\Omega_{1,\text{sink}}^{\max} + \Omega_{2,\text{sink}}^{\max} + \Omega_{3,\text{sink}}^{\min}}, \qquad C_{3,\text{sink}}^{\max} = \frac{\Omega_{3,\text{sink}}^{\max}}{\Omega_{1,\text{sink}}^{\min} + \Omega_{2,\text{sink}}^{\min} + \Omega_{3,\text{sink}}^{\max}} \tag{3.25}$$

In 2003, El-Halwagi et al. addressed the task of mapping the feasibility region from the property domain to the cluster domain. Although the region represents an infinite number of feasible points, the developed technique required no enumeration. The feasibility region is first overestimated by simply using the minimum and maximum values of the clusters to place bounds on the region by the six line segments shown in Figure 3.5. Although the overestimated region does not define the true feasibility region,

it narrows down the search space and guarantees that no feasible point will exist outside it. Subsequently, the six cluster points defined by equations 3.23-3.25 are plotted as a point on each of the six line segments. Since the six points are part of the true feasibility region, any mixture of the points will also be part of the true region. The connection of the six points defined the underestimated region. According to the findings by El-Halwagi et. al (2003) and Eden (2003), the feasibility region is defined by six unique points, and their findings are summarized in Rule 1 below.

Rule 1: *Expressing property constraints as a Feasibility Region*

- The boundary of the true feasibility region can be accurately represented by no more than six linear segments.

- When extended, the linear segments of the boundary of the true feasibility region constitute three convex hulls (cones) with their heads lying on the three vertices of the ternary cluster diagram.

- The six points defining the boundary of the true feasibility region are determined a priori and are characterized by the following values of dimensionless operators (see Figure 3.6)

$$\left(\Omega_1^{min},\Omega_2^{min},\Omega_3^{max}\right)\left(\Omega_1^{min},\Omega_2^{max},\Omega_3^{max}\right)\left(\Omega_1^{min},\Omega_2^{max},\Omega_3^{min}\right)$$
$$\left(\Omega_1^{max},\Omega_2^{max},\Omega_3^{min}\right)\left(\Omega_1^{max},\Omega_2^{min},\Omega_3^{min}\right)\left(\Omega_1^{max},\Omega_2^{min},\Omega_3^{max}\right)$$

The feasibility region boundary analysis described above provides the exact expression for the feasibility region a priori and without enumeration (El-Halwagi et al., 2003; Eden, 2003).
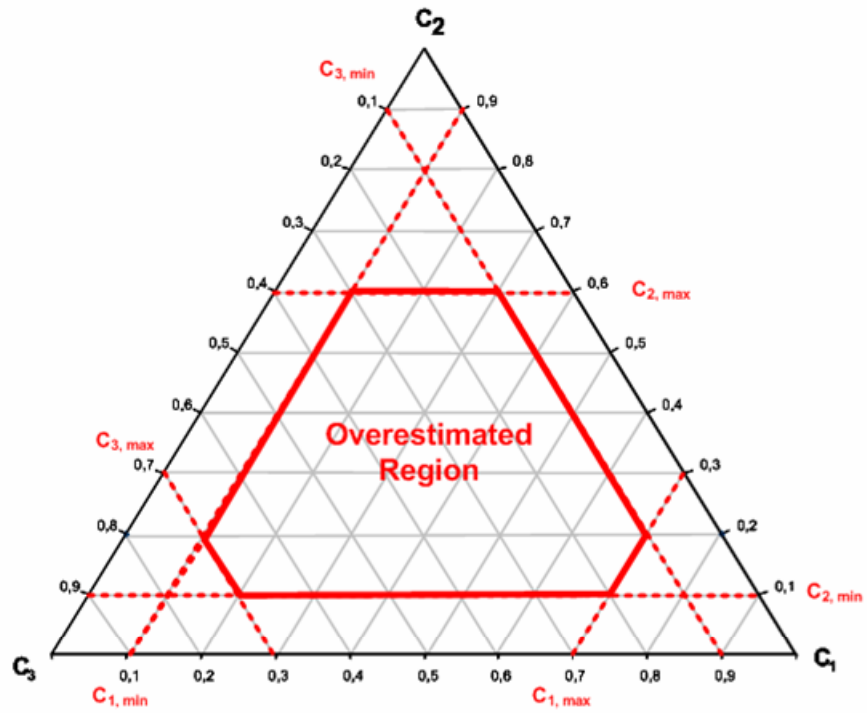
Figure 3.5: Overestimation of feasibility region
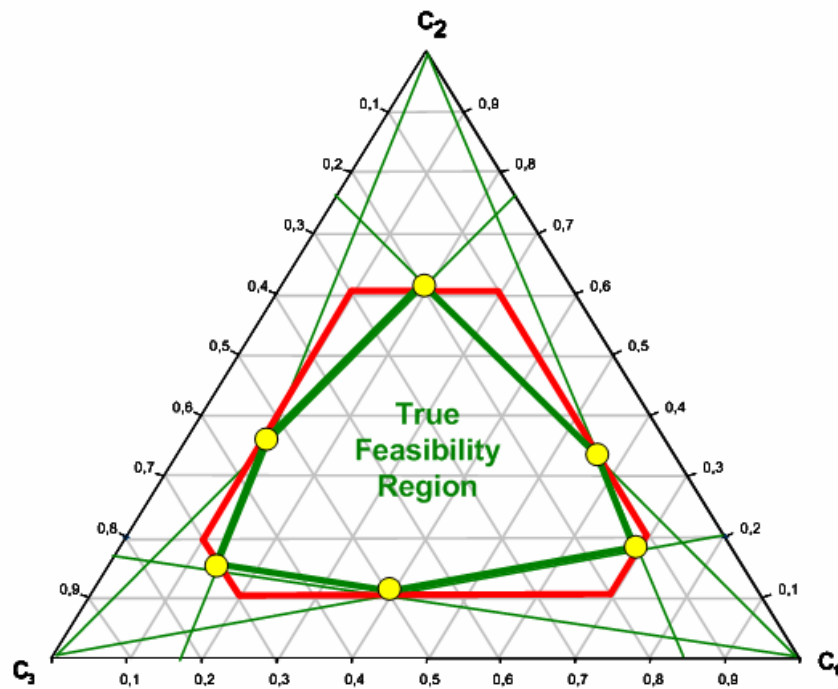
Figure 3.6: True feasibility region of a sink.

3.2. Molecular Property Clusters

3.2.1. Group Contribution

To provide a methodology for handling molecular design problems, the property integration framework is extended to include Group Contribution Methods (GCM), which allow for prediction of physical properties (e.g. boiling and melting temperature, enthalpy, and heat of vaporization), from structural information. As stated in Section 2.9, initially the Group Contribution Methods were based on the contributions of first order groups that make up the molecule (Joback and Reid, 1983), then to increase the accuracy of the predicted properties work by Constantinou and Gani (1994) and later by Marrero and Gani (2001), estimate properties utilizing first order, second order, and third order groups which use first order groups as building blocks. Understanding that the goals of this research is to develop the first implementation of a property based molecular algorithm that can handle systematic generation of formulations in response to property needs, the prediction accuracy of the first-order GCM is sufficient. Once the proposed framework is established, implementation of higher order GCM to enhance the accuracy of the selected property models will be explored. For now, the general group contribution model equation used to predict properties is:

$$\overbrace{f(X)}^{\text{LHS}} = \overbrace{\sum_i N_i C_i}^{\text{RHS}}$$

(3.26)

C_i is the contribution of the first-order group type i which occurs N_i times, and $f(X)$ is a function of property X. Table 3.2 presents ten main properties predicted using GCM (Constantinou and Gani, 1994; Constantinou et al., 1995; and Marrero and Gani,

2001). The left hand side (LHS) of equation 3.26 for each property X is shown in the table 3.2. The universal constants e.g. t_{mo}, t_{bo} etc. are part of the general model and their values for the various properties are listed in table 3.3. Only the first order group contribution terms are listed for the right hand side (RHS) of equation 3.26, there is data available for second and third order terms as mentioned previously (Constantinou and Gani, 1994; Marrero and Gani, 2001). The group contribution property data used by this method has been determined by regression using a large data bank of more than 2000 compounds collected at CAPEC-DTU, see Appendix A (Marrero and Gani, 2001).

Property (X)	LHS of Eq. 3.26 Function $f(X)$	RHS of Eq. 3.26 1st order GC term
Normal melting point (T_m)	$\exp\left(\dfrac{T_m}{t_{mo}}\right)$	$\sum_i N_i T_{m1_i}$
Normal boiling point (T_b)	$\exp\left(\dfrac{T_b}{t_{bo}}\right)$	$\sum_i N_i T_{b1_i}$
Critical temperature (T_c)	$\exp\left(\dfrac{T_c}{t_{co}}\right)$	$\sum_i N_i T_{c1_i}$
Critical pressure (P_c)	$\left(P_c - P_{c1}\right)^{-0.5}$	$\sum_i N_i P_{c1_i}$
Critical volume (V_c)	$V_c - V_{co}$	$\sum_i N_i V_{c1_i}$
Standard Gibbs energy[1] (G_f)	$G_f - G_{fo}$	$\sum_i N_i G_{f1_i}$
Standard enthalpy formation[1] (H_f)	$H_f - H_{fo}$	$\sum_i N_i H_{f1_i}$
Standard enthalpy vaporization[1] (H_v)	$H_v - H_{vo}$	$\sum_i N_i H_{v1_i}$
Standard enthalpy fusion (H_{fus})	$H_{fus} - H_{fuso}$	$\sum_i N_i H_{fus1_i}$
Liquid molar volume[1] (V_l)	$V_l - d$	$\sum_i N_i v_{1_i}$

[1]Properties predicted at 298K

Table 3.2: Property functions for Group Contribution Methods

Universal Constants	Value
t_{mo}	102.425 K
t_{bo}	204.359 K
t_{co}	181.128 K
p_{cl}	1.3705 bar
v_{co}	4.35 cm^3/mol
g_{fo}	-14.828 kJ/mol
h_{fo}	10.835 kJ/mol
h_{vo}	6.829 kJ/mol
h_{fuso}	-2.806 kJ/mol
d	0.01211 m^3/kmol

Table 3.3: Listed values of GCM universal constants

3.2.2. Bridging the Gap between Process and Molecular Design

By combining property clustering techniques and first order group contribution methods (GCM), a systematic methodology is obtained that facilitates simultaneous consideration of process and molecular design. In the same manner that Eden (2003) reformulated the process design problem as two reverse problems, the process design will be solved in terms of property values with the design targets set as constraints. Figure 3.7 describes the general flow of information from process design to molecular design and back again; where the output of the process design algorithm will be a set of property values. These values are the property targets for the molecular design problem. The molecular design algorithm developed in this thesis and described in the following sections will systematically generate molecular formulations to satisfy the property targets/constraints identified by solving the process design problem.
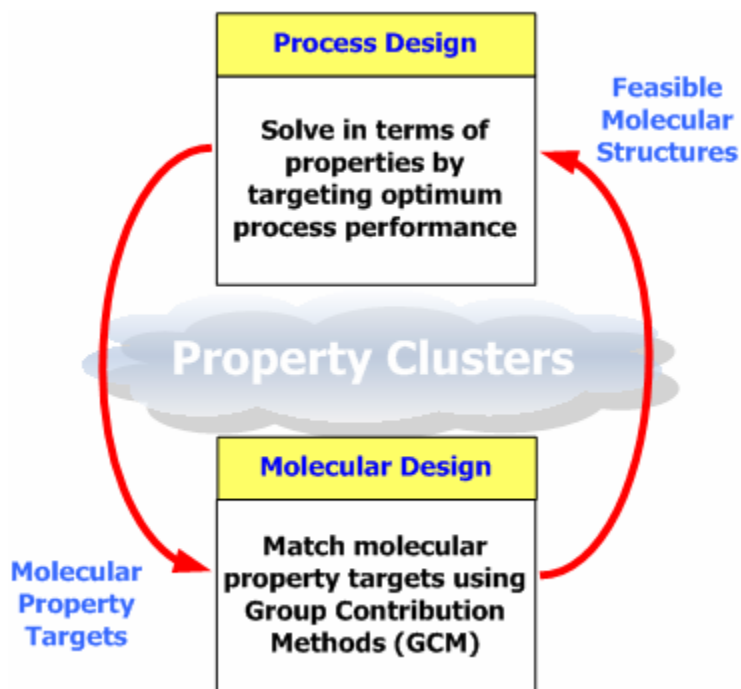
Figure 3.7: Property driven approach to integrated process and molecular design

3.2.3. Molecular Property Operators

Extending the property clustering technique to include GCM for molecular design, introduces the need for molecular property operators. Like the original operators, their formulation must be such that it still allows for simple linear additive rules for combining the groups, which can be described by the following:

$$\psi_j^M (P_j) = \sum_{g=1}^{N_g} n_g \cdot P_{jg} \tag{3.27}$$

In equation 3.27, $\psi_j^M (P_j)$ is the molecular property operator of the jth property. The molecular property operator describes the functional relationship of the group

contribution property equations in a manner so that the RHS of the equations is always in the form of a summation of the number of each group (n_g) multiplied by the contribution to property j from group g (P_{jg}). Some properties are not predicted directly from group contribution methods, but are estimated as functions of other properties that can be predicted using GCM, e.g. vapor pressure (VP) can not be estimated directly, however it can be estimated from the boiling point, which is a property described by GCM, as shown in equations 3.28 and 3.29 (Sinha and Achenie, 2001).

$$\log VP = 5.58 - 2.7 \left(\frac{T_{bp}}{T} \right)^{1.7}$$

(3.28)

$$\psi^M (T_{bp}) = \exp\left(\frac{T}{t_{bo}} \right) = \sum_{g=1}^{N_g} n_g \cdot t_{b_g}$$

(3.29)

Where, T and t_{bo} are the chosen condensing temperature and the group contribution boiling temperature constants, respectively.

Notice that the property operator can be very complex, but molecular formulation on the ternary diagram is still simple because the property operators obey simple linear additive rules. Next, the molecular property operators can be converted to clusters according to the procedures developed for the original property clusters, see Section 3.1.2 (Shelley and El-Halwagi, 2000; Eden et al., 2004; El-Halwagi et al., 2004).

Since properties can have various functional forms and units, the molecular property operators like process property operators are normalized into a dimensionless form by dividing by a reference operator. This reference is appropriately chosen such

that the resulting dimensionless properties are all of the same order of magnitude. The

normalized property operator for group g is given as:

$$\Omega^M_{jg} = \frac{\psi^M_{jg}(P_{jg})}{\psi^{ref}_j(P_j)} \tag{3.30}$$

An Augmented Property index AUP^M for each group g is defined as the summation of all

the NP dimensionless property operators, $\Omega_{jg}{}^M$:

$$AUP^M_g = \sum_{j=1}^{NP} \Omega^M_{jg} \tag{3.31}$$

Molecular fragment g's property cluster $C_{jg}{}^M$ for property j is defined as the ratio of the

normalized molecular property operator and the $AUP_g{}^M$:

$$C^M_{jg} = \frac{\Omega^M_{jg}}{AUP^M_g} \tag{3.32}$$

3.2.4. Conservation Rules for Molecular Clusters

Visualization of the molecular design problem is very valuable to this

methodology. To ensure that the molecular clusters are conserved, they have to posses

both intra- and inter-molecular conservation. Similar to the intra-stream conservation

rule for processes, the intra-molecular conservation rule requires that the sum of

individual clusters C_{jM} for each molecular formulation M must sum to unity as shown in

equation 3.33. This is proven by the summing of all cluster values for all j properties for

molecule M in equation 3.32 and substituting the AUP definition (see equation 3.34).

82

$$\sum_{j=1}^{N_C} C^M_{\ j} = 1 \tag{3.33}$$

$$\sum_{j=1}^{NP} C^M_{\ j} = \frac{\sum_{j=1}^{NP} \Omega^M_{\ j}}{AUP^M} = \frac{AUP_M}{AUP_M} = 1 \tag{3.34}$$

The inter-molecular conservation rule for adding molecular groups or fragments on the ternary diagram is derived analogous to inter-stream conservation. The general additive rule for molecular operators (equation 3.27) is normalized by a reference value and the definition of the dimensionless molecular operator (equation 3.30) is substituted to yield the following mixing rule:

$$\Omega^M_{\ j,mix} = \sum_{g=1}^{N_g} n_g \cdot \Omega_{jg} \tag{3.35}$$

Inter-molecular conservation requires that the individual molecular cluster of mixing two groups $C^M_{j,mix}$ is conserved. For two groups each possessing their own individual cluster values, lever-arm rules like equation 3.36 are needed to allow for easy determination of the mixture cluster value for each property j.

$$C^M_{\ j,mix} = \sum_{g=1}^{N_g} \beta_g \cdot C_{jg} \tag{3.36}$$

The definition of molecular property cluster given in equation 3.32 applies to any cluster including molecular fragments, therefore the cluster of a mixture of two molecular groups or fragments is:

$$C^M_{\ j,mix} = \frac{\Omega^M_{\ j,mix}}{AUP^M_{mix}} \tag{3.37}$$

It is crucial to validate the inter-molecular conservation rule. First the mixing rule for the dimensionless property operator is inserted into equation 3.37. Next, the definition of a molecular fragment cluster is rearranged and substituted. This proves the inter-molecular conservation rule according to equation 3.38- 3.40.

$$C_{j,mix}^{M} = \frac{\sum_{g=1}^{N_g} n_g \cdot \Omega_{jg}^{M}}{AUP_{mix}^{M}} = \frac{\sum_{g=1}^{N_g} n_g \cdot C_{jg}^{M} \cdot AUP_{g}^{M}}{AUP_{mix}^{M}} \tag{3.38}$$

$$\beta_g = \frac{n_g \cdot AUP_{g}^{M}}{AUP_{mix}^{M}} \tag{3.39}$$

$$C_{j,mix}^{M} = \sum_{g=1}^{N_g} \beta_g \cdot C_{jg}^{M} \tag{3.40}$$

3.3. Visual Molecular Design using Property Clusters

The conversion of property data to cluster values outlined in Section 3.1.2 for process design was developed by Eden et al. (2004). The conversion of molecular property data to cluster values follows a similar procedure as given in Table 3.4.

Step	Description	Equation
1	Calculate molecular property operators	3.27
2	Calculate dimensionless molecular property values	3.30
3	Calculate molecular AUP indices	3.31
4	Calculate molecular cluster values for each formulation	3.32
5	Plot the points on the ternary cluster diagram	--

Table 3.4: Calculation of cluster^{M} values from GCM predicted property data

The primary visualization tool from the mass integration framework, the source-sink mapping, which is discussed in Section 2.5.2 is utilized in the molecular synthesis framework. In the original cluster formulation for process design, mixing of two sources is a straight line, i.e. the mixing operation can be optimized using lever-arm analysis. Analogously, combining or "mixing" two molecular fragments in the molecular cluster domain follows a straight line (an illustrative example is given in Figure 3.8 below). Design and optimization rules have been developed for property based process design problems (Eden et al., 2004; El-Halwagi et al., 2004), and in the following similar rules are presented for property based molecular design problems.
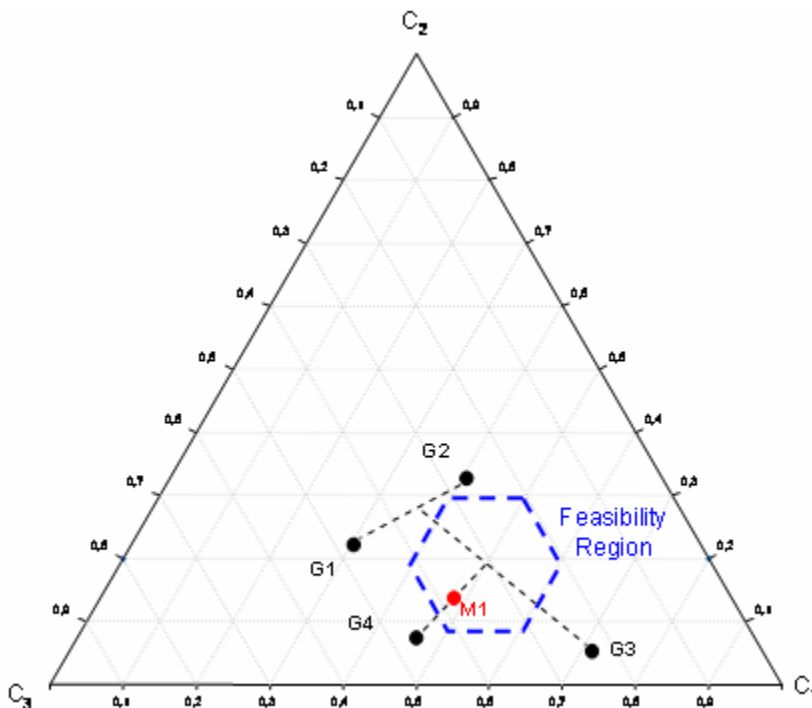
Figure 3.8: Group addition on ternary cluster diagram.

Rule 2: Two groups, G1 and G2, are added linearly on the ternary diagram, where the visualization arm β_l, describes the location of G1-G2 molecule.

$$\beta_1 = \frac{n_1 \cdot AUP_1}{n_1 \cdot AUP_1 + n_2 \cdot AUP_2} \tag{3.41}$$

Rule 3: More groups can be added as long as the Free Bond Number (*FBN*) is not zero.

$$FBN = \left[\sum_{g=1}^{N_g} n_g \cdot FBN_g \right] - \left[\sum_{g=1}^{N_g} n_g - 1 \right] - 2 \cdot NO_{Rings} \tag{3.42}$$

FBN is the free molecular bond number of the formulation, n_g is the number of occurrences of group g, FBN_g is the unique free bond number associated with group g, and NO_{Rings} is the number of rings in the formulation.

Rule 4: Location of the final formulation is independent of the order of group addition. The location of the formulation is unique, and is only based on the number of each group in the molecule.

For example, consider Butyl methyl ether ($C_5H_{12}O$); it is made up of the following groups: CH_3, CH_2, and CH_3O. Constructing this molecule on the ternary cluster

diagram, using three chosen properties, can be done in a variety of ways. However, regardless of the sequence in which the groups are combined, the resulting molecule (CH_3O-CH_2-CH_2-CH_2-CH_3) is located at the same unique point. To make sure Rule 2 is satisfied each molecular fragment's Free Bond Number (*FBN*) is placed within brackets on the ternary diagram. As proof of concept, a random feasibility region (see Figure 3.9A and 3.9B) is represented by the dotted region; expressing the targeted approach of building molecular formulation to satisfy a set of given property constraints. Looking at Figure 3.9A, the starting point is CH_3O then adding three CH_2 fragments then CH_3. Figure 3.9B starts with CH_3, then adds three CH_2 molecules, and finally CH_3O. Both paths shown, as well as many others end up at the same point, hence *the location of each molecular formulation is unique and independent of group addition path*.

> **Rule 5**: For completeness, the final formulation must not have any free bonds, i.e. *FBN* has to be equal to zero.

Given a completed molecular formulation, three conditions must be satisfied for the designed molecule to be a valid solution to the process and molecular design problem. Rules 5 and 6 are the necessary conditions, while rule 8 is the sufficient condition:

> **Rule 6**: The cluster value of the formulation must be contained within the feasibility region of the sink on the ternary molecular cluster diagram.
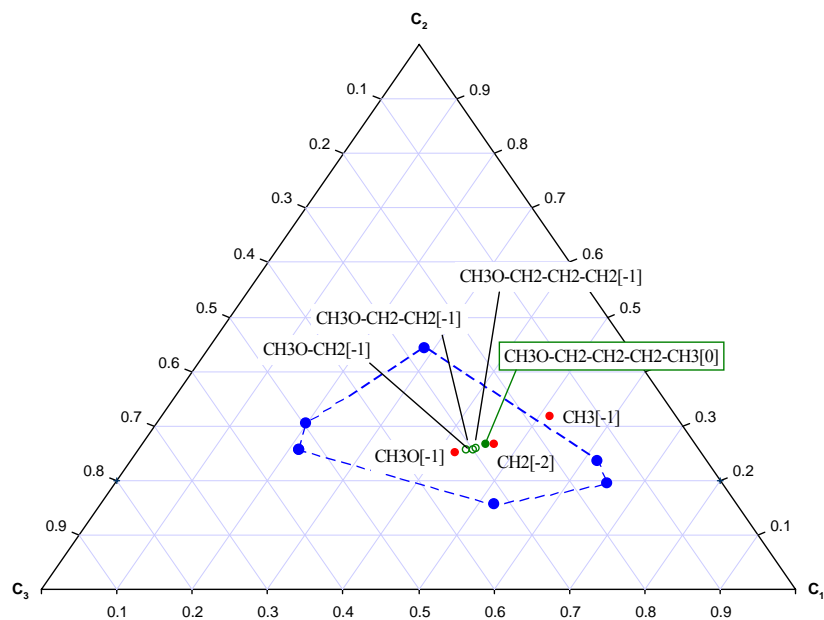
Figure 3.9A: Group addition path A for formulation of Butyl methyl ether.
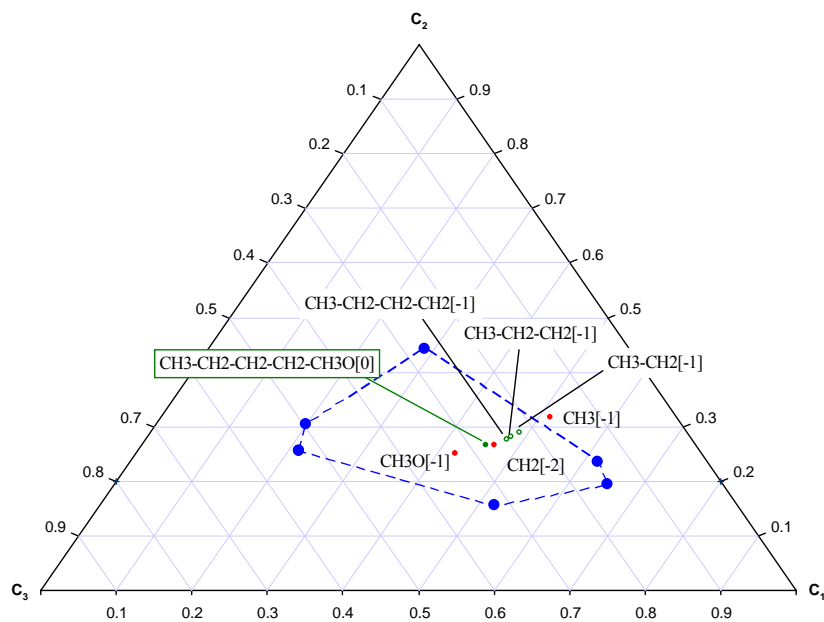
Figure 3.9B: Group addition path B for formulation of Butyl methyl ether.

Rule 7:   The *AUP* value of the designed molecule must be within the range of the target. If the *AUP* value falls outside the range of the sink, the designed molecule is not a feasible solution.

Rule 8:   For the designed molecule to match the target properties, the *AUP* value of the molecule has to match the *AUP* value of the sink at the same cluster location. And in the case where the design problem included Non-GC properties, those properties must be back calculated for the designed molecule using the appropriate corresponding GC property, and those values have to match the target Non-GC properties.

Now that the process and molecular design problems are both described in terms of clusters, a unifying framework exists for simultaneous solution of property driven design problems. This is important because as mentioned earlier the clustering technique reduces the dimensionality of both problems, thus it is possible to visually identify the solutions, which is a significant advantage of this approach.

Figure 3.10 highlights the flow of information in the molecular property cluster framework for product and process design problems. The framework requires property data as input to the algorithm, but whether the data is dictated by the process, in terms of performance requirements, or from product design requirements, does not affect the methodology behind this algorithm. Once the property targets are identified, they set the problem constraints and the selection of property operators to represent the target. The

availability of GC property models shape how the property is used in the methodology. If models are available, then (primary) property operators are formulated directly, otherwise empirical equations are used to correlate the secondary property to the primary. Having formulated the design problem, it is then mapped onto the molecular ternary cluster diagram. The property constraints are represented by a region, and the group building blocks as discrete points. Visual synthesis is performed by combining molecular fragments, followed by screening of the formulation using the developed necessary and sufficient conditions. The result are candidate molecules that posses the pre-determined property targets.
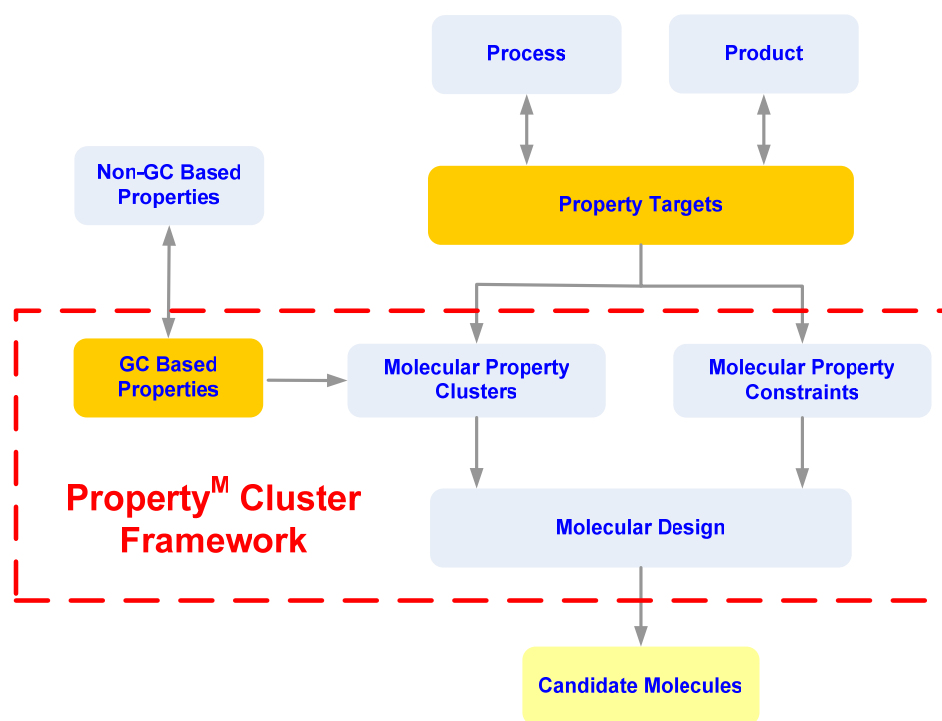
Figure 3.10: Molecular property cluster framework.

The contributions of the developed algorithm in this work are two-fold. First, the developed molecular design methodology bridges the gap between process and molecular design via incorporation of property clusters into its CAMD strategy. Second, for systems that can be sufficiently described using three properties or functionalities, the molecular synthesis problem is solved visually by mixing molecular fragments on a ternary diagram using simple lever-arm rules.

Application examples utilizing the developed techniques are presented in Chapters 4 and 5.

3.4. Algebraic Property Clustering Technique for Molecular Design

As stated previously, the ability to synthesize molecules within the clustering domain is key to bridging the gap between process and molecular design, however utilizing the visualization approach limits the application range to cases that can be expressed using three properties. It is recognized that not all design problems can be described by just three properties. For property integration through componentless design of processes, Qin et al. (2004) introduced an algebraic approach to overcome this bottleneck, by taking advantage of the mathematical structure of the property clusters. Presented here is an analogous algebraic method that expands the application range of the molecular property clustering technique. Here we will further exploit the advantages of the linear additive rules of the molecular operators to setup the design problem as a set of linear algebraic equations.

Problem Statement: Synthesize molecular formulations, given a set of molecular building blocks (first order groups from GCM) represented by n_g and a set of property performance requirements/constraints that is described by:

$$P_{ij}^{lower} \leq P_{ij} \leq P_{ij}^{upper} \qquad (3.43)$$

Where i, is the index for the molecular formulation, and j is the index of properties. The property constraints can be expressed in terms of the normalized property operators by combining the mixing rules for operators (equation 3.27) with the corresponding reference values.

$$\Omega_j^{\min} \leq \Omega_{ij} \leq \Omega_j^{\max} \qquad (3.44)$$

Recall the generalized dimensionless additive rule for a given property j and n_g molecular groups is written as:

$$\Omega_j = \sum_{g=1}^{N_g} n_g \cdot \Omega_{jg} \qquad (3.35)$$

The substitution of equation 3.35 into the inequality expression given by equation 3.44 generates the following:

$$\Omega_j^{min} \le \sum_{g=1}^{N_g} n_g \cdot \Omega_{jg} \le \Omega_j^{max} \qquad (3.45)$$

Thus each property constraint can be expressed as a set of inequality expressions, which are the basis for the algebraic approach. These sets of equations will help place bounds on the feasibility region, referred to as the sink. Because each property can be expressed in terms of two inequalities, each property can be combined with another property in two ways. In the original visualization approach for the molecular design framework, the bounds on three properties can be represented by a set of six points (Eden et al, 2004; Qin et al., 2004). Similarly, for systems made up of four properties, Ω_1-Ω_4, each with a lower and upper limit, the bounds on the feasibility region can be described by eight points. These points are determined by the following (Eljack et al., 2007a):

Rule 9: Each property constraint is translated into the inequality expression from equation 3.45, and then split into two equations, one for minimum (min) and one for maximum (max).

$$\Omega_j^{min} \le \sum_{g=1}^{N_g} n_g \cdot \Omega_{jg} \qquad \sum_{g=1}^{N_g} n_g \cdot \Omega_{jg} \le \Omega_j^{max} \qquad (3.46)$$

Hence there will be *2NP* (number of properties) inequality equations that constitute the main set. The *AUP* values for these set of equations will be calculated in order to determine the *AUP* range of the sink.

Rule 10: From the main set of equations, *2NP* subsets will be generated. Each subset will contain an equation for each of the properties used to describe the system.

For a four property system, there will be 8 inequality equations for the original set, from which eight subsets will be developed. Each subset will be made up of four equations and only one of the two inequalities used to describe each property will be used in each subset. For the normalized operators of the system (Ω_1, Ω_2, Ω_3, Ω_4) the following combinations from the original set should be used to generate the eight subsets of equations:

$$
\begin{array}{ll}
(\Omega_1^{max},\Omega_2^{min},\Omega_3^{min},\Omega_4^{min}) & (\Omega_1^{min},\Omega_2^{max},\Omega_3^{max},\Omega_4^{max}) \\
(\Omega_1^{min},\Omega_2^{max},\Omega_3^{min},\Omega_4^{min}) & (\Omega_1^{max},\Omega_2^{min},\Omega_3^{max},\Omega_4^{max}) \\
(\Omega_1^{min},\Omega_2^{min},\Omega_3^{max},\Omega_4^{min}) \quad , \quad & (\Omega_1^{max},\Omega_2^{max},\Omega_3^{min},\Omega_4^{max}) \\
(\Omega_1^{min},\Omega_2^{min},\Omega_3^{min},\Omega_4^{max}) & (\Omega_1^{max},\Omega_2^{max},\Omega_3^{max},\Omega_4^{min})
\end{array}
\qquad (3.47)
$$

As stated earlier the subsets of equations are used to consider all possible ways the properties can be combined with each other to place bounds on the feasibility regions.

Rule 11: The generated subsets of equations constitute the property constraints. In addition, structural constraints such as non-negativity constraints for the contribution of each group and a limit on the size of a molecular formulation need to be included (equation 3.48) and a possible limit on the length of a molecular formulation (equation 3.49):

$$n_g \geq 0 \qquad g = \{1,\ldots,N_g\} \tag{3.48}$$

$$\sum_{g=1}^{N_g} n_g \leq NF \tag{3.49}$$

Rule 12: For this algorithm a limit on the number of first order group fragments (NF) will also need to be specified ahead of design. To ensure that all valences in a molecule are satisfied, the following equation is used to place another structural constraint on the design problem.

$$\text{FBN} = \left[\sum_{g=1}^{N_g} n_g \cdot \text{FBN}_g\right] - 2 \cdot \left[\sum_{g=1}^{N_g} n_g - 1\right] \tag{3.50}$$

Each group g has a free bond number (FBN) associated with it (e.g. CH_3 has FBN = 1, CH_2 has FBN=2). It should be noted that equation (3.50) only takes non-cyclical

compounds into account, as does the algebraic approach. However, further studies are looking at how to include them within the framework.

Now that the main concepts behind this methodology have been established, an outline of the algebraic technique is given by Table 3.5.

The proposed technique lacks visualization aspects; however, it has provided important contributions:

- Lowers the complexity of the design problem by setting up the design problem as a set of linear algebraic equality and inequality equations.

- It expanded the application range of the recently introduced molecular clustering technique to enable handling of problems requiring more than three properties.

The algebraic approach opens a new area of research that would concentrate on developing tools directed at incorporating this algebraic method with other mathematical design approaches, i.e. MILP or LP optimization methods.

Step	Description	Equation
1	Transform given property data into molecular property operator terms	3.27
2	Express property constraints as inequalities forming the main set of inequality equations	3.43 – 3.44
3	Determine the *AUP* range of the sink	3.31
4	Develop the subsets of inequality equations following Rule 10	--
5	Generate the structural constraints	3.48 – 3.50
6	Find the solution to each subset of linear inequality equations along with the structural constraint equations in order to determine the min and max n_g of each group g. This is done with the objective being: first minimize the *AUP* of each subset and then to maximize the *AUP* of each subset. This step can be solved using various programs: MATLAB, Visual C++, etc. For the examples shown in this chapter, Microsoft Excel was used.	--
7	If the *AUP* values of each subset do not fall within the *AUP* range of the sink, those solutions are excluded. Then the range of valid n_g values should satisfy all remaining solutions. Thus if one solution gives *g1* between 3 and 6 and another between 2 and 10 then the true range that will satisfy all constraints is 3-6.	--
8	Solutions for n_g will not always be integer values, thus the solutions are rounded up for minimum values and rounded down for maximum values. This step can be bypassed by placing another constraint on the problem where n_1, n_2... n_g are defined as integer values.	--
9	Generate all the feasible formulations and perform the final checks that all property constraints are satisfied	--

Table 3.5: Outline of algebraic molecular cluster approach.

3.4.1. Proof of Concept Example

To highlight the different aspects of this new algebraic molecular clustering method, a simple design problem is presented. *Problem statement:* Given a system described by critical volume (V_c), heat of vaporization (H_v) and heat of fustion (H_{fus}) and the following molecular fragments as building blocks: CH_2 and OH, identify molecular formulations that will satisfy the following performance requirements:

$$310 \leq V_c \, (cm^3/mol) \leq 610 \qquad 90 \leq H_v \, (kJ/mol) \leq 120$$

$$20 \leq H_{fus} \, (kJ/mol) \leq 64 \qquad 450 \leq T_b \, (K) \leq 560 \qquad (3.51)$$

g	Group	*FBN*	V_c (cm^3/mol)	H_v (kJ/mol)	H_{fus} (kJ/mol)	T_b (K)
1	CH_2	2	56.28	4.91	2.64	0.9225
2	OH	1	30.61	24.21	4.79	3.21

Table 3.6: Property data for each molecular group.

The Group Contribution (GC) property data of the molecular groups is given in Table 3.6. In addition, the additive rules for the molecular operators of the targeted properties are represented by equation 3.52 (Constantinou and Gani, 1994; Marrero and Gani, 2001). The formulation of the operators from GC property models is outlined in the original molecular clustering framework (Section 3.2.3; Eljack et al., 2006).

$$V_c - v_{c0} = \sum_{g=1}^{N_g} n_g \cdot v_{c1} \qquad\qquad H_v - h_{v0} = \sum_{g=1}^{N_g} n_g \cdot h_{v1}$$

<div align="right">(3.52)</div>

$$H_{fus} - h_{fus0} = \sum_{g=1}^{N_g} n_g \cdot h_{fus1} \qquad \exp\left(\frac{T_b}{t_{bo}}\right) = \sum_{g=1}^{N_g} n_g \cdot t_{b1}$$

Other constraints are placed on the problem, i.e. the maximum length of the molecule can not exceed 15 fragments and no cyclical compounds should be formed.

Given equations 3.51 and 3.52, and the information in Table 3.6, the data for the four properties: critical volume, heat of vaporization, heat of fusion and boiling temperature (1, 2, 3, 4) can be transformed to Ω_1, Ω_2, Ω_3, Ω_4 using the normalized property operator definition (equation 3.30) along with the following reference values (20, 1.0, 0.5, 7.0), respectively. The same reference values are also used to convert the group data given in Table 3.6. These values were selected in order to keep the operators in the same order of magnitude. The resulting Ω values for all four property constraints are shown in Table 3.7. The *AUP* range of the feasibility region (sink) was calculated to be 141.19 – 273.27.

	Ω_{Vc}	Ω_{Hv}	Ω_{Hfus}	Ω_{Tb}
Ω^{min}	15.105	78.26	45.612	1.291
Ω^{max}	30.102	108.26	4.133	2.213

<div align="center">Table 3.7: Calculated Ω for the given property constraints.</div>

Next, the provided data along with equation 3.46 are used to generate the main set of linear inequality equations, from which eight subsets are generated. The equations involved in subset one according to equation 3.47 are provided below in equation 3.53. The remaining 7 subsets are generated in the same way. Finally the structural constraints are given in equation 3.54 and 3.55.

$$2.81 \cdot g_1 + 1.53 \cdot g_2 \leq 30.10$$
$$4.91 \cdot g_1 + 24.21 \cdot g_2 \geq 78.26$$
$$5.28 \cdot g_1 + 9.571 \cdot g_2 \geq 45.61 \tag{3.53}$$
$$0.131 \cdot g_1 + 0.459 \cdot g_2 \geq 1.29$$

$$g_1 \geq 0 \qquad , \qquad g_2 \geq 0 \qquad , \qquad g_1 + g_2 \leq 15 \tag{3.54}$$

$$[g_1 \cdot FBN_1 + g_2 \cdot FBN_2] - 2 \cdot [g_1 + g_2 - 1] = 0 \tag{3.55}$$

The results from solving the subsets equations are summarized in Table 3.8. The solutions to the minimization problem of subsets 2, 5, 7 and 8 are excluded because their *AUP* values are outside the *AUP* range of the feasibility region. The results show that HO-$(CH_2)_7$-OH, HO-$(CH_2)_8$-OH, and HO-$(CH_2)_9$-OH are the formulations that satisfy all of the property and structural constraints. The true physical properties for the three candidate molecules were back calculated from the operator values of the solution.

Subset	g_1+g_2	g_1	g_2	Objective	FBN	Ω_1	Ω_2	Ω_3	Ω_4	AUP
1	8.1	6.1	2	min	0	20.2	78.3	51.2	1.7	151.4
	11.6	9.6	2	max	0	30.1	95.6	69.9	2.2	197.8
~~2~~	~~7~~	~~5~~	~~2~~	~~min~~	~~0~~	~~17.2~~	~~73.1~~	~~45.6~~	~~1.6~~	~~137.4~~
	14.2	12.2	2	max	0	37.4	108.3	83.5	2.5	231.6
3	8.1	6.1	2	min	0	20.2	78.3	51.2	1.7	151.4
	15	13	2	max	0	39.6	112.3	87.8	2.6	242.3
4	8.1	6.1	2	min	0	20.2	78.3	51.2	1.7	151.4
	15	13	2	max	0	39.6	112.3	87.8	2.6	242.3
~~5~~	~~6.3~~	~~4.3~~	~~2~~	~~min~~	~~0~~	~~15.1~~	~~69.4~~	~~41.7~~	~~1.5~~	~~127.8~~
	11.8	9.8	2	max	0	30.7	96.7	71.0	2.2	200.6
6	8.1	6.1	2	min	0	20.2	78.3	51.2	1.7	151.4
	11.6	9.6	2	max	0	30.1	95.6	69.9	2.2	197.8
~~7~~	~~7~~	~~5~~	~~2~~	~~min~~	~~0~~	~~17.2~~	~~73.1~~	~~45.6~~	~~1.6~~	~~137.4~~
	11.6	9.6	2	max	0	30.1	95.6	69.9	2.2	197.8
8	~~4.8~~	~~2.8~~	~~2~~	~~min~~	~~0~~	~~11.0~~	~~62.3~~	~~34.1~~	~~1.3~~	~~108.8~~
	11.6	9.6	2	max	0	30.1	95.6	69.9	2.2	197.8

Table 3.8. Result of solving to the molecular synthesis problem

In this section, an algebraic technique for molecular design based on molecular property clusters has been introduced. Using the developed concepts of molecular property operators (Section 3.2.3), this algebraic approach extends the application range of the original methodology to include more than three properties. The molecular design problem is solved to identify all possible formulations within the design space given a set of molecular building blocks as well as property and structural constraints. The linearity of the molecular operators plays an important role as it helps in lowering the complexity

of the design problem. The design problem is formulated as a set of linear algebraic equations. A simple example that had constraints in terms of four properties was solved successfully using this technique. The developed algebraic approach can be applied to problems that require both a higher number of properties as well as additional groups. The resulting optimization problems are simply larger, but would still consist of linear algebraic equations, thus lowering the complexity from a MINLP to a LP problem (Eljack et al., 2007a).

4. Molecular Synthesis Application Examples

4.1. Example 1 – Aniline Extraction Solvent Design

Liquid-liquid extraction involves the extraction of a substance from one liquid phase to another based on solution preferences. The success of the extraction is dependent on the immiscibility of the two liquid and the component's affinity for one substance over the other. Often one of the liquid phases is an aqueous solution and the other is an organic solvent. The selectivity of a suitable solvent is essential, so that the solute in the bulk solution (aqueous phase) has more affinity towards the added solvent, allowing for mass transfer of the solute from the bulk solution to the solvent.

In this molecular design application example, an aqueous solution containing Aniline is investigated. It is required to remove Aniline from solution in order to achieve product specifications (Eden et al., 2002)

4.1.1. Problem statement

Identify a solvent that will extract aniline from an aqueous solution. The required solvent characteristics include: immiscible with water, its solubility in water must be below that of Aniline's, and there needs to be a difference between the boiling points of the solvent and aniline to allow for the regeneration of the solvent after extraction. All the property data and molecular groups given as starting building blocks are summarized in Table 4.1.

Property	Lower Bound	Upper Bound	Molecular Building Blocks	
T_b (K)	350	431	CH_3	CH_2
H_v (kJ/mol)	36.7	46.8	CH_2CO	CH_3CO
V_m (cm^3/mol)	115	180	CH_2O	CH_3O
R^{ij} (MPa$^{1/2}$)	-	24		

Table 4.1: Property data and molecular groups for aniline design problem

4.1.2. Molecular Synthesis

The system is described by boiling temperature (T_b), heat of vaporization (H_v), molar volume (V_m) and solubility parameter (R^{ij}). The visual tool only allows for three properties, thus the first three properties are used in the design of the solvent and the last is chosen as a screening criteria. Property operators and the corresponding reference values required to transform the design data into molecular property clusters are given by equations 4.1-4.3.

$$\exp\left(\frac{T}{t_{bo}}\right) = \sum_{g=1}^{N_g} n_g \cdot t_{b1} \qquad T_{b, ref} = 0.1 \ K \qquad (4.1)$$

$$\Delta H_v - h_{vo} = \sum_{g=1}^{N_g} n_g \cdot h_{v1} \qquad H_{v, ref} = 0.5 \ kJ/mol \qquad (4.2)$$

$$V_m - d = \sum_{g=1}^{N_g} n_g \cdot v_1 \qquad V_{m, ref} = 2.0 \ cm^3/mol \qquad (4.3)$$

Solubility is measured as a function of the interactions between two substances (i,j). According to Hansen (1967), in a three-dimensional space solubility of component i is a function of non-polar (δ_d^i), polar (δ_p^i) and hydrogen-bonding (δ_h^i) parameters. Solubility of component i in j is considered feasible if the radius of interaction (R^{ij}) of i is found within that of j, $R^{ij} \leq R^j$. According to Barton (1985), the solubility parameter can be calculated according to the following (see Appendix B):

$$R^{ij} = \sqrt{4\left(\delta_d^i - \delta_d^j\right)^2 + \left(\delta_p^i - \delta_p^j\right)^2 + \left(\delta_h^i - \delta_h^j\right)^2} \qquad (4.4)$$

According to the Hoftyzer and Van Krevelen method, Hansen solubility parameters, based on group contributions, may be predicted using the following set of equations (Van Krevelen , 1976):

$$\delta_d = \frac{\sum F_{di}}{V_m} \qquad\qquad \delta_p = \frac{\sqrt{\sum F_{pi}^2}}{V_m} \qquad\qquad \delta_h = \sqrt{\frac{\sum E_{hi}}{V_m}} \qquad (4.5)$$

F_{di}, F_{ji}, E_{hi} are contributions from group i for calculating dispersion, polar, and hydrogen component solubility parameters, respectively (see Appendix B). Note that the molar volume (V_m) of a molecule is estimated by group contribution methods (Constantinou et al., 1995).

$$V - d = \sum_i g_i \cdot v_{1i} \qquad (4.6)$$

The bounded search space (sink) represented by the dotted line in Figure 4.1 is determined by six unique points, according to Rule 1. Figure 4.2 illustrates the molecular fragments included in the solvent synthesis. Eight different molecules are formulated (M1-M8) (see Figure 4.3). All of the candidates are structurally sound molecules whose locus lies within the sink. Next, the *AUP* values of the candidates are checked to see if they lay within the *AUP* range of the sink 182.4 - 215.9. Candidates M7 and M8 fail to satisfy this condition (see Table 4.2); the remaining molecules, M1-M6, satisfy all other criteria.
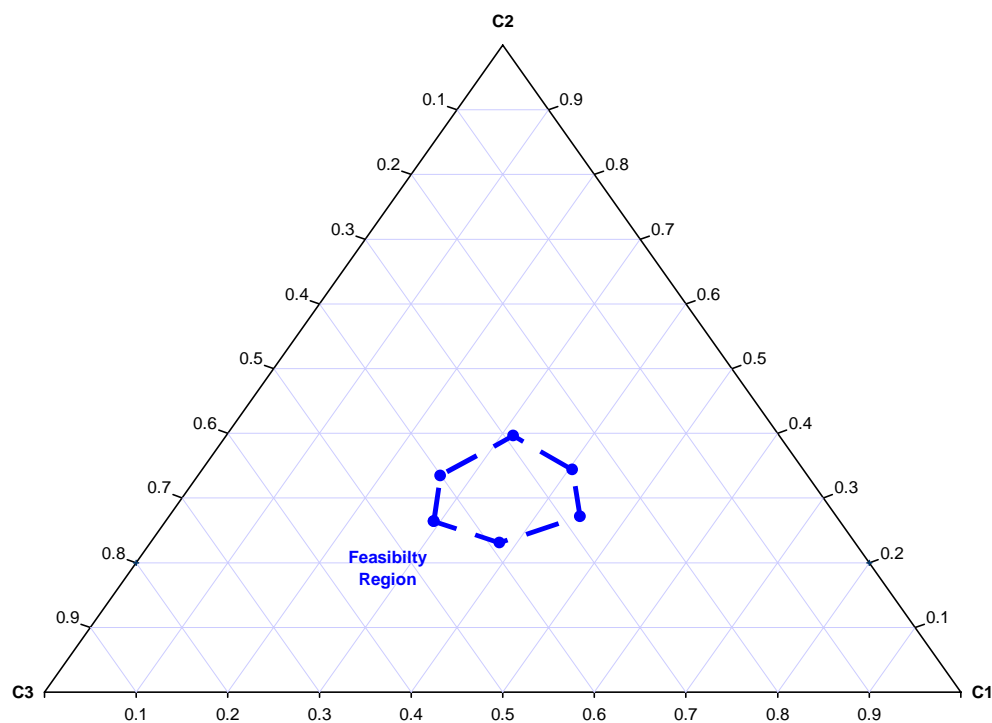
Figure 4.1: Feasibility region for aniline extraction solvent
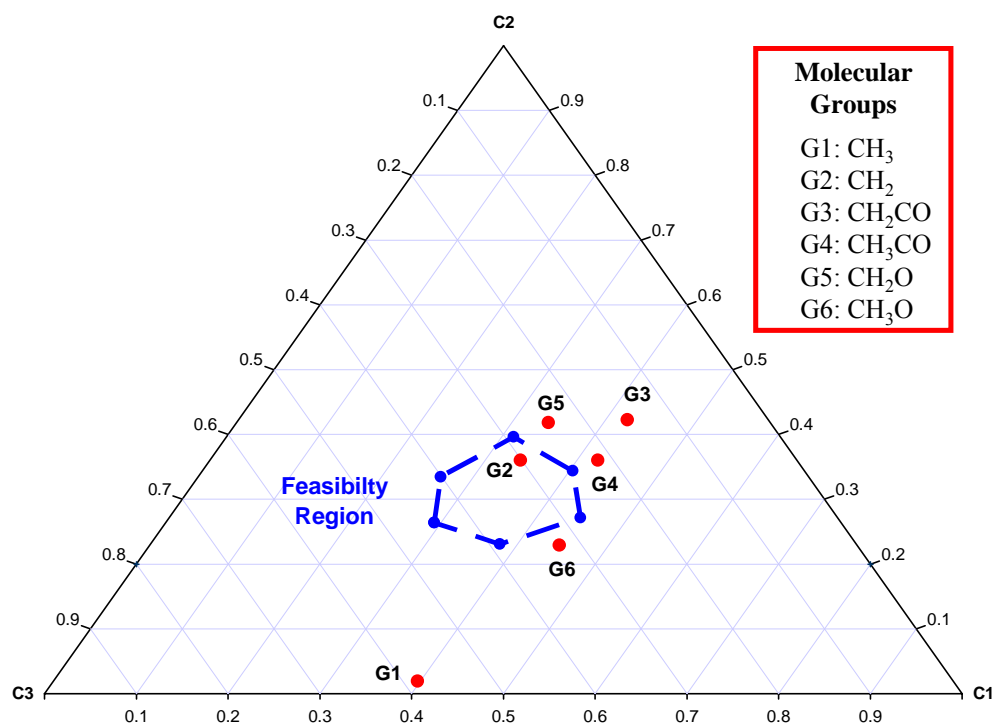
Figure 4.2: Aniline extraction solvent synthesis problem
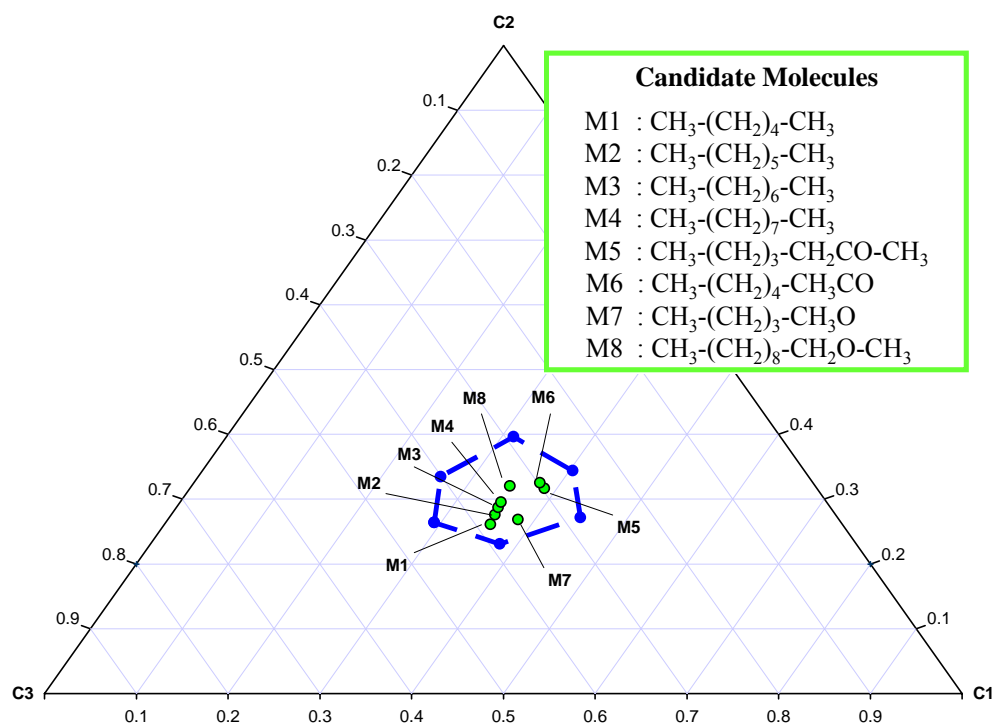
Figure 4.3: Candidates formulated for aniline extraction solvent

Formulations	AUP	T_b (K)	H_v (kJ/mol)	V_m (cm³/mol)	R^{ij} (MPa$^{1/2}$)
M1	153.8	347.2	31.8	130.0	15.5
M2	181.0	379.1	36.7	146.4	15.3
M3	208.3	406.6	41.6	162.9	15.1
M4	235.5	430.9	46.5	179.3	15.0
M5	218.4	436.0	46.3	141.8	14.9
M6	215.7	428.7	46.8	140.4	9.6
~~M7~~	~~310.6~~	~~486.0~~	~~61.4~~	~~218.8~~	~~10.7~~
~~M8~~	~~154.6~~	~~363.1~~	~~32.5~~	~~120.2~~	~~11.6~~

Table 4.2: Candidate solvents for aniline extraction

The final step is to determine the solubility parameter values for each of the designed molecules to address screening criterion constraints. The calculation of the solubility parameters and all required information is provided in Appendix B. Table 4.2 summarizes the four property values of the designed molecules. The solubility for all candidates is below that of aniline ($R^{ij} \leq 24$ MPa$^{1/2}$), thus the generated candidates satisfy the screening criteria. For verification, the property values predicted by the algorithm which are based on the contributions of the individual molecular fragments are checked against experimental values, see Table 4.3. The predicted values for the 1st order group contribution properties boiling temperature (T_b) and molar volume (V_m) are within 98% of the experimental values, the heat of vaporization (H_v) fall within 85% range, and those for Hansen's solubility parameter within 95 %. The precision of the property prediction method (1st order GCM) is sufficient for the design tools developed here as the method is intended as a first-cut conceptual design approach to determining feasible candidates.

The precision of the predicted properties plays a role in subsequent steps of validation as additional analysis and screening is certainly needed to refine the list of candidates and to rule out impractical alternatives.   In addition, the accuracy of the predicted properties is only dependent on the group contribution models and is not a reflection of the presented molecular clustering algorithm.

Property	[1] T_b (K)	[1] H_v (kJ/mol)	[1] V_m (cm^3/mol)	[2] R^{ij} ($MPa^{1/2}$)
Solvents	Predicted Values			
n-hexane	347.2	31.8	130.0	15.5
n-heptane	379.1	36.7	146.4	15.3
n-octane	406.6	41.6	162.9	15.1
n-nonane	430.9	46.5	179.3	15.0
	Experimental Values			
n-hexane	342.2	37.6	130.5	14.9
n-heptane	371.6	42.6	147.4	15.3
n-octane	398.9	45.9		15.5
n-nonane	423.7	50.5	179.7	15.6
	Predicted Value Percent Error – compared to experimental			
n-hexane	1.5%	15.4%	0.4%	4.1%
n-heptane	2.0%	13.8%	0.7%	0.1%
n-octane	1.9%	9.3%	-	2.5%
n-nonane	1.7%	7.8%	0.2%	4.0%

Experimental data obtained from the following references:

[1] *Perry's Chemical Engineering Handbook (Green, 1997)*

[2] *Handbook of Solubility Parameters (Barton, 2000)*

Table 4.3: Accuracy of predicted properties values

Formulations 2-heptanone (M6) and 4-heptanone (M5), from the provided list of candidates, posses the same number of atoms ($C_7H_{14}O$), although the molecules are synthesized from different first order fragments. The molecular makeup of 2-heptanone

involves CH_3, CH_2, and CH_3CO; while 4-heptanone includes CH_3, CH_2 and CH_2CO. Hence, the molecular clustering technique is able to synthesize isomers; but in order to differentiate between the isomers in terms of predicted properties, the inclusion of 2nd and 3rd order groups is critical. Section 6.2 focuses on this issue as part of the future directions for advancing the molecular clustering methodology.

4.2. Example 2 - Blanket Wash Solvent Design

Lithographic printers are used to print a variety of products including books and newspapers. Offset presses in industry transfer the printed image from a plate to a rubber or plastic blanket and then to the paper or other medium being used. The produced quality images are greatly dependent on the cleanliness of the blanket. Blanket washes, consisting of a variety of solvents, are used to remove ink, paper dust, and other debris from the blanket cylinders. They are generally petroleum-based solvents that consist of volatile organic compounds (VOCs), which are found in the printing ink as well. Reasonably, there is a lot of concern regarding the effects of such solvents on the environment as well as the direct effect on human health.

Blanket solvents are designed to address specific needs: Minimal drying time, liquid at room temperature, low vapor pressure (*VP*), and to dissolve the ink, solubility (R_{ij}) of the solvent is an important factor. Such demands on product performance can be described using properties. The drying time is related to the heat of vaporization (H_v), and the state of the solvent at room temperature is directly related to melting (T_m) and boiling (T_b) temperatures.

4.2.1. Problem Statement

The objective of this case study is to design a blanket wash solvent for a phenolic resin printing ink, specifically '*Super Bakacite ® 1001, Reichold'*. Formulations are designed from a bank consisting of 7 possible groups, with a maximum formulation length of 7 groups. The design of the solvent involves the properties and constraints listed in Table 4.4. Sinha and Achenie (2001) solved this design problem as a mixed-integer non-linear programming problem (MINLP). In this paper, the problem is solved using the developed molecular property clusters (Eljack and Eden, 2007).

Property	Lower Bound	Upper Bound
Hv (kJ/mol)	20	60
Tb (K)	350	400
Tm (K)	150	250
VP (mmHg)	100	---
R^{ij}	0	19.8

Table 4.4: Property constraints for blanket wash solvent

4.2.2. Property Prediction (GCM)

Visualization of the design problem on the ternary diagram, dictates the use of only three properties. In this approach heat of vaporization, boiling and melting temperatures are used, with vapor pressure and solubility used as final screening properties. First order group contribution (GCM) equations are used to predict the first three properties (Constantinou and Gani, 1994):

$$\Delta H_v - h_{vo} = \sum_i g_i \cdot h_{vi} \tag{4.7}$$

$$T_b = t_{bo} \cdot \exp\left(\sum_i g_i \cdot t_{bi}\right)$$

(4.8)

$$T_m = t_{mo} \cdot \exp\left(\sum_i g_i \cdot t_{mi}\right)$$

(4.9)

Group contribution (GCM) does not include vapor pressure in its bank of properties. Vapor pressure is predicted using the McGowon Hovarth Equation, as a function of boiling and operating temperatures (Sinha and Achenie, 2001):

$$\log VP \ (mmHg) = 5.58 - 2.7\left(\frac{T_{bp}}{T}\right)^{1.7}$$

(4.10)

The effectiveness of the designed solvent is greatly dependent on its power to dissolve the ink, i.e. it is dependent on the solubility power of the designed solvent. The interactions between phenolic resin molecules (solute) with the solvent molecules are very important in this design problem. Solubility is determined according to Hansen parameters (see equations 4.4 – 4.6, Appendix B).

4.2.3. Molecular Property Operators

Now the properties used to achieve the targets are heat of vaporization (H_v), boiling (T_b) and melting temperature (T_m). Notice that only these properties are used initially, again this is to be able to visualize the design problem on the ternary diagram, however an algebraic and optimization based approach to solve molecular design

problems with more than three properties has recently been introduced by Eljack and Eden (2007). The other properties, vapor pressure and solubility are non-group contribution properties, and thus will be used post molecular synthesis to screen the designed solvents.

The property operators derived from equations 4.7-4.9 and their reference values are summarized in Table 4.5. Notice again that RHS of the equations allow/exhibit linear additive rules.

Property	LHS of equation Ψ_j^M	RHS of equation 1st order GC expression	Reference values
Standard Heat of Vaporization	$\Delta H_v - h_{vo}$	$\sum_{g=1}^{N_g} n_g \cdot h_v$	20
Normal Boiling Temperature	$\exp\left(\dfrac{T}{t_{bo}}\right)$	$\sum_{g=1}^{N_g} n_g \cdot t_b$	7
Melting Temperature	$\exp\left(\dfrac{T}{t_{mo}}\right)$	$\sum_{g=1}^{N_g} n_g \cdot t_m$	7

Table 4.5: Property operators for blanket wash solvent problem.

4.2.4. Molecular Synthesis

The problem is visualized by converting the property targets to cluster values following the methodology described in Section 3.2. The property constraints are represented as a feasibility region, which is identified according to the feasibility rules highlighted in Section 3.1.5. The resulting ternary diagram is shown in Figure 4.4, where the dotted lines represent the feasibility region for the solvent design.
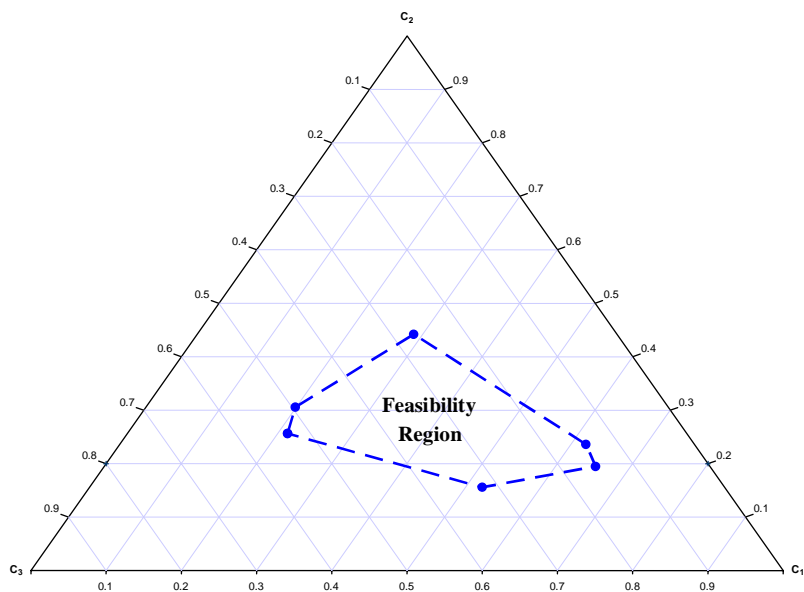
Figure 4.4: Feasibility region for blanket wash solvent problem.
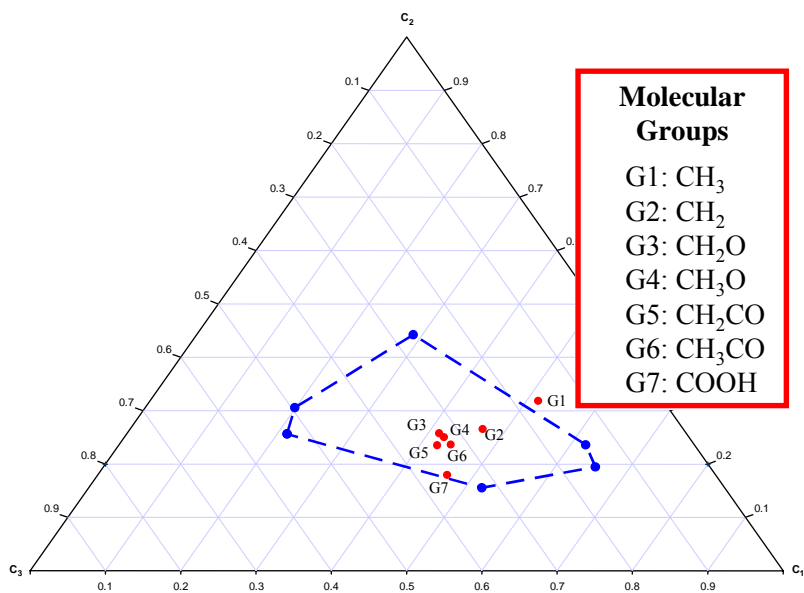
Figure 4.5: Blanket wash solvent synthesis problem

Notice that even though some of the property operators formulated earlier are very complex, molecular synthesis on the ternary diagram is still simple because these operators obey simple linear additive rules. It should be noted again, that the location of the formulated molecules is independent of group addition path. The molecules that will be designed in this domain can be made up of seven chemical groups. Carboxyl and methyl groups are amongst the selection. All the groups used in this synthesis problem are presented on Figure 4.5. It should be emphasized that these are the same fragments used by Sinha and Achenie (2001) to solve their MINLP problem.

A number of candidate molecules, M1-M11, are formulated on the ternary diagram, see Figure 4.6, however to exhaust all possible combinations of molecular building blocks to provide a complete list of candidates requires the development of a software implementation. The validity of the designed formulations is satisfied only after satisfying the conditions summarized by Rules 4-8 in Section 3.2. The cluster values of the designed molecules are checked to make sure that they lie within the sink. The values of the augmented property index (AUP) of the designed molecule must lie within the AUP range of the sink; which in this example ranged from 2.28-5.09, see Table 4.6. It can be seen that molecules M9-M11 fail to satisfy this condition.

The final necessary and sufficient conditions is that the property values of the remaining new formulations must lie within the upper and lower constraints placed on the molecular design problem, which includes the Non-GC properties. The property values for the new formulations are back calculated using the methodology outlined earlier in Section 3.2. The remaining formulated solvents satisfy the necessary and sufficient conditions. Consequently, M1-M8 are the final valid formulations shown in Figure 4.7.

The candidate molecules M1-M7 identified visually in this work corresponds to the solutions found by the MINLP approach used by Sinha and Achenie (2001). Formulation $M1$ is a cyclical molecule. Such molecules can be excluded ahead of design by simply placing another constraint on the problem. Molecules $M2$, $M3$ and $M4$ are ethers and $M7$ is known as methyl ethyl ketone (MEK), commonly found in printing inks (Sinha and Achenie, 2001). The key here is that blanket wash solvents are usually ketones or ethers, these aforementioned formulations are common components in commercial blanket wash solvents (United States Environmental Protection Agency (EPA), 1997). The final valid formulation, $M8$ is heptane and although the property values match the targets; it is not an ideal solvent for this case because it is highly flammable (Material Safety Data Sheet, 2006).

Formulations	AUP	H_v (kJ/mol)	T_b (K)	T_m (K)	VP (mmHg)	R^{ij} (MPa$^{1/2}$)
M1	3.20	33.91	359.14	201.24	1117.85	10.88
M2	3.08	33.99	355.34	189.86	1240.95	15.39
M3	3.10	34.67	364.49	183.38	963.57	12.83
M4	3.17	35.81	363.09	186.26	1001.85	18.09
M5	3.47	36.15	370.61	211.95	811.54	15.82
M6	3.61	36.74	382.51	216.49	578.05	11.31
M7	3.17	35.10	354.80	193.13	1259.28	16.84
M8	3.28	38.31	379.07	175.55	638.03	19.77
M9	6.79	68.87	457.92	286.76	56.69	13.83
M10	7.79	78.17	494.54	297.68	16.55	12.68
M11	7.83	74.75	535.55	292.08	3.86	11.33

Table 4.6: Candidate blanket wash solvents.
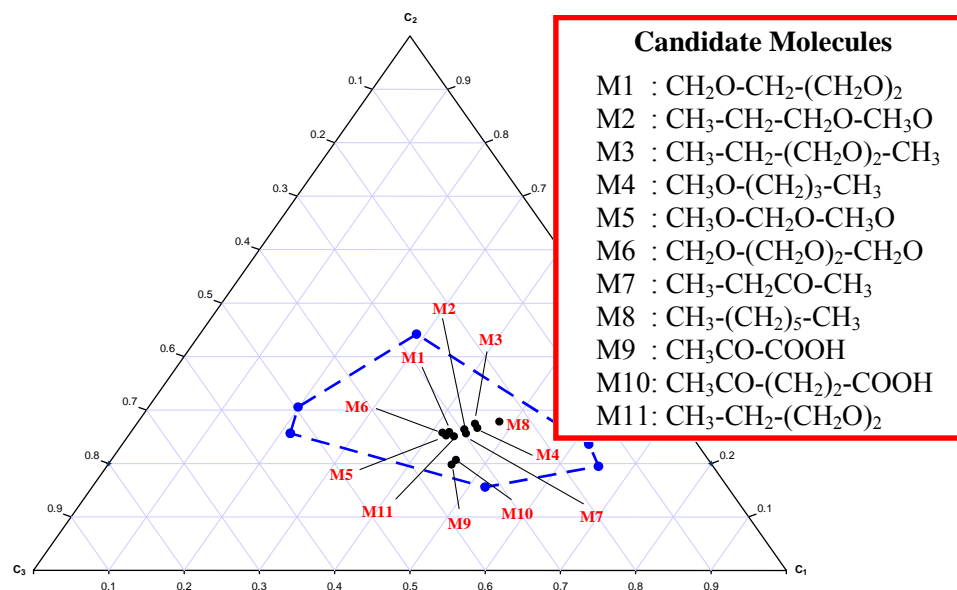
Figure 4.6: Candidate formulations for blanket wash solvent
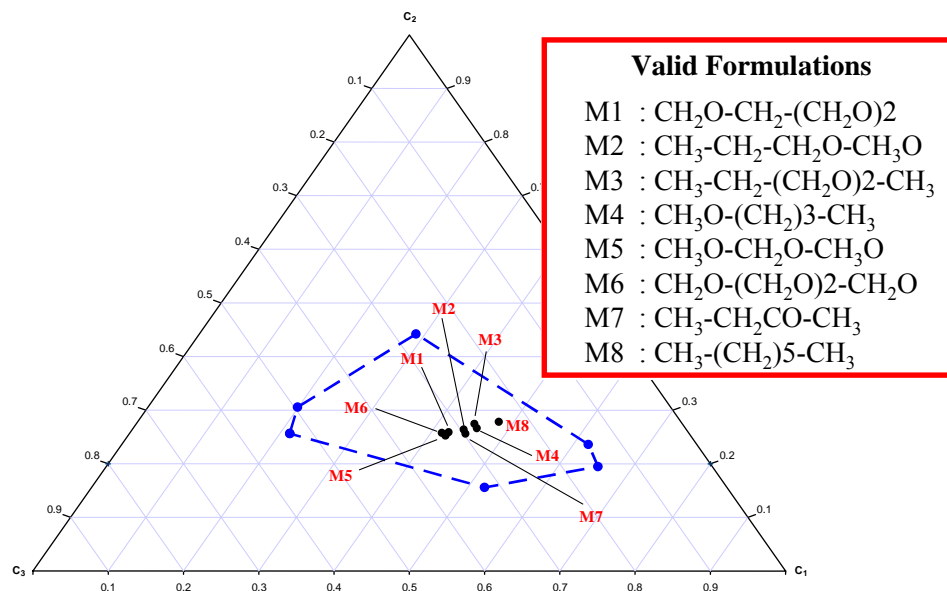
Figure 4.7: Valid formulations for blanket wash solvents.

4.3. Summary

A significant result of the developed methodology is that for problems that can be satisfactorily described by just three properties, the molecular design problem is solved visually on a ternary diagram, irrespective of how many molecularly fragments are included in the search space. This solvent design case study also showed how regardless of the group addition path chosen, the final location of a designed formulation on the molecular ternary diagram is unique.

5. Simultaneous Process and Molecular Design

The molecular clustering technique was initially developed as a means of providing a bridge that will facilitate the flow of information between process and molecular design algorithms. In this chapter two application examples are solved used the developed simultaneous technique.

5.1. Application Example 1 - Metal Degreasing Process

A case study is presented here to show the merits of using the simultaneous approach via GCM and property clusters. Figure 5.1 illustrates a metal degreasing facility that consists of an absorber and a degreaser. The fresh resources are in the form of two organic solvent streams (Shelley and El-Halwagi, 2000). The off-gas Volatile Organic Compounds (*VOC*s) are byproducts from the degreasing unit, and the current treatment of this stream is flaring. Such treatment methods lead to economic loss and environmental pollution (Eden, 2003).
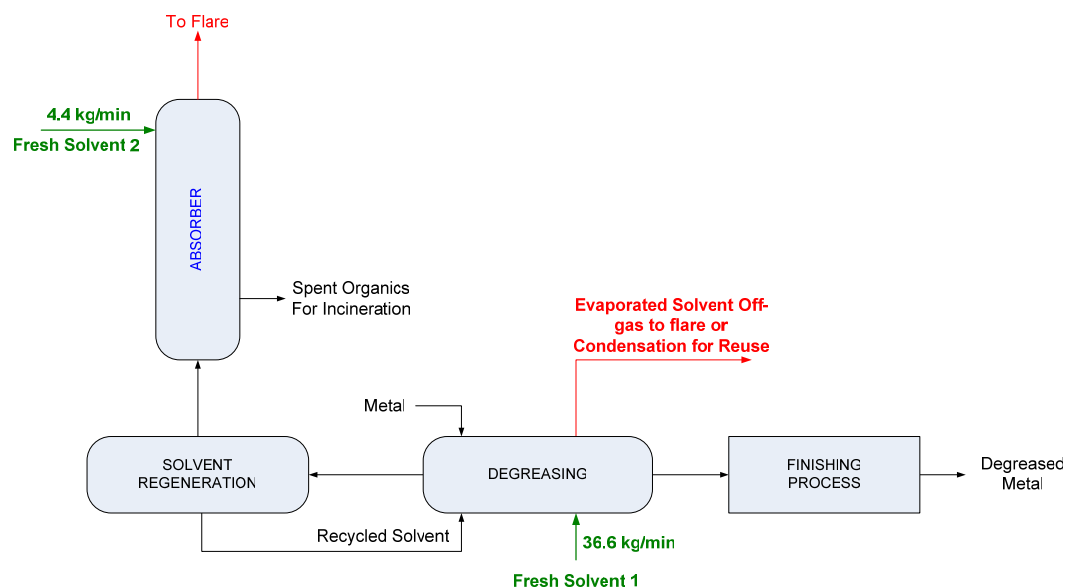
Figure 5.1: Original metal degreasing process.

In this case study, the objective is to explore the possibility of condensing the off gas VOCs, to (1) optimize the use of fresh solvent and (2) to simultaneously identify candidate solvents for the degreaser (see Figure 5.2). Three properties are examined to determine the suitability of a given organic process fluid for use in the degreaser:

- Sulfur content (S) - for corrosion considerations, expressed as weight percent.

- Molar Volume (V_m) - for hydrodynamic and pumping aspects.

- Vapor Pressure (VP) – for volatility, makeup and regeneration.

The synthesized solvents will be pure components; thus the sulfur content of these streams will be zero, as no sulfur containing groups will be included in the fragment search space.
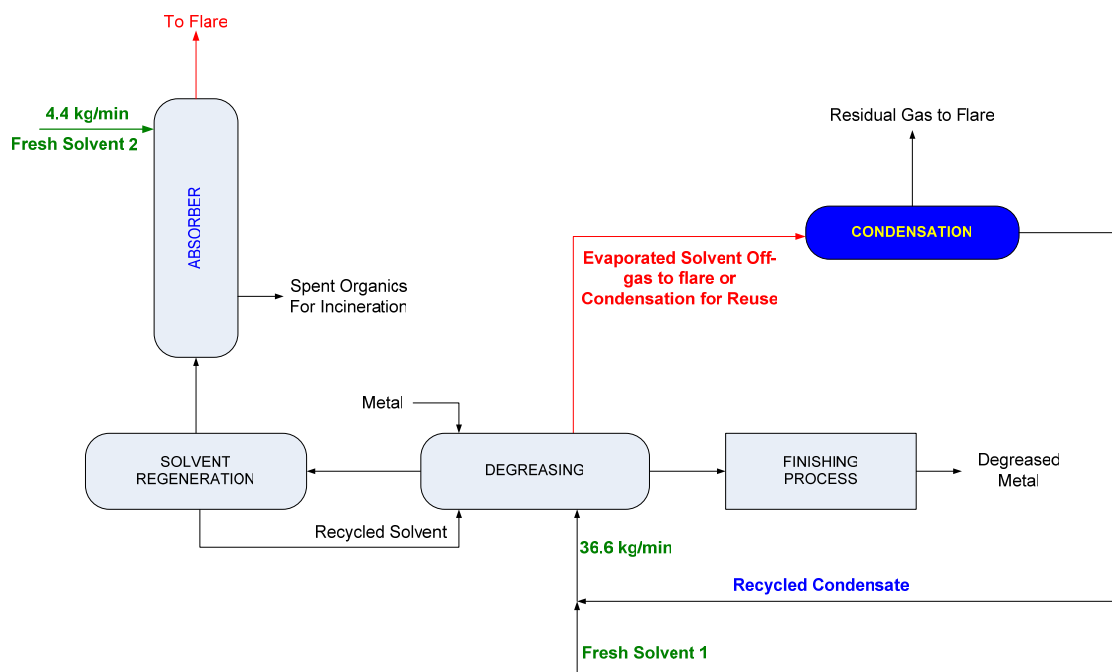
Figure 5.2: Metal degreasing process after property integration

5.1.1. Process Design

The constraints on the inlet streams to the degreaser are given in Table 5.1:

Property	Lower Bound	Upper Bound
S (%)	0.00	1.00
V_m (cm^3/mol)	90.09	487.80
VP (mmHg)	1596	3040
T_b (K)	430.94	463.89
Flow rate (kg/min)	36.6	36.8

Table 5.1: Degreaser feed constraints

The process property operator mixing rules needed to describe the system are given by the following equations:

$$S_M = \sum_{s=1}^{Ns} x_s \cdot S_s \qquad , \qquad S_{ref} = 0.5 \text{ wt\%} \qquad (5.1)$$

$$V_{m_M} = \sum_{s=1}^{Ns} x_s \cdot V_{ms} \qquad , \qquad V_m{}^{ref} = 80 \text{ cm}^3/\text{mol} \qquad (5.2)$$

$$VP_M^{1.44} = \sum_{s=1}^{Ns} x_s \cdot VP_s^{1.44} \quad , \qquad VP^{1.44, \, ref} = 760 \text{ mmHg} \qquad (5.3)$$

Samples of the off-gas were taken, and then condensed at various temperatures ranging from 400-550 K, providing measurements of the three properties as well as the flowrate of the condensate (Shelley and El-Halwagi, 2000). The data for the degreaser unit and for the VOC condensate are converted to cluster values according to cluster methodology developed by Eden et al. (2004) and discussed in Section 3.1.2 (see Figure 5.3). The degreaser property constraints are translated into a feasibility region according to the procedure highlighted in Section 3.1.5.

Now that the problem has been mapped to the property domain and visualized on the ternary diagram, some additional constraints are placed on the process: The condensation temperature is set to 500K and the fresh synthesized solvents have no sulfur containing groups. By fixing the condensation temperature at 500K, the locus of possible solvents is bound by straight lines between the condensate and points A and B (see Figure 5.4). This adheres to the first constraint. Applying the second constraint on the process (no sulfur in fresh solvent), shows that the cluster solution to the degreaser

problem corresponds to all points between points A and B on the C_2-C_3 axis (Eljack et al., 2007b).
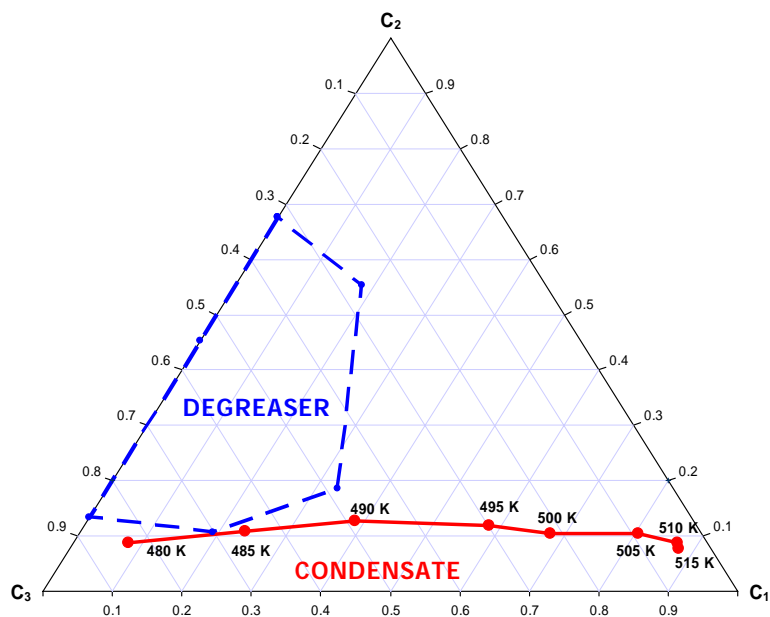
Figure 5.3: Metal degreasing problem in process design
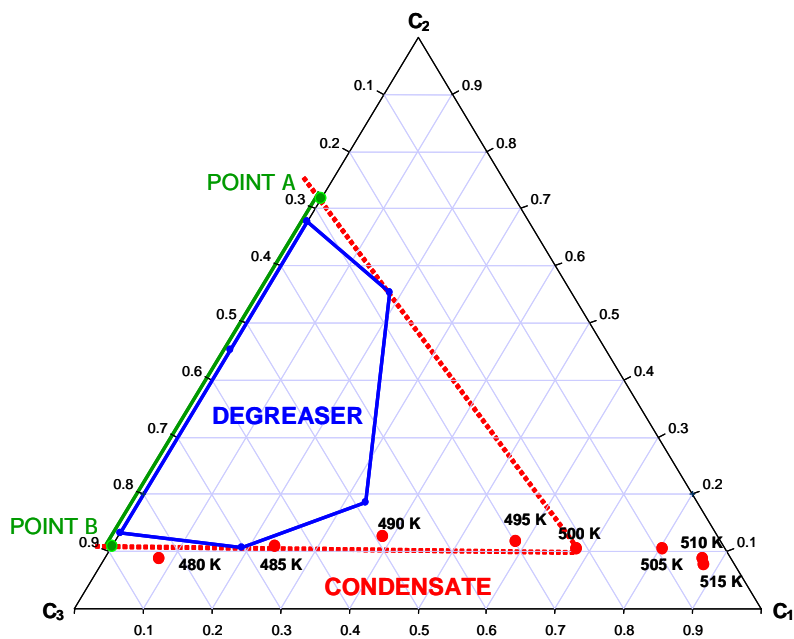
Figure 5.4: Property targets of solvent for maximum condensate recycle.

5.1.2. Molecular Design: Fresh Solvent Synthesis

Once all the constraints have been taken into account and the property targets for molecular formulations have been fixed by the process design problem the second phase of this case study begins.

The cluster values associated with points A and B from the clustering diagram in Figure 5.4, are translated to physical property values using the methodology developed by Shelley and El-Halwagi (2000) and Eden (2003). These property targets obtained from solving the process design problem are now the upper and lower property constraints placed on the solvent/molecular design problem, see Table 5.2.

The zero sulfur constraint placed on the problem provides an extra degree of freedom. So a heat of vaporization constraint is now placed on the fresh solvent problem. Now the properties used to describe the problem are heat of vaporization (H_v), boiling temperature (T_b) and molar volume (V_m). Notice that boiling temperature is used instead of vapor pressure since there is no direct group contribution method for predicting vapor pressure. However, according to equation 4.10, vapor pressure is a function of boiling temperature. Hence, the vapor pressure property constraints are converted to boiling temperature upper and lower limits. All of the property constraints on the molecular design problem are now shown in Table 5.3.

	S (%)	VP (mmHg)	V_m (cm^3/mol)
Point A	0.00	1825.4	720.8
Point B	0.00	3878.7	102.1

Table 5.2: Property constraints obtained from process design problem.

Property	Lower Bound	Upper Bound
H_v (kJ/kg)	50	100
VP (mmHg)	1825.4	3878.7
V_m (cm^3/mol)	90.1	720.8
T_b (K)	418.01	457.16

Table 5.3: Revised property constraints for fresh solvent synthesis.

The physical properties are predicted using the following 1st order group contribution equations:

$$\Delta H_v = h_{vo} + \sum_i n_i \cdot h_{v1_i} \tag{5.4}$$

$$V_m = d + \sum_i n_i \cdot v_{1_i} \tag{5.5}$$

$$T_{bo} = t_{bo} \cdot \ln \sum_i n_i \cdot t_{b1} \tag{5.6}$$

The property operators derived from the above equations and their reference values are summarized in Table 5.4. Notice again that RHS of the equations exhibit linear additive rules.

Property	LHS of equation Ψ_j^M	RHS of Equation 1st order GC Expression	Reference values
Standard Heat of Vaporization	$\Delta H_v - h_{vo}$	$\sum_{g=1}^{N_g} n_g \cdot h_{v1}$	20
Molar Volume	$V_m - d$	$\sum_{g=1}^{N_g} n_g \cdot v_1$	100
Normal Boiling Temperature	$\exp\left(\dfrac{T}{t_{bo}}\right)$	$\sum_{g=1}^{N_g} n_g \cdot t_{b1}$	7

Table 5.4: Property operators needed for molecular synthesis

The problem is visualized by converting the property targets to cluster values following methodology described in Table 3.4. The property constraints are represented as a feasibility region, as outlined in Section 3.1.4. The resulting ternary diagram is shown in Figure 5.5, where the dotted lines represent the feasibility region in the molecular design domain.

The molecules to be designed can be made up of eight chemical groups. Carboxyl, methyl, and amine groups are amongst the selection. All the groups used in the molecular synthesis problem are shown in Figure 5.5.

Notice that even though some of the property operators formulated earlier are very complex, molecular synthesis on the ternary diagram is still simple because these operators obey simple linear additive rules. Seven candidates, M1-M7, are formulated for this solvent design problem (see Figure 5.6). However, the validity of the formulations is satisfied only after satisfying the conditions summarized by Rules 4-8 in Section 3.3. The cluster values of the designed molecules, M1-M7, are checked to make

sure that they lie within the sink. The values of the augmented property index of each designed molecule must lie within the *AUP* range of the sink; in the degreaser case study the *AUP* range of the sink is 4.22-12.65, see Table 5.5. It is seen that molecules M5 and M6 fail to satisfy this condition.
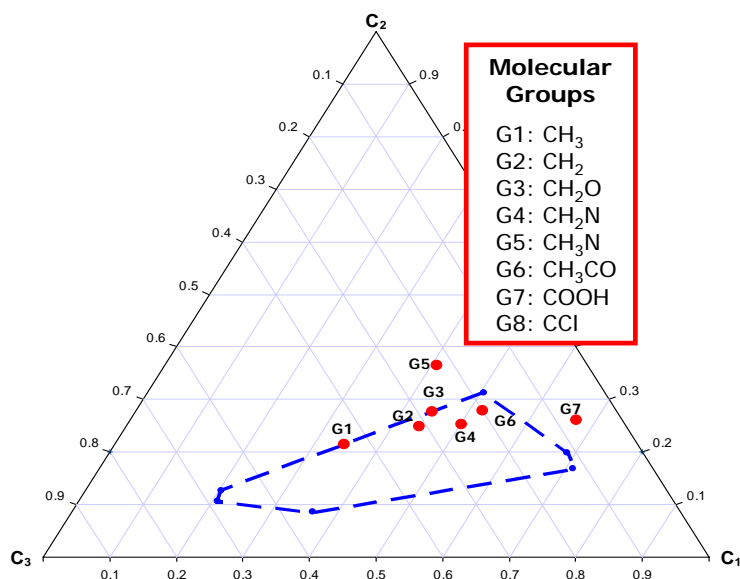
Figure 5.5: Metal degreasing solvent problem.

The final necessary and sufficient condition is that the property values of the new formulations must lie within the upper and lower constraints placed on the molecular design problem, which includes the Non-GC property constraints. The property values for the new formulations are back calculated using the methodology outlined in Section 3.3. Molecule M3 did not match the heat of vaporization property range of the sink; and although M7 satisfies the three GC properties, H_v, V_m and T_b, it fails to satisfy the Non-GC property for vapor pressure.
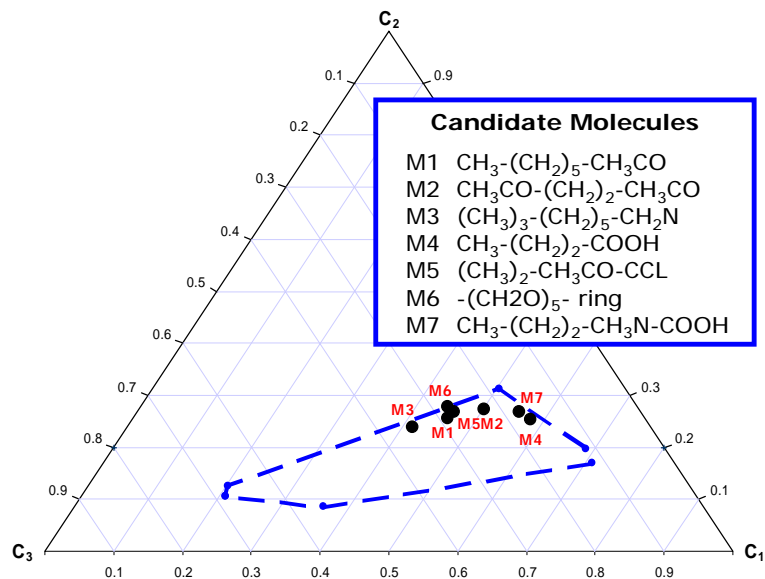
Figure 5.6: Candidate metal degreasing solvents.

Formulation	AUP	T_b (K)	H_v (kJ/mol)	V_m (cm^3/mol)	VP (mmHg)
M1	5.06	450.58	53.19	156.85	2078.98
M2	4.71	448.54	54.13	118.03	2163.90
M3	5.11	437.29	49.35	189.41	2692.07
M4	4.86	438.97	63.29	93.39	2606.12
M5	4.02	413.20	43.88	121.14	4241.48
M6	4.19	428.11	44.22	127.66	3208.12
M7	5.71	485.01	70.24	112.52	1037.99

Table 5.5: Candidate molecules for metal degreasing problem.

Consequently, M1, M2, and M4 are the final valid formulations. After searching the ICAS database (CAPEC 2006), M1, M2 and M4 correspond to 2-octanone, 2,5-hexadione, and butanoic acid respectively. The valid molecular structures are shown in

Figure 5.7. The three candidates are mapped back to the process design framework to identify the formulation that will maximize recycle of condensate at 500K (Eljack et al., 2007b). Using lever arm analysis, 19.36 kg/min of fresh 2,5-hexadione will allow for a maximum condensate flow rate of 17.44 kg/min.
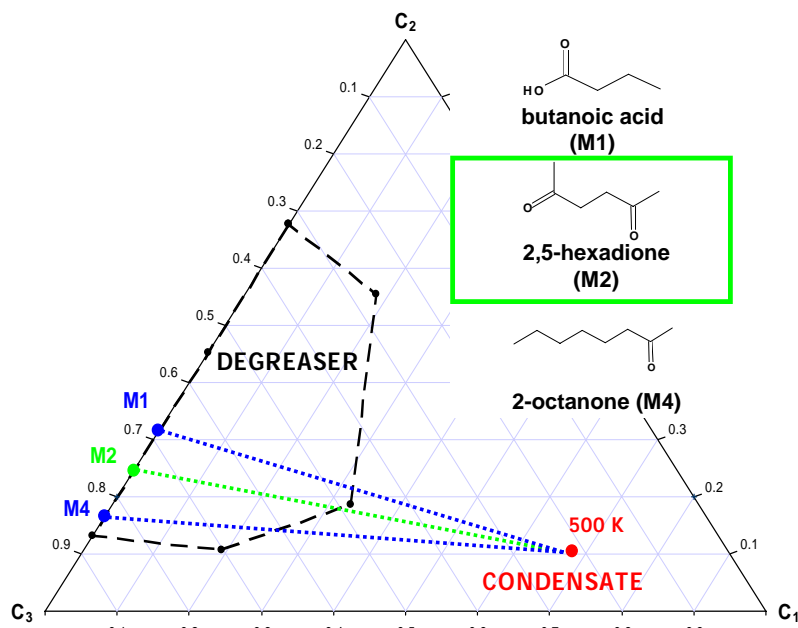
Figure 5.7: Selection of metal degreasing solvent.

5.1.3. Summary

This case study illustrates a systematic property based framework for simultaneous solution of process and molecular design problems. Using property clusters, the process design problem is solved to identify the property targets corresponding to desired process performance. The molecular design problem is solved to generate structures that match these targets. The approach provides a unifying framework that uses the physical properties to interface the process and molecular design problems.

5.2. Application Example 2 – Gas Purification

5.2.1. Problem Statement

A current gas treatment process uses fresh methyl diethanol amine, MDEA, (HO-$(CH_2)_4$-CH_3N-OH) and two other recycled process sources ($S1$, and $S2$) as a feed into the acid gas removal unit (AGRU). Another process stream, $S3$, currently a waste stream could be recycled as a feed if mixed with a fresh source to allow the mixed stream properties to match the sink (Kazantzi, 2006, Kazantzi et al., 2007). The property and flowrate data for all streams ($S1$, $S2$ and $S3$) and the sink are summarized in Table 5.6.

Design objectives and requirements: identify a solvent that will replace MDEA as a fresh source and that will maximize the flowrate of all available sources ($S1$, $S2$ and $S3$). The solvent must then posses similar characteristics to that of MDEA and thus the molecular building blocks are limited to OH, CH_3N and CH_2. The designed solvent should be a diol in order to posses MDEA characteristics. The sink performance requirements are functions of critical volume (V_c), heat of vaporization (H_v) and heat of fusion (H_{fus}).

Property	Lower Bound	Upper Bound	$S1$	$S2$	$S3$
V_c (cm^3/mol)	530	610	754	730	790
H_v (kJ/mol)	100	115	113	125	70
H_{fus} (kJ/mol)	20	40	15	15	20
Flowrate (kmol/hr)	300		50	70	30

Table 5.6: Property data for gas purification example

5.2.2. Process Design

The first step in implementing the simultaneous clustering approach requires the transformation of all process sources and sinks from the property domain to the cluster, using equations 3.2, 3.4-6. The process property operator mixing rules for the three properties critical volume, heat of vaporization and heat of fusion (ψ_1, ψ_2, ψ_3) are defined by the following equations:

$$V_{cM} = \sum_{s=1}^{Ns} x_s \cdot V_{cs} \qquad\qquad , V_{c,\,ref} = 2.5 \; cm^3/mol \qquad\qquad (5.7)$$

$$H_{vM} = \sum_{s=1}^{Ns} x_s \cdot H_{vs} \qquad\qquad , H_{v,\,ref} = 0.35 \; kJ/mol \qquad\qquad (5.8)$$

$$H_{fus\,M} = \sum_{s=1}^{Ns} x_s \cdot H_{fus\,s} \qquad\qquad , H_{fus,\,ref} = 0.10 \; kJ/mol \qquad\qquad (5.9)$$

Boundary constraints of the sink will be determined according to Rule 1 by six unique points seen as FP1-FP6 on Figure 5.8; while the sources are represented by discrete points. Notice the lumped source (S_L) point on the diagram; it represents the mixture property value of the three recycle streams ($S1$, $S2$, $S3$); the resulting data is shown in Table 5.7.
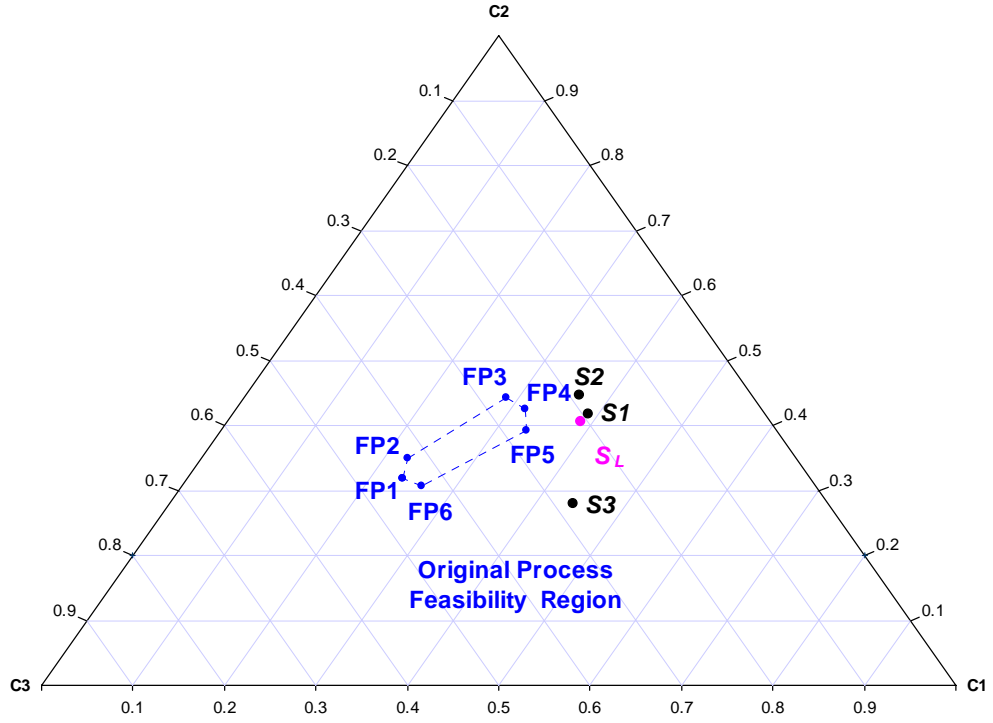
Figure 5.8: Gas purification process – feasibility regions and streams

Source	V_c cm^3/mol	H_v kJ/mol	H_f kJ/mol	Flowrate kmol/hr	Ω_1	Ω_2	Ω_3	AUP
S1	754	113	15	50	301.6	322.9	150	774.5
S2	730	125	15	70	292	357.1	150	799.1
S3	790	70	20	30	316	200.0	200	716.0
Lumped source (S_L)	750	110	16	150	300	314.3	160	774.3

Table 5.7: Mixture property data of lumped source (S_L)
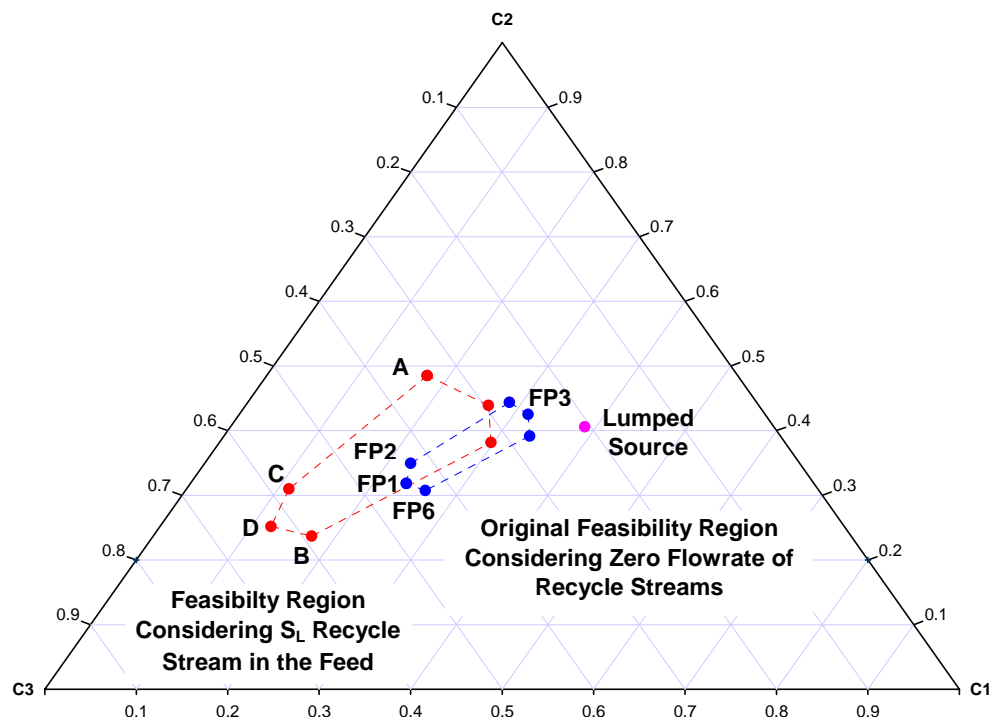
Figure 5.9: New feasibility region – reflects mixture/blend design constraints

The synthesis of new molecules is dependent on the process constraints; and it will be designed as a blend/mixture formulation. Two streams will be recycled to the process sink, the lumped source (S_L) at 150 kmol/hr and the newly designed solvent at a rate of 150 kmol/hr in order to fulfill the 300 kmol/hr flowrate constraint of the sink. In Figures 5.8 and 5.9, there are two feasibility regions. The first reflect the sink's original property demands as seen in Table 5.6, and the second is the newly defined search space that integrates the process requests to have the new designed molecule mix with the lumped source stream at a fractional flowrate contribution (x_L) of 0.5.
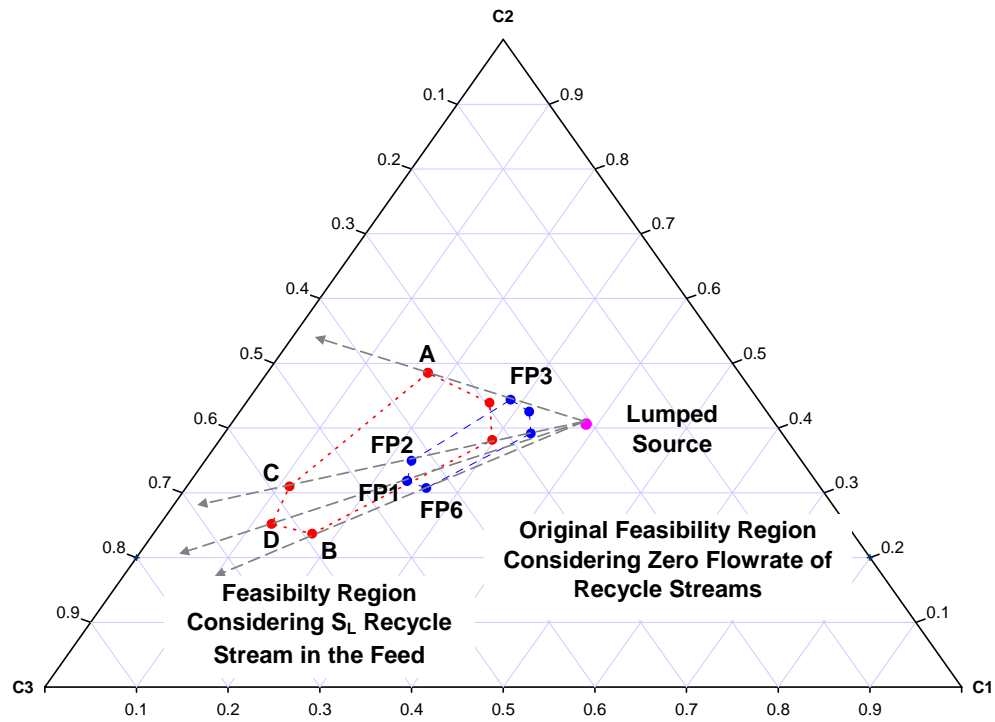
Figure 5.10: Identification of mixture (new) feasibility region
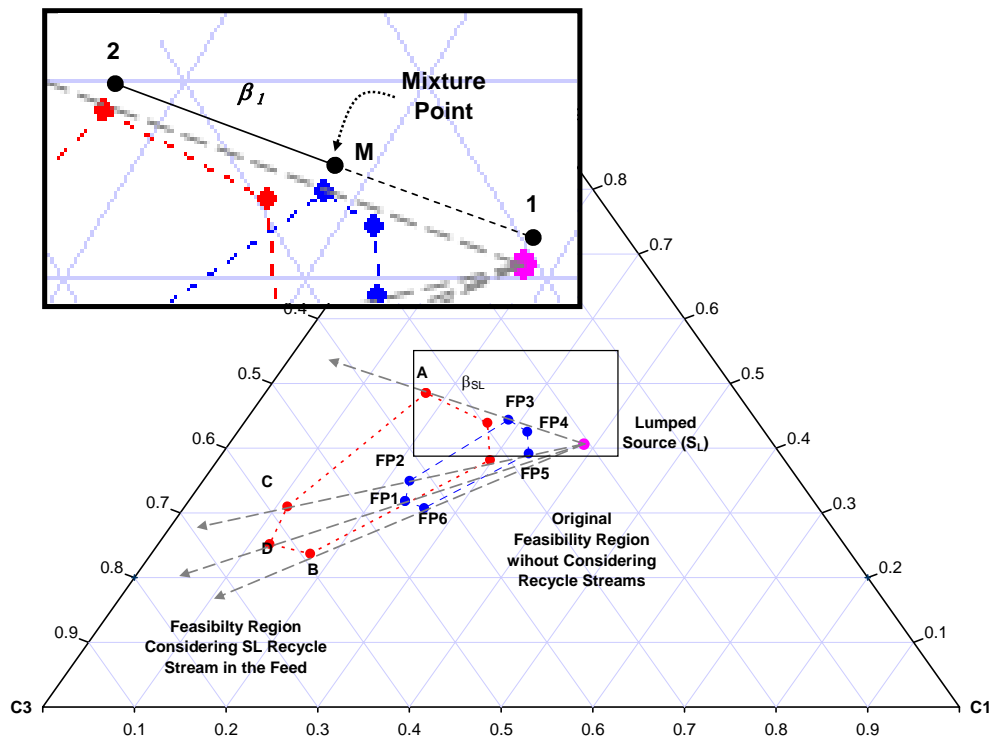
Figure 5.11: New feasibility region – Gas Purification Example

The mixed feasibility region points (A, B, C and D) can be easily determined using lever arm analysis. Taking advantage of the visual aid, it is easily determined that the new region is bounded by points [FP4, FP3, A, C, D, B, FP6 and FP5]. Points A – D are the only unknown points, the remaining are already established. The cluster values of points A – D are calculated using the lever-arm rules (Section 3.1.2). An example step by step calculation is shown here for the case of determining point A. For generalization, the line segment connecting points S_L and A in Figure 5.11 has been magnified, with points S_L and A, shown as points *1* and *2* respectively in the magnification. FP3 on the line marks the location of the mixture point, now represented by *M*. The cluster values for points *1* and *M* are given in Table 5.8

The mixture point *M* also marks the location of the relative cluster arm β_1, in the magnification. Given that, x_1, AUP_1 and AUP_M are known; equation 3.13 is used to calculate the value of β_1.

$$\beta_s = \frac{x_s \cdot AUP_s}{AUP_M}$$ (3.13)

Next, the cluster values (C_{12}, C_{22} and C_{32}) for point *2* on Figure 5.11 are calculated according to the cluster conservation rule. Expanding equation 3.12, results in the following:

$$C_{jM} = \sum_{s=1}^{N_s} \beta_s \cdot C_{js}$$ (3.12)

$$C_{1M} = \beta_1 \cdot C_{11} + (1 - \beta_1) \cdot C_{12}$$
$$C_{2M} = \beta_1 \cdot C_{21} + (1 - \beta_1) \cdot C_{22} \qquad (5.10)$$
$$C_{3M} = \beta_1 \cdot C_{31} + (1 - \beta_1) \cdot C_{32}$$

The steps outlined above are used to determine the remaining points B – D (see Table 5.8). The six cluster points and their respective property values that bound the new feasibility region are summarized in Table 5.9. The property values are back calculated from the property operator expressions and reference values (equations 5.7 – 5.9) given in Section 5.2.2.

Hence, the new property requirements specified by the process needs are back calculated from the determined cluster values and are now identified as the upper and lower bounds on the three properties (see Table 5.10); and used as input to the molecular design algorithm.

Points	V_c	H_v	H_{fus}	Ω_1	Ω_2	Ω_3	AUPs	C_1	C_2	C_3	ΣC_{js}	Xcc	Ycc
Lumped Source (*1*)	750	110	16	300.0	314.3	160	774.3	0.3875	0.4059	0.2066	1.0	0.590	0.406
PT 3 on Feasibility (*M*)	530	115	20	212	328.6	200	740.6	0.2863	0.4437	0.2701	1.0	0.508	0.444
Point A (*2*)	310	120	24	124	342.9	240	706.9	0.1754	0.4850	0.3395	1.0	0.418	0.485
	x	β											
	0.5	0.522											
PT 6 on Feasibility(*M*)	610	100	40	244	285.7	400	929.7	0.2624	0.3073	0.4302	1.0	0.416	0.307
Point B (*2*)	470	90	64	188	257.1	640	1085.1	0.1732	0.2370	0.5898	1.0	0.292	0.237
	x	β											
	0.5	0.4164											
PT 2 on Feasibility (*M*)	530	115	40	212	328.6	400	940.6	0.2254	0.3493	0.4253	1.0	0.400	0.349
Point C (*2*)	310	120	64	124	342.9	640	1106.9	0.1120	0.3098	0.5782	1.0	0.267	0.310
	x	β											
	0.5	0.411											
PT 1 on Feasibility (*M*)	530	100	40	212	285.7	400	897.7	0.2362	0.3183	0.4456	1.0	0.395	0.318
Point D (*2*)	310	90	64	124	257.1	640	1021.1	0.1214	0.2518	0.6267	1.0	0.247	0.252
	x	β											
	0.5	0.431											

Table 5.8: Calculation data for new feasiblity region

H_{fus}	Ω_1	Ω_2	Ω_3	AUP_s	C_1	C_2	C_3	ΣC_{js}	Xcc	Ycc
40	244	285.7	400	929.7	0.262	0.307	0.43	1	0.416	0.307
20	244	285.7	200	729.7	0.334	0.392	0.274	1	0.530	0.392
20	244	328.6	200	772.6	0.316	0.425	0.259	1	0.528	0.425
20	212	328.6	200	740.6	0.286	0.444	0.27	1	0.508	0.444
24	124	342.9	240	706.9	0.175	0.485	0.34	1	0.418	0.485
64	124	342.9	640	1106.9	0.112	0.310	0.578	1	0.267	0.310
64	124	257.1	640	1021.1	0.121	0.252	0.627	1	0.247	0.252
64	188	257.1	640	1085.1	0.173	0.237	0.59	1	0.292	0.237

Table 5.9: New Feasibility Region Data

Property	LL	UL
V_c	310	610
H_v	90	120
H_{fus}	20	64

Table 5.10: Determined property constraints for molecular design algorithm

5.2.3. Molecular Design

Property models for the three functionalities (V_c, H_v, and H_{fus}) are available in the bank of group contribution models and have been used in the formulation of the corresponding molecular property operators (ψ^M_1, ψ^M_2, ψ^M_3), see table 5.11. The molecular feasibility region for the design problem has been plotted on Figure 5.12. The molecular building blocks given as input into the algorithm are represented by the discrete points on the same plot.

Having the molecular synthesis problem represented visually, all that remains is to proceed with molecular addition of groups until molecular candidates are generated (M1-M6), whose locus falls within the sink, this satisfies the first feasibility condition (Rules 5-6) (Figure 5.13). For complete validation of the designed formulations all remaining conditions must be satisfied; the *AUP* of the formulations all fall within the *AUP* range of the sink, determined to be 172-306. The candidate formulations M1 and M2 failed to satisfy AUP constraint (see table 5.12). Hence, M3-M6 are the only molecules that satisfy all the necessary and sufficient conditions. As a final check the designed formulations are mapped back to the process design level and as seen on Figure 5.14, all the formulations fall within designated design space.

j	Property (X)	GC Property Model	Property Operator	ψ_{ref}
1	V_c	$V_c - V_{co} = \sum_i n_g \cdot V_{c1_i}$	$\sum_i n_g \cdot V_{c1_i}$	20
2	H_v	$H_v - H_{vo} = \sum_i n_g \cdot H_{v1_i}$	$\sum_i n_g \cdot H_{v1_i}$	1
3	H_{fus}	$H_{fus} - H_{fuso} = \sum_i n_g \cdot H_{fus1_i}$	$\sum_i n_g \cdot H_{fus1_i}$	0.5

Table 5.11: Property operators for gas purification molecular synthesis
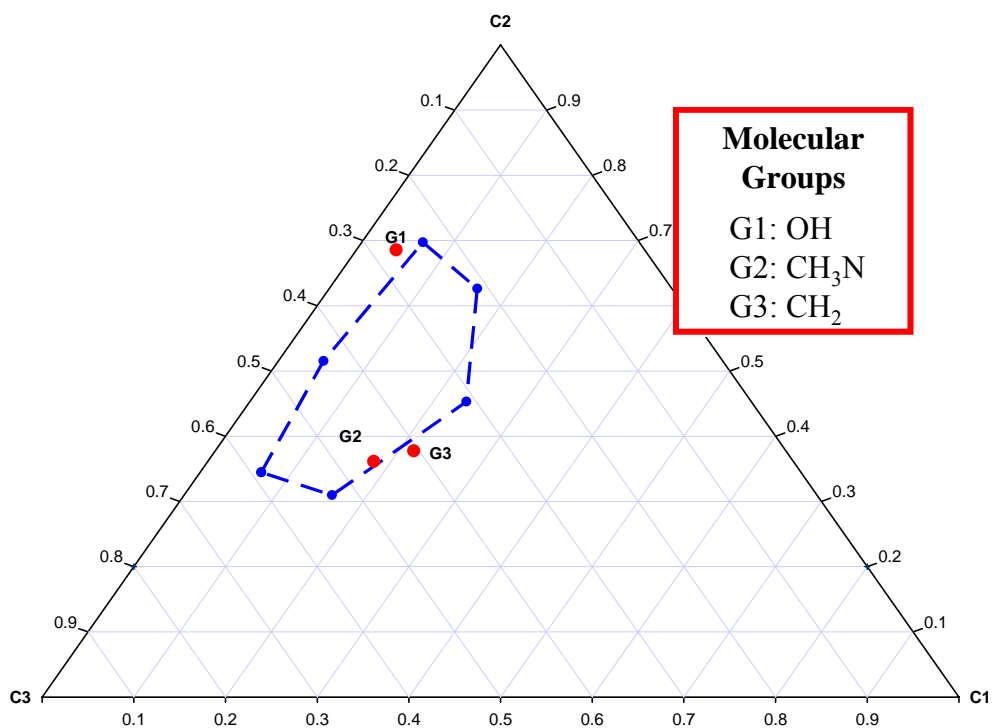
Figure 5.12: Molecular synthesis of gas purification solvent
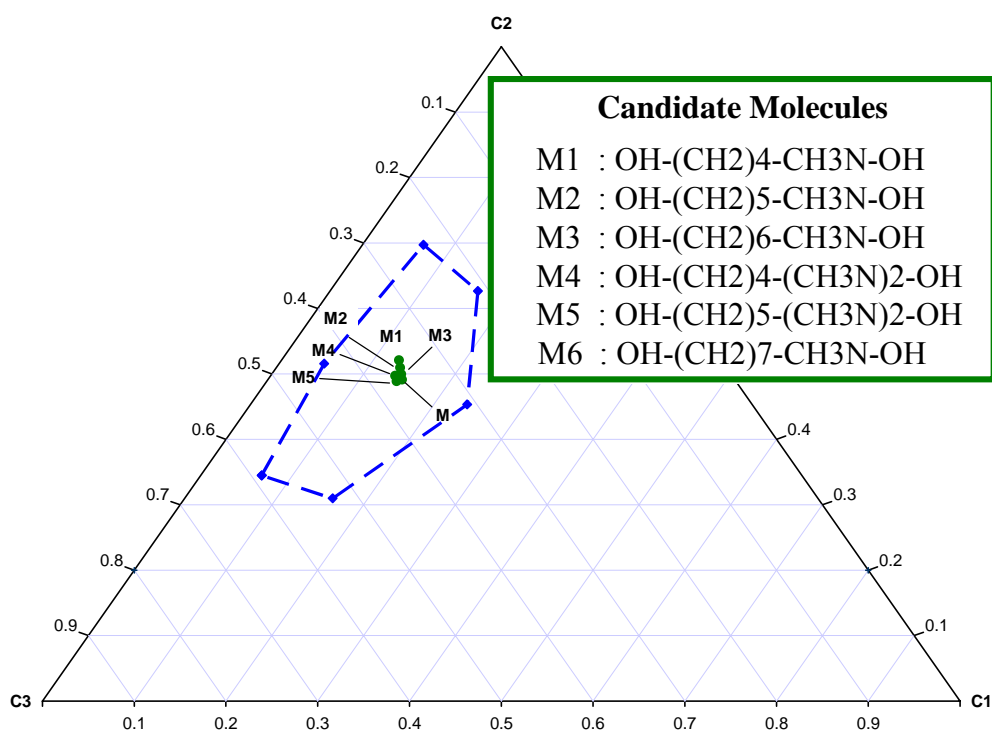
Figure 5.13: Candidate molecules for gas purification solvent

Candidates	AUP	Vc (cm³/mol)	Hv (kJ/mol)	Hf (kJ/mol)
*M*1	~~148.9~~	~~389.23~~	~~89.294~~	~~23.33~~
*M*2	~~161.9~~	~~445.51~~	~~94.204~~	~~25.969~~
*M*3	174.9	501.79	99.114	28.608
*M*4	175.2	484.17	98.787	29.338
*M*5	188.2	540.45	103.697	31.977
*M*6	187.9	558.07	104.024	31.247

Table 5.12: Candidate property data for gas purification solvent
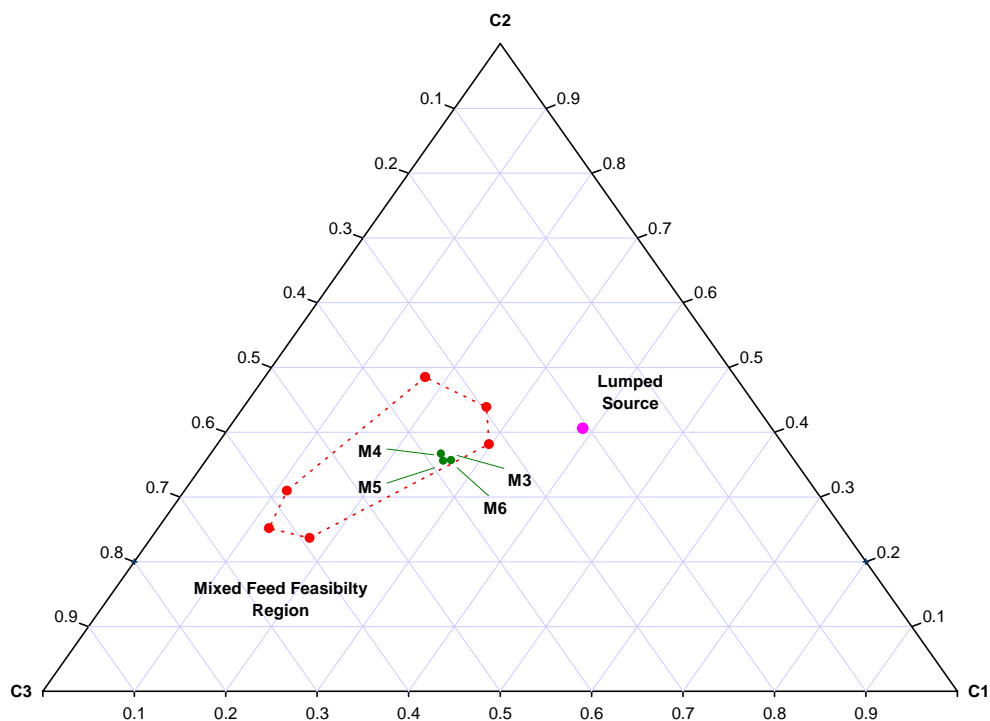
Figure 5.14: Verification of candidate molecules in process domain

5.2.4. Summary

The solution of the gas purification study has illustrated the simultaneous approach and its ability to transfer information from the process level to the molecular level and back again. Formulation of the process design problem on the ternary diagram enabled the decomposition of the problem by identifying the optimal feasibility region through the simple use of lever arm analysis. The framework allows for the complete integration of process sources and sinks for identification of property targets. The methodology facilitates the flow of information from the process domain to the molecular domain and back without the need for extensive calculation.

6. Conclusions and Future Work

6.1. Achievements

The main achievement of this work is the development of the Molecular Property Cluster algorithm – a property based framework that allows for the systematic synthesis of molecular formulations based on molecular fragments. The method is capable of simultaneously considering both process and molecular design needs. In that sense it is a truly integrated approach. Developed within the property clustering platform, it has established a systematic means of formulating molecular property operators, which helps in lowering the complexity of the design problem. In this work the property clustering technique has been combined with first-order Group Contribution Methods (GCM) to produce a systematic methodology capable of handling property design targets and synthesizing molecular options to satisfy them. Current integrated solution strategies like mathematical optimization struggle with limitations on flexibility of the property models. The problem becomes too complex, which makes it difficult to achieve convergence. In this approach the complex nature of the property models is hidden within the formulation of the molecular operators. The concept of linearizing non-linear functionalities aided in handling the convergence limitations due to complex property models. The development of the molecular operators was key in bridging the gap between the previously decoupled design problems.

It is a targeting approach that sets up the design problem as a reverse problem formulation, where the property performance requirements in this approach are obtained directly from the process clustering algorithm, as established by Eden (2003). The process algorithm was developed within the clustering platform; the process design problem is solved in terms of the constitutive variables (properties) and the generated solutions are also in terms of the constitutive variables. Once again, the reverse approach plays an important role; here it allows for the solution of process design problems in the property domain without having to commit to any components ahead of design. The process needs (solution in terms of functionalities) are now the input to the molecular design algorithm.

Like the original property cluster operators used for processes, the formulation of the molecular property operators allows for simple linear additive rules of the individual molecular groups that make up the formulation. A systematic methodology to convert property data and constraints into molecular cluster data has been presented. Furthermore, a significant contribution of the developed methodology is that for problems that can be adequately described by just three properties, the process and molecular design problems are solved visually and simultaneously on ternary diagrams, irrespective of how many process streams or molecular fragments are included in the search space. First the process design problem is visualized on a process ternary cluster diagram, where the clusters are formulated according to the process operator mixing rules. Next, the process design algorithm identifies the optimal design in terms of process property clusters which are then converted to physical property data. The solution to the process design problem provides the property constraints used as input to

the molecular design algorithm. Using the molecular cluster rules developed in this work, the data is converted to molecular property targets. Next, the molecular property targets are plotted as a feasibility region on the molecular ternary cluster diagram. The set of molecular groups used as input to the algorithm are plotted as points on the ternary diagram. In regards to selection of groups for the molecular synthesis, all available groups can be included if no restrictions or constraints are placed on the design problem (e.g. if only alcohols are desired the –OH would be included as one of the groups in the list of molecular fragments). The synthesis of candidate molecules is achieved in accordance with the necessary and sufficient conditions developed in this work. The rules describe how groups can be visually added on the diagram; and how the location of the final formulation is independent of group addition path. Once the molecular formulation is completed there are checks to guarantee the validity of the design molecule. The cluster value of the formulation must be contained within the feasibility region of the sink on the molecular ternary cluster diagram. The *AUP* value of the designed molecule must be within the range of the target. If the *AUP* value falls outside the range of the sink, the designed molecule is not a feasible solution. The aforementioned conditions are necessary but the sufficient condition is that for the designed molecule to match the target properties, the *AUP* value of the molecule has to match the *AUP* value of the sink at the same cluster location. And in the case where the design problem included Non-GC properties, those properties must be back calculated for the designed molecule using the appropriate corresponding GC property, and those values have to match the target Non-GC properties. The developed concepts have been illustrated through various application examples.

Although only those problems that can be described by three properties are covered by the visualization approach, the proposed molecular clustering methodology is capable of handling as many properties as needed to describe the system. In such cases, the visualization tool will no longer be available but the design problem is still simplified. The algebraic molecular clustering approach is used to formulate the design problem, with the molecular operators as the basis, therefore the dimensionality and complexity of the problem is significantly lowered from a MINLP to a LP. The molecular design problem is formulated as a set of equality and inequality equations to place bounds on the search space, while structural and non-structural constraints are also considered in the formulation. A proof of concept example has been solved to highlight the merits of the approach.

The Molecular Property Cluster algorithm has proven to be a powerful tool in the simultaneous consideration of molecular and process design problems. The methodology can also be used independently for just molecular synthesis, e.g. solvent design as in the provided cases of the blanket wash and the gas purification solvents.

The significant achievement of the methods presented in this thesis is the development of a systematic framework that enables a property based visual representation of the molecular synthesis problem. Molecular formulations are synthesized on the ternary diagram using lever-arm additive rules. The method enables synthesizing molecules systematically based on molecular fragments to satisfy the specific property needs of the process. The visual tool gives the designer a guide to which groups to include in the synthesis and those that will not help in satisfying the target performance requirements. For cases that require more than three properties, the

algebraic molecular clustering approach succeeds in lowering the dimensionality of the design problem from mixed integer non-linear program (MINLP) to that of a linear program (LP). The molecular property clusters are the key to bridging the gap between process and molecular design, thus allowing for a truly integrated design.

6.2. Future Directions

The work presented in this thesis has resulted in a vital tool for the areas of process and molecular integration, as well as molecular synthesis. Recognizing that the field of property clustering is fairly new, there is still a lot of work that needs to be covered. The property clustering techniques for process and molecular design were developed to aid in those cases where conventional component based algorithms fail. Several issues need to be addressed in order to increase the application range of this innovative approach.

6.2.1. Property Model Development

The molecular property clustering methods developed in this work are based on the molecular property operator formulations, which are functions of additive mixing rules according to the available GC property models. As long as models are available, the presented molecular synthesis rules are valid. Hence, to take advantage of the useful aspects of this methodology in molecular synthesis and design, additional efforts should be devoted to expanding the availability of group contribution properties, because that would translate to a wider scope of industrial applications for these techniques. In the

case of simultaneous process and molecular design, efforts need to concentrate on developing new property operator mixing rules with the same goal in mind, to have a well established bank of physical properties available. For example, properties like glass transition temperature for polymers already have mixing rules available but other properties like Knoop hardness and degree of polymerization still need to be developed.

6.2.2. Defining the Search Space

In synthesizing molecular formulations, property targets as well as molecular fragments are used as a part of the input information for the algorithm. All possible fragments can be included as part of the synthesis problem. The ternary cluster diagram, that is used to visualize the design problem, can be used to help eliminate infeasible molecular fragments that do no help to reach the targeted feasibility region. Thus, the visualization tool is used to help in narrowing down the search space and in turn the synthesis problem is simplified. There will be a need for the development of automatic systems to guide how molecular fragments should be excluded as well as how to narrow down the search space without risking excluding optimal candidates.

6.2.3. Expanding the Application Range

This thesis has shown how the development of the molecular clustering methods can help bridge the gap between process and molecular design. Although the introduced approach can be utilized independently for molecular design, it can also be used in conjunction with other algorithms. The molecular cluster algorithm developed here can be used in combination with a wide variety of other process synthesis and optimization

tools such as the property based pinch analysis developed by Kazantzi and El-Halwagi (2005). The synthesis problem is reformulated in terms of properties having environmental impact (e.g. toxicity). In such cases, valid empirical equations that could link the environmental properties to those of GCM are needed. Once those are identified, this methodology could be used to directly target those waste minimization criteria resulting in the synthesis of environmentally benign chemicals. Chemicals that might have been overlooked if only using the traditional dependence on laboratory experience.

The development of an algebraic method for the formulation of the molecular design problem is significant. Although visualization is no longer a viable tool, solution of the process design problem is achieved by solving a set of linear algebraic equality and inequality equations as a result of a constraint reduction approach. Efforts should concentrate on developing simultaneous algebraic techniques for solving process and molecular design problems. The process algebraic methods have been developed by Qin et al. (2004) and the algebraic formulation of the molecular design problem is introduced here. Therefore, all that remains is an outline of the merged approach.

References

Achenie, L. E. K. and M. Sinha (2004). "The design of blanket wash solvents with environmental considerations." Advances in Environmental Research 8(2): 213-227.

Achenie, L. E. K., R. Gani, V. Venkatasubramanian, Eds. (2003). Computer Aided Molecular Design: Theory and Practice. Computer Aided Chemical Engineering, 12, Elsevier.

Albanese, J. (2004). Optimizing Formulas by Experimental Design. The NY Chapter Society of Cosmetic Chemists' newsletter "Cosmetiscope".

Anderson, M. and P. Whitcomb (1996). "Optimize your Process Optimization Efforts." Chemical Engineering Progress 12: 51-60.

Barnicki, S. D. and J. J. Siirola (2004). "Process synthesis prospective" Computers & Chemical Engineering 28(4): 441.

Barton, A. F. (1985). Handbook of Solubility Parameters and other Cohesion Parameters. Boca Raton, Florida, CRC Press.

Box, G., W. Hunter, and J. Hunter (1978). Statistics for Experimenters. New York, Wiley.

Brignole, E. A. and M. Cismondi (2003). Molecular Design - Generation & Test Methods. Computer Aided Molecular Design: Theory and Practice. L. E. K. achenie, R. Gani and V. Venkatasubramanian, Elsevier. 12: 23-41.

Burke, J. (1984). Solubility Parameters: Theory and Application. The Book and Paper Group Annual: The American Institute for Conservation Annual Meeting.

CAPEC (2006). ICAS database. CAPEC, Technical University of Denmark, Denmark.

Cerda, J., A. W. Westerberg, D. Mason, B. Linnhoff (1983). "Minimum utility usage in heat exchanger network synthesis A transportation problem." Chemical Engineering Science 38(3): 373.

Constantinou, L. and R. Gani (1994). "New group contribution method for estimating properties of pure compounds." AIChE Journal 40(10): 1697-1710.

Constantinou, L., K. Bagherpour, R. Gani, J. Kein, D. Wu (1996). "Computer aided product design: problem formulations, methodology and applications." Computers & Chemical Engineering 20(6-7): 685-702.

Constantinou, L., R. Gani, J. O'Connell (1995). "Estimation of the acentric factor and the liquid molar volume at 298 K using a new group contribution method." Fluid Phase Equilibria 103(1): 11-22.

Constantinou, L., S. Prickett, and M. Mavrovouniotis (1993). "Estimation of thermodynamic and physical properties of acyclic hydrocarbons using the ABC

approach and conjugation operators." Industrial & Engineering Chemistry Research 32(8): 1734-1746.

Cornell, J. (1990). Experiments with Mixtures. New York, John Wiley and Sons Inc.

CRC Handbook of Chemistry and Physics (1980). Boca Raton, FL, CRC Press.

Cussler, E. L. and G. D. Moggridge (2001). Chemical Product Design. New York, Cambridge University Press.

d'Anterroches, L. and R. Gani (2005). "Group contribution based process flowsheet synthesis, design and modelling." Fluid Phase Equilibria 228-229: 141-146.

d'Anterroches, L., R. Gani, P. Harper, and M. Hostrup (2005). Design of Molecules, Mixtures and Processses through a Novel Group Contribution Method. 7th World Congress of Chemical Engineering, Glasgow, UK.

Doyle, S.J. and R. Smith (1997). "Targeting Water Reuse with Multiple Contaminants," Trans. Inst. Chem. E., B75: 181.

Dunn, R. and G. Bush (2001). "Usiing Process Integration Technology for Cleaner Production." Journal of Cleaner Production 9: 1-23.

Duvedi, A. P. and L. E. K. Achenie (1996). "Designing environmentally safe refrigerants using mathematical programming." Chemical Engineering Science 51(15): 3727.

Eden, M. R. (2003). Property Based Process and Product Synthesis and Design. CAPEC, Department of Chemical Engineering, Technical University of Denmark. Ph.D Thesis.

Eden, M. R., P. M. Harper, R. Gani, and S. Jorgensen (2002). "Design of Separation Process for Synthesis A/S - Separation of Aniline from Water". Lyngby, CAPEC, Technical University of Denmark.

Eden, M. R., S. B. Jørgensen, R. Gani and, M. El-Halwagi (2003). "Reverse Problem Formulation based Techniques for Process and Product Design." Computer Aided Chemical Engineering, 15A.

Eden, M. R., S. B. Jorgensen, R. Gani, and M. El-Halwagi (2004). "A novel framework for simultaneous separation process and product design." Chemical Engineering and Processing 43(5): 595-608.

El-Halwagi, M. (1997). Pollution Prevention Through Process Integration: Systematic Design Tools. San Diego, CA, Academic Press.

El-Halwagi, M. (2006). Process Integration. Process Systems Engineering. Amsterdam, Academic Press. 7.

El-Halwagi, M. M. and H. D. Spriggs (1996). "An integrated approach to cost and energy efficient pollution prevention". Fifth World Congress of Chemical Engineering, San Diego, USA.

El-Halwagi, M. M. and H. D. Spriggs (1998). "Solve Design Puzzles with Mass Integration." Chemical Engineering Progress 94: 25-44.

El-Halwagi, M. M. and V. Manousiouthakis (1989). "Synthesis of mass exchange networks." AIChE Journal 35(8): 1233-1244.

El-Halwagi, M. M., I. M. Glasgow, X. Qin, M. R. Eden (2004). "Property integration: Componentless design techniques and visualization tools." AIChE Journal 50(8): 1854-1869.

Eljack F.T., Eden M.R., Kazantzi V., El-Halwagi M.M. (2007a): "Molecular Design via Molecular Property Cluster - An Algebraic Approach", Computer Aided Chemical Engineering (accepted)

Eljack F.T., Eden M.R., Kazantzi V., El-Halwagi M.M. (2007b): "Simultaneous Process and Molecular Design - A Property Based Approach", AIChE Journal 53(5), 1232-1239.

Eljack F.T.., Eden M.R. (2007): "A Visual Approach to Molecular Design using Property Clusters and Group Contribution", *Computers and Chemical Engineering* (accepted)

Eljack, F. T., A. F. Abdelhady, M. Eden, F. Gabriel, X. Qin, and M. El-Halwagi (2005). Targeting optimum resource allocation using reverse problem formulations and property clustering techniques. Computers & Chemical Engineering 29(11-12): 2304-2317.

Eljack, F. T., M. R. Eden, V. Kazantzi, M. M. El-Halwagi (2006). "Property Clustering and Group Contribution for Process and Molecular Design". Computer Aided Chemical Engineering 21, Elsevier.

EPA, United States Environmental Protection Agency (1997). "Printing/Publishing Industry." www.epa.gov/region02/p2/printer.htm.

Floudas, C. A. (1995). Nonlinear and Mixed-Integer Optimization. New York, Oxford University Press.

Foo, D. C. Y., V. Kazantzi, M.M. El-Halwagi, Z. Abdul Manan (2006). "Surplus diagram and cascade analysis technique for targeting property-based material reuse network." Chemical Engineering Science 61(8): 2626.

Franklin, J. L. (1949). "Prediction of Heat and Free Energies of Organic Compounds." Industrial & Engineering Chemistry Research 41: 1070-6.

Friedler, F., L. T. Fan, L. Kalotai, A. Dallos (1998). "A combinatorial approach for generating candidate molecules with desired properties based on group contribution." Computers & Chemical Engineering 22(6): 809.

Gani, Rafiqul, J. Perregaard, and H. Johansen (1990). "Simulation Strategies for Design and Analysis of Complex Chemical Processes." Trans. I. Chem. E., vol. 68A: 407-417.

Gani, R. (2001). Computer aided process/product synthesis and design: Issues, needs and solution approaches". AIChE Annual Meeting, Reno.

Gani, R. and E. N. Pistikopoulos (2002). "Property modelling and simulation for product and process design." Fluid Phase Equilibria 194-197: 43-59.

Gani, R. and J. P. O'Connell (2001). "Properties and CAPE: from present uses to future challenges." Computers & Chemical Engineering 25(1): 3.

Gani, R., B. Nielsen, A. Fredenslund (1991). "A group contribution approach to computer-aided molecular design." AIChE Journal 37(9): 1318-1332.

Gani, R., L. Achenie, and V. Venkatasubramanian (2003). Challenges and Opportunities for CAMD. Computer Aided Molecular Design: Theory and Practice. L. Achenie, R. Gani and V. Venkatasubramanian,eds.. Amsterdam, Elsevier. Computer Aided Chemical Engineering 12: 357.

Garrison, G. W., A. A. Hamad, M. M. El-Halwagi. (1995). "Synthesis of waste interception networks". AIChE Annual Meeting, Miami.

Gundersen, T. and L. Naess (1988). "The synthesis of cost optimal heat exchanger networks: An industrial review of the state of the art." Computers & Chemical Engineering 12(6): 503.

Harper P. M. and R. Gani, (1999). "CAMD and Solvent Design: From Group Contribution to Molecular Encoding", AIChE Annual Meeting 1999, Dallas, TX.

Harper, P. M. (2000). A Multi-Phase, Multi-Level Framework for Computer Aided Molecular Design. Ph.D. Thesis, CAPEC, Department of Chemical Engingeering, Technical University of Denmark.

Harper, P. M. and R. Gani (2000). "A multi-step and multi-level approach for computer aided molecular design." Computers & Chemical Engineering 24 (2-7): 677-683.

Harper, P. M., R. Gani, P. Kolar, T. Ishikawa (1999). "Computer-aided molecular design with combined molecular modeling and group contribution." Fluid Phase Equilibria 158-160: 337.

Hohmann, E. (1971). Optimum Networks for Heat Exchange. Ph.D. Thesis, University of South California, Los Angeles.

Holland, J. H. (1975). Adaptation in Neural and Artifical Systems. Ann Arbor, Univeristy of Michigan Press.

Hostrup, M. (2002). Integrated Approaches to Computer Aided Molecular Design. Ph.D. Thesis, Computer Aided Process Engineering Center (CAPEC), Technical University of Denmark.

Hostrup, M., P. M. Harper, and R. Gani. (1999). "Design of environmentally benign processes: integration of solvent design and separation process synthesis." Computers & Chemical Engineering 23(10): 1395-1414.

Hovarth, A. L. (1992). Molecular Design. Amsterdam, Elsevier.

Jalowka, J. and T. Daubert (1986). "Group Contribution Method to Predict Critical Temperature and Pressure of Hydrocarbons." Industrial & Engineering Chemistry Process Design and Development 25(4): 139.

Joback, K. (2006). "Molecular Knowledge Systems, Inc.- Designing Better Chemical Products." from www.molknow.com.

Joback, K. G. and G. Stephanopoulos (1995). "Searching in Spaces of of Discrete Solutions: The Design of Molecules Possessing Desired Physical Properties". Advances in Chemical Engineering, 21, Academic Press.

Joback, K. G. and R. C. Reid (1983). "Estimation of Pure-Component Properties from Group Contributions." Chemical Engineering Communication 57: 233.

Kazantzi V., Qin X., El-Halwagi M.M., Eljack F.T., Eden M.R. (2007): "Simultaneous Process and Molecular Design through Property Clustering Techniques", *Industrial & Engineering Chemistry Research* (published online April 14, 2007)

Kazantzi, V. and M. M. El-Halwagi (2005). "Targeting material reuse via proeprty integration." Chemical Engineering Progress 101(8): 28-37.

Kazantzi, V., Harell, D., Gabriel, F., Qin, X., El-Halwagi, M.M. (2004a). "Property-based integration for sustainable development". Computer-Aided Process Engineering 14 ,Elsevier, pp. 1069–1074.

Kazantzi, V., Qin, X., Gabriel, F., Harell, D., and El-Halwagi, M.M. (2004b). "Process modification through visualization and inclusion techniques for property based integration". In: Floudas, C.A., Agrawal, R. (Eds.), Proceedings of the Sixth Foundations of Computer Aided Design (FOCAPD). CACHE Corp., pp. 279–282.

Kirkpatrick, S., C. D. G. Jr., M. Vecchi (1983). "Optimization by Simulated Annealing." Science 220: 671-680.

Kreglewsi, A. and B. Zwolinski (1961). "A New Relation for Physical Properties of n-Alkanes and n-Alkyl Compounds." Journal of Physical Chemistry 65(6): 1050.

Linke, P. and A. Kokossis (2002). "Simultaneous Synthesis and Design of Novel Chemicals and Chemical Process Flowsheets. Computer Aided Chemical Engineering 10: 115-121.

Linnhoff, B. and E. Hindmarsh (1983). "The pinch design method for heat exchanger networks." Chemical Engineering Science 38(5): 745.

Linnhoff, B., D. Townsen, D. Boland, G. Hewitt, B. Thomas, A. Guy, R. Marsland (1982). A User Guide on Process Integration for the Efficient Use of Energy. Rugby, UK, Institute of Chemical Engineers.

Lydersen, A. (1955). Estimation of Critical Properties of Organic Compounds. Ph.D. University of Wisconsin, Madison, WI.

Lyman, W., W. Reehl, and D. Rosenblatt (1990). Handbook of Chemical Property Estimation Methods. American Chemical Society, Washington D.C.

Marcoulaki, E. C. and A. C. Kokossis (1998). "Molecular design synthesis using stochastic optimisation as a tool for scoping and screening." Computers & Chemical Engineering 22(Supplement 1): S11-S18.

Marrero, J. and R. Gani (2001). "Group-contribution based estimation of pure component properties." Fluid Phase Equilibria 183-184: 183-208.

Material Safety Data Sheets (2006). www.msdsonline.com. MSDSonline $^{®}$

Nielsen, J., M. Hansen, et al. (1996). "Heat Exchanger Network Modeling Framework for Optimal Design and Retrofitting." Computers & Chemical Engineering 20: S249-S254.

Odele, O. and S. Macchietto (1993). "Computer aided molecular design: a novel method for optimal solvent selection." Fluid Phase Equilibria 82: 39.

Ourique, J. E. and A. S. Telles (1998). "Computer-Aided Molecular Design with Simulated Annealing and Molecular Graphs." Computers & Chemical Engineering 22: S615-S618.

Papoulias, S. A. and I. E. Grossmann (1983). "A structural optimization approach in process synthesis--II: Heat recovery networks." Computers & Chemical Engineering 7(6): 707.

Pistikopoulos, E. N. and S. K. Stefanis (1998). "Optimal solvent design for environmental impact minimization." Computers & Chemical Engineering 22(6): 717.

Pretel, E. J., P. A. López, S. Bottini, E. Brignole (1994). "Computer-aided molecular design of solvents for separation processes." AIChE Journal 40(8): 1349-1360.

Qin, X., F. Gabriel, D. Harell, M. M. El_Halwagi (2004). "Algebraic Techniques for Proeprty Integration via Componenetless Design." Indusrial & Engineering Chemistry Research 43: 3792-3798.

Reid, R. C., J. M. Prausnitz, B. Poling (1987). The Properties of Gases and Liquids. New York, McGraw-Hill.

Shelley, M. D. and M. M. El-Halwagi (2000). "Component-less design of recovery and allocation systems: a functionality-based clustering approach." Computers & Chemical Engineering 24(9-10): 2081-2091.

Shenoy, U. (1995). Heat exchange network synthesis: process optimization by energy and resource analysis. Houston, Gulf Publishing Company.

Sherali, H.D. and W.P. Adams (1999). A Reformulation-Linearization Technique for Solving Discrete and Continuous Non-Convex Problems. Dordrecht, Kluwer Academic Publishers.

Sinha, M. and L. E. K. Achenie (2001). "Systematic design of blanket wash solvents with recovery considerations." Advances in Environmental Research 5(3): 239-249.

Smith, R. (2004) Processing integration extends its reach. ChemialProcessing.com: The digital resoruce of chemical processing magazine Volum 359.

Srinivas, B. K. (1997). An overview of Mass Integration and its Application to Process Development, GE Research & Development Center.

Srinivas, B. K. and M. M. El-Halwagi (1993). "Optimal design of pervaporation systems for waste reduction." Computers & Chemical Engineering 17(10): 957-970.

Stats-Ease (1999). *Design Expert* 6.0. State-Ease inc.

Takama, N., T. Kuriyama, K. Shiroko and T. Umeda (1980). "Optimal Water Allocation in a Petroleum Refinery," Computers and Chemcical Engineering, Vol. 4, p. 251.

Teja, A. S., R. J. Lee, D. Rosenthal, and M. Anselme (1990). "Correlation of the critical properties of alkanes and alkanols." Fluid Phase Equilibria 56: 153.

Tsonopoulos, C. (1987). "Critical Constatns of Normal Alkanes from Methane to Polyethylene." AIChE Journal 33(12): 2080-2083.

Tsonopoulos, C. and Z. Tan (1993). "The Critical Constants of Normal Alkanes From Methane to Polyethylene: II. Application of the Flory Theory." Fluid Phase Equilibria 83: 127.

Vaidyanathan, R. and M. El-Halwagi (1994). "Global optimization of nonconvex nonlinear programs via interval analysis." Computers & Chemical Engineering 18(10): 889-897.

Van Krevelen, D. W. (1990). Properties of polymers. Amsterdam, Elsevier.

Van Krevelen, D. W. and P. J. Hoftyzer (1976). Properties of Polymers: Their Estimation and Correlation with Chemical Structure. Amsterdam, Elsevier Scientific Publishing.

Venkatasubramanian, V., K. Chan, J. Caruthers (1994). "Computer-aided molecular design using genetic algorithms." Computers & Chemical Engineering 18(9): 833.

Wang, Y. P. and R. Smith (1994). "Wastewater minimisation." Chemical Engineering Science 49(7): 981.

Whitting, W. B. and Y. Xin (1999). Sensativity and uncertaintity of process simulation to thermodynamics data and models: case studies. American Institute of Chemical Engineering spring meeting, Houston.

Appendices

Appendix A: Group Contribution

A.1: 1st order GC Data

Constantinou and Gani (1994) estimate properties of pure organic compounds from their 1st and 2nd order groups. They have provided property models for the following properties:

- Normal boiling and melting temperatures

- Critical pressure, critical volumes and critical temperature

- Standard enthalpy of vaporization and standard Gibbs energy, and standard enthalpy of formation

The general group contribution model equation used to predict properties is described by equation 3.26. The left hand side (LHS) of the equations represents property functionality and the right hand side (RHS) is the property contribution of each group. Universal constants that are included in the property models are listed in Table A.1, the GC property models are summarized in Table A.2 and all data for the first order groups is listed in Table A.3 (Marrero and Gani, 2001)

$$\overbrace{f(X)}^{\text{LHS}} = \overbrace{\sum_i N_i C_i}^{\text{RHS}}$$

(3.26)

Universal Constants	Value
t_{mo}	102.425 K
t_{bo}	204.359 K
t_{co}	181.128 K
p_{c1}	1.3705 bar
v_{co}	4.35 cm^3/mol
g_{fo}	-14.828 kJ/mol
h_{fo}	10.835 kJ/mol
h_{vo}	6.829 kJ/mol
h_{fuso}	-2.806 kJ/mol
D	0.01211 m^3/kmol

Table A.1: Listed values of GCM universal constants

Property (X)	LHS of Eq. 3.26 Function $f(X)$	RHS of Eq. 3.26 1st order GC term
Normal melting point (T_m)	$\exp\left(\dfrac{T_m}{t_{mo}}\right)$	$\sum_i N_i T_{m1_i}$
Normal boiling point (T_b)	$\exp\left(\dfrac{T_b}{t_{bo}}\right)$	$\sum_i N_i T_{b1_i}$
Critical temperature (T_c)	$\exp\left(\dfrac{T_c}{t_{co}}\right)$	$\sum_i N_i T_{c1_i}$
Critical pressure (P_c)	$\left(P_c - P_{c1}\right)^{-0.5}$	$\sum_i N_i P_{c1_i}$
Critical volume (V_c)	$V_c - V_{co}$	$\sum_i N_i V_{c1_i}$
Standard Gibbs energy[1] (G_f)	$G_f - G_{fo}$	$\sum_i N_i G_{f1_i}$
Standard enthalpy formation[1] (H_f)	$H_f - H_{fo}$	$\sum_i N_i H_{f1_i}$
Standard enthalpy vaporization[1] (H_v)	$H_v - H_{vo}$	$\sum_i N_i H_{v1_i}$
Standard enthalpy fusion (H_{fus})	$H_{fus} - H_{fuso}$	$\sum_i N_i H_{fus1_i}$

[1]Properties predicted at 298K

Table A.2: Property functions for Group Contribution Methods

	Group	Example	T_{m1i}	T_{b1i}	T_{c1i}	P_{c1i}	V_{c1i}	G_{f1i}	H_{f1i}	H_{v1i}	H_{fus1i}
1	CH₃	n-Tetracontane (2)	0.6953	0.8491	1.7506	0.018615	68.35	2.878	−42.479	0.217	1.660
2	CH₂	n-Tetracontane (38)	0.2515	0.7141	1.3327	0.013547	56.28	8.064	−20.829	4.910	2.639
3	CH	2-Methylpentane (1)	−0.3730	0.2925	0.5960	0.007259	37.50	8.254	−7.122	7.962	0.134
4	C	2,2-Dimethylbutane (1)	0.0256	−0.0671	0.0306	0.001219	16.01	16.413	8.928	10.730	−1.232
5	CH₂=CH	1-Hexene (1)	1.1728	1.5596	3.2295	0.025745	111.43	95.738	57.509	4.031	1.268
6	CH=CH	2-Hexene (1)	0.9460	1.5597	3.0741	0.023003	98.43	92.656	69.664	9.456	4.441
7	CH₂=C	2-Methyl-1-butene (1)	0.7662	1.3621	2.7717	0.021137	91.40	85.107	61.625	8.602	2.451
8	CH=C	2-Methyl-2-butene (1)	0.1732	1.2971	2.5666	0.019609	83.89	88.691	81.835	14.095	3.032
9	C=C	2,3-Dimethyl-2-butene (1)	0.3928	1.2739	2.6391	0.014114	90.66	93.119	95.710	19.910	2.616
10	CH₂=C=CH	1,2-Butadiene (1)	1.7036	2.6840	5.4330	0.035483	143.57	229.906	198.840	11.310	7.076
11	CH₂=C=C	3-Methyl-1,2-butadiene (1)	1.5453	2.4014	4.8219	0.029678	146.36	226.710	208.490	*****	7.435
12	CH=C=CH	2,3-Pentadiene (1)	1.2850	2.5400	*****	*****	*****	*****	*****	*****	6.000
13	CH≡C	1-Pentyne (1)	2.2276	1.7618	3.7897	0.014010	84.60	230.029	224.902	6.144	−1.548
14	C≡C	3-Decyne (1)	2.0516	1.6767	4.5870	0.010888	74.66	216.013	228.282	12.540	6.128
15	aCH	Benzene (6)	0.5860	0.8365	2.0337	0.007260	42.39	26.732	12.861	3.683	1.948
16	aC fused with aromatic ring	Naphthalene (2)	1.8955	1.7324	5.4979	0.003564	35.71	20.379	20.187	6.631	0.845
17	aC fused with non-aromatic subring	Indane (2)	1.2065	1.1995	3.1058	0.006512	34.65	33.912	30.768	6.152	1.095
18	aC except as above	Benzophenone (1)	0.9176	1.5468	4.5344	0.012859	26.47	23.331	24.701	6.824	−0.531
19	aN in aromatic ring	Pyridine (1)	2.0438	1.3977	4.0954	−0.003339	36.47	89.902	70.862	9.420	2.555
20	aC–CH₃	Toluene (1)	1.0068	1.5653	3.4611	0.020907	97.33	24.919	−19.258	8.279	2.969
21	aC–CH₂	Ethylbenzene (1)	0.1065	1.4925	2.9003	0.018082	87.19	31.663	4.380	11.981	0.948
22	aC–CH	Cumene (1)	−0.5197	0.8665	1.9512	0.011795	73.51	30.393	18.440	13.519	−1.037
23	aC–C	tert-Butylbenzene (1)	−0.1041	0.5229	0.8576	0.011298	67.20	40.127	35.297	16.912	−2.856
24	aC–CH=CH₂	Styrene (1)	1.2832	2.4308	5.7861	0.030637	134.69	114.531	77.863	*****	4.013
25	aC–CH=CH	1-Propenylbenzene (1)	1.7744	2.9262	6.5062	0.026282	128.84	111.216	88.084	*****	8.274
26	aC–C=CH₂	α-Methylstyrene (1)	1.2612	2.1472	4.9967	0.026371	110.74	115.728	90.927	*****	3.324
27	aC–C≡CH	Phenylacetylene (1)	1.7495	2.3057	6.4572	0.019507	112.08	263.205	257.448	*****	2.514
28	aC–C≡C	1-Phenyl-1-propyne (1)	*****	2.7341	*****	*****	*****	*****	*****	*****	*****
29	OH	1,4-Butanediol (2)	2.7888	2.5670	5.2188	−0.005401	30.61	−144.051	−178.360	24.214	4.786
30	aC–OH	Phenol (1)	5.1473	3.3205	9.3472	−0.008788	50.77	−131.327	−164.191	34.099	8.427
31	COOH	1,5-Pentanedioic acid (2)	7.4042	5.1108	14.6038	0.009885	90.66	−337.090	−389.931	17.002	10.692
32	aC–COOH	Benzoic acid (1)	12.4296	6.0677	15.4515	0.017100	119.10	−312.422	−361.249	*****	14.649
33	CH₃CO	2-Butanone (1)	2.9588	3.1178	7.0058	0.025227	127.99	−120.667	−180.604	15.195	8.062
34	CH₂CO	3-Pentanone (1)	2.5232	2.6761	5.7157	0.019619	112.79	−120.425	−163.090	19.392	8.826
35	CHCO	2,4-Dimethyl-3-pentanone (1)	1.1565	2.1748	4.4743	0.012487	97.16	−116.799	−139.909	20.350	7.205
36	CCO	2,2,4,4-Tetramethyl-3-pentanone (1)	1.0638	1.7287	*****	*****	*****	*****	*****	*****	*****
37	aC–CO	Acetophenone (1)	2.9157	3.4650	9.4806	0.011007	90.69	−91.812	−106.965	25.036	4.852
38	CHO	1-Hexanal (1)	3.0186	2.5388	5.8013	0.010204	71.08	−100.882	−130.816	12.370	11.325
39	aC–CHO	Benzaldehyde (1)	2.4744	3.5172	9.4795	0.019633	122.91	−80.222	−107.159	*****	7.273
40	CH₃COO	Butyl acetate (1)	2.1657	3.1228	6.3179	0.033812	148.91	−306.733	−387.458	19.342	7.910
41	CH₂COO	Methyl Butyrate (1)	1.6329	2.9850	5.9619	0.026983	132.89	−298.332	−364.204	21.100	9.479
42	CHCOO	Ethyl isobutyrate (1)	1.0668	2.2869	4.7558	0.021990	125.52	−301.414	−352.057	24.937	9.317
43	CCOO	Ethyl 2,2-dimethylpropionate (1)	0.3983	1.6918	*****	*****	*****	*****	*****	23.739	*****
44	HCOO	Propyl formate (1)	2.0223	2.5972	5.6064	0.015249	93.29	−276.878	−327.678	15.422	8.115

Table A.2: 1st order Groups and their contributions (Marrero and Gani, 2001)

45	aC–COO	Methyl benzoate (1)	1.3348	3.1952	6.7311	0.018948	105.53	−291.662	−307.727	25.206	8.149
46	aC–OOCH	Phenyl formate (1)	*****	0.4621	*****	*****	*****	*****	*****	*****	*****
47	aC–OOC	Phenyl acetate (1)	4.8044	3.0854	*****	*****	*****	*****	*****	*****	5.875
48	COO except as above	Ethyl acrylate (1)	1.5038	2.1903	4.7346	0.013087	81.17	−299.803	−331.397	*****	10.573
49	CH$_3$O	Methyl butyl ether (1)	1.3643	1.7703	3.4393	0.020084	88.20	−90.329	−156.062	5.783	5.089
50	CH$_2$O	Di-*n*-butyl ether (1)	0.8733	1.3368	2.4217	0.017954	74.03	−105.579	−152.239	9.997	4.891
51	CH–O	sec-Butyl ether (1)	0.2461	0.8924	0.7889	0.014487	60.06	−101.207	−147.709	14.620	4.766
52	C–O	*tert*-Butylether (1)	−0.4446	0.4983	0.2511	0.005613	52.96	−92.804	−121.608	13.850	2.458
53	aC–O	Methyl phenyl ether (1)	1.3045	1.8522	3.6588	0.005115	47.27	−83.354	−101.783	16.151	−0.118
54	CH$_2$NH$_2$	Ethylamine (1)	3.2742	2.7987	8.1745	0.011413	117.52	68.812	−10.703	15.432	13.482
55	CHNH$_2$	sec-Butylamine (1)	30.8394	2.0948	4.2847	0.013049	76.36	61.452	0.730	16.048	6.283
56	CNH$_2$	*tert*-Butylamine (1)	11.7400	1.6525	2.8546	0.010790	80.01	55.202	2.019	17.257	*****
57	CH$_3$NH	Dimethylamine (1)	2.4034	2.2514	4.5529	0.015863	77.04	88.512	24.740	11.831	4.490
58	CH$_2$NH	Dipropylamine (1)	1.7746	1.8750	3.2422	0.020482	95.15	88.874	23.610	13.067	7.711
59	CHNH	Diisopropylamine (1)	1.7577	1.2317	2.0057	0.005329	99.16	73.101	21.491	14.048	2.561
60	CH$_3$N	Methyldiethylamine (1)	0.9607	1.3841	3.0106	0.021186	94.94	125.906	55.024	9.493	6.008
61	CH$_2$N	Triethylamine (1)	0.0442	1.1222	2.1673	0.027454	74.05	121.247	65.331	12.636	1.756
62	aC–NH$_2$	Aniline (1)	3.9889	3.8298	10.2155	0.005335	81.40	66.470	17.501	23.335	6.542
63	aC–NH	*N*-methyl aniline (1)	1.4837	2.9230	8.4081	−0.005596	86.37	98.195	53.274	23.026	0.624
64	aC–N	*N,N*-dimethyl aniline (1)	1.7618	2.1918	5.8536	−0.000838	108.39	143.280	115.606	22.249	−2.576
65	NH$_2$ except as above	Cyclobutylamine	3.3478	2.0315	4.7420	0.000571	63.39	42.687	−8.556	13.425	6.158
66	CH=N	Acetaldazine (2)	8.8492	1.5332	*****	*****	*****	*****	*****	*****	*****
67	C=N	Ketazine (2)	1.4621	1.4291	*****	*****	*****	*****	*****	*****	*****
68	CH$_2$CN	Propionitrile (1)	2.5760	4.5871	12.9827	0.036523	133.62	134.997	99.245	21.923	7.303
69	CHCN	Isobutyronitrile (1)	2.1393	3.9774	8.4309	0.029034	134.73	142.475	151.390	24.963	9.464
70	CCN	2,2-Dimethylpropionitrile (1)	3.3807	2.8870	5.8829	0.024654	120.74	142.295	124.770	24.967	4.166
71	aC–CN	Benzonitrile (1)	5.1346	4.1424	10.4124	0.020978	119.08	162.175	148.968	*****	6.788
72	CN except as above	Acrylonitrile (1)	3.2747	3.0972	8.1381	0.024346	94.91	130.986	124.917	16.639	6.867
73	CH$_2$NCO	Ethyl isocyanate (1)	4.2256	3.4891	*****	*****	*****	*****	*****	*****	*****
74	CHNCO	Isopropyl isocyanate (1)	*****	3.1220	*****	*****	*****	*****	*****	*****	*****
75	CNCO	*tert*-Butyl isocyanate (1)	9.1492	*****	*****	*****	*****	*****	*****	*****	*****
76	aC–NCO	Phenyl isocyanate (1)	2.2327	3.1853	6.5884	0.025065	141.24	*****	*****	*****	*****
77	CH$_2$NO$_2$	1-Nitropropane (1)	3.2131	4.5311	10.9507	0.021056	157.57	25.783	−65.620	29.640	10.989
78	CHNO$_2$	2-Nitropropane (1)	0.7812	3.8069	9.5487	0.014899	143.36	16.407	−60.750	29.173	*****
79	CNO$_2$	2-Methyl-2-nitropropane (1)	5.6280	3.3059	*****	*****	*****	*****	*****	*****	−4.187
80	aC–NO$_2$	Nitrobenzene (1)	4.3531	4.5750	12.1243	0.018311	133.06	57.352	−22.931	24.863	7.572
81	NO$_2$ except as above	Nitrocyclohexane (1)	3.0376	3.2069	*****	*****	*****	*****	*****	*****	6.302
82	ONO	Butyl nitrite (1)	*****	1.8896	*****	*****	*****	*****	*****	*****	*****
83	ONO$_2$	*n*-Butyl nitrate (1)	2.5974	3.2656	*****	*****	*****	*****	*****	*****	9.353
84	HCON(CH$_2$)$_2$	Diethylformamide (1)	*****	5.8779	*****	*****	*****	*****	*****	*****	*****
85	HCONHCH$_2$	Ethylformamide (1)	*****	7.4566	*****	*****	*****	*****	*****	46.490	*****
86	CONH$_2$	Butyramide (1)	13.2124	6.5652	25.1184	0.001467	138.71	−127.512	−201.369	44.240	16.840
87	CONHCH$_3$	Methylacetamide (1)	5.4720	5.0724	20.5590	0.023455	190.71	−102.912	−203.069	*****	17.429
88	CONHCH$_2$	Ethylacetamide (1)	5.8825	6.6810	*****	*****	*****	*****	−183.613	52.723	*****
89	CON(CH$_3$)$_2$	Dimethylacetamide (1)	4.1720	6.0070	15.4603	0.043090	244.71	−56.412	−188.069	38.290	11.553
90	CONCH$_3$CH$_2$	Methylethylacetamide (1)	*****	*****	*****	*****	*****	*****	−48.210	*****	*****
91	CON(CH$_2$)$_2$	Diethylacetamide (1)	*****	5.0664	*****	*****	*****	*****	*****	*****	*****
92	CONHCO	Diacetamide (1)	9.1763	7.6172	*****	*****	*****	*****	*****	*****	*****
93	CONCO	Methyldiacetamide	3.2657	5.6487	*****	*****	*****	*****	*****	*****	*****

Table A.2: Cont'd

	Group	Example	T_{m1i}	T_{b1i}	T_{c1i}	P_{c1i}	V_{c1i}	G_{f1i}	H_{f1i}	H_{v1i}	H_{fus1i}
94	aC–CONH$_2$	Benzamide	12.8071	8.3775	*****	*****	*****	*****	*****	*****	16.811
95	aC–NH(CO)H	N-phenylformamide (1)	5.6631	7.3497	19.8979	0.023447	162.08	−44.595	−125.052	*****	8.658
96	aC–N(CO)H	N-methyl-N-phenylmethanamide (1)	3.3602	5.1373	*****	*****	*****	*****	*****	*****	*****
97	aC–CONH	N-methylbenzamide (1)	6.5160	7.5850	*****	*****	*****	*****	*****	*****	10.959
98	aC–NHCO	N-(2-methylphenyl)acetamide (1)	9.8204	7.4955	*****	*****	*****	*****	*****	*****	4.370
99	aC–NCO	Phenylmethylacetamide (1)	7.2552	*****	*****	*****	*****	*****	*****	*****	*****
100	NHCONH	N,N'-dimethylurea (1)	9.3110	8.9406	*****	*****	*****	*****	*****	*****	9.862
101	NH$_2$CONH	Methylurea (1)	14.2020	16.3539	*****	*****	*****	*****	*****	*****	12.845
102	NH$_2$CON	N,N-dimethylurea (1)	13.0856	2.0796	*****	*****	*****	*****	*****	*****	10.958
103	NHCON	Trimethylurea (1)	8.4447	7.1529	*****	*****	*****	*****	*****	*****	12.098
104	NCON	Tetramethylurea (1)	3.5041	4.1459	*****	*****	*****	*****	*****	*****	9.557
105	aC–NHCONH$_2$	Phenylurea (1)	13.4695	5.7604	*****	*****	*****	*****	*****	*****	16.703
106	aC–NHCONH	N,N'-diphenylurea	23.2570	1.1633	*****	*****	*****	*****	*****	*****	18.460
107	NHCO except as above	N-chloroacetamide (1)	3.0882	*****	*****	*****	*****	*****	*****	*****	*****
108	CH$_2$Cl	1-Chlorobutane (1)	1.9253	2.6364	6.2561	0.021419	112.12	−19.484	−65.056	11.754	6.353
109	CHCl	2-Chloropropane (1)	1.0224	2.0246	4.3756	0.015640	100.78	−31.933	−65.127	12.048	*****
110	CCl	2-Chloro-2-methylpropane (1)	1.8424	1.7049	3.7063	0.009187	87.01	−37.848	−62.881	16.597	−0.082
111	CHCl$_2$	1,1-Dichloroethane (1)	2.5196	3.3420	7.8956	0.028236	159.79	−24.214	−80.812	17.251	6.781
112	CCl$_2$	2,2-Dichloropropane (1)	3.6491	2.9609	*****	*****	*****	*****	*****	20.473	1.823
113	CCl$_3$	1,1,1-Trichloroethane	4.4493	3.9093	8.8073	0.036746	204.71	−44.122	−105.369	20.550	3.492
114	CH$_2$F	1-Fluorobutane (1)	1.5597	1.5022	3.3179	0.023315	87.71	−180.212	−227.469	8.238	7.139
115	CHF	2-Fluorobutane (1)	1.1289	1.3738	2.6702	0.020040	78.08	−228.239	−261.901	*****	3.917
116	CF	2-Fluoro-2-methylpropane (1)	2.5398	1.0084	2.1633	−0.010120	*****	*****	*****	6.739	*****
117	CHF$_2$	1,1-Difluoroethane (1)	2.1689	2.2238	3.5702	0.031524	102.71	−411.239	−463.901	*****	7.011
118	CF$_2$	Perfluorohexane (4)	0.1312	0.5142	0.8543	0.018572	95.09	*****	*****	1.621	*****
119	CF$_3$	Hexafluoroethane (2)	1.4828	1.1916	1.7737	0.048565	108.85	−615.333	−673.875	7.352	2.526
120	CCl$_2$F	Tetrachloro-1,2-difluoroethane (2)	3.2035	2.5053	5.1653	0.037948	171.04	−249.020	−306.765	8.630	3.114
121	HCClF	1-Chloro-1,2,2,2-tetrafluoroethane (1)	*****	2.0542	*****	*****	*****	*****	*****	*****	*****
122	CClF$_2$	1,2-Dichlorotetrafluoroethane (2)	1.7510	1.7227	3.0593	0.041641	146.01	−396.814	−458.074	8.086	2.156
123	aC–Cl	Chlorobenzene (1)	1.7134	2.0669	5.7046	0.016033	92.67	1.985	−17.002	11.224	4.435
124	aC–F	Hexafluorobenzene (6)	0.9782	0.7945	1.5491	0.014037	54.36	−141.306	−160.965	3.965	2.003
125	aC–I	Iodobenzene (1)	2.1905	3.7739	12.4470	0.014403	131.08	91.505	95.048	*****	2.814
126	aC–Br	Bromobenzene (1)	2.4741	2.8414	8.4199	0.010199	104.12	42.977	38.917	14.393	5.734
127	I⁻ except as above	Iodoethane (1)	1.9444	3.1778	8.5775	−0.004637	104.28	43.910	47.632	14.171	6.103
128	Br⁻ except as above	Bromoethane (1)	1.7641	2.4231	4.5036	−0.001460	77.99	5.528	−1.703	9.888	4.826
129	F⁻ except as above	Benzyl fluoride (1)	1.2308	0.8504	0.8976	0.012034	24.62	−182.973	−201.968	*****	3.096
130	Cl⁻ except as above	Ethyl chloroacetate (1)	1.5454	1.5147	4.0947	0.007923	57.77	−29.876	−46.963	*****	5.181
131	CHNOH	Propionaldehyde oxime (1)	3.9813	4.5721	*****	*****	*****	*****	*****	*****	*****
132	CNOH	Diethyl ketoxime (1)	3.5484	4.0142	*****	*****	*****	*****	*****	*****	*****
133	aC–CHNOH	Phenyl oxime (1)	10.5579	*****	*****	*****	*****	*****	*****	*****	*****
134	OCH$_2$CH$_2$OH	2-Ethoxyethanol (1)	2.3651	4.8721	10.4579	0.025986	159.33	−233.335	−343.903	31.493	8.454
135	OCHCH$_2$OH	2-Ethoxy-1-propanol (1)	*****	4.2329	*****	*****	*****	*****	*****	*****	*****
136	OCH$_2$CHOH	1-Methoxy-2-propanol (1)	1.5791	3.6653	*****	0.018783	147.66	−239.423	−333.385	*****	12.594
137	–O–OH	tert-Butylhydroperoxide (1)	4.8181	3.1669	5.8307	−0.002815	58.01	−75.568	−125.111	*****	*****
138	CH$_2$SH	Ethanethiol (1)	2.2992	3.1974	7.7300	0.017299	105.68	27.469	−8.021	16.815	10.068
139	CHSH	2-Propanethiol (1)	0.9704	2.5910	5.8527	0.008968	109.36	27.030	3.510	17.098	4.266
140	CSH	2-Methyl-2-propanethiol (1)	4.2329	2.0902	4.6431	0.005118	94.01	27.338	12.589	18.397	−0.623

Table A.2: Cont'd

141	aC–SH	Benzenethiol (1)	2.8464	3.2675	9.5115	0.010086	95.08	48.905	41.648	17.413	4.513
142	–SH (except as above)	Cyclohexanethiol (1)	0.9600	2.3323	7.7987	0.006399	57.89	15.818	11.339	9.813	5.829
143	CH$_3$S	Dimethylsulfide (1)	1.7150	2.9892	6.9733	0.018013	122.03	35.845	−3.337	14.296	7.497
144	CH$_2$S	Diethylsulfide (1)	1.0063	2.6524	6.4871	0.015254	106.60	42.684	21.492	16.965	4.096
145	CHS	Diisopropylsulfide (1)	0.7892	2.0965	*****	*****	*****	*****	*****	19.038	*****
146	CS	di-*tert*-Butylsulfide (1)	1.1170	1.6412	*****	*****	*****	*****	*****	19.996	*****
147	aC–S–	Phenyl methyl sulfide (1)	0.9646	2.9731	*****	*****	*****	*****	*****	*****	*****
148	SO	Dimethyl sulfoxide (1)	5.3663	6.2796	19.8953	−0.005534	82.36	−52.231	−71.050	*****	13.403
149	SO$_2$	Dimethyl sulfone (1)	7.0778	7.0976	17.2586	−0.000784	89.95	−257.608	−305.498	*****	17.748
150	SO$_3$ (sulfite)	Dimethyl sulfite (1)	*****	3.9199	8.6910	0.004240	115.80	*****	−430.833	*****	*****
151	SO$_3$ (sulfonate)	Dimethyl sulfonate (1)	5.8426	6.7785	*****	*****	*****	*****	*****	*****	*****
152	SO$_4$ (sulfate)	Dimethyl sulfate (1)	3.6976	5.5627	18.9366	−0.027208	144.58	−519.853	−621.412	*****	*****
153	aC–SO	Phenyl methyl sulfoxide (1)	3.9911	6.1185	*****	*****	*****	*****	*****	*****	*****
154	aC–SO$_2$	Diphenyl sulfone (1)	5.2948	8.4333	*****	*****	135.47	−314.643	−370.493	*****	3.281
155	PH (phosphine)	Dimethylphosphine (1)	*****	2.0536	*****	*****	*****	*****	*****	*****	*****
156	P (phosphine)	Trimethylphosphine (1)	*****	1.0984	*****	*****	*****	*****	*****	*****	*****
157	PO$_3$ (phosphite)	Triethylphosphite (1)	1.0306	2.7900	*****	*****	*****	*****	*****	*****	*****
158	PHO$_3$ (phosphonate)	Dimethylphosphonate (1)	*****	5.6433	*****	*****	*****	*****	*****	*****	*****
159	PO$_3$ (phosphonate)	Trimethylphosphonate (1)	*****	4.5468	*****	*****	*****	*****	*****	*****	*****
160	PHO$_4$ (phosphate)	Diethylphosphate (1)	2.7461	5.1567	*****	*****	*****	*****	*****	*****	*****
161	PO$_4$ (phosphate)	Trimethylphosphate (1)	2.0330	3.7657	16.9914	−0.029036	85.59	*****	−1060.325	*****	*****
162	aC–PO$_4$	Triphenylphosphate (1)	−1.7840	2.3522	*****	*****	*****	*****	−1005.161	*****	4.256
163	aC–P	Triphenylphosphine (1)	0.2337	2.9272	38.6148	−0.126108	−142.79	*****	72.339	*****	−5.654
164	CO$_3$ (carbonate)	Diethylcarbonate (1)	3.6593	2.8847	6.6804	0.007235	93.56	−447.186	−516.282	21.613	8.363
165	C$_2$H$_5$O	Ethyl oxirane (1)	1.3135	2.8451	6.6418	0.021238	125.43	11.149	−52.241	*****	*****
166	C$_2$H$_2$O	2,2-Dimethyl oxirane (1)	*****	2.6124	6.0159	0.010678	194.36	1.890	−51.390	*****	*****
167	C$_2$O	Trimethyl oxirane (1)	*****	2.2036	*****	*****	*****	*****	*****	*****	*****
168	CH$_2$ (cyclic)	Cyclopentane (5)	0.5699	0.8234	1.8815	0.009884	49.24	13.287	−18.575	3.341	1.069
169	CH (cyclic)	Methylcyclopentane (1)	0.0335	0.5946	1.1020	0.007596	44.95	6.107	−12.464	6.416	2.511
170	C (cyclic)	1,1-Dimethylcyclohexane (1)	0.1695	0.0386	−0.2399	0.003268	33.32	−0.193	−2.098	7.017	−0.921
171	CH=CH (cyclic)	Cyclobutene (1)	1.1936	1.5985	3.6426	0.013815	83.91	86.493	59.841	7.767	1.185
172	CH=C (cyclic)	1-Methylcyclopentene (1)	0.4344	1.2529	3.5475	0.010576	70.98	67.056	64.295	7.171	2.559
173	C=C (cyclic)	1,2-Dimethylcyclopentene (1)	0.3048	1.1975	*****	*****	*****	*****	*****	*****	*****
174	CH$_2$=C (cyclic)	Methylene cyclohexane (1)	0.2220	1.5109	4.4913	0.019101	83.96	*****	*****	*****	5.351
175	NH (cyclic)	Cyclopentimine (1)	3.4814	2.1634	5.9726	−0.003678	51.80	72.540	23.138	13.700	8.655
176	N (cyclic)	*N*-methylpyrrolidine (1)	0.6040	1.6541	4.3905	−0.001179	31.41	83.779	65.622	*****	0.269
177	CH=N (cyclic)	Imidazole (1)	5.5779	6.5230	*****	*****	*****	*****	*****	*****	3.993
178	C=N (cyclic)	2-Methyl-1H-imidazole (1)	6.6382	6.6710	*****	*****	*****	*****	*****	*****	*****
179	O (cyclic)	Tetrahydropyran (1)	1.3828	1.0245	2.7409	−0.000387	17.69	−114.062	−137.353	6.877	3.806
180	CO (cyclic)	Cyclobutanone (1)	3.2119	2.8793	12.6396	−0.000207	57.38	−156.672	−180.166	17.124	6.137
181	S (cyclic)	2-Methyl-thiophene (1)	1.6023	2.3256	5.5523	0.001540	45.45	12.020	15.453	12.262	5.170
182	SO$_2$ (cyclic)	Cyclobutadiene sulfone (1)	6.1006	*****	24.3995	0.002487	96.66	−241.601	−283.839	*****	9.934

Table A.2: Cont'd

A.2: Molar Volume GC data

The group contribution model for molar volume by Constantinou and Gani (1995) was developed to include 1st and 2nd order groups, however the molecular property clusters presented in this work only considered first order groups, see equation A.1.

$$V_m - d = \sum_{g=1}^{Ng} n_g \cdot v_{mg} \qquad (A.1)$$

The first order group contribution data for molar volume is provided in Table A.4.

Group	v_m	Group	v_m
CH$_2$	0.01641	CHC1	0.02663
CH	0.00711	CCl	0.02020
C	-0.00380	CHC12	0.04682
CH 2 CH	0.03727	CCl 2	****
CH CH	0.02692	CCl 3	0.06202
CH2~	0.02697	ACC1	0.02414
CH C	0.01610	CH2NO 2	0.03375
C C	0.00296	CHNO 2	0.02620
CH2-C=CH	0.04340	ACNO2	0.02505
ACH	0.01317	CH,SH	0.03446
AC	0.00440	I	0.02791
ACCH 3	0.02888	Br	0.02143
ACCH~	0.01916	CH C	****
ACCH	0.00993	C C	0.01451
OH	0.00551	ACF	0.01727
ACOH	0.01133	CI (C C)	0.01533
CH3CO	0.03655	HCON(CH2) 2	****
CH2CO	0.02816	CF 3	****
CHO	0.02002	CF2	****
CH3 COO	0.04500	CF	****
CH2COO	0.03567	COO	0.01917
HCOO	0.02667	CC12 F	0.05384
CH30	0.03274	HCC1F	****
CH20	0.02311	CC1F 2	0.05383
CH-O	0.01799	F	****
FCH20	0.02059	CONH 2	****
CH2NH 2	0.02646	CONHCH 3	****
CHNH 2	0.01952	CONHCH 2	****
CH3NH	0.02674	CON(CH 3)2	0.05477

Table A.4: 1st order groups and their V_m contributions (Constantinou et al., 1995)

Group	v_m	Group	v_m
CH2NH	0.02318	CONCHsCH 2	****
CHNH	0.01813	CON(CH2)2	****
CH 3 N	0.01913	C2H502	0.04104
CH2 N	0.01683	C2H402	****
ACNH 2	0.01365	CH3S	0.03484
CsH4N	0.06082	CH2 S	0.02732
CsH3N	0.05238	CHS	****
CHzCN	0.03313	C4H3 S	****
COOH	0.02232	C4H2S	****
CH 2C1	0.03371		

Table A.4: Cont'd

Appendix B: Solubility Estimation Method

Hansen (1976) assumed that solubility is a function of the non-polar (δ_d), polar (δ_p) and hydrogen bonding (δ_h) contribution to the cohesive energy. These solubility parameters can be determined from molecular make-up according to equation 4.5 (van Krevelen, 1990).

$$\delta_d = \frac{\sum F_{di}}{V_m} \qquad \delta_p = \frac{\sqrt{\sum F_{pi}^2}}{V_m} \qquad \delta_h = \sqrt{\frac{\sum E_{hi}}{V_m}} \tag{4.5}$$

F_{di}, F_{ji}, and E_{hi} values for a selection of molecular building blocks are listed in Table B.1. The molar volume (V_m) calculations are predicted using group contribution equation (4.6) and the tabulated group contribution parameters were published by Constantinou et al. (1995). Hildebrandt parameters were originally a measure of cohesive energy density $(cal/cm^3)^{1/2}$ and a newer form that conforms to the Standard International (SI) units is in terms of cohesive pressure $(MPa)^{1/2}$. According to Burke (1984) the conversion between the two units is as follows:

$$\delta (MPa)^{1/2} = 2.0455 \cdot \delta \left(\frac{cal}{cm^3} \right)^{1/2} \tag{B.1}$$

Barton (1985) determined the solute-solvent constraint, R^{ij}, for synthesized molecules according to equation 4.5, where i is the solute and j is the solvent; and for the aniline solvent application example, the calculated solubility parameter are listed in Table B.2 In this case study i represents the designed formulations (M1-M8) and j is for aniline (CAS 62-53-3), whose Hansen's solubility parameters are obtained from the ICAS data bank, and are listed in Table B.2.

$$R^{ij} = \sqrt{4\left(\delta_d^i - \delta_d^j\right)^2 + \left(\delta_p^i - \delta_p^j\right)^2 + \left(\delta_h^i - \delta_h^j\right)^2} \tag{4.4}$$

g	F_{di} $J^{1/2} *cm^{3/2}/mol$	F_{pi} $J^{1/2} .cm^{3/2}/mol$	E_{hi} J/mol
CH_3-	420	0	0
-CH_2-	270	0	0
-CH_2O-	520	400	3000
CH_3CH_2-	690	0	0
CH_3O-	520	400	3000
-CH_2CO	560	770	2000
-CH_3CO	710	770	2000
-O-	100	400	3000
-CO-	290	770	2000

Table B.1: Parameters for estimation of Hansen solubility (van Krevelen, 1990)

CH$_3$	CH$_2$	CH$_2$O	CH$_3$O	CH$_2$CO	CH$_3$CO	V_m m^3/mol	δ_d MPa$^{1/2}$	δ_p MPa$^{1/2}$	δ_H MPa$^{1/2}$	R^{ij} MPa$^{1/2}$
2	4					130.03	14.77	0	0	15.50
2	5					146.44	14.95	0	0	15.29
2	6					162.85	15.11	0	0	15.11
2	7					179.26	15.23	0	0	14.98
1	3			1		141.78	12.63	5.43	3.76	14.88
1	4				1	140.44	15.74	5.48	3.77	9.65
2	8	1				218.78	16.09	1.83	3.70	10.73
1	3		1			120.22	14.56	3.33	5.00	11.58

solubility parameters (ICAS, 2006): δ_d = 19.32, δ_p = 7.86 and δ_H = 9.78 MPa$^{1/2}$

Table B.2: Solubility calculations for candidate solven