A Property Based Design Approach for Simultaneous Optimization of Product and Process Needs
Proceedings of European Congress of Chemical Engineering (ECCE-6)
Copenhagen, 16-20 September 2007

A Property Based Design Approach for Simultaneous Optimization of Product and Process Needs

F. T. Eljack[a,b], C. C. Solvason [a], N. Chemmangattuvalappil [a], M. R. Eden[a]

[a]*Department of Chemical Engineering, Auburn University, Auburn, AL, USA*
[a]*Department of Chemical Engineering, Qatar University, Doha, Qatar*

Abstract

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. The method is capable of simultaneously considering both process and molecular design needs. In that sense it is a truly integrated approach. Visualization aids in deciding which groups to include in the synthesis and which to omit as they will not help in achieving the performance requirements. For cases that require more than three properties, this paper presents an algebraic molecular clustering approach that enables lowering the dimensionality to a simple linear program (LP) molecular operators to setup the design problem as a set of linear algebraic equations.

Keywords: Property clusters, molecular design, algebraic approach

1. Introduction

The terms product synthesis and design designate problems involving identification and selection of compounds or mixtures that are capable of performing certain tasks or possess certain physical 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. However, the performance requirements for the desired component are usually dictated by the process and thus the identification of the desired component properties should be driven by the desired process performance. Numerous contributions have been made in the areas of molecular synthesis and Computer Aided Molecular Design (CAMD) e.g. Harper and Gani (2000) , Marcoulaki and Kokossis (1998); however in order to utilize these techniques the desired component properties must be specified ahead of design. Doing so may lead to suboptimal designs, as the property targets for a new compound inherently will be dictated by the process, where it is to be employed. With molecular design techniques, the desired target properties are required input to the solution algorithm. Once again these decisions are made ahead of design and are usually based on qualitative process knowledge and/or experience, hence again risking a sub-optimal design. Thus there is a critical need for systematic methodologies capable of addressing both problems simultaneously, i.e. identify the target properties of a new compound from the process design problem and subsequently synthesize molecular structures that match the targets.

.

2. Property Integration Framework

Introduction of the property integration framework by Shelley and El-Halwagi (2000) allows for representation of process and products from a properties perspective. Utilizing this methodology enables identification of the desired component properties by targeting optimum process performance without committing to any component during the solution step (Eden et al., 2003). For systems that can be described by only three properties, visualization of the problem as well as identification of the solution is achieved using a ternary diagram. The identified property targets can then be used as inputs for solving a molecular design problem, which returns the corresponding components (see Figure 1).

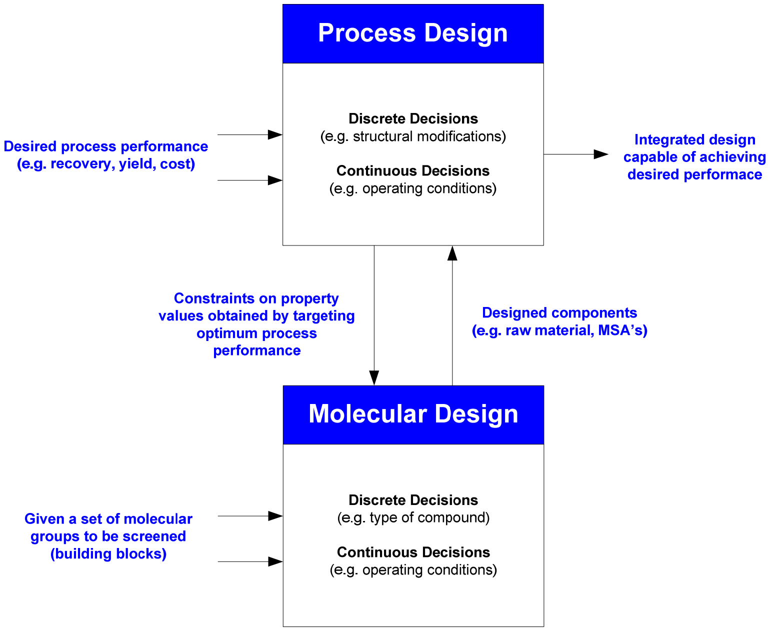

Figure 1: Simultaneous approach to solving process and molecular design problems

To provide a unifying methodology for handling process and molecular design problems, the property integration framework is extended to include Group Contribution Methods (GCM), which allow for prediction of physical properties from structural information. Initially the methods were based on the first order groups (Joback and Reid, 1983); however, work by Constantinou and Gani (1994) and by Marrero and Gani (2001) extended the methods to include second and third order groups to increase the accuracy of the predicted properties. For this first unified framework only first order GC methods are considered, but higher order estimation will be included later.

By combining property clustering techniques and first order group contribution methods (GCM), a systematic methodology is obtained that facilitates simultaneous consideration of property performance requirements of the designed component as well as process and molecular constraints.

2.1 Property Operators

The basis for the property clustering technique is the use of property operators. Property clusters are conserved surrogate properties that are functions of non-conserved properties. They are obtained by mapping property relationships into a low dimensional domain, thus allowing for visualization of the problem [3]. 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 et al., 2004). It is clear that 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 similar fashion (recall that visualization is only feasible for systems that can be described just by three properties). Extending this technique to include GCM for molecular design, introduces molecular property operators. Like the original process operators, their formulation must be such that it still allows for simple linear additive rules of the groups, which can be described by the following:

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

In equation 1, ψ^M_j (P_j) is the molecular property operator of the j^{th} property. The molecular property operator describes the functional relationship of 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 GCM, 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 2 and 3 (Sinh and Achenie, 2001).

$$\log VP = 5.58 - 2.7\left(\frac{T_{bp}}{T}\right)^{1.7} \tag{2}$$

$$\psi^M(T_{bp}) = \exp\left(\frac{T}{t_{bo}}\right) = \sum_{g=1}^{N_g} n_g \cdot t_{b_g} \tag{3}$$

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 are forced to obey simple linear additive rules (Shelley and El-Halwagi, 2000; Eden et al., 2004; El-Halwagi et al., 2004).

Since the properties can have various functional forms and units, the molecular property operators like process property operators are normalized into 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 is given as:

$$\Omega^{M}_{ji} = \frac{\psi^{M}_{j}(P_{ji})}{\psi^{ref}_{j}(P_{ji})} \tag{4}$$

An Augmented Property index AUP^{M} for each molecule (M) is defined as the summation of all the NP dimensionless property operators, (Ω^{M}):

$$AUP^{M} = \sum_{j=1}^{NP} \Omega^{M}_{j} \tag{5}$$

And the property cluster C_j for property j is defined as the ratio of the Ω^{M} and AUP^{M}:

$$C_{j} = \frac{\Omega^{M}_{j}}{AUP^{M}} \tag{6}$$

2.2 Visualization of the Problem

The conversion of property data to cluster values for process design has been developed by Eden et al. (2004). Similarly the conversion of molecular property data to cluster values follows the outline in Table 1.

Step	Description	Equation
1	Calculate molecular property operators	1
2	Calculate dimensionless molecular property values	4
3	Calculate molecular AU P indices	5
4	Calculate ternary cluster values for each formulation	6
5	Plot the points on the ternary cluster diagram	--

Table 1: Calculation of cluster^{M} values from GCM predicted property data

The primary visualization tool from the mass integration framework is the source-sink mapping (El-Halwagi, 1997). This tool is utilized in the molecular synthesis framework. In the original cluster formulation for process design (described by three properties), mixing of two sources is a straight line on the ternary diagram, 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 2). Design and optimization rules have been developed for property based process design problems (Eden et al., 2004; El-Halwagi et al., 2004; Eljack et al., 2005), and in the following similar rules are presented for property based molecular design problems (Eljack et al., 2006):

Rule 1: Two groups, G1 and G2, are added linearly on the ternary diagram, where the visualization arm β_1, describes the location of G1-G2 molecule.

$$\beta_1 = \frac{n_1 \cdot AUP_1}{n_1 \cdot AUP_1 + n_2 \cdot AUP_2} \qquad (7)$$

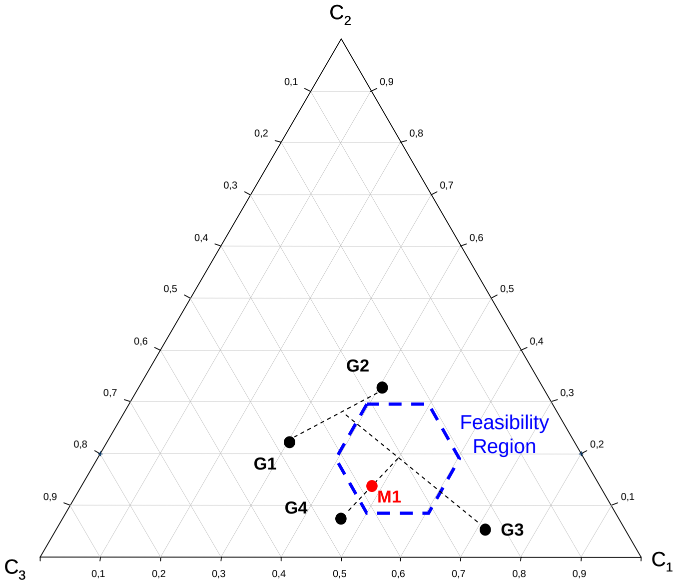

Figure 2: Example of visual molecular synthesis.

Rule 2: More groups can be added as long as the Free Bond Number (FBN) is not zero.

$$FBN = \left[\sum_{g=1}^{N_g} n_g \cdot FBN_g\right] - 2 \cdot \left[\sum_{g=1}^{N_g} n_g - 1\right] - 2 \cdot NO_{Rings} \qquad (8)$$

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 3: Location of the final formulation is independent of the order of group addition.

Rule 4: 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 7 is the sufficient condition.

Rule 5: The cluster value of the formulation must be contained within the feasibility region of the sink on the ternary molecular cluster diagram.

Rule 6: 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 7: 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 property.

2.3 Feasibility region boundaries

Constraints placed on both process and molecular design problems are represented by a feasibility region on the ternary diagram. These property values are mapped as a feasibility region defined by six unique points, according to the findings of El-Halwagi et al. (2004), and they are summarized in Rule 8.

Rule 8: *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

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

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. In addition, the clustering technique reduces the dimensionality of both problems, thus for systems that can be described by just three properties it is possible to visually identify the solutions, which is a significant advantage of this approach. The integrated clustering methodology for the simultaneous solution of process and molecular design is summarized in Figure 3.

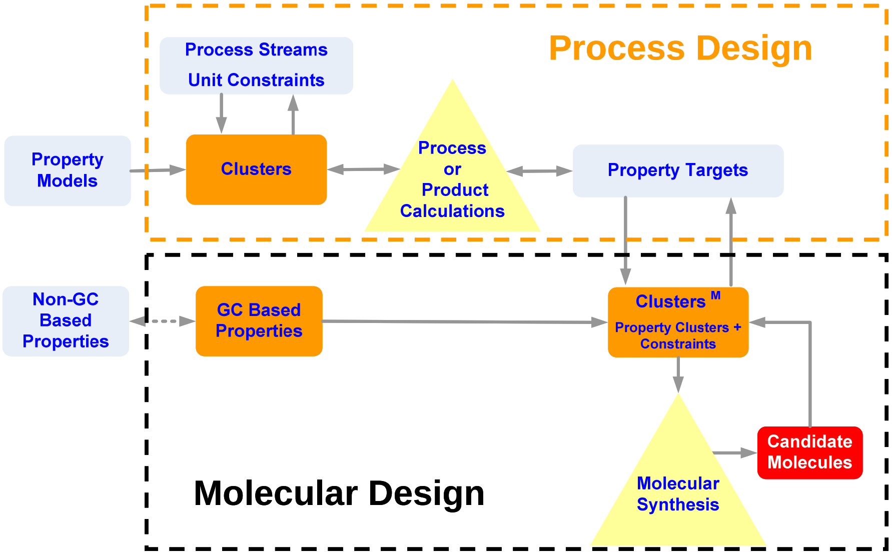

Figure 3: Outline of simultaneous approach to process and molecular design.

3. Degreaser Case Study 1

A case study is discussed here to show the merits of using the simultaneous approach to solving process and molecular design problems via GCM and property clusters. Figure 4 illustrates a metal degreasing facility, with an absorber and degreaser. The process fresh resources are in the form of two organic solvent streams (Shelley and El-Halwagi, 2000). The off-gas Volatile Organic Compounds (VOCs) are a byproduct from the degreasing unit, and the current treatment of this stream is flaring. The problem with such a treatment is it leads to economic loss and environmental pollution (Eden, 2003).

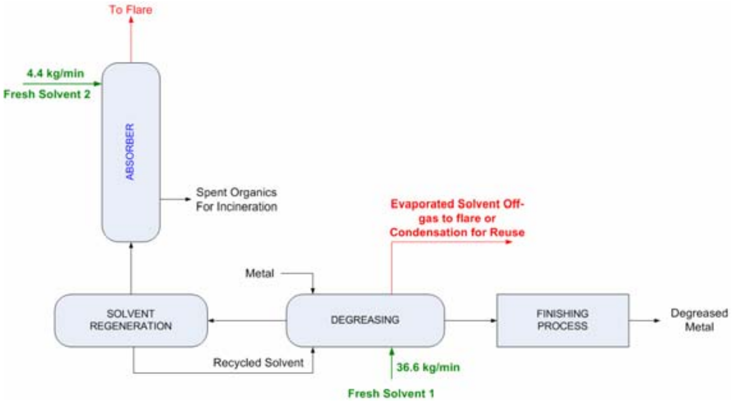

Figure 4: Schematic representation of original metal degreasing process

In this case study, the objective is to explore the possibility of condensing the off gas VOCs, to (1) minimize the use of fresh solvent and (2) to simultaneously identify candidate alternative solvents for the degreaser (See Figure 5). Three properties are examined to determine the suitability of a given organic process fluid for use in the degreaser:

- Sulfur content (S) - for corrosion consideration, expressed as weight percent.
- Molar Volume (V_m) - for hydrodynamic and pumping aspects.
- Vapor Pressure (VP) – for volatility, makeup and regeneration.

The solvents that will be synthesized will be pure component fluids; hence the sulfur content of these streams will be zero.

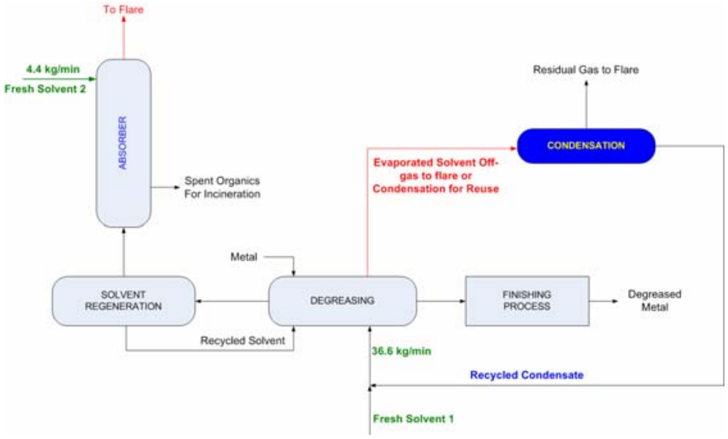

Figure 5: Schematic representation of metal degreasing process after property integration

3.1 Process Design

The constraints on the inlet streams to the degreaser are given in Table 2.

Property	Lower Bound	Upper Bound
S (%)	0.00	1.00
Vm (cm^3/mol)	90.09	487.80
VP (mmHg)	1596	3040
Tb (K)	430.94	463.89
Flow rate (kg/min)	436.6	36.8

Table 2: Feed constraint to the degreaser

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

$$S_M = \sum_{s=1}^{Ns} x_s \cdot S_s \qquad , S^{ref} = 0.5\ wt\% \tag{9}$$

$$V_{m_M} = \sum_{s=1}^{Ns} x_s \cdot V_{m\,s} \qquad , V_m^{ref} = 80\ cm^3/mol \tag{10}$$

$$VP_M^{1.44} = \sum_{s=1}^{Ns} x_s \cdot VP_s^{1.44} \qquad , VP^{1.44,ref} = 760\ mmHg \tag{11}$$

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 condensate of the VOCs, are converted to cluster values according to cluster methodology developed by Eden et al. (2004), see Figure 6. The degreaser property constraints are translated as a feasibility region.

Now that the problem has been mapped to the property domain and visualized on the ternary diagram, some constraints are placed on the process: the condensate temperature is set to 500K and the fresh synthesized solvents are sulfur free. 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 7) Using lever arm analysis, between the degreaser feasibility region and the condensate recycle stream, point A and point B are now identified on Figure 7. 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.

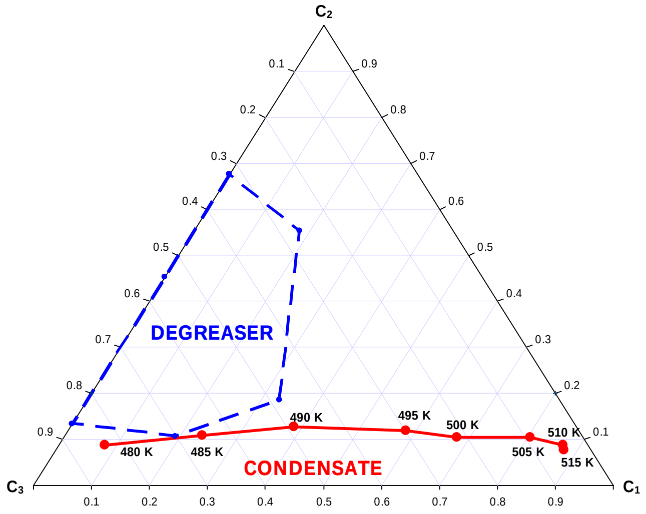

Figure 6: Ternary representation of metal degreaser problem in process design

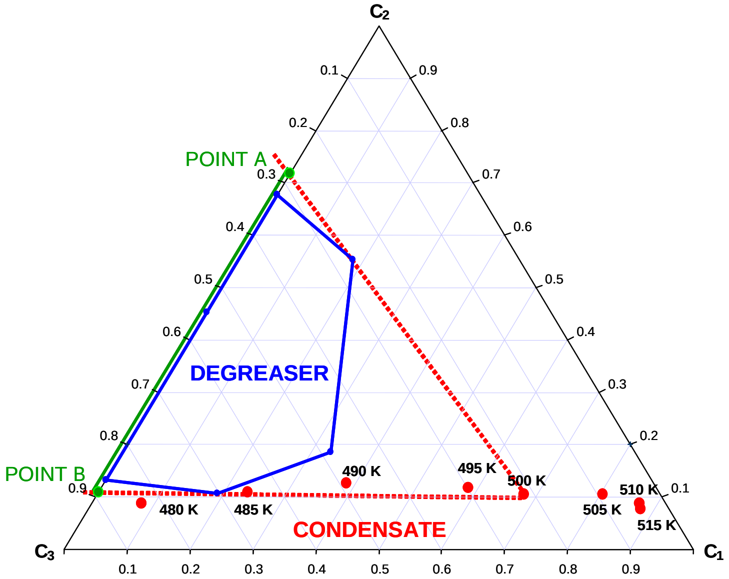

Figure 7: Identifying property targets of the fresh solvent needed to maximize condensate recycle

3.2 Molecular Design - Fresh Solvent Synthesis

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

The cluster values associated with points A and B from the clustering diagram in Figure 7, are translated to physical property values using the methodology developed by Shelley and El-Halwagi (2000) and Eden et al. (2004). These property targets obtained from solving our process design problem are now the upper and lower property constraints placed on the solvent/molecular design problem, see Table 3.

	S (%)	**VP** (mmHg)	**V_m** (cm3/mol)
Point A	0.00	1825.4	720.8
Point B	0.00	3878.7	102.1

Table 3: Corresponding property values for cluster values obtained from process design

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 (Achenie and Sinha, 2004). However according to equation (2) 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 4.

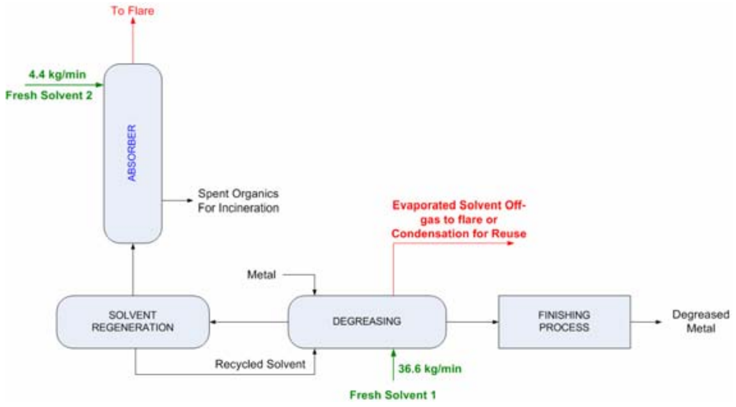

Figure 4: Schematic representation of original metal degreasing process

The physical properties are predicted using the following 1st order group contribution equations (Constantinou and Gani, 1994; Constantinou et al., 1995):

$$\Delta H_v = h_{vo} + \sum_i n_i \cdot h_{v1_i} \qquad (12)$$

$$V_m = d + \sum n_i \cdot v_{1_i} \qquad (13)$$

$$T_{bo} = t_{bo} \cdot \ln \sum_i n_i \cdot t_{b1} \qquad (14)$$

The property operators derived from the above equations and their reference values are summarized in Table 5. Notice that RHS of the equation allows for the linear additive rules.

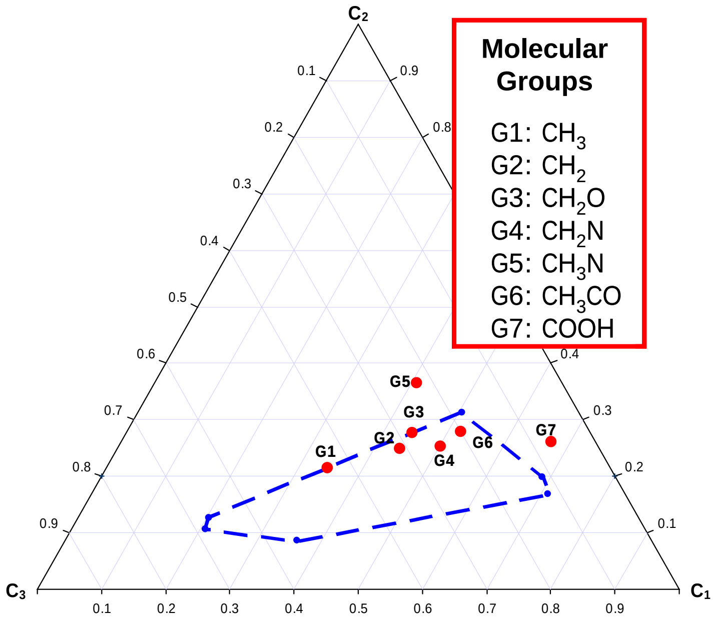

Figure 8: Ternary diagram used to represent molecular synthesis problem

The problem is visualized by converting the property targets to cluster values following the methodology described in Table 1. The targeted properties are represented by a feasibility region, which has been identified as outlined in section 2.2. The resulting ternary diagram is shown in Figure 8, where the dotted lines represent the feasibility region in the molecular design domain.

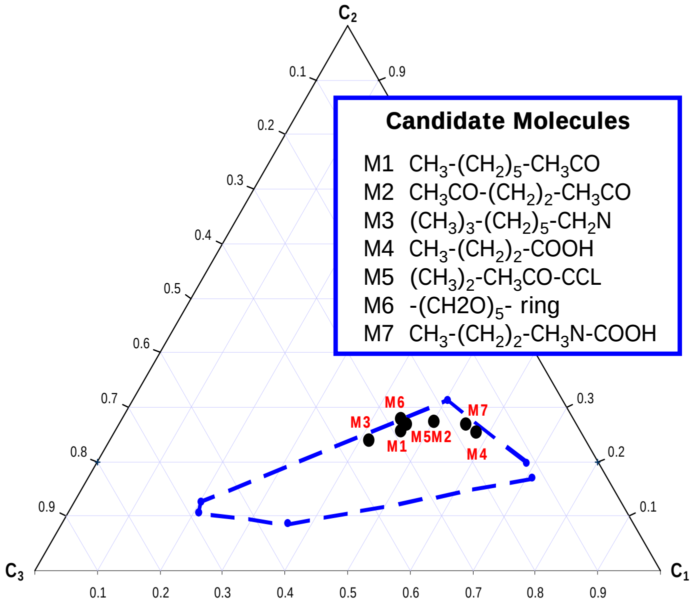

Figure 9: Candidate formulated molecules

The design problem also states that 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 8 and were chosen to highlight the visual aspects of this clustering technique. The selection of building blocks is considered a pre-design step. Any other constraints on the molecular design problem can be imposed at this point.

Translating health and environmental concerns to the level of molecular synthesis can be achieved via exclusion or inclusion of certain molecular groups, such as chloro compounds and amines or by excluding the formation of cyclical compounds. The latter is translated to the design problem by simply expressing the NO_{Rings} to be zero.

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 are forced to obey simple linear additive rules. Seven candidates, M1-M7, are formulated for this solvent design problem (see Figure 9). However, the validity of the formulations is satisfied only after satisfying conditions summarized by Rules 4-8 in Section 2.2. The cluster values of the designed molecules, M1-M7, are checked to make sure that they lie within that of the sink. The values of the augmented property index of the designed molecule must lie within the AUP range of the sink; in the degreaser case study the AUP of the sink ranged from 4.22-12.65, see Table 6. It is seen that molecules M5 and M6 fail to satisfy this condition.

Formulation	AUP	T_b (K)	H_v (kJ/mol)	V_m (cm3/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 4: Molecular candidate AUP and property value

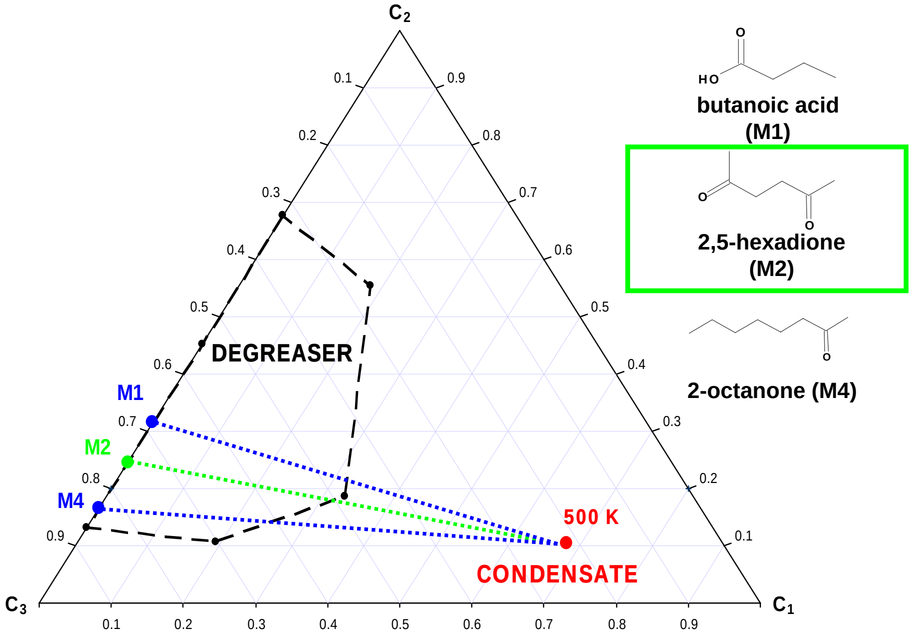

Figure 10: Candidate molecular solutions obtained to satisfy optimized process design targets

The final necessary and sufficient condition is 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 earlier in Section 2.2. Molecule M3 fails to satisfy the property condition in the molecular domain; 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.

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-hexanedione, and butanoic acid respectively. The valid molecular structures are shown in Figure 10. The three candidates are mapped back to the process design framework to identify the formulation that will maximize recycle of condensate at 500K. Using lever arm analysis, 19.36 kg/min of fresh solvent 2,5-hexadione will allow for maximum condensate flow rate of 17.44 kg/min.

4. Gas Purification Case Study 2

A current gas treatment process uses fresh methyl diethanol amine, MDEA, ($HO-(CH_2)_4-CH_3N-OH$) and two other recycled process sources (*S*1, and *S*2) as a feed (See Figure 11). Another process stream, *S*3, currently a waste stream could be recycled as a feed if mixed with a fresh source to allow the mixed stream properties to match that of the process sink (Kazantzi et al., 2007). The property and flowrate data for all streams (*S*1, *S*2 and *S*3) and the sink are summarized in Table 7.

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 (*S*1, *S*2 and *S*3), see Figure 12. 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}).

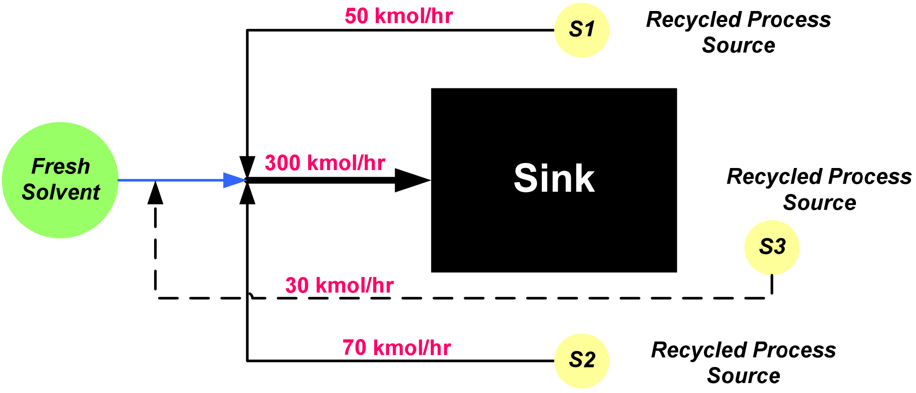

Figure 11: Schematic Representation of original gas treatment unit

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 7: Property data for gas purification example

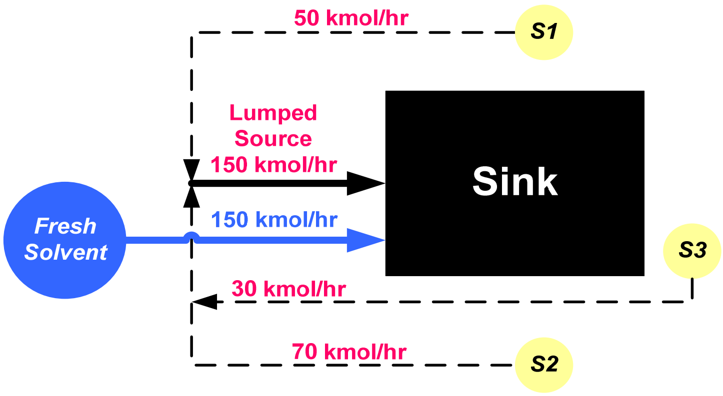

Figure 12: Schematic representation of gas treatment unit after integration

4.1 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 (Eden, 2003). 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:

$$V_{c\,M} = \sum_{s=1}^{Ns} x_s \cdot V_{c\,s} \qquad , V_{c,\,ref} = 2.5\ cm^3/mol \tag{15}$$

$$H_{v\,M} = \sum_{s=1}^{Ns} x_s \cdot H_{v\,s} \qquad , H_{v,\,ref} = 0.35\ kJ/mol \tag{16}$$

$$H_{fus\,M} = \sum_{s=1}^{Ns} x_s \cdot H_{fus\,s} \qquad , H_{fus,\,ref} = 0.10\ kJ/mol \tag{17}$$

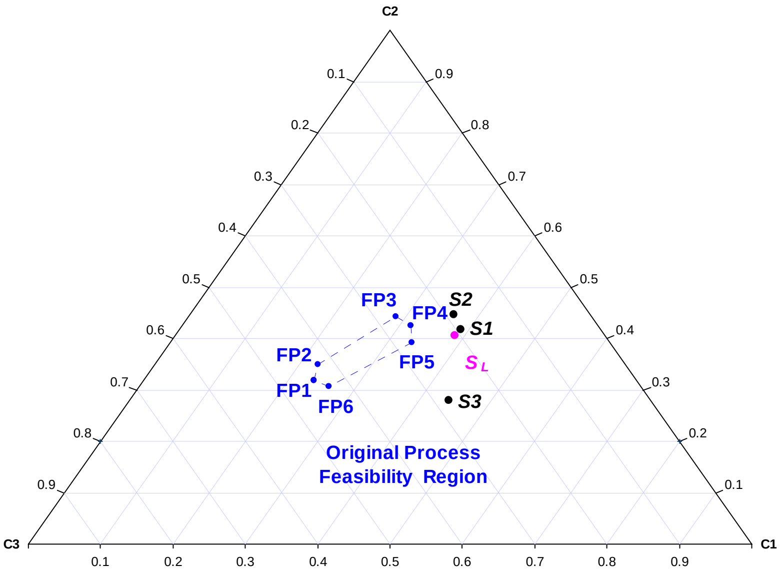

Figure 13: Gas purification process – feasibility regions and streams

Source	V_c cm3/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 8: Mixture proeprty data of lumped source (S_L)

Boundary constraints of the sink will be determined according to Rule 8 by six unique points seen as FP1-FP6 on Figure 13; 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 8.

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 13 and 14, there are two feasibility regions. The first reflect the sink's original property demands as seen in Table 7, 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.

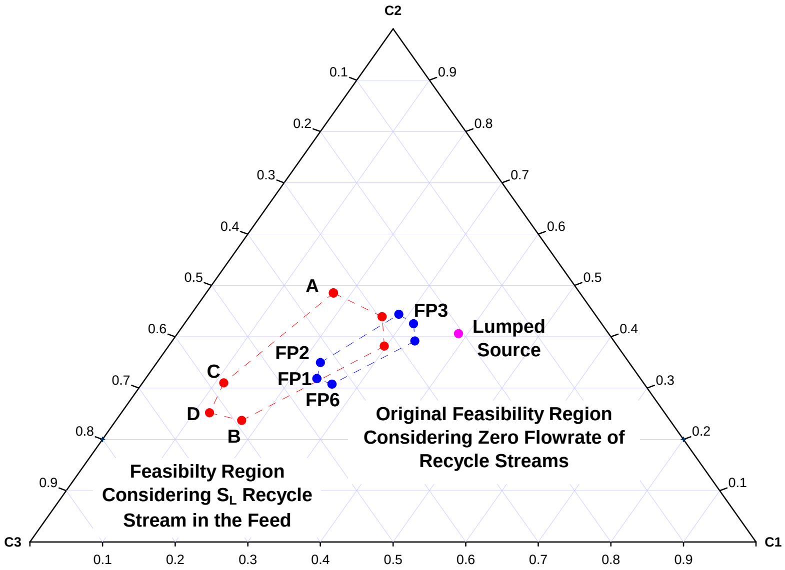

Figure 14: New feasibility region – reflects mixture/blend design constraints

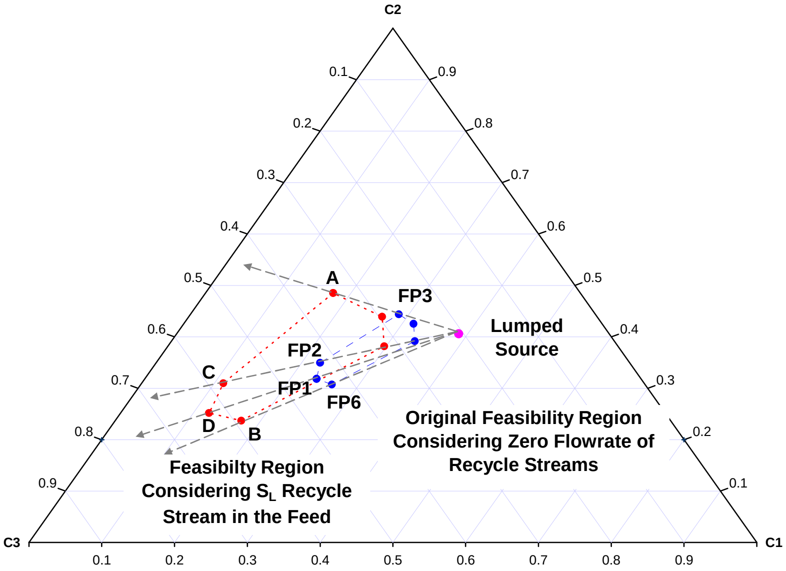

Figure 15: Identification of mixture (new) feasibility region

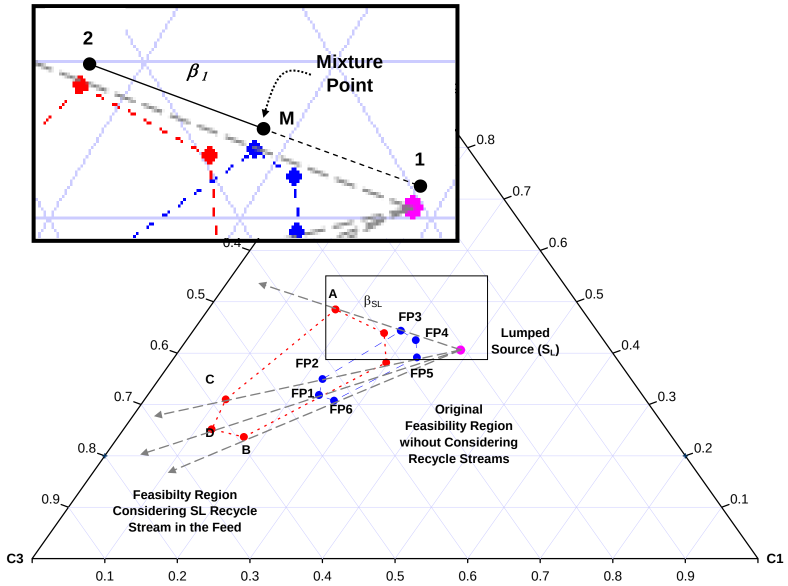

Figure 16: New feasibility region – Gas Purification Example

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 (Eden, 2003). 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 16 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 on Table 9.

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 18 is used to calculate the value of β_1.

$$\beta_s = \frac{x_s \cdot AUP_s}{AUP_M} \tag{18}$$

Next, the cluster values (C_{12}, C_{22} and C_{32}) for point *2* on Figure 16 are calculated according to the cluster conservation rule. Expanding equation 19, results in the following:

$$C_{jM} = \sum_{s=1}^{N_s} \beta_s \cdot C_{js} \tag{19}$$

$$\begin{aligned} C_{1M} &= \beta_1 \cdot C_{11} + (1-\beta_1)\cdot C_{12} \\ C_{2M} &= \beta_1 \cdot C_{21} + (1-\beta_1)\cdot C_{22} \\ C_{3M} &= \beta_1 \cdot C_{31} + (1-\beta_1)\cdot C_{32} \end{aligned} \tag{20}$$

The steps outlined above are used to determine the remaining points B – D (see Table 9). The six cluster points and their respective property values that bind the new feasibility region are summarized in Table 10. The property values are back calculated from the property operator expressions and reference values (equations 15-17).

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 11); 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 9: Calculation data for new feasiblity region

New Feasibility PT	V_c	H_v	H_{fus}	Ω_1	Ω_2	Ω_3	AUP_s	C_1	C_2	C_3	ΣC_{js}	Xcc	Ycc
Point A	310	120	24	124	343	240	707	0.262	0.307	0.43	1	0.418	0.485
Point C	310	120	64	124	343	640	1107	0.334	0.392	0.274	1	0.267	0.310
Point D	310	90	64	124	257	640	1021	0.316	0.425	0.259	1	0.247	0.252
Point B	470	90	64	188	257	640	1085	0.286	0.444	0.27	1	0.292	0.237
Point F	823	148	36	329	422	356	685	0.175	0.485	0.34	1	0.488	0.381
Point E	735.30	170.01	32.70	294	486	327	771	0.112	0.310	0.578	1	0.485	0.439

Table 10: New Feasibility Region Data

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

Table 11: Determined property constraints for molecular design algorithm

4.2 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 ($\psi^M{}_1$, $\psi^M{}_2$, $\psi^M{}_3$), see table 12. The molecular feasibility region for the design problem has been plotted on Figure 17. 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 18). 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 154 - 257. The candidate formulation M1 failed to satisfy the lower limit of H_v property (see Table 13). Hence, M2-M5 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 19, 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 12: Property operators for purificiation molecular synthesis

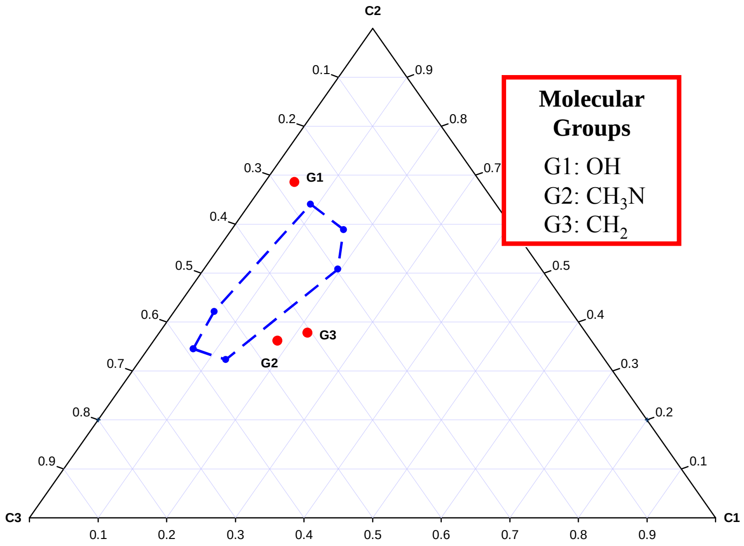

Figure 17: Molecular synthesis of gas purification solvent

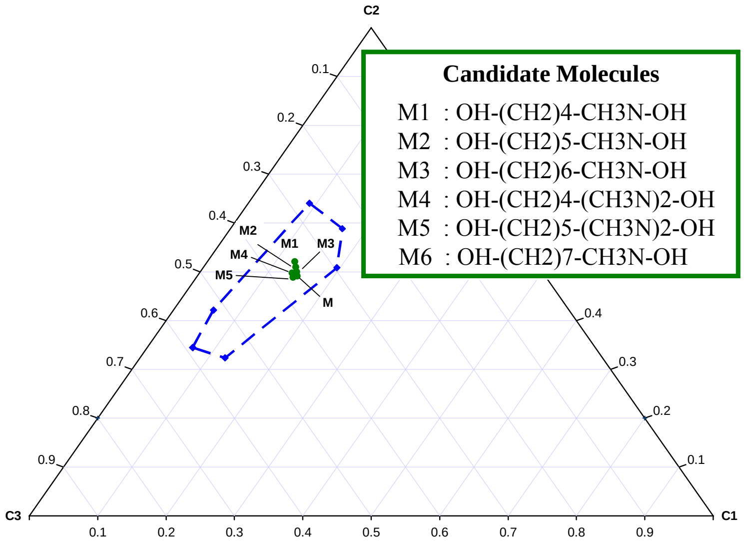

Figure 18: Candidate molecules for gas purification solvent

Candidates	V_c (cm3/mol)	H_v (kJ/mol)	H_{fus} (kJ/mol)
$M1$	389.23	~~89.294~~	23.33
$M2$	445.51	94.204	25.969
$M3$	501.79	99.114	28.608
$M4$	484.17	98.787	29.338
$M5$	540.45	103.697	31.977
$M6$	558.07	104.024	31.247

Table 13: Candidate property data for gas purification solvent

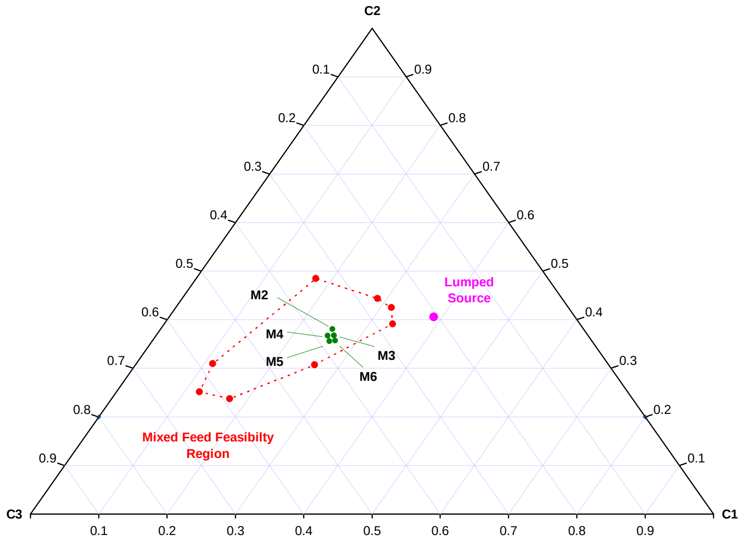

Figure 19: Verification of candidate molecules in process domain

5. 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.

5.1 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:

$$P_{ij}^{lower} \leq P_{ij} \leq P_{ij}^{upper} \qquad (21)$$

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 4) with the corresponding reference values.

$$\Omega_j^{\min} \leq \Omega_{ij} \leq \Omega_j^{\max} \qquad (22)$$

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

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

The substitution of equation 1 into the inequality expression given by equation 22 generates the following:

$$\Omega_j^{\min} \leq \sum_{g=1}^{N_g} n_g \cdot \Omega_{jg} \leq \Omega_j^{\max} \qquad (23)$$

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., 2007):

Rule 9: Each property constraint is translated into the inequality expression from equation 23, and then split into two equations, one for minimum (min) and one for maximum (max).

$$\Omega_j^{\min} \le \sum_{g=1}^{N_g} n_g \cdot \Omega_{jg} \qquad\qquad \sum_{g=1}^{N_g} n_g \cdot \Omega_{jg} \le \Omega_j^{\max} \tag{24}$$

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:

$$\begin{matrix} (\Omega_1^{\max},\Omega_2^{\min},\Omega_3^{\min},\Omega_4^{\min}) & & (\Omega_1^{\min},\Omega_2^{\max},\Omega_3^{\max},\Omega_4^{\max}) \\ (\Omega_1^{\min},\Omega_2^{\max},\Omega_3^{\min},\Omega_4^{\min}) & & (\Omega_1^{\max},\Omega_2^{\min},\Omega_3^{\max},\Omega_4^{\max}) \\ (\Omega_1^{\min},\Omega_2^{\min},\Omega_3^{\max},\Omega_4^{\min}) & , & (\Omega_1^{\max},\Omega_2^{\max},\Omega_3^{\min},\Omega_4^{\max}) \\ (\Omega_1^{\min},\Omega_2^{\min},\Omega_3^{\min},\Omega_4^{\max}) & & (\Omega_1^{\max},\Omega_2^{\max},\Omega_3^{\max},\Omega_4^{\min}) \end{matrix} \tag{25}$$

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

26) and a possible limit on the length of a molecular formulation (equation 27):

$$n_g \geq 0 \qquad g = \{1, \ldots, N_g\} \tag{26}$$

$$\sum_{g=1}^{N_g} n_g \leq NF \tag{27}$$

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.

$$\mathrm{FBN} = \left[\sum_{g=1}^{N_g} \mathrm{n_g} \cdot \mathrm{FBN_g}\right] - 2 \cdot \left[\sum_{g=1}^{N_g} \mathrm{n_g} - 1\right] \tag{28}$$

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 (28) 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 14.

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	4
2	Express property constraints as inequalities forming the main set of inequality equations	21 - 22
3	Determine the *AUP* range of the sink	5
4	Develop the subsets of inequality equations following Rule 10	--
5	Generate the structural constraints	26 - 28
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 $n_1, n_2 \ldots n_g$ are defined as integer values.	--
9	Generate all the feasible formulations and perform the final checks that all property constraints are satisfied	--

Table 14: Outline of algebraic molecular cluster approach.

5.2 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 (Eljack, 2007):

$$310 \leq V_{\mathrm{c}}\,(\mathrm{cm^3/mol}) \leq 610 \qquad 90 \leq H_{\mathrm{v}}\,(\mathrm{kJ/mol}) \leq 120$$

$$20 \leq H_{\mathrm{fus}}\,(\mathrm{kJ/mol}) \leq 64 \qquad 450 \leq T_{\mathrm{b}}\,(\mathrm{K}) \leq 560 \tag{29}$$

g	Group	*FBN*	V_{c} ($\mathrm{cm^3/mol}$)	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 15: Property data for each molecular group.

The Group Contribution (GC) property data of the molecular groups is given in Table 15. In addition, the additive rules for the molecular operators of the targeted properties are represented by equation 30 (Constantinou and Gani, 1994; Marrero and Gani, 2001). The formulation of the operators from GC property models is outlined in the molecular clustering framework (Section 2.2; Eljack et al., 2006).

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

$$H_{fus} - h_{fus0} = \sum_{g=1}^{N_g} n_g \cdot h_{fus1} \qquad \exp\left(\frac{T_b}{t_{bo}}\right) = \sum_{g=1}^{N_g} n_g \cdot t_{b1} \tag{30}$$

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 29 and 30, and the information in Table 15, 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 4) 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 15. 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 16. The *AUP* range of the feasibility region (sink) was calculated to be 141.19 – 273.27.

Next, the provided data along with equation 24 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 25 are provided below in equation 31. The remaining 7 subsets are generated in the same way. Finally the structural constraints are given in equation 32 and 33.

	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

Table 16: Calculated Ω for the given property constraints.

$$\begin{aligned} & 2.81 \cdot g_1 + 1.53 \cdot g_2 \leq 30.10 \\ & 4.91 \cdot g_1 + 24.21 \cdot g_2 \geq 78.26 \\ & 5.28 \cdot g_1 + 9.571 \cdot g_2 \geq 45.61 \\ & 0.131 \cdot g_1 + 0.459 \cdot g_2 \geq 1.29 \end{aligned} \tag{31}$$

$$g_1 \geq 0 \quad , \quad g_2 \geq 0 \quad , \quad g_1 + g_2 \leq 15 \tag{32}$$

$$\left[g_1 \cdot FBN_1 + g_2 \cdot FBN_2 \right] - 2 \cdot \left[g_1 + g_2 - 1 \right] = 0 \tag{33}$$

The results from solving the subsets equations are summarized in Table 17. 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 17. Result of solving to the molecular synthesis problem

6. Conclusions

In this paper, a systematic property based framework for simultaneous solution of process and molecular design problems has been presented. The recently introduced property integration framework has been extended to include group contribution methods for molecular design. 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.

A significant result of the developed methodology is that for problems that can be satisfactorily described by just three properties, the process and molecular design problems are solved visually and simultaneously on a ternary diagram, irrespective of how many molecularly fragments are included in the search space.

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.

References

Achenie LEK and Sinha M (2004). Advances in Environmental Research; 8(2): 213-227.

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

Constantinou L and Gani R (1994). AIChE Journal; 40(10): 1697-1710.

Constantinou L, Gani R and O'Connell J (1995). Fluid Phase Equilibria; 103(1): 11-22.

Eden MR (2003). CAPEC, Department of Chemical Engineering.Technical University of Denmark: Ph.D.

Eden MR, Jorgensen SB, Gani R, and El-Halwagi MM. (2004). Chemical Engineering and Processing; 43(5): 595-608.

Eden MR, Jørgensen SB, Gani R and El-Halwagi MM (2003). Computer Aided Chemical Engineering, 15A.

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

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

El-Halwagi MM, Glasgow IM, Qin X and Eden MR (2004) AIChE Journal 50; (8): 1854-1869.

Eljack FT, Eden MR, Kazantzi V and El-Halwagi M (2006). ESCAPE 16 Conference; Garmisch-Paternich, Germany: Elsevier.

Eljack FT, Abdelhady AF, Eden MR, Gabriel F, Qin X and El-Halwagi MM (2005). Computers & Chemical Engineering; 29(11-12): 2304-2317.

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

Eljack, F. T. (2007). A Property Based Approach to Integrated Process and Molecular Design. Chemical Engineering. Auburn, Auburn University. Ph.D.

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.

Harper PM and Gani R. (2000). Computers & Chemical Engineering: 24(2-7): 677-683.

Joback KG and Reid RC (1983). 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)

Marcoulaki EC and Kokossis AC (1998). Computers & Chemical Engineering, 22(Supplement 1): S11-S18.

Marrero J and Gani R. (2001). Fluid Phase Equilibria; 183-184: 183-208.

Shelley MD and El-Halwagi MM (2000). Computers & Chemical Engineering. ; 24(9-10): 2081-2091.

Sinha M and Achenie LEK (2001). Advances in Environmental Research; 5(3): 239-249.