

# AN ALGEBRAIC APPROACH FOR SIMULTANEOUS SOLUTION OF PROCESS AND MOLECULAR DESIGN PROBLEMS

S. Bommareddy, N. G. Chemmangattuvalappil, C. C. Solvason and M. R. Eden\*

Department of Chemical Engineering, Auburn University, AL, 36849, USA.  
E-mail: edenmar@auburn.edu

(Submitted: December 11, 2009 ; Revised: July 28, 2010 ; Accepted: July 30, 2010)

**Abstract** - The property integration framework has allowed for simultaneous representation of processes and products from a properties perspective and thereby established a link between molecular and process design problems. The simultaneous approach involves solving two reverse problems. The first reverse problem identifies the property targets corresponding to the desired process performance. The second reverse problem is the reverse of a property prediction problem, which identifies the molecular structures that match the targets identified in the first problem. Group Contribution Methods (GCM) are used to form molecular property operators that will be used to track properties. Earlier contributions in this area have worked to include higher order estimation of GCM for solving the molecular design problem. In this work, the accuracy of the property prediction is further enhanced by improving the techniques to enumerate higher order groups. Incorporation of these higher order enumeration techniques increases the efficiency of property prediction and thus the application range of the group contribution methods in molecular design problems. Successful tracking of properties is the key in applying the reverse problem formulation for integrated process and product design problems. An algebraic technique has been developed for solving process and molecular design problems simultaneously. Since both process and molecular property operators target the same optimum process performance, the set of inequality expressions can be solved simultaneously to identify the molecules that meet the desired process performance. Since this approach is based on an algebraic algorithm, any number of properties can be tracked simultaneously.

**Keywords:** Property Operators; Reverse Problem Formulation; Molecular Design.

## INTRODUCTION

Molecular design is an important activity that helps in obtaining the best performance from a process. Numerous contributions have been made in the field of Computer-Aided Molecular Design (CAMD). Many of these methods include the use of Group Contribution Methods (GCM) which utilize tables comprised of various molecular fragments/groups and their contribution towards a property of the molecule. Higher order groups are also given in these tables to better explain the change in the contribution of a group towards a property due to its neighboring groups. Employing systematic algebraic techniques to design molecules based on GCM decreases the permutations of groups that need to be checked if they have a valid

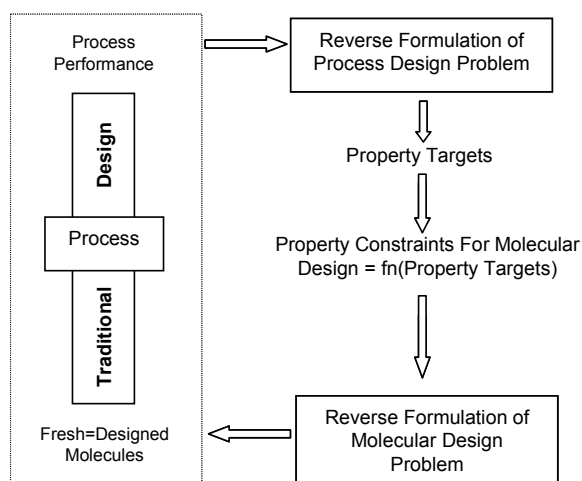
molecule hidden in them. Algorithms to identify the molecules that meet the process targets have been developed by many research groups, including Marcoulaki and Kokossis (1998), Harper *et al.* (1999), Harper and Gani (2000), Eljack *et al.* (2007) and Chemmangattuvalappil *et al.* (2009). In this contribution, the focus is to introduce methods that improve the efficiency of solving the molecular design problem as part of the simultaneous solution of process and product design problems. Earlier methods either did not efficiently incorporate higher groups within the algorithm or did not incorporate the contribution of higher order groups during the initial stage, thus increasing the number of combinations that need to be checked for whether the molecule leading to given process performance is structurally sound.

---

\*To whom correspondence should be addressed

This is an extended version of the manuscript presented in the PSE 2009 - 10th International Symposium on Process Systems Engineering, 2009, Salvador, Brazil, and published in Computer Aided Chemical Engineering, vol. 27, p. 1096-1100.

Incorporating higher order groups at a later stage in the algorithm may also lead to a situation where some potential combinations are omitted without being considered in further stages of the algorithm. There is a definite need for solving process and molecular design problems together, as the solution space is limited if these problems are solved separately. For example, the properties of fresh material added to a process depend on existing recycle streams within the process. Solving the process design problems alone would require committing to specific raw materials well in advance to secure a solution. Hence, when process and product design problems are solved together, it will be possible to design molecules that meet the desired process performance. The identification of optimal molecule(s) corresponding to optimum process performance is a challenging issue. Molecular design subject to process constraints through Property Integration and Group Contribution Methods is one possible solution to overcome this limitation. The concept of reverse problem formulation (RPF) has helped formulate integrated process-product design problems without leading to complex MINLP formulations by insightful decoupling of the constitutive equations from the process model (Eden *et al.*, 2003a). RPF enables design of novel molecules and solution of process design problems without commitment to specific components during the solution step (see Figure 1). Techniques have been developed by Eden *et al.* (2003a, 2003b) for the identification of property targets corresponding to the optimum process performance using a visual approach. An algebraic approach based on a two-stage generate and test methodology would be an efficient extension to these methods, particularly for large problems that normally are prone to combinatorial explosion.



**Figure 1:** Simultaneous Solution using Reverse Problem Formulation

## PROPERTY INTEGRATION

Processes and molecules can be represented by their properties through the Property Integration framework (Shelley and El-Halwagi, 2000; Eden *et al.*, 2003a; Eljack *et al.*, 2007). The key point lies in targeting the optimal property ranges of the desired molecules for a given process performance. These property ranges serve as the input to the molecular design problem which returns the possible molecules. To enable this simultaneous solution approach, Group Contribution Methods (Marrero and Gani, 2001) that assist in predicting the properties of molecules based on their structural characteristics were incorporated into the Property Integration framework.

### Process Property Operators

Properties form the basis of the performance of many process units and so they should be considered as design variables for a given process/product performance. As properties are not generally conserved, a function is tailored for an original property to find its conserved surrogate (Shelley and El-Halwagi, 2000). These functions, called the Process Property Operators, are often non-linear, but they are tailored to add the property contributions from different streams linearly to obtain the conserved property surrogates corresponding to the properties of the product stream:

$$\phi_j(P_j) = \sum_i^{N_{\text{sources}}} x_i \phi_j(P_{ji}) \quad (1)$$

where  $\phi_j(P_{ji})$  is the property operator of the  $j^{\text{th}}$  property  $P_{ji}$  of stream  $i$ ,  $x_i$  is the fractional contribution and  $N_{\text{sources}}$  is the total number of streams. For instance, density does not exhibit a linear mixing rule; however, the reciprocal value of density follows a linear mixing rule. Hence, the Property Operator for density would be its inverse:

$$\frac{1}{\rho} = \frac{1}{\rho_1} + \frac{1}{\rho_2} + \frac{1}{\rho_3} \quad (2)$$

### Group Contribution Method for Property Estimation

The group contribution method (GCM) provides a convenient way to relate molecular structures to properties. In GCM, the property function  $f(Y)$  of a compound is estimated as the sum of the property contributions of all the molecular groups present in the structure (Marrero and Gani, 2001):

$$f(Y) = \sum_f N_f C_f + \sum_s N_s C_s + \sum_t N_t C_t \quad (3)$$

$N_f$ ,  $N_s$  and  $N_t$  are the number of first, second and third order groups and  $C_f$ ,  $C_s$  and  $C_t$  are their respective property contributions. The property function is complex, but property contributions from each molecular group are forced to obey simple linear additive rules. Higher order groups in the group contribution tables helps in describing the molecules better because they account for proximity effects.

### Molecular Property Operators

The operator expressions for molecular fragments differ from that of process streams. However, since they represent the same properties, it is possible to relate them. Inclusion of GCM for molecular design introduces molecular property operators. The molecular property operators are tailored in such a way that the right-hand side 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  $g$ , ( $P_{jg}$ ). The molecular property operator  $\phi_j^M$  is given by (Eljack et al., 2007):

$$\phi_j^M(P_{j,m}) = \sum_{g=1}^{N_g} n_g P_{j,g} \quad (4)$$

## SIMULTANEOUS PROCESS AND MOLECULAR DESIGN

Traditionally process and molecular design problems have been solved separately with little or no feedback between each other. Since both design problems aim at satisfying a given process performance, they can be connected from a property perspective to bridge the gap between process and molecular design problems. Visualization approaches have been developed (Eljack *et al.*, 2007; Kazantzi *et al.*, 2007), but, due to their inability to handle more than three properties efficiently in these componentless design problems, an algebraic approach is needed. Also, scaling down data to fit into a graphical representation may prevent it from capturing data at the right level of detail (Qin *et al.*, 2004). An algebraic molecular design approach using higher order groups has been developed (Chemangattuvalappil *et al.*, 2009), but its application range could be improved if

the accuracy of the property prediction is enhanced by better techniques for enumerating higher order groups. The technique incorporated in this work aims at reducing the number of combinations of groups that should be checked for forming a molecule with the required properties. This new method is significantly less computationally intensive compared to the previous geometric methods of molecular design. The number of permutations of first order groups can be decreased by introducing higher order groups before permuting the first order groups for each possible combination of them. Since the higher order groups are introduced at the design level itself and not in the screening level, we can be sure that the probability of missing a potential molecule is minimized compared to other methods. Since the possibility of non-existence of each higher order group is considered while generating the combinations, structural isomers can be generated to some extent.

### General Problem Statement

The problem is to identify molecules that meet a given process performance and environmental restrictions defined in terms of properties.

### Process Design Model

The process design problem aims at finding the admissible property ranges in its conservative surrogate form [ $\phi_{jf}^{\text{lower}}$ ,  $\phi_{jf}^{\text{upper}}$ ] and feed flowrate  $F_f$  that meets the given process performance without committing to any components.

Process performance may be given by the upper and lower limits of different properties. Consider a process sink which processes a mixture of sources  $i$ , such that  $\{i = 1, 2, 3, \dots, N_{\text{sources}}\}$ . Each source entering the sink has a flow rate  $F_i$  and property  $P_{ji}$ . For a sink with a permissible flowrate  $F$ , the performance can be defined by:

$$P_j^{\text{lower}} \leq P_j \leq P_j^{\text{upper}} \quad (5)$$

Since the properties may not be conserved, they are reformulated in terms of their conserved surrogates by the application of the property operators:

$$\phi_j^{\text{lower}}(P_j) \leq \phi_j(P_j) \leq \phi_j^{\text{upper}}(P_j) \quad (6)$$

where  $\phi_j$  is the property operator surrogate of property  $j$  that can be found by the linear sum of all streams and feed to the sink:

$$\phi_j(P_j) = \sum_i^{N_{\text{sources}}} x_i \phi_{ji}(P_{ji}) + x_f \phi_{jf}(P_{jf}) \quad (7)$$

Hence, the property operator value for the fresh feed stream is calculated as,

$$\phi_{jf}(P_{jf}) = \frac{1}{x_f} \left[ \phi_j(P_j) - \sum_i^{N_{\text{sources}}} x_i \phi_{ji}(P_{ji}) \right] \quad (8)$$

where,

$$x_i = \frac{F_i}{F} \quad (9)$$

and the feed flowrate to the sink is given by:

$$F_f = F - \sum_i^{N_{\text{sources}}} F_i \quad (10)$$

It is evident that, for each property, there will be two inequality expressions: one for the minimum value and the other for the maximum value for a given sink (Qin *et al.*, 2004). Solving the model to find the upper and lower property limits of the fresh feed into the sink would provide the input to the molecular design problem.

### Molecular Design Model

When designing molecules to serve a specific purpose, some constraints are needed. The methodology developed employs a generate and test technique where all combinations are generated at two levels: first order combinations and higher order group combinations; these are gradually cut down by efficiently applying the constraints in a definite order. This methodology basically prevents the problem from being addressed as an MINLP problem. The constraints include structural and some specific properties depending on the process requirements. The structural constraints that are to be imposed are:

1. The Free Bond Number of the final molecule is zero to ensure that the number of bonds attached to each group is equal to its valance and hence eliminate the possibility of any free bonds left in the molecule.

The mathematical expression for the Free Bond Number (FBN), which is the number of free bonds in each acyclic molecular string, is (Eljack *et al.*, 2007):

$$\text{FBN} = \sum_{g=1}^{N_g} n_g \text{FBN}_g - 2 \left( \sum_{g=1}^{N_g} n_g - 1 \right) - 2N_r \quad (11)$$

where  $\text{FBN}_g$  is the free bond number associated with group  $g$ , and  $N_r$  is the number of rings found in the structure.

2. The following expressions have been developed to ensure the existence of a meaningful molecule:

$$n_g \geq 0 \text{ (acyclic)} \quad n_g \geq 3 \text{ (ring)} \quad (12)$$

Specific constraints are imposed to ensure that the molecule will result in the desired process performance, i.e., physical property constraints, environmental constraints and the constraints from the process model need to be specified. Environmental and process constraints are tracked using molecular property operators. The output of the process model can be written as:

$$\phi_{jf}^{\text{lower}}(P_{jf}) \leq \phi_{jf}(P_{jf}) \leq \phi_{jf}^{\text{upper}}(P_{jf}) \quad (13)$$

Molecular property operators are expressed as a function of the above conserved property surrogates of non-conserved properties to obtain the input range of process constraints for molecular design problem:

$$\phi_j^M(P_{j,m}) = f(\phi_{jf}(P_{jf})) \quad (14)$$

Hence,

$$\phi_j^{\text{M,lower}}(P_{j,m}) \leq \phi_j(P_{j,m}) \leq \phi_j^{\text{M,upper}}(P_{j,m}) \quad (15)$$

Along with this input of process constraints, similar constraints for physical, environmental properties and preselected groups are combined with the structural constraints shown below:

$$N_f \leq N_{\text{max}}, n_g \geq 0 \text{ (acyclic)}, n_g \geq 3 \text{ (ring)} \quad (16)$$

The number of first order groups that could possibly be present in the designed candidate molecules is maximized. The reason behind maximizing the number of these groups is to ensure that no potential molecule is left out of the search space.

The properties of the groups participating in the above model are found using group contribution methods (Marrero and Gani, 2001), (Stefanis *et al.*, 2004).

$$\phi_j^M(P_{j,mf}) = \sum_{f=1}^{N_{fs}} n_f P_{j,f} \quad (17)$$

### Enumeration of Higher Order Groups

In order to increase the accuracy of property based molecular design techniques, the effects of higher order molecular groups need to be considered while designing molecules. Higher order groups give a better description of molecules and differentiation of structural isomers is possible using such groups. Different combinations of first order groups are generated and for each combination whose FBN is zero, the following methodology is used to find the maximum number of higher order groups in each. If cyclic molecules are included during the generation of combinations of first order groups, a maximum number of constituent groups must be imposed (since the efficiency of GCM is not great for high molecular weight molecules). The following rules should be considered for the generation of molecules:

#### Rule 1

Higher order groups can only be formed from complete molecular fragments. For instance, to form the higher order group CH (CH<sub>3</sub>) CH (CH<sub>3</sub>), there must be two CH and two (CH<sub>3</sub>) groups. It is not possible to consider a CH (CH<sub>3</sub>) group as half of a higher order group.

#### Rule 2

If any of the higher order groups completely overlap some other higher order group, only the larger group must be chosen in order to prevent redundant description of the same molecular fragment.

So, if (k : n) is the set of first order groups that are the building blocks of one higher order group, h, then (n<sub>k</sub> : n<sub>n</sub>) is the number of those first order groups present in the molecule, η is the number of occurrences of one particular first order group in a selected higher order group, and n<sub>h</sub> is the number of possible higher order groups from those first order groups, then:

$$n_h = \text{Min} \left( \frac{n_k}{\eta_k} : \frac{n_n}{\eta_n} \right) \quad (18)$$

n<sub>h</sub> must be rounded down to the nearest integer number according to Rule 1. Moreover, two higher order groups of the same kind can even share first order groups just like higher order groups of different kind. For instance, 2 OH and 1 CH group can form 2CHOH groups. Hence the possibility of sharing of available first order groups participating in the given higher order group is considered. To account for these additions to the count of higher order groups by sharing of groups, the methodology in Figure 2 is used if n<sub>h</sub> ≥ 0. Higher order groups can be found by the algebraic equations given by Eq. (18) but, due to above mentioned reasons, a step by step method has to be used to ensure that no higher order groups are wrongly omitted.

**Example:** Higher order group considered:

CH<sub>3</sub>-CH-CH-CH<sub>3</sub>

Number of CH<sub>3</sub> groups: 4

Number of CH groups: 3

Initially n<sub>h</sub> = Int [ Min (  $\frac{3}{2} : \frac{4}{2}$  ) ] = 1

Free bonds on CH<sub>3</sub>=0, CH=2,

CH=2, CH<sub>3</sub>=0

Remaining CH<sub>3</sub> groups = 2,

CH groups = 1

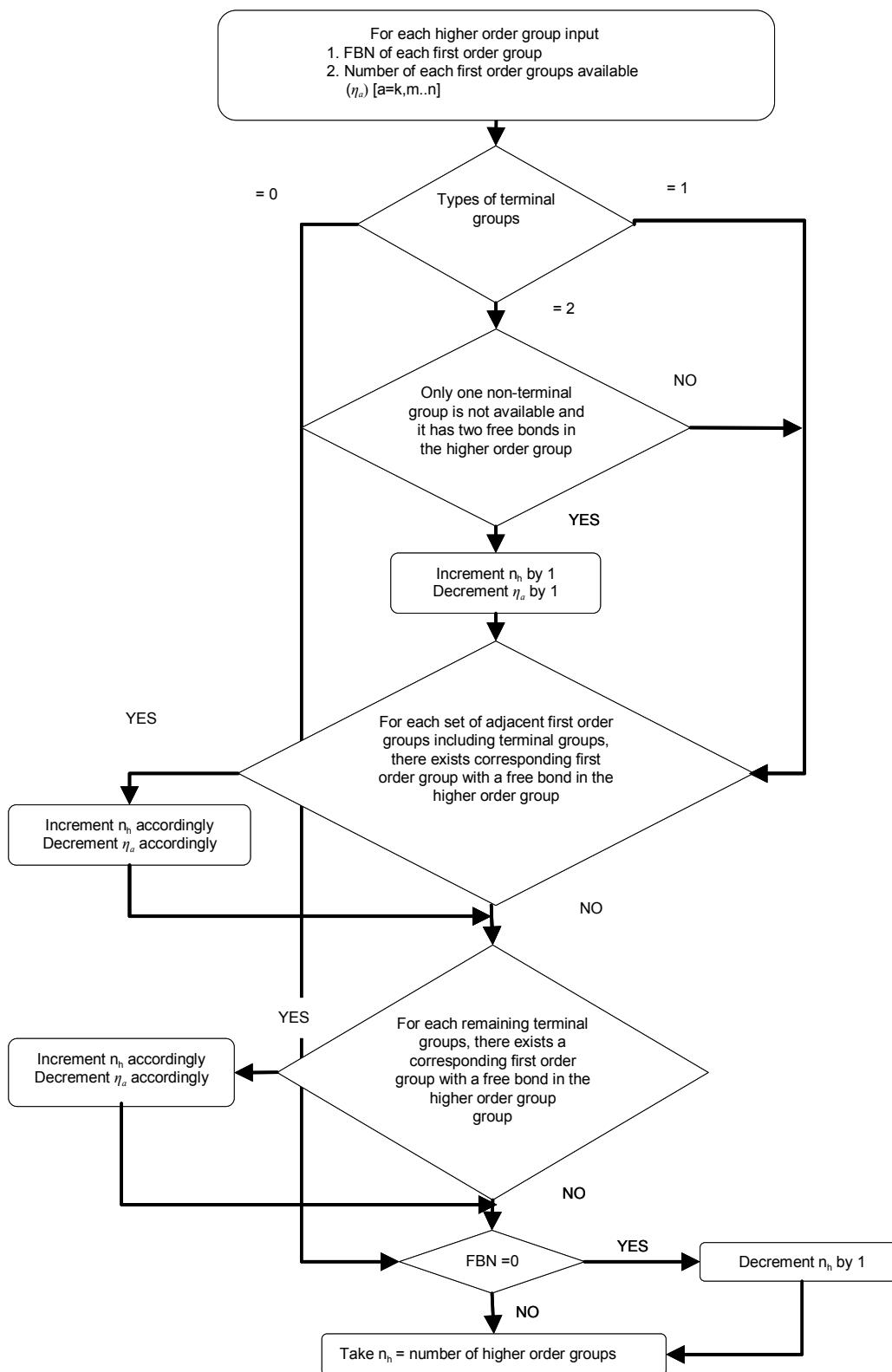
Addition of 1 CH and 1 CH<sub>3</sub> groups to a CH group with one of its free bonds leads to formation of one more CH<sub>3</sub>-CH-CH-CH<sub>3</sub> group. Addition of a CH<sub>3</sub> group to a CH group is not allowed since (CH<sub>3</sub>)<sub>2</sub>CH is another higher order group. Since no remaining first order groups lead to formation of higher order groups, the maximum number of higher order group CH<sub>3</sub>-CH-CH-CH<sub>3</sub> for a given first order combination with 4 CH<sub>3</sub> and 3 CH groups is taken to be 3. This is one kind of scenario that can be considered by using the methodology in Figure 2.

If  $\phi_j^M(P_{j,mh})$  is the property contribution from the higher order groups, it is calculated as:

$$\phi_j^M(P_{j,mh}) = \sum_{h=1}^{N_h} n_h P_{j,h} \quad (19)$$

The property for molecule *i* can now be estimated as (Chemangattuvalappil *et al.*, 2009):

$$\phi_j^M(P_{j,m}) = \phi_j^M(P_{j,mf}) + \phi_j^M(P_{j,mh}) \quad (20)$$



**Figure 2:** Procedure for enumerating higher order groups

Since all the combinations of all first order groups, whose maximum occurrence is taken as the limit, are considered, no potential molecule is left out. Introduction of the higher order groups prior to this stage in the molecular design model would lead to a non-linear model and does not increase the accuracy of the model significantly. So, finding the maximum of these higher order groups from different combinations of first order groups that obey the structural constraints would serve the purpose of finding potential molecules for a given process performance. Hence, all possible combinations of the numbers  $[0, n_g]$  and  $[0, n_h]$  are generated subject to the following constraint:

FBN = 0 with the inclusion of the possibility of the number of rings in the case of cyclic molecules.

The above method of generation enables the identification of structural isomers to some extent as the possibility of nonexistence of each higher order group is considered. The possible molecules are screened out by checking if the combination of all the groups satisfies all structural, property and process constraints. Rule 2 indicates that a higher order group is not completely overlapped by any other higher order group. Hence, after screening out the molecules based on the above constraints, an extra condition of whether any second order group is completely overlapped by another is checked before the final screening.

#### STEPWISE PROCEDURE FOR ALGEBRAICALLY SOLVING A MOLECULAR DESIGN PROBLEM

1. Transform the required property range of the process sink into sets of Maxima and Minima of their respective surrogates using the Process Property Operators.
2. Solve the reverse formulated process model to find the range of fresh feed properties into the sink and these may serve as the initial input to the molecular design problem.
3. Transform the required property range into sets of Maxima and Minima using the molecular property operators and their corresponding functions in GCM.
4. Select the first order groups to form the candidate molecules based on the nature of the final product.
5. Using the contributions of each molecular fragment, find the maximum number of each first order group by solving the molecular design model.

6. Enumerate all possible combinations of first order groups whose FBN is zero.

7. Evaluate all the possible numbers of higher order groups and generate all the combinations of first and higher order groups using the methodology represented by the flowsheet in Figure 2.

8. Screen out the molecules based on the structural and specific constraints while also considering Rule 2 of the higher order enumeration technique.

#### CASE STUDY

A gas treatment process uses fresh methyl diethanol amine, MDEA,  $(\text{HO}-(\text{CH}_2)_2-(\text{CH}_3)\text{N}-(\text{CH}_2)_2-\text{OH})$  and two other recycled process sources ( $S_1$ , and  $S_2$ ) as a feed into the acid gas removal unit. 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 the sink (Kazantzi et al., 2007). Design objectives and requirements are to find 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  $S_3$ ). The following three properties are considered: critical volume ( $V_c$ ), heat of vaporization ( $H_v$ ) and heat of fusion ( $H_{\text{fus}}$ ). Additionally, two thermal constraints are imposed on the synthesized molecules to make sure that the designed molecule will remain in the liquid state under the process conditions and to prevent excessive solvent losses via evaporation. Also, the boiling point constraint ensures that the solvent's flammability is checked, as the parameter used in the context of flammability, the flash point, directly depends on boiling point (Affens, 1966). An additional environmental constraint is imposed to check the toxicity level of solvents used.  $\text{LC}_{50}$  is the parameter that measures the limiting concentration of material to which test organisms can be exposed. The higher the  $\text{LC}_{50}$  value, the less toxic the substance is. The octanol/water partition coefficient ( $K_{\text{ow}}$ ) is defined as the ratio of a chemical's concentration in the octanol phase to its concentration in the aqueous phase of a two-phase octanol/water system. It represents the tendency of the chemical to partition between an organic phase (e.g., a fish, soil) and an aqueous phase. The  $\text{LC}_{50}$  value is related to the  $K_{\text{ow}}$  value (Konemann, 1981) and, hence, a constraint on the  $\log K_{\text{ow}}$  value places a check on the toxicity. A constraint of  $\log K_{\text{ow}} < 4$  is imposed based on studies of Reducing the Environmental impact of Acid gas Control Technologies.

The process property targets and flow rate data for all streams ( $S_1$ ,  $S_2$  and  $S_3$ ) and the sink are summarized in Table 1.

**Table 1: Property data for gas purification**

$P_i$	Property Bounds on Sink	$S_1$	$S_2$	$S_3$
$V_c$	529-610	754	730	790
$H_v$	100-115	113	125	70
$H_{fus}$	20-24	15	15	20
Flowrate	300	50	70	30

These selected process properties combine linearly and hence the property operator of each is the property itself. The Process Property Operators of the above properties are given by:

$$V_c = x_f V_{c,f} + x_{S_1} V_{c,S_1} + x_{S_2} V_{c,S_2} + x_{S_3} V_{c,S_3} \quad (21)$$

$$H_v = x_f H_{v,f} + x_{S_1} H_{v,S_1} + x_{S_2} H_{v,S_2} + x_{S_3} H_{v,S_3} \quad (22)$$

$$H_{fus} = x_f H_{fus,f} + x_{S_1} H_{fus,S_1} + x_{S_2} H_{fus,S_2} + x_{S_3} H_{fus,S_3} \quad (23)$$

$$x_i = \frac{F_i}{F}, \quad i \in (f, S_1, S_2, S_3) \quad (24)$$

Given the values above, the property range of critical volume ( $V_c$ ), heat of vaporization ( $H_v$ ) and heat of fusion ( $H_{fus}$ ) are found. These are utilized as the process property constraints in the molecular model. The results of the process model are given in Table 2.

**Table 2: Property surrogates of Feed**

$P_i$	Lower Limit	Upper Limit
$V_c$	307.98	469.98
$H_v$	90.008	120.008
$H_{fus}$	24.00	32.00

The property operators for the molecular target properties and the functions that link them to the process property operators are given below:

$$T_b^M = \exp\left(\frac{T_b}{t_{b0}}\right) = \sum_{g=1}^{N_g} n_g T_{bg} \quad , \quad t_{b0} = 222.543 \quad (25)$$

$$T_m^M = \exp\left(\frac{T_m}{t_{m0}}\right) = \sum_{g=1}^{N_g} n_g T_{mg} \quad , \quad t_{m0} = 147.45 \quad (26)$$

$$V_c^M = \sum_{g=1}^{N_g} n_g V_{cg} \quad , \quad V_c^M = V_c - 7.95 \quad (27)$$

$$H_v^M = \sum_{g=1}^{N_g} n_g H_{vg} \quad , \quad H_v^M = H_v - 11.733 \quad (28)$$

$$H_{fus}^M = \sum_{g=1}^{N_g} n_g H_{fusg} \quad , \quad H_{fus}^M = H_{fus} + 2.806 \quad (29)$$

$$\log K_{ow} - K_{ow0} = \sum_{g=1}^{N_g} n_g \log K_{owg} \quad , \quad (30)$$

$$K_{ow0} = 0.097$$

The property operator values on the process side of the problem are calculated using the equations on the right side of every equation above and these values are tailored using the equations to the left to be used by the molecular design model.

Additional physical and environmental property constraints are imposed on the feed. Table 4 gives the property target inputs to the molecular design model.

**Table 4: Molecular Property Operator values**

$P_i$	Lower Limit	Upper Limit
$T_m^M$		9.375
$T_b^M$	13.16	
$V_c^M$	300.03	462.03
$H_v^M$	78.275	108.275
$H_{fus}^M$	26.806	34.806
$\log K_{ow}$		4

Now, the following first order groups have been considered for molecular design:

$CH_3$ ,  $CH_2$ ,  $CH$ ,  $CH_3O$ ,  $CH_2O$ ,  $CHO$ ,  $CH_2NH_2$ ,

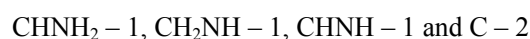
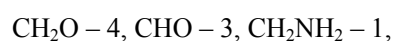
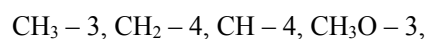
$CHNH_2$ ,  $CH_2NH$ ,  $CHNH$  and  $C$

Property data for these selected groups are taken from the Group Contribution methods by Marrero and Gani (2001) and Stefanis *et al.* (2004). Now, inequality expressions for each property are formulated. The number of preselected first order molecular groups is maximized subject to the specific constraints mentioned in Table 4. The reason behind



maximizing these groups is to ensure that no potential molecule is left out. The variations in the property values caused by inclusion of the higher order groups will be considered in the later stages. To insure water solubility and to reduce vapor pressure, the molecule must have two or more –OH groups. To limit the extent of corrosion, only one amino group is allowed to be in the amine (N in the amino group either connects to H or C). Finally, to limit detrimental effects of direct exposure to the solvent, tertiary amines are ruled out in this case study.

The maximum values are as follows:



The possible higher order groups from the above first order groups are found and the maximum number of each higher order groups using the methodology explained by Figure 2 of each combination of first order groups whose FBN is zero is found. Their contribution to the molecular property operator values are found using Eq. (19). All possible combinations of both first and higher order groups are generated.

The above method of generation enables the identification of isomers as the possibility of nonexistence of each higher order groups is considered. The possible molecules are screened out by checking if the combination of all the groups satisfies all structural, property and process constraints. After screening out the molecules based on the above constraints, an extra condition of whether any second order group is completely overlapped by another is checked before the final screening. The final list of candidate molecular structures that should be investigated further is given in Table 5 below.

**Table 5: Candidate molecular structures**

S. No	Molecule
1	OH-CH(OCH <sub>3</sub> )-CH(CH <sub>2</sub> NH <sub>2</sub> )-OH
2	(OH) <sub>2</sub> -CH-CH(OCH <sub>3</sub> )-CH <sub>2</sub> NH <sub>2</sub>
3	(OH) <sub>2</sub> (CH <sub>2</sub> NH <sub>2</sub> )C-CH <sub>2</sub> -OCH <sub>3</sub>
4	OH-CH <sub>2</sub> -O-CH <sub>2</sub> -CH(OH)-CH <sub>2</sub> NH <sub>2</sub>
5	(OH) <sub>2</sub> CH-CH <sub>2</sub> -OCH <sub>2</sub> -CH <sub>2</sub> NH <sub>2</sub>
6	OH-NH-CH <sub>2</sub> -CH(CH(OCH <sub>3</sub> ))(CH <sub>3</sub> )-CH <sub>2</sub> -OH
7	OH-NH-CH <sub>2</sub> -CH(OH)-CH(CH <sub>3</sub> )-CH <sub>2</sub> -O-CH <sub>3</sub>
8	OH-CH(CH <sub>2</sub> NH <sub>2</sub> )-CH <sub>2</sub> -CH <sub>2</sub> -O-CH(OH)-CH <sub>3</sub>
9	(OH) <sub>2</sub> (CH <sub>3</sub> )C-CH <sub>2</sub> NH-CH <sub>2</sub> -OCH <sub>2</sub> -CH <sub>3</sub>

The generated molecular structures have properties consistent with the property limits of the sink; however hemiacetals, carbinolamines and germinal diols are often unstable and as such would need to be further tested for their applicability.

The objective of this case study was to illustrate the generation of molecules that satisfy certain property ranges and since the second step of the simultaneous design considers the elimination of each possible higher order group, success in detection of isomers is reached to some extent. The same methodology can be used to generate cyclic molecules by including the constraint on the number of rings in each combination of groups.

## CONCLUSIONS

In this work, an algebraic method has been developed that solves process and molecular design problems simultaneously to find molecules that would meet the needs of the process. Higher order group contribution has been included in the molecular design model. Enhanced enumeration techniques of higher order groups increase the application range and reliability of the molecular design step. The incorporation of these techniques into the solution method at the design level ensures that no potential molecule is omitted. The solution method is efficient, but the accuracy of the property prediction will depend only on the accuracy of the property models used. Here, an algebraic approach has been followed which decreases the combinatorial size of the problem significantly and also ensures that any number of properties can be tracked.

## NOMENCLATURE

FBN	Free bond number
FBN <sub>g</sub>	Free bond number associated with group g
F	Allowable flow rate into the sink
F <sub>i</sub>	Flow rate into source i
GCM	Group Contribution Method
H <sub>fus</sub>	heat of fusion
h <sub>fus0</sub>	Adjustment parameter used in the estimation of heat of fusion
H <sub>v</sub>	Standard heat of vaporization at 298 K
h <sub>v0</sub>	Adjustment parameter used in the estimation of heat of vaporization
K <sub>ow</sub>	Octanol/water partition coefficient
N <sub>f</sub>	Total number of first order molecular groups

$N_h$	Total number of higher order molecular groups
$N_r$	Total number of rings in a molecule
$N_{sources}$	Total number of sources into the sink
$n_f$	Number of functional groups in the main chain
$n_g$	Molecular group
$P_{j,g}$	Contribution of property j from group g
$P_{j,m}$	Property j of molecule m
$P_{ji}$	Property j of stream i
$T_b$	Boiling point
$t_{b0}$	Adjustment parameter used in the estimation of boiling temperature
$T_m$	Melting point
$t_{m0}$	Adjustment parameter used in the estimation of melting temperature
$x_i$	Fractional contribution
$V_c$	Critical volume
$v_{c0}$	Adjustment parameter used in the estimation of critical volume
$k, m \dots n$	first order groups

### Greek Symbols

$\phi_j^M$	Molecular property operator of the $j^{\text{th}}$ property
$\phi_j$	Process property operator of the $j^{\text{th}}$ property
$\eta_k, \eta_m \dots \eta_n$	Number of occurrences of first order groups

### REFERENCES

- Chemmangattuvalappil, N. G., Solvason. C. C., Eljack. F. T., Eden. M. R., A Novel Algorithm for Molecular Synthesis Using Enhanced Property Operators. *Computers & Chemical Engineering*, 33, (3), 636-643 (2009).
- Eden, M. R., Jørgensen, S. B., Gani, R., El-Halwagi, M. M., Reverse Problem Formulation based Techniques for Process and Product Design. *Computer Aided Chemical Engineering* 15A, 451-456 (2003a).
- Eden, M. R., Jørgensen, S. B., Gani, R., El-Halwagi, M.M., Property Cluster Based Visual Technique for Synthesis and Design of Formulations. *Computer Aided Chemical Engineering* 15B, 1175-1180 (2003b).
- Eljack, F. T., Eden, M. R., Kazantzi, V., El-Halwagi, M. M., Simultaneous Process and Molecular Design - A Property Based Approach. *AIChE Journal*, 53, (5), 1232-1239 (2007).
- Harper, P. M., Gani, R., Kolar, P., Ishikawa, T., Computer-aided Molecular Design with Combined Molecular Modelling and Group Contribution. *Fluid Phase Equilibria*, 158-160, 337-347 (1999).
- Harper, P. M., Gani, R., A Multi-step and Multi-level approach for Computer aided Molecular Design. *Computers & Chemical Engineering*, 24 (2-7), 677-683 (2000).
- Kazantzi, V., Qin, X., El-Halwagi, M. M., Eljack, F. T., Eden, M. R., Simultaneous Process and Molecular Design through Property Clustering Techniques. *Industrial & Engineering Chemistry* 46(10), 3400-3409 (2007).
- Marcoulaki, E. C., Kokossis, A. C., Molecular design synthesis using stochastic optimization as a tool for scoping and screening. *Computers & Chemical Engineering*, 22, (1), S11-S18 (1998).
- Marrero, J., Gani, R., Group Contribution Based Estimation of Pure Component Properties. *Fluid Phase Equilibria*, 183-184, 183-208 (2001).
- Qin, X., Gabriel, F., Harell, D., El-Halwagi, M. M., Algebraic Techniques for Property Integration via Componentless Design. *Ind. Eng. Chem. Res.*, 43, 3792-3798 (2004).
- Shelley, M. D., El-Halwagi, M. M., Component-less Design of Recovery and Allocation systems: A Functionality-based Clustering Approach. *Computers & Chemical Engineering*, 24, 2081-2091 (2000).
- Affens, W. A., Flammability properties by hydrocarbons fuels. *J. Chem Eng. Data*, 11, 197-202 (1966).
- Konemann, H., Quantitative Structure-Activity relationships in fish toxic studies. 1. Relationship for 50 industrial chemicals. *Toxicology* 19, 209-221 (1981).
- Stefanis, E., Constantinou, L., Panayiotou, C., A Group-Contribution Method for predicting pure component properties of Biochemical and safety interest. *Ind. Eng. Chem Res.*, 43, 6253-6261 (2004).