Automatic Synthesis of Thermodynamically Feasible Reaction Clusters

Konstantinos Holiastos and Vasilios Manousiouthakis
Chemical Engineering Dept., University of California, Los Angeles, CA 90095

Reaction clusters serve to carry out reactions whose thermodynamic yield is too low for practical large-scale application. They therefore play an important role in the economics of the process and can decisively control its environmental impact. Reaction cluster synthesis must address such economic and environmental considerations. This was successfully accomplished in this work, which is the first employing an optimization technique for this task. The problem formulation ensures that all mass balances are satisfied, that all reactions are thermodynamically feasible, and that they follow certain user-defined characteristics of a generalized chemical reaction. In addition, given a set of chemical species, all clusters are identified that can represent this reaction; the ones found first are the most promising candidates. The formulation is flexible, accommodating a different number of reactions per cluster and allowing specification of many attributes by the user.

Introduction

Chemical reactions are an essential step in the production of most chemicals. Although there are often numerous reactions that could be employed to form a desired compound, only some of them are physically realizable. Indeed, for its practical implementation in an industrial setting, a reaction must fulfill a number of criteria, such as process safety, environmental protection, and economic viability.

Safety considerations favor nontoxic, nonflammable reactants and byproducts. Environmental considerations suggest reducing byproduct formation and avoiding the use of hazardous compounds. Finally, economic viability is burdened by the costs of reactants, byproduct separation and disposal, and overall process operations. Any desired or projected profits necessitate significant product generation, which in turn specifies an acceptable reaction yield.

Central to the determination of reaction yield are thermodynamics, since they provide an upper bound on a reaction's attainable yield. If a process has a yield, as specified by thermodynamics, that is greater than or equal to the yield ensuring economic feasibility, then the corresponding reaction is called *thermodynamically feasible*. Next, the relationship between the thermodynamic yield and the reaction free energy is outlined.

Correspondence concerning this article should be addressed to V. Manousiouthakis.

A reaction proceeds to form products when its free energy (ΔG) is less than zero, while equilibrium is reached when $\Delta G = 0$. At that point, the final composition of products and reactants (and by extension the yield) can be obtained through the equilibrium constant $K(T)$, which is a function of $\Delta G^0(T)$, the free energy of the reaction under standard conditions at temperature T. Often the yield is smaller than that allowed for profitable operation of the given reaction. In essence, thermodynamics prevent its successful implementation. However, given the relation between K and ΔG^0, as well as the dependence of each on T, a methodology can be developed to overcome the thermodynamic feasibility barrier for any reaction under consideration.

In this methodology an attempt is first made to improve thermodynamic yield by varying process conditions. If this is unsuccessful, the cluster synthesis technique presented here can be employed to produce a thermodynamically feasible reaction cluster. Both steps just presented and any associated terminology are discussed next.

The free energy is a function of process conditions. Most commonly, temperature is the variable with which it is manipulated, and this will be the case in our work. The aim is to find a temperature at which the reaction becomes thermodynamically feasible. Often, however, the required temperature is too high. The process will then be too expensive to maintain. Also, the participating species are subject to decomposi-

tion. Thus, whenever the required temperature is outside some previously defined range of allowable operating temperatures, the chosen reaction cannot be used.

If there is no operational temperature at which the reaction can be carried out, it may be implemented through a reaction cluster, that is, a set of reactions (constitutive reactions) whose sum is the overall reaction. To ensure substantial yield on the overall reaction, each reaction step within the cluster must be thermodynamically feasible. Such a cluster is called a *thermodynamically feasible reaction cluster.*

As an illustrative example, consider an overall reaction that is infeasible within a temperature range:

$$A+B \rightarrow C+D. \tag{1}$$

Consider also the two-reaction cluster

$$A+E \rightarrow F+D \tag{2}$$

$$B+F \rightarrow E+C, \tag{3}$$

which yields the overall reaction. If these reactions (Eqs. 2 and 3) are thermodynamically feasible for part of the operational temperature range, then this cluster is a thermodynamically feasible reaction cluster.

This work deals with the *complete* and *systematic* generation of such thermodynamically feasible reaction clusters for thermodynamically infeasible reactions Holiastos and Manousiouthakis (1996). The synthesis task can be stated as: Given an overall reaction that is thermodynamically infeasible within a temperature range, according to a thermodynamic feasibility criterion, a set of chemical species, and their free energies as a function of temperature; determine, within the same T range, *all* thermodynamically feasible reaction clusters that achieve this overall reaction, and rank them based on a chosen performance objective.

In the present work this synthesis task is for the first time formulated as an optimization problem in which a variety of performance objectives can be considered, such as the total profit (cost) of chemicals produced (consumed), and safety or environmental considerations.

Reaction cluster synthesis was first studied by May and Rudd (1976). In their development, a compound or linear combination of compounds was chosen to represent a common difference. A number of reactions were then selected that contain that common difference. May and Rudd named these reactions half reactions. (In this work they are called primary reactions because here a half reaction is defined in a different manner.) To create a cluster reaction, two of the primary reactions are subtracted such that the common difference cancels. Because any reaction is considered infeasible if its ΔG is larger than 0, two primary reactions can only be matched to give a cluster reaction if their free energies differ by less than 0. Continuing this process, always using one primary reaction in generating two consecutive cluster reactions, so that its species cancel, a reaction cluster is formed. An easy way to visualize and automate the procedure is to create plots of the free energy of all primary reactions with temperature. The allowable matches are then easily identified in this graphical approach.

Stephanopoulos and coworkers (Rotstein et al., 1982) were the first to develop an algebraic method for carrying out this task, and extended its application to open reaction networks, that is, groups of reactions that lead to the products given the reactants but do not sum to a particular overall reaction. To aid the generation of such networks several properties were developed, useful in screening alternative paths.

The use of groups as building blocks for molecular species (Gani et al., 1991) has also been applied in reaction pathway synthesis. Fornari and Stephanopoulos (1994a,b) investigated the use of groups and considered the additional constraints of gross added value (worth of all products minus that of all reactants), role specification (the function of some species in the reaction network), and demand and supply in order to aid screening of possible solutions.

Crabtree and El-Halwagi (1995) studied the synthesis of environmentally acceptable reactions. Their problem formulation is a mixed-integer nonlinear program with stoichiometric, thermodynamic, and envionmental constraints. The latter place an upper bound on the concentrations of all species in the process.

An algorithm for the synthesis of biochemical pathways was presented by Mavrovouniotis et al. (1992). Subsequently (Mavrovouniotis, 1995) employed duality theory to study whether such pathways are feasible. If they are not, the point of infeasibility within the pathway was identified.

Seressiotis and Bailey (1988) developed a software system, called MPS, that utilizes an enzyme and a substance database to obtain and analyze possible biochemical reaction pathways with certain characteristics.

Valdèz-Perèz presented (1992, 1993, 1994) an algorithm to generate open reaction pathways that produce experimentally observed compounds given initial ones. The algorithm is able to conjecture species by specifying their chemical formula.

Many investigators focus their research on determination of reaction mechanisms, which are essentially reaction clusters whose constitutive equations are elementary steps. Happel and Sellers (1989) and Happel et al. (1990) investigated a system defined by species whose role as reactants or products of the overall reaction is specified, and by conjectured elementary reactions that can potentially participate in a mechanism. Then a direct mechanism is defined as one that does not contain a subset of elementary reactions that is a mechanism for the overall reaction. Similarly, a direct overall reaction is defined as one, composed of system species, that does not reduce to the sum of two distinct subreactions, also composed of system species. Then a procedure was presented whereby all possible direct reactions and their corresponding direct mechanisms are obtained.

Mavrovouniotis and Stephanopoulos (1992) and Mavrovouniotis (1992) in a development related to that of Happel and Sellers take into account the directionality of intermediate steps.

Many of these approaches to obtain reaction mechanisms can be used—with little or no modification—for the generation of reaction clusters. Alternatively, the formulation presented here could be employed to postulate reaction mechanisms just by adjusting several constraint parameters.

In a reaction cluster any chemical species may potentially appear. For example, the Solvay cluster has the overall reaction

$$2NaCl+CaCO_3 \rightarrow Na_2CO_3+CaCl_2, \tag{4}$$

but contains ammonia and other nitrogeneous species. In contrast, for a reaction mechanism the list of potential participating species includes radicals, stable intermediates, and catalyst complexes. These must contain atoms found in the overall reaction and the known catalyst. Furthermore, stable intermediate species, radicals, and catalyst complexes can be conjectured based on experience and experimental evidence, and may be collected to form the database. The database required for the determination of a reaction mechanism is therefore considerably smaller than the one required for a reaction cluster, leading to great efficiency when the problem formulation presented here is used to generate reaction mechanisms.

Thermodynamic Feasibility

Thermodynamic feasibility for reactions and clusters will now be precisely defined, once some necessary background information is presented.

To examine the relation between ΔG and the direction in which the reaction proceeds, consider the total differential of the internal energy U and the free energy G (Denbigh, 1981; Pitzer, 1961):

$$dU = TdS - PdV + \sum_i \mu_i dn_i \tag{5}$$

$$dG = -SdT + VdP + \sum_i \mu_i dn_i, \tag{6}$$

where μ_i is the chemical potential of species i and dn_i is the amount of i transferred to or from the system.

Now consider a closed, multicomponent system in which a chemical reaction occurs (Reid, 1960). If the reaction is represented as

$$\sum_i^m \gamma_i \Gamma_i \rightarrow \sum_i^n \delta_i \Delta_i, \tag{7}$$

an infinitesimal spontaneous small advancement $d\xi > 0$ for the reaction will cause a change of $dn_i = -\gamma_i d\xi$ for the reactants and $dn_i = \delta_i d\xi$ for the products. Substituting in Eq. 5,

$$dU = TdS - PdV - \left(\sum_i \mu_{\gamma_i}\gamma_i - \sum_i \mu_{\delta_i}\delta_i\right)d\xi \Rightarrow \tag{8}$$

$$-\sum_i \mu_i dn_i = \left(\sum_i \mu_{\gamma_i}\gamma_i - \sum_i \mu_{\delta_i}\delta_i\right)d\xi = TdS - dU - PdV =$$

$$TdS - dQ \geq 0, \tag{9}$$

where the first and second laws of thermodynamics were used. Substituting into Eq. 6, and considering T, P to be constant,

$$dG = -\left(\sum_i \mu_{\gamma_i}\gamma_i - \sum_i \mu_{\delta_i}\delta_i\right)d\xi. \tag{10}$$

The quantity $dG < 0$ remains negative, as the reaction occurs spontaneously at constant temperature and pressure, and consequently $\Delta G < 0$ for any such reaction. When $\Delta G = 0$, according to Eq. 10 the reaction is at equilibrium.

The quantity $\Delta G^0(T)$ is the standard free energy of the reaction, defined as the free energy the reaction would have at the particular temperature T if all products and all reactants had unit activities. It is possible to relate the system's equilibrium position to $\Delta G^0(T)$. At equilibrium (Denbigh, 1981)

$$\Delta G^0(T) = -RT \ln K(T), \tag{11}$$

where $K(T)$ is the equilibrium constant, and is only a function of temperature. Note that when $\Delta G^0(T)$ is negative, $K(T)$ is greater than one. This is interpreted as an equilibrium favoring the products.

The standard state only fixes activities, not temperatures. At constant temperature, Eq. 11 correctly predicts an increase in $K(T)$ with decreasing $\Delta G^0(T)$. However, with a change in temperature the dependence of $K(T)$ on $\Delta G^0(T)$ is more complicated because now it is $\Delta G^0(T)/T$ that determines the change of $K(T)$ with temperature, not just $\Delta G^0(T)$.

One can imagine a particular reaction where $\Delta G^0(T)$ decreases with increasing T. If it so happens that a relatively large change in temperature is needed for a small change in free energy, then the fraction $\Delta G^0(T)/T$ will increase (move closer to zero) even though the numerator decreases. According to Eq. 11, this translates to a decrease of the equilibrium constant.

Under all circumstances where this effect may exist, the equilibrium constant will still be greater than one as long as $\Delta G^0(T) < 0$. This is because $\ln K(T)$ must be positive. The products will always be favored to some extent. It is desirable that the definition of thermodynamic feasibility be flexible enough to specify the minimum value of $K(T)$ that every thermodynamically feasible reaction must have. Let us require that $K(T)$ must be greater than some value e^z with $z > 0$, so that $K(T)$ is always more than or equal to one. We then have

$$K(T) \equiv e^{-\Delta G^0(T)/RT} \geq e^z \Leftrightarrow -\frac{\Delta G^0(T)}{RT} - z \geq 0 \Leftrightarrow \Delta G^0(T) + zRT \leq 0. \tag{12}$$

We may now precisely define a thermodynamically feasible reaction as one satisfying the relation

$$\Delta G^0(T) + zRT \leq 0 \tag{13}$$

for a given $z > 0$.

A thermodynamically feasible reaction cluster is now more precisely defined as one whose constitutive reactions are thermodynamically feasible. Such a definition ensures that the cluster as a whole can effectively carry out the overall reaction because each of its steps has a high thermodynamic yield.

The essence of a thermodynamically feasible cluster is well captured by a diagram such as the one in Figure 1 depicting the two-reaction cluster given in Eqs. 1–3.

The ordinate is the thermodynamic feasibility criterion, while the abscissa is temperature. The overall reaction is above the $\Delta G^0 + zRT = 0$ line for the whole temperature range from T_L to T_U, signifying that it is thermodynamically infeasible. However, the plots for the two constitutive reactions are thermodynamically feasible for at least part of the

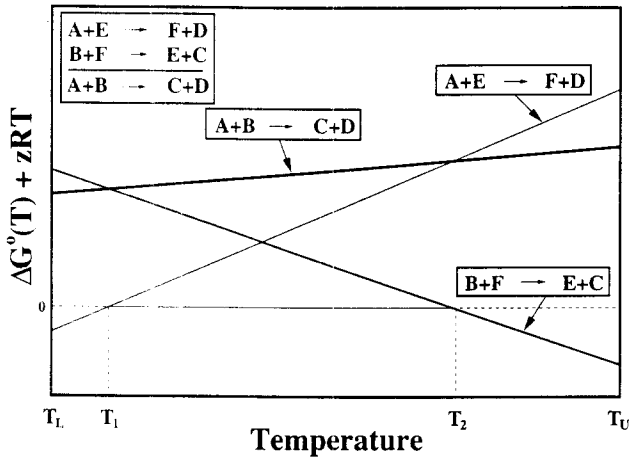

Figure 1. Pictorial representation of a thermodynamically feasible two reaction cluster.

temperature range. This cluster therefore presents a way to effectively carry out the overall infeasible reaction. If the reaction in Eq. 2 is run between the temperatures T_L and T_1, and the reaction in Eq. 3 between T_2 and T_U, the overall reaction (Eq. 1) is effectively carried out, and the thermodynamic barrier is overcome.

Allowable Chemical Reactions

Chemical reactions, as they occur in nature and as they are written in a chemical equation, possess specific charcteristics. A generated reaction that exhibits these is called an *allowable reaction*, and its precise definition is necessary for the development of an effective formulation of the cluster synthesis problem.

As a motivation for the definition of an allowable reaction let us investigate the manner in which it takes place. It turns out that the path followed is rarely evident from the reaction stoichiometry. Most often there is a reaction mechanism that describes how products are formed from reactants. A reaction mechanism is composed of elementary reactions (elementary steps). An elementary reaction is one that depicts a chemical transformation as it truly happens at the atomic scale. An elementary reaction with one, two, or three reactants is called unimolecular, bimolecular, and termolecular, respectively.

The formation or breakage of bonds that accompanies an elementary chemical reaction is, in quantum mechanical terms, a change in the orbital configuration of the reactant(s) that results in the product(s). For bimolecular and higher-order reactions it follows that such changes can only take place if the orbitals of the two or more reacting species come sufficiently close to one another. More importantly, they must approach each other at the correct orientation. Solely on statistical grounds, these two requirements are more easily satisfied for a bimolecular reaction than for a termolecular one. Three molecules are less likely to collide than two; additionally, it is less likely that the three molecules will collide at the correct orientation. Bimolecular elementary reactions are therefore more common than termolecular ones. Fourth-or higher-order elementary reactions are not known.

It is now possible to enumerate the general characteristics of an elementary reaction. It has integer coefficients, and may have up to three reactants. Two or more of these can be the same type of molecule. In general there will be up to three products. That will definitely be the case if the elementary reactions can be shown to be reversible.

In the spirit of what constitutes an elementary reaction, we now define the characteristics of an allowable chemical reaction. It has integer coefficients, with a specified upper bound, and a limited number of species per half reaction. These requirements lead to several constraints, called variable constraints, because they depend on the definition of an acceptable reaction.

The variable constraints allow significant control over the form of the resulting chemical equations. They dictate the following features on every cluster reaction:

- It may have no more than s species per half reaction.
- Its coefficients can only attain specific values.
- The sum of the coefficients in each half reaction must range between 0 and δ_j^μ for the products and between 0 and γ_j^μ for the reactants, where j is the reaction number within the cluster.
- It is thermodynamically feasible.

The first requirement reflects the observation that a reaction containing an excessive number of species does not accurately represent a real process. Normally, the number of species present in a half reaction ranges from one to five. The second constraint only allows specific values for the coefficients. These are integers from zero to some maximum value. The third constraint is another tool in shaping the reaction to its final desirable form. It forces—if necessary—some of the coefficients not to be at their upper bounds, without imposing limits on any particular species.

Mathematical Description

The problem formulation will generate thermodynamically feasible reaction clusters that are composed of allowable reactions, as they were defined in the previous section. Let there be N^R reactions in the cluster, each denoted by the index j and taking place at temperature T_j. Also let δ_{ij} and γ_{ij} denote the coefficient of species i in the product and reactant side of the jth reaction, respectively.

Most constraints pertain to allowable reactions. These may therefore be used as a unit in any problem where generation of chemical reactions is desired, to ensure acceptable results.

Elemental mass balances

Let there be N^S species in the database, and N^A types of atoms among them. For each reaction there are as many atomic balance constraints as there are elements. The atom information (the number and kind of each atom in each species) is stored in a matrix B whose rows represent elements and whose columns represent species. Element B_{ki} is then the number of atoms of type k in species i.

The mass-balance constraints are

$$\sum_{i=1}^{N_S} B_{ki}(\delta_{ij} - \gamma_{ij}) = 0 \quad k \in \{1,2,\ldots,N^A\} \quad j \in \{1,2,\ldots,N^R\}. \tag{14}$$

Species constraints

The single type of constraint pertaining to reaction clusters as a whole and not to reactions alone is the species constraint, which forces the sum of the constitutive reactions to be the overall one. The net appearance of a species is equal to its net appearance in the overall reaction. The form of a species constraint is

$$\sum_{j=1}^{N_R} (\delta_{ij} - \gamma_{ij}) = \nu_i \quad i \in \{1,2,\ldots,N^S\}, \tag{15}$$

where ν_i is the coefficient of species i in the overall reaction, positive if a product, negative if a reactant, and zero if it does not appear. There are as many such constraints as items in the database. In a sense they are balances over the species. It is therefore necessary to use a balanced overall reaction if all of the constitutive reactions are to be balanced.

Thermodynamic feasibiity constraints

To calculate the free energy of a reaction, the free energies of formation (ΔG_f) of its participating species are employed.

It is assumed that ΔG_f (and consequently ΔG) varies with temperature in a linear manner. The assumption of linear dependence of ΔG_f on T is good, and its theoretical basis lies in the compensation effect (Benson, 1976). In any reaction a change in enthalpy (H) is usually accompanied by an associated change in entropy (S). The compensation effect describes a quantitative relationship that is often observed between these two quantities. The total energy of the reactants is utilized in part to keep order and structure among atoms of a molecule as well as between molecules. If a reaction is exothermic ($\Delta H < 0$), energy is lost in the form of heat, and thus less energy is available to maintain the order of the products at the same level as that of the reactants. This implies that $\Delta S < 0$. In a similar manner an increase in enthalpy will likely result in a corresponding increase in entropy. Looking at the fundamental thermodynamic equation relating ΔG, ΔH, and ΔS,

$$\Delta G = \Delta H - T\Delta S, \tag{16}$$

the effects on ΔG due to changes in enthalpy and entropy cancel to a large extent, resulting in a variation of the free energy much smaller than that of either enthalpy or entropy. It is because of this in-tandem variability that a linear expression is adequate to represent the free-energy–temperature relationship.

Denoting as $\Delta G_{f,i}(T)$ the free energy of formation as a function of temperature for species i, use of the compensation effect yields

$$\Delta G_{f,i}(T) = a_i T + b_i, \tag{17}$$

where a_i, b_i are constants particular to species i, and T is any temperature within the range for which a_i, b_i are determined. For any cluster reaction under consideration it follows

$$\Delta G_j^0(T_j) = \sum_{i=1}^{N^S} \delta_{ij} \Delta G_{f,i}(T_j) - \sum_{i=1}^{N^S} \gamma_{ij} \Delta G_{f,i}(T_j)$$

$$j \in \{1,2,\ldots,N^R\}, \tag{18}$$

where T_j is the temperature at which the reaction is carried out. Substituting for $\Delta G_{f,i}(T_j)$ from Eq. 17

$$\Delta G_j^0(T_j) = \sum_{i=1}^{N^S} (a_i T_j + b_i)(\delta_{ij} - \gamma_{ij}) \quad j \in \{1,2,\ldots,N^R\}. \tag{19}$$

Substituting Eq. 19 in Eq. 13, we obtain the thermodynamic feasibility constraint, applicable to any reaction, that will be employed in the formulation:

$$\sum_{i=1}^{N^S} (a_i T_j + b_i)(\delta_j - \gamma_j) + zRT \le 0$$

$$j \in \{1,2,\ldots,N^R\}. \tag{20}$$

This constraint is nonlinear if the reaction temperature T is not specified.

Sum of coefficients constraints

These variable constraints place upper (δ_j^u, γ_j^u) and lower (δ_j^l, γ_j^l) bounds on the sum of coefficients in each half reaction. They are

$$\delta_j^l \le \sum_{i=1}^{N^S} \delta_{ij} \le \delta_j^u \quad j \in \{1,2,\ldots,N^R\} \tag{21}$$

$$\gamma_j^l \le \sum_{i=1}^{N^S} \gamma_{ij} \le \gamma_j^u \quad j \in \{1,2,\ldots,N^R\}. \tag{22}$$

The interplay between these and other variable constraints are discussed in a separate section.

Allowed coefficient values constraints

By definition, allowable reactions have integer coefficients. The allowed coefficient values constraints specify the integer values a coefficient may attain. These constraints are variable since the user chooses all allowed values according to what is desired or acceptable. The form of the constraints is

$$\delta_{ij}(\delta_{ij} - \phi_1)\cdots(\delta_{ij} - \phi_n) = 0 \quad i \in \{1,2,\ldots,N^S\}$$

$$j \in \{1,2,\ldots,N^R\} \tag{23}$$

$$\gamma_{ij}(\gamma_{ij} - \phi_1)\cdots(\gamma_{ij} - \phi_n) = 0 \quad i \in \{1,2,\ldots,N^S\}$$

$$j \in \{1,2,\ldots,N^R\}, \tag{24}$$

where ϕ_1, $\phi_2,\ldots,\phi_n$ are the allowed integer values. These constraints are implemented through a branch-and-bound procedure.

Bounds of coefficients

There is an upper bound δ_{ij}^{u} and γ_{ij}^{u} for each δ_{ij} and γ_{ij}, respectively. The bounds take the form

$$0 \le \delta_{ij} \le \delta_{ij}^{u} \quad i \in \{1,2,\ldots,N^S\} \quad j \in \{1,2,\ldots,N^R\} \quad (25)$$

$$0 \le \gamma_{ij} \le \gamma_{ij}^{u} \quad i \in \{1,2,\ldots,N^S\} \quad j \in \{1,2,\ldots,N^R\}. \quad (26)$$

Species limit constraints

These variable constraints provide for a maximal number of s species per half reaction. Let there be a binary variable η_{ij} per-product coefficient, and one θ_{ij} per-reactant coefficient, such that

$$\eta_{ij}(\eta_{ij}-1)=0 \quad i \in \{1,2,\ldots,N^S\} \quad j \in \{1,2,\ldots,N^R\} \quad (27)$$

$$\theta_{ij}(\theta_{ij}-1)=0 \quad i \in \{1,2,\ldots,N^S\} \quad j \in \{1,2,\ldots,N^R\}. \quad (28)$$

Then, let their values be defined according to

$$0 \le \delta_{ij} \le \delta_{ij}^{u}\eta_{ij} \quad i \in \{1,2,\ldots,N^S\} \quad j \in \{1,2,\ldots,N^R\} \quad (29)$$

$$0 \le \gamma_{ij} \le \gamma_{ij}^{u}\theta_{ij} \quad i \in \{1,2,\ldots,N^S\} \quad j \in \{1,2,\ldots,N^R\}. \quad (30)$$

This ensures that whenever γ_{ij} or δ_{ij} is not zero, its corresponding binary variable will be one. The species limit constraint is then enforced by

$$\sum_{i=1}^{N^S} \eta_{ij} \le s \quad j \in \{1,2,\ldots,N^R\} \quad (31)$$

$$\sum_{i=1}^{N^S} \theta_{ij} \le s \quad j \in \{1,2,\ldots,N^R\}. \quad (32)$$

Here s is assumed constant for all half reactions, but may be specified for each one individually if desired. These constraints are enforced through a modification of the branch-and-bound procedure.

Bounds on reaction temperatures

The operational temperature range under which the constitutive reactions are allowed to operate is specified with

$$T_L \le T_j \le T_U \quad j \in \{1,2,\ldots,N^R\}, \quad (33)$$

where T_L is the lowest and T_U is the highest allowable temperature.

Objective

The objective is the weighted sum of the squares of the coefficients of all species in all cluster reactions:

$$\sum_{i=1}^{N^S}\sum_{j=1}^{N^R}\left(d_i\delta_{ij}^2 + c_i\gamma_{ij}^2\right). \quad (34)$$

Here, d_i and c_i are the weighing coefficients for δ_{ij} and γ_{ij}, respectively, which will select clusters based on economics, safety, or other characteristics. A quadratic objective is chosen because it ensures uniqueness of solution for the subproblems encountered during the employed branch-and-bound procedure. This allows *all* thermodynamically feasible clusters to be identified.

Problem formulation

The resulting problem formulation consists of the constraints and equations just described and will be referred to as Problem $P1$. It is

$$\min \sum_{i=1}^{N^S}\sum_{j=1}^{N^R}\left(d_i\delta_{ij}^2 + c_i\gamma_{ij}^2\right) \quad (P1) \quad (34)$$

$$\text{st} \sum_{i=1}^{N^S} B_{ki}(\delta_{ij}-\gamma_{ij}) = 0 \quad k \in \{1,2,\ldots,N^A\} \; j \in \{1,2,\ldots,N^R\} \quad (14)$$

$$\sum_{j=1}^{N^R}(\delta_{ij}-\gamma_{ij}) = \nu_i \quad i \in \{1,2,\ldots,N^S\} \quad (15)$$

$$\sum_{i=1}^{N^S}(a_iT_j + b_i)(\delta_{ij}-\gamma_{ij}) + zRT_j \le 0 \quad j \in \{1,2,\ldots,N^R\}. \quad (20)$$

$$\delta_j^l \le \sum_{i=1}^{N^S}\delta_{ij} \le \delta_j^u \quad j \in \{1,2,\ldots,N^R\} \quad (21)$$

$$\gamma_j^l \le \sum_{i=1}^{N^S}\gamma_{ij} \le \gamma_j^u \quad j \in \{1,2,\ldots,N^R\} \quad (22)$$

$$\delta_{ij}(\delta_{ij}-\phi_1)\cdots(\delta_{ij}-\phi_n)=0 \quad i \in \{1,2,\ldots,N^S\} \quad j \in \{1,2,\ldots,N^R\} \quad (23)$$

$$\gamma_{ij}(\gamma_{ij}-\phi_1)\cdots(\gamma_{ij}-\phi_n)=0 \quad i \in \{1,2,\ldots,N^S\} \quad j \in \{1,2,\ldots,N^R\} \quad (24)$$

$$0 \le \delta_{ij} \le \delta_{ij}^{u} \quad i \in \{1,2,\ldots,N^S\} \quad j \in \{1,2,\ldots,N^R\} \quad (25)$$

$$0 \le \gamma_{ij} \le \gamma_{ij}^{u} \quad i \in \{1,2,\ldots,N^S\} \quad j \in \{1,2,\ldots,N^R\} \quad (26)$$

$$\eta_{ij}(\eta_{ij}-1)=0 \quad i \in \{1,2,\ldots,N^S\} \quad j \in \{1,2,\ldots,N^R\} \quad (27)$$

$$\theta_{ij}(\theta_{ij}-1)=0 \quad i \in \{1,2,\ldots,N^S\} \quad j \in \{1,2,\ldots,N^R\} \quad (28)$$

$$0 \le \delta_{ij} \le \delta_{ij}^{u}\eta_{ij} \quad i \in \{1,2,\ldots,N^S\} \quad j \in \{1,2,\ldots,N^R\} \quad (29)$$

$$0 \le \gamma_{ij} \le \gamma_{ij}^{u}\theta_{ij} \quad i \in \{1,2,\ldots,N^S\} \quad j \in \{1,2,\ldots,N^R\} \quad (30)$$

$$\sum_{i=1}^{N^S}\eta_{ij} \le s \quad j \in \{1,2,\ldots,N^R\} \quad (31)$$

$$\sum_{i=1}^{N^S}\theta_{ij} \le s \quad j \in \{1,2,\ldots,N^R\} \quad (32)$$

$$T_L \le T_j \le T_U \quad j \in \{1,2,\ldots,N^R\}. \quad (33)$$

Problem $P1$ is an MINLP. All γ_{ij}, δ_{ij} are integers, while T_j is a continuous variable, appearing in the nonlinear constraint (Eq. 20). The integer constraints are shown in (Eqs. 23–32). Next, two problem properties are presented that allow the MINLP to be converted into N^R-1 integer problems (IP). In these problems all nonlinear terms are removed by specifying the temperature at which each reaction takes place. The integer constraints and quadratic objective do not change.

Problem Properties

The first property states that any feasible solution of Problem $P1$ may be written with all reaction temperatures at either T_L or T_U.

Property 1. Consider an optimal solution δ_{ij}^*, γ_{ij}^*, T_j^* for Problem $P1$ with $T_j^* \in [T_L, T_U]$ for all j. Then another feasible solution δ_{ij}', γ_{ij}', T_j' exists such that the objective value remains the same with $\gamma_{ij}' = \gamma_{ij}^*$ and $\delta_{ij}' = \delta_{ij}^*$, while the temperature T_j' satisfies $(T_j' - T_L)(T_j' - T_U) = 0$ for all j.

Proof. It is first necessary to show that any single reaction j with a specific temperature $T_j^* \in [T_L, T_U]$ is also feasible with a T_j' at either T_L or T_U. There are two cases to consider:

• *Case i:*

$$T_j^* = T_U \quad \text{or} \quad T_L.$$

We simply need to take $T_j' = T_j^*$.

• *Case ii:*

$$T_j^* \in (T_U, T_L).$$

Since the thermodynamic constraint is satisfied,

$$\Delta G_j^* + zRT_j^* = \sum_{i=1}^{N_S} (a_i T_j^* + b_i)(\delta_{ij} - \gamma_{ij}) + zRT_j^* \le 0. \quad (35)$$

Since this is the only constraint containing temperature, we can perturb T and evaluate the effect of the change in the whole formulation by examining only this same constraint.

There are two possibilities:

• $\dfrac{d\Delta G_j^*}{dT} + zR \ge 0.$

Let $T_j' = T_L$ and $\delta_{ij}' = \delta_{ij}^*$, $\gamma_{ij}' = \gamma_{ij}^*$ for all i and j.
Then, $\Delta G_j' + zRT_L \le \Delta G_j^* + zRT_j^* \le 0.$

• $\dfrac{d\Delta G_j^*}{dT} + zR < 0$

Let $T_j' = T_U$ and $\delta_{ij}' = \delta_{ij}^*$, $\gamma_{ij}' = \gamma_{ij}^*$ for all i and j.
Then, $\Delta G_j' + zRT_U < \Delta G_j^* + zRT_j^* \le 0.$

Therefore, if a reaction is feasible somewhere in the allowed temperature range, then it is feasible at one of the interval edges, and thus the thermodynamic feasibility constraints can only be made easier to satisfy. It follows that any cluster can be represented with all N^R reaction temperatures at their bounds. In addition,

$$\delta_{ij}' = \delta_{ij}^* \quad i \in \{1,2,\ldots,N^S\} \quad j \in \{1,2,\ldots,N^R\}$$

$$\gamma_{ij}' = \gamma_{ij}^* \quad i \in \{1,2,\ldots,N^S\} \quad j \in \{1,2,\ldots,N^R\}$$

and the coefficients for the whole cluster remain the same. Because temperature only appears in the thermodynamic constraints, all other constraints are still satisfied while the objective value remains the same. Therefore the new solution is also an optimal solution.

The property just proved provides a means of transforming the problem formulation into an IP. Since the temperature of a cluster reaction can only have one of two values, T_j is set to either T_U or T_L. This removes the only nonlinearity from the constraints.

Because there are N^R reactions per cluster, it is not known *a priori* which ones should be set to which temperature in order to capture the global optimum. Property 2 provides an answer to this question and significantly reduces the complexity of the problem.

Property 2. The global optimum of Problem $P1$ is the minimum of N^R-1 IPs that arise from Problem $P1$ when the first i reaction temperatures are set to T_L and the last N^R-i to T_U, for each $i = 1,2,\ldots,N^R-1$.

Proof. Consider an optimum solution vector $\mathbf{x}$ for an IP in which the first i reactions take place at $T_j = T_L$. Consider also a perturbation $\mathbf{x}'$ of $\mathbf{x}$, which results from the reordering of the first i reactions among each other and of the last N^R-i reactions among each other. Then $\mathbf{x}'$ is a feasible vector and represents the same cluster as $\mathbf{x}$. Furthermore, since the objective function coefficients d_i, c_i do not depend on j, it also holds

$$\nu' = \nu, \quad (36)$$

where ν' and ν are the objective values of $\boldsymbol{x}'$ and $\boldsymbol{x}$, respectively. Thus $\boldsymbol{x}'$ is also an optimum solution. Because $\boldsymbol{x}$ and $\boldsymbol{x}'$ represent the same cluster, to improve computational speed we need only generate solutions that have all their low-temperature reactions first.

The following temperatures for the N^R reactions should then be employed for each of the N^R-1 IPs that need to be solved:

1 reaction at T_L	N^R-1 reactions at T_U
2 reactions at T_L	N^R-2 reactions at T_U
...	...
N^R-2 reactions at T_L	2 reactions at T_U
N^R-1 reactions at T_L	1 reaction at T_U

Integer Problem Formulation

Properties 1 and 2 allow the removal of the nonlinear thermodynamic feasibility constraints, since the temperature is now set to T_L or T_U for each reaction. Because the allowed coefficient values and species limit constraints are implemented through a modified branch-and-bound procedure, there are no nonlinear constraints implemented within the algorithm. The modified formulation involves solution of a sequence of IPs, and the optimal answer can be found in a finite number of steps.

The new problem formulation, Problem $P2$, is

$$\min \quad \{c_h\} \qquad h \in \{1,2,\ldots,N^R-1\} \quad (P2) \qquad (37)$$

$$c_h = \min \sum_{i=1}^{N^S} \sum_{j=1}^{N^R} \left(d_i \delta_{ij}^2 + c_i \gamma_{ij}^2 \right) \quad h \in \{1,2,\ldots,N^R - 1\} \quad (38)$$

$$\text{st} \quad \sum_{i=1}^{N^S} B_{ki} (\delta_{ij} - \gamma_{ij}) = 0 \qquad k \in \{1,2,\ldots,N^A\}$$

$$j \in \{1,2,\ldots,N^R\} \quad (14)$$

$$\sum_{j=1}^{N^R} (\delta_{ij} - \gamma_{ij}) = \nu_i \quad i \in \{1,2,\ldots,N^S\} \quad (15)$$

$$\sum_{i=1}^{N^S} (a_i T_j + b_i)(\delta_{ij} - \gamma_{ij}) + zRT_j \le 0 \quad j \in \{1,2,\ldots,N^R\} \quad (20)$$

$$\delta_j^l \le \sum_{i=1}^{N^S} \delta_{ij} \le \delta_j^u \quad j \in \{1,2,\ldots,N^R\} \quad (21)$$

$$\gamma_j^l \le \sum_{i=1}^{N^S} \gamma_{ij} \le \gamma_j^u \quad j \in \{1,2,\ldots,N^R\} \quad (22)$$

$$\delta_{ij}(\delta_{ij} - \phi_1)\cdots(\delta_{ij} - \phi_n) = 0 \quad i \in \{1,2,\ldots,N^S\}$$

$$j \in \{1,2,\ldots,N^R\} \quad (23)$$

$$\gamma_{ij}(\gamma_{ij} - \phi_1)\cdots(\gamma_{ij} - \phi_n) = 0 \quad i \in \{1,2,\ldots,N^S\}$$

$$j \in \{1,2,\ldots,N^R\} \quad (24)$$

$$0 \le \delta_{ij} \le \delta_{ij}^u \quad i \in \{1,2,\ldots,N^S\} \quad j \in \{1,2,\ldots,N^R\} \quad (25)$$

$$0 \le \gamma_{ij} \le \gamma_{ij}^u \quad i \in \{1,2,\ldots,N^S\} \quad j \in \{1,2,\ldots,N^R\} \quad (26)$$

$$\eta_{ij}(\eta_{ij} - 1) = 0 \quad i \in \{1,2,\ldots,N^S\} \quad j \in \{1,2,\ldots,N^R\} \quad (27)$$

$$\theta_{ij}(\theta_{ij} - 1) = 0 \quad i \in \{1,2,\ldots,N^S\} \quad j \in \{1,2,\ldots,N^R\} \quad (28)$$

$$0 \le \delta_{ij} \le \delta_{ij}^u \eta_{ij} \quad i \in \{1,2,\ldots,N^S\} \quad j \in \{1,2,\ldots,N^R\} \quad (29)$$

$$0 \le \gamma_{ij} \le \gamma_{ij}^u \theta_{ij} \quad i \in \{1,2,\ldots,N^S\} \quad j \in \{1,2,\ldots,N^R\} \quad (30)$$

$$\sum_{i=1}^{N^S} \eta_{ij} \le s \quad j \in \{1,2,\ldots,N^R\} \quad (31)$$

$$\sum_{i=1}^{N^S} \theta_{ij} \le s \quad j \in \{1,2,\ldots,N^R\} \quad (32)$$

$$T_j = T_L \quad j \in \{1,2,\ldots,h\} \quad (39)$$

$$T_j = T_U \quad j \in \{h+1, h+2, \ldots, N^R\}. \quad (40)$$

Problem Features

The cluster generation technique presented here has several features which make it a very powerful and versatile tool. These features stem from the types of constraints incorporated in (*P2*) as much as from their associated parameter values.

Obtaining all solutions

In this optimization approach, the problem formulation (Problem *P2*) produces the best reaction cluster based on a set of performance criteria and constraints. In practice it is desirable to consider more than one cluster. It is necessary, therefore, that the employed synthesis method be able to deliver either a number or all feasible clusters.

The modified branch-and-bound technique employed here accomplishes just that. Once the optimal cluster is found, the algorithm continues on remaining branches until the next best cluster is produced. This may continue until the first n solutions are found or until all branches are exhausted. The latter eventuality implies that all clusters have been found.

Use of a strictly convex quadratic objective function ensures that solutions are not lost because at each (feasible) branch there corresponds a unique optimum.

Constitutive reaction final form

The variable constraints have been incorporated in the problem formulation to afford control over the eventual form of the reactions. The species limit constraints (Eqs. 27–32) prevent the appearance of a large number of species in a half reaction that would otherwise only be limited by the sum of coefficients upper bound (Eqs. 21 and 22). The allowed coefficient value constraints (Eqs. 23 and 24) provide desirable values for each species, while the sum of coefficients constraint prevents, if desired, the possibility that all species are at or close to their upper bounds by forcing some coefficients to be smaller than ϕ_n without imposing bounds on any particular one.

Parameter value selection determines allowed reaction types. For example, if only reactions of the type

$$\gamma_1 A + \gamma_2 B \rightarrow \delta_1 C + \delta_2 D$$

are desired, where 0, 1, and 2 are the only allowed coefficient values, they can be obtained by the following parameter settings: $\phi_1 = 1$, $\phi_2 = \phi_n = 2$, $s = 2$, $\delta_j^u = \gamma_j^u = 4$, $\delta_j^l = \gamma_j^l = 1$. Note that reactions of the type

$$\mathrm{A+B+C+D \rightarrow E+F+G+H}$$

are excluded due to the species limit constraint. If instead one desires to exclude the reaction

$$\mathrm{2A+2B \rightarrow 2C+2D},$$

the bounds δ_j^u and γ_j^u must simply be set to 3.

Overall reaction

An overall reaction does not have to conform to the definition of an allowable reaction; it can have more than s reactant or product species, and its coefficients can be other than those specified in the allowed coefficient values constraints. Thus, the method presented here has the capacity to generate acceptable clusters for a great variety of infeasible reactions.

Objective

As was mentioned earlier, the objective is also useful in reaction cluster generation. It can reflect the total cost associated with use of the chemicals, and may include cost of

purchase and material loss due to purge streams and separation processes.

Alternate objective functions may be used, which reflect reactivity, toxicity, or environmental safety of the participating species. The objective can also be used to shift the focus of generated solutions to those containing intermediates with certain characteristics. For example, one might be interested in generation of reaction clusters incorporating nitrogen-containing species, or might want to favor certain desirable molecular groups. Finally, it is possible that an objective reflects more than one of the previously mentioned quantities.

When all clusters are desired, the objective merely affects the order in which they appear, the ones generated first being the best candidates for the task.

Case Study

Problem $P2$ was implemented with a database of fifty species for the overall reaction of the decomposition of water,

$$2H_2O \rightarrow 2H_2 + O_2$$

a highly infeasible and very important reaction from a clean energy generation point of view (Chao, 1974). The thermodynamic data used were the least-squares linear fits of the free energy of formation vs. temperature relationships of the species in the database. The fits were calculated by the thermodynamic relation (Eq. 16)

$$\Delta G = \Delta H - T\Delta S,$$

where ΔH and ΔS were given, for a certain temperature range, as curve fits based on experimental data. The free-energy linear fits are a good representation of the actual data. Table 1 lists the species used as well as the parameters a_i (slope) and b_i (intercept) for the linear $\Delta G_{f,i}$ relationship (Eq. 17).

Table 1. Standard Free Energies of Formation $\Delta G_{f,i}$ (T) (kJ/mol)

	$\Delta G_{i,f}(T) = a_i T + b_i \quad T$(K)				
Species	Slope (a_i)	Intercept (b_i)	Species	Slope (a_i)	Intercept (b_i)
$Br_2(g)$	−0.32526	96.098	CH_3NO_3	−0.57308	120.740
$BrCl(g)$	−0.31784	77.958	CH_3OH	−0.41221	−43.118
C_2Cl_2O	−0.44474	−3.042	$CH_4(g)$	−0.32620	54.509
$C_2H_2(g)$	−0.33923	347.710	CHBrO	−0.36191	−42.778
$C_2H_2Cl_2$	−0.51173	195.025	CHClO	−0.38986	−54.593
C_2H_2O	−0.39220	137.752	H_2O_2	−0.36042	−25.071
C_2H_3Br	−0.36850	123.200	$CN(g)$	−0.27411	494.647
$C_2H_3Cl_3$	−0.47259	−88.819	C_2HCl_5	−0.63411	16.616
C_2H_3Cl	−0.37943	84.409	$CO(g)$	−0.26548	−54.204
$C_2H_4(g)$	−0.40480	220.351	$CO_2(g)$	−0.32387	−300.201
$C_2H_4Br_2$	−0.46442	26.072	Cl_2	−0.29995	63.012
C_2H_5Br	−0.40372	−4.374	C_2Cl_4	−0.60026	200.458
C_2H_5Cl	−0.41918	−33.344	H_2	−0.19530	54.015
C_2H_5OH	−0.55450	14.154	H_2O	−0.27542	−166.447
C_2H_6	−0.43084	79.732	HBr	−0.26472	19.268
$C_2H_5NH_2$	−0.63935	396.532	$HNO_3(g)$	−0.46729	191.292
$C_3H_5NO_2$	−0.68298	362.175	$HCl(g)$	−0.25125	−37.714
C_3H_6O	−0.59735	266.525	N_2	−0.25844	55.956
CH_2Br_2	−0.37005	40.689	$N_2O(g)$	−0.33123	175.921
C_3H_8	−0.60167	202.033	$N_2O_4(g)$	−0.54584	212.821
$CH_2O(g)$	−0.34178	−6.595	$N_2O_5(g)$	−0.63529	250.630
CH_3Br	−0.31490	−1.833	NH_3	−0.30678	56.694
$CH_3Cl(g)$	−0.39016	56.083	$NO(g)$	−0.28026	148.167
CH_3NH_2	−0.44391	163.702	$NO_2(g)$	−0.34609	122.333
CH_3NO_2	−0.49927	127.550	O_2	−0.27601	59.625

Table 2. Parameter Values for the Decomposition of Water Case Study

$N^S = 50$	$N^R = 2$
$d_i = 1\ \forall i$	$c_i = 1\ \forall i$
$\phi_1 = 1$	$\phi_2 = \phi_n = 2$
$\delta_i^u = 4$	$\gamma_j^u = 4$
$\delta_j^l = 1$	$\gamma_j^l = 1$
$z = 0.0$	s varied
T_L = varied	T_U = varied

The formulation featured two rections per cluster, while the parameter values were as shown in Table 2.

The objective had equal weights for every species in the cluster, and the problem was solved for several values of s, on several temperature ranges. The issue of whether all species were thermally stable at the temperatures used in this example was not examined.

With $s = 4$ the species limit constraint was nonbinding because a maximum of four species was allowed by the constraints on the sum of the coefficients. For this case, there exist 4 thermodynamically feasible clusters in the temperature range 400 K to 1600 K. All of these had four species in at least one of their half reactions. They are:

$$C_2H_3Cl_3 + C_2H_3Cl \rightarrow C_2H_2 + C_2Cl_4 + 2H_2$$

$$T = 1600\ K \quad \Delta G^0_{1600\,K} = -104.3\ \text{kJ}$$

$$C_2H_2(g) + C_2Cl_4 + 2H_2O \rightarrow C_2H_3Cl_3 + C_2H_3Cl + O_2$$

$$T = 400\ K \quad \Delta G^0_{400\,K} = -15.1\ \text{kJ}.$$

$$2C_2H_3Cl_3 \rightarrow C_2H_2Cl_2 + C_2Cl_4 + 2H_2$$

$$T = 1600\ \text{K} \quad \Delta G^0_{1600\,K} = -210.7\ \text{kJ}$$

$$C_2H_2Cl_2 + C_2Cl_4 + 2H_2O \rightarrow 2C_2H_3Cl_3 + O_2$$

$$T = 400\ \text{K}\ \Delta G^0_{400\,K} = -3.9\ \text{kJ}$$

$$C_2H_3Cl_3 + C_2H_5Cl + H_2O \rightarrow C_2H_2Cl_2 + C_2H_5OH + Cl_2 + H_2$$

$$T = 1600\ \text{K} \quad \Delta G^0_{1600\,K} = -16.0\ \text{kJ}$$

$$C_2H_2Cl_2 + C_2H_5OH + Cl_2 + H_2O \rightarrow C_2H_3Cl_3 + C_2H_5Cl + H_2 + O_2$$

$$T = 400\ \text{K} \quad \Delta G^0_{400\,K} = -2.9\ \text{kJ}.$$

$$2C_2H_3Cl_3 + H_2O \rightarrow C_2H_5OH + Cl_2 + C_2Cl_4 + H_2$$

$$T = 1600\ \text{K} \quad \Delta G^0_{1600\,K} = -11.3\ \text{kJ}$$

$$C_2H_5OH + Cl_2 + C_2Cl_4 + H_2O \rightarrow 2C_2H_3Cl_3 + H_2 + O_2$$

$$T = 400\ \text{K} \quad \Delta G^0_{400\,K} = -49.7\ \text{kJ}.$$

When the value of $s = 2$ was employed (allowing only two species per half reaction) there were no acceptable clusters in the temperature range [400, 1600]. Extrapolating the thermodynamic data to larger temperature intervals, the first acceptable cluster occurs at the temperature range [700, 3600], and is

$$2CO_2 \rightarrow 2CO + O_2 \quad T = 3600 \text{ K } \Delta G^0_{3600\,K} = -21.7 \text{ kJ}$$

$$2H_2O + 2CO \rightarrow 2CO_2 + 2H_2 \quad T = 700 \text{ K } \Delta G^0_{700\,K} = -20.6 \text{ kJ}.$$

Because the data were extrapolated beyond the range specified by the linear fits, this result should be viewed with caution. Nevertheless, it demonstrates the invention capabilities of the proposed synthesis technique. The aforementioned cluster was not included in Chao (1974).

Conclusions

The complete and automatic synthesis of all thermodynamically feasible reaction clusters that exist within a given set of species was formulated as an optimization problem. Central to the structure of the formulation is the thermodynamic feasibility of each cluster reaction, conservation of mass, as well as the definition and generation of allowable reactions.

The formulation is flexible, allowing for a given number (N^R) of reactions per cluster. The user may specify the number of participating species per half reaction, as well as the allowed values that chemical coefficients can attain. Reactions can be further controlled by forcing some coefficients to be away from their maximal values. A minimum value for the equilibrium constant can be specified in all reactions, thereby allowing the overall reaction to be carried out at an economically profitable yield.

Acknowledgments

Financial support for this work was provided by NSF Award CTS9528653, GER9554570 as well as by U.S. Department of Education Grant P200A40732-95. We gratefully acknowledge the help of Prof. Selim Senkan, of the UCLA Chemical Engineering Department, who provided the database on thermodynamic properties of species. We also thank Guillermo Pont and Christopher Kim, also in the UCLA Department of Chemical Engineering, for fruitful discussions at the initial stages of the project.

Literature Cited

Benson, S., *Thermochemical Kinetics*, Wiley, New York (1976).

Chao, R., "Thermochemical Water Decomposition Processes," *Ind. Eng. Chem. Prod. Res. Develop.*, **13**(2), 94 (1974).

Crabtree, E., and M. El-Halwagi, "Synthesis of Environmentally Acceptable Reactions," *AIChE Symp. Ser.*, **90**(303), 117 (1995).

Denbigh, K., *The Principles of Chemical Equilibrium with Applications in Chemistry and Chemical Engineering*, 4th ed., Cambridge Univ. Press, Cambridge, England (1981).

Fornari, T., and G. Stephanopoulos, "Synthesis of Chemical Reaction Paths: Economic and Specification Constraints," *Chem. Eng. Commun.*, **129**, 159 (1994a).

Fornari, T., and G. Stephanopoulos, "Synthesis of Chemical Reaction Paths: The Scope of Group Contribution Methods," *Chem. Eng. Commun.*, **129**, 135 (1994b).

Gani, R., B. Nielsen, and A. Fredenslund, "A Group Contribution Approach to Computer-Aided Molecular Design," *AIChE J.*, **37**(9), 1318 (1991).

Happel, J., and P. Sellers, "Characterization of Complex Systems of Chemical Reactions," *Chem. Eng. Commun.*, **83**, 221 (1989).

Happel, J., P. Sellers, and M. Otarod, "Mechanistic Study of Chemical Reaction Systems," *Ind. Eng. Chem. Res.*, **29**, 1057 (1990).

Holiastos, K., and V. Manousiouthakis, "Automatic Synthesis of Thermodynamically Feasible Reaction Clusters," Annu. Meeting of the American Institute of Chemical Engineers, Chicago (1996).

Mavrovouniotis, M., and G. Stephanopoulos, "Synthesis of Reaction Mechanisms Consisting of Reversible and Irreversible Steps. 1. A Synthesis Algorithm," *Ind. Eng. Chem. Res.*, **31**, 1625 (1992).

Mavrovouniotis, M., "Synthesis of Reaction Mechanisms Consisting of Reversible and Irreversible Steps. 2. Formalization and Analysis of the Synthesis Algorithm," *Ind. Eng. Chem. Res.*, **31**, 1637 (1992).

Mavrovouniotis, M., G. Stephanopoulos, and G. Stephanopoulos, "Synthesis of Biochemical Production Routes," *Comput. Chem. Eng.*, **16**(6), 605 (1992).

Mavrovouniotis, M., "Duality Theory for Thermodynamic Bottlenecks in Bioreaction Pathways," *Chem. Eng. Sci.*, **5**(9), 1495 (1995).

May, D., and D. Rudd, "Development of Solvay Clusters of Chemical Reactions," *Chem. Eng. Sci.*, **31**, 59 (1976).

Pitzer, K., and L. Brewer, *Thermodynamics*, 2nd ed., McGraw-Hill, New York (1961).

Reid, C. E., *Principles of Chemical Thermodynamics*, Reinhold, New York (1960).

Rotstein, E., D. Resasco, and G. Stephanopoulos, "Studies on the Synthesis of Chemical Reaction Paths: I", *Chem. Eng. Sci.*, **37**(9), 1337 (1982).

Seressiotis, A., and J. Bailey, "MPS: An Artificially Intelligent Software System for the Analysis and Synthesis of Metabolic Pathways," *Biotechnol. Bioeng.*, **31**, 587 (1988).

Valdez-Perez, R., "Algorithm to Test the Structural Plausibility of a Proposed Elementary Reaction," *J. Comput. Chem.*, **14**(12), 1454 (1993).

Valdez-Perez, R., "Algorithm to Infer the Structures of Molecular Formulas within a Reaction Pathway," *J. Comput. Chem.*, **15**(11), 1266 (1994).

Valdez-Perez, R., "Algorithm to Generate Reaction Pathways for Computer-Assisted Elucidation," *J. Comput. Chem.*, **13**(9), 1079 (1992).

Manuscript received Mar. 12, 1997, and revision received Sept. 22, 1997.