Computer Generated Pyrolysis Modeling: On-the-Fly Generation of Species, Reactions, and Rates

Linda J. Broadbelt, Scott M. Stark, and Michael T. Klein*

Center for Catalytic Science and Technology, Department of Chemical Engineering, University of Delaware, Newark, Delaware 19716

The development of an integrated system for the computer generation of kinetic models is described. Required input is the structure of the reactants, the reaction rules, and the parameters of a structure/property kinetics correlation. The algorithm transforms this information into reactant/product relationships, i.e., the reaction network, species properties, rate constants, and the FORTRAN code corresponding to the governing species' balance equations, and offers a solution capability. Graph theory is exploited to represent the constituent atoms of a molecule to allow determination of species' uniqueness, implement chemical reactions, and identify reaction products. Special attention was devoted to improved algorithm efficiencies, the handling of ring systems, and "on-the-fly" quantum chemical calculations. This general approach is described in using ethane and cyclohexane pyrolysis case studies. The increase in the number of equations and number of components for ethane pyrolysis was exponential with the carbon number of allowed species.

Introduction

Environmentally unacceptable waste streams, complex hydrocarbon feedstocks, and mixtures of polymeric wastes are examples from a wide array of potential reactants that share the common feature of consisting of thousands of reactive molecules. The reactions of these complex mixtures are also subject to ever-increasing scrutiny as to not only performance but also environmental properties. This has motivated the development of models and modeling approaches for the reactions of complex feeds.

The forces requiring such molecular detail are emerging at a time when analytical and computational hardware provide unprecedented opportunity to measure and model molecules. Modern analytical methods indicate the existence of at least $O(10^5)$ different molecules in the mixtures. The sheer size of the thus-implied modeling problem engenders a conflict between the need for molecular detail and the formulation and solution of the model. Traditional deterministic models would comprise an extremely large $O(10^5)$ number of species and therefore differential equations. Thus the formulation, let alone solution, of a single model would be formidable. Moreover, improvement through iteration and reformulation of a model becomes extremely difficult after several years' investment time. Clearly the use of the computer to not only solve but also formulate the model would be an important achievement that would allow the modeler to focus on the basic chemistry and rules of the model.

This work was aimed at developing such an automated capability to model development. The hope was that the modeler would be called upon for only three components: (1) the structure of the reactants; (2) the rules by which the reactants and product species react; and (3) the parameters of a structure/reactivity relation (e.g., Polanyi relation) that would provide an estimate of reaction rate constants. The computer would take this information and create reactant/product relationships, i.e., the reaction mechanism, calculate species' properties, estimate rate constants, write the FORTRAN code corresponding to the governing ODE's, and offer a solution capability. The related previous literature, described below, provided a foundation in terms of powerful portions of the integrated package we sought. Especially significant was the concept of a molecular graph and the related matrix representations and operations that effect chemical reactions. Our goal was to extend this foundation in both breadth and depth to arrive at the totally automated product.

This work is described as follows. The background section reviews the significant prior work that forms the foundation for our extensions. This is followed by a description of the present approach in terms of ethane and cyclohexane pyrolysis case studies. Special attention is devoted to the improved algorithm efficiencies, the handling of ring systems, and the use of "on the fly" quantum chemical calculations. Results are presented in terms of the algorithm performance as a function of carbon number and pyrolysis rules.

Background

The most comprehensive example of computer-generated reaction networks is Mobil's (Quann and Jaffe, 1992) elegant structure-oriented lumping (SOL) approach to catalytic hydroprocessing modeling. This pathways-level model describes individual molecules as a vector of structural increments. The application of reaction rules (e.g., ring saturation, sulfur removal) to the feedstock changes the set of structural vectors and thus builds the reaction network. This automated process allows for the routine revision of modeling concepts (rules) and thus-implied networks. The present work was aimed at realization of what could be called the "SOL concept" for mechanistic models of hydrocarbon pyrolysis, with explicit accounting of the radical intermediates. Having the structures of both molecules and intermediates explicit in turn would allow for "on-the-fly" computational quantum calculations for the estimation of rate constants.

The computer generation of reaction species, properties, and mechanisms relies heavily on three capabilities concerning the atomic connectivity of species: (1) the representation of constituent atoms and their chemical environment in a molecule; (2) the implementation of chemical reaction and determination of reaction products; and (3) the determination of the uniqueness of chemical species. The graph theory representation of molecules is ideally suited for providing these capabilities.

Graph Theory Description of Molecules. Graph theory emerged in 1736 through Euler's application of mathematics to show that traversing the four connected regions of the city of Konigsberg required multiple

© 1994 American Chemical Society

crossings of at least one of the seven bridges spanning its river. This problem helped define a graph as an abstract entity of a finite set of vertices and edges and a set of rules describing the edge/vertex connectivity.

Brown (Biggs et al., 1976) described a chemical species as a graph with the use of letters, enclosed in circles, to represent atoms and connectivity lines to represent single and multiple bonds. This is essentially the notation that is used today. This powerful representation of a chemical species as a graph provides access to the mathematics and associated algorithms of graph theory.

Two of these are especially useful in the computer generation of reaction mechanisms. The first permits the quantitative representation of the connectivity of graphs. This enables the determination of the number of distinct molecules a graph comprises. The second allows for the determination of uniqueness within a population of graphs. In a chemical sense, this distinguishes molecules to the level of isomeric detail. The powerful computer algorithms representing the realization of these graph theory concepts are the underlying foundation for the computer generation of reactions and reaction mechanisms. The vital link is the representation of chemical species as graphs.

Matrix Representation of Species and Reactions. The implementation of these graph theory concepts and algorithms requires efficient representation and encoding of the graphs. The adjacency matrix (Tarjan, 1977) for a graph G is the n by n matrix $\mathbf{M} = (m_{ij})$ with elements 0 and 1, such that $m_{ij} = 1$ if (v_i, v_j) is an edge of G or a connection between vertices (or atoms) of G and $m_{ij} = 0$ otherwise. This adjacency matrix representation of chemical species is sparse because of the small, fixed valences of atoms. The more efficient adjacency structure of a graph is a set of lists, one for each vertex. The list for a vertex v contains all the vertices connected to it. The bond and electron (BE) matrix (Ugi et al., 1979) provides description of not only the connectivity of a molecule but also its formal electronic state. The diagonal element, ii, of the BE matrix gives the number of nonbonded valence electrons of atom i, and off-diagonal entries, ij, provide the connectivity and bond order of atoms i and j.

The BE matrix is well suited for description of chemical reactions. This is because the number of atoms actually affected in a chemical reaction is small. The BE submatrix comprising only those atoms is small and dense. This allows addition of a reaction matrix (Ugi et al., 1979) to the BE submatrix comprising the affected atoms to be a simple and efficient matrix operation. In short, the transformation $\mathbf{R} + \mathbf{B} = \mathbf{E}$, i.e., the addition of the reaction matrix $\mathbf{R}$ to the reactant submatrix $\mathbf{B}$, yields a product submatrix $\mathbf{E}$ composed of the same atoms with altered connectivity. The product submatrix $\mathbf{E}$ can then be incorporated back into the overall BE matrix and adjacency structure to represent the entire product molecule(s). In this way, the optimal storage of molecule information in the form of the adjacency list is combined with the facile implementation of reaction by simple addition of small, dense matrices.

Species' Differentiation and Uniqueness. The connectivity of graphs specifies the identity and uniqueness of individual reaction products. Algorithms that exist for determining groups of connected vertices, i.e., the components, of a graph employ a depth-first search to visit successively each of the graph's vertices (Tarjan, 1972). Each vertex of a connected component is labeled with the same identifying number. An individual product comprises vertices of the same identifying number.

The unique labeling of each product graph provides species differentiation. This raises the concept of graph isomorphism. Cayley's (1874) (Bonchev and Rouvray, 1991) early work in the elucidation of paraffin isomers is an early example of the representation of molecules as graphs (specifically tree structures) to determine species isomorphism. The appropriate algorithm for the determination of graph isomorphism is dependent upon the graph type. Tree structures, where the number of edges is one less than the number of vertices (i.e., no rings), allow efficient $O(N)$ determination of isomorphism. The computer science literature (Tarjan, 1977; Foulds, 1992; Hopcroft and Tarjan, 1972; Corneil and Gotlieb, 1970) includes more general algorithms that can handle the majority of known chemical species, including ring structures.

Network Generators. Various combinations of these graph theory concepts have appeared in reaction network generation algorithms. Clymans and Froment (1984) used the concept of adjacency matrices, or Boolean relation matrices, in the computer generation of the reaction paths in the thermal cracking of normal and branched paraffins. In some cases, reaction paths were revealed by the computationally demanding operation of raising the adjacency matrix to a power n. For example, the molecules that could undergo a 1,5-isomerization and the atoms involved in such a transformation were identified by raising the adjacency matrix to the fourth power. Isomorphism was not discussed. This work was extended by Hillewaert et al. (1988) by the addition of a self-learning scheme that significantly reduced computational demand.

Chevalier et al. (1988) used rule-oriented programming with LISP as a support system to generate oxidation reaction mechanisms. The canonically sorted tree structures allowed identification of equivalent molecules.

DiMaio and Lignola (1992) constructed oxidation reaction mechanisms through the addition of reaction matrices to BE matrices. Their approach to the isomorphism issue was based on the notion of a graph invariant (Foulds, 1992). The permutation-invariant eigenvalues of the BE matrix revealed equivalent BE matrix representations of the same species. However, matrices with the same eigenvalues are not necessarily equal (Foulds, 1992; Greenberg, 1988). The calculation of eigenvalues is $O(N^3)$, whereas determination of isomorphic graph trees is $O(N)$ and of isomorphic planar graphs is $O(N \log N)$.

Summary of Present Capability. The present work was built on the foundation of the approaches and algorithms noted above to broaden the concept of computer-generated reaction modeling. The goal was to automate the transformation of the graph representation of reactants into FORTRAN code for the ordinary differential equations representing the reaction mechanism. This was to include not only reaction paths but also rate constants. The literature clearly provides both guidance and specific modules. However, many of these are burdened by algorithm- and problem-specific limitations. For example, cyclic structures were not addressed, and the description of molecules as BE matrices and the associated determination of species' uniqueness did not allow for unambiguous identification of species. Moreover, the computational demands of the algorithms were often extreme. Most significantly, an integrated package including computational quantum chemistry was not available. However, this impressive literature suggested that the realization of general, computer-generated reaction predictions was possible. This motivated the present work, where further attention was paid to the systematic generation and identification of species, the specification
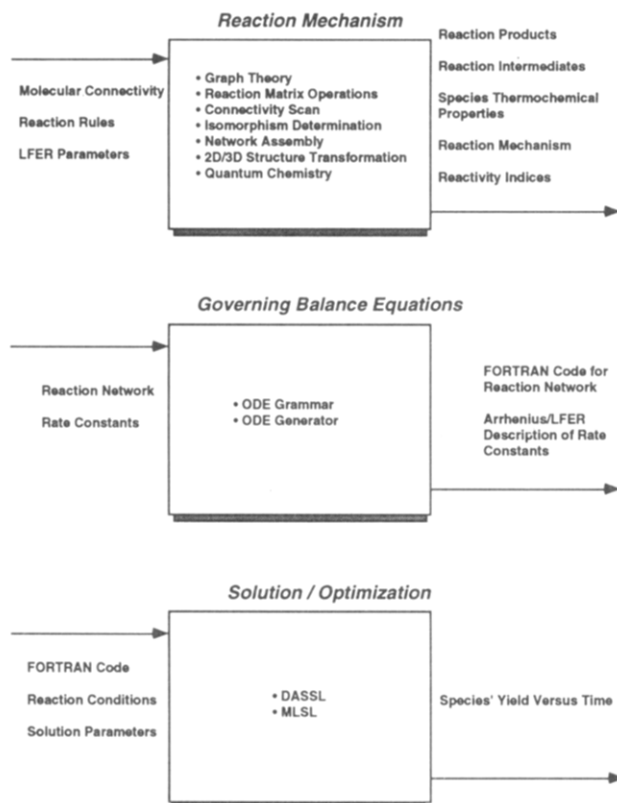
Figure 1. Vision of integrated computer-generated reaction modeling through the development of module models.

of quantitative mechanism parameters, and the integration of species and reaction generation to yield quantitative reaction mechanism predictions.

Present Approach to Computer-Generated Pyrolysis Modeling

Overview. The present logic for computer-generated pyrolysis modeling is illustrated in Figure 1. The first module is development of the reaction mechanism and rate parameters. The required input for this consists of the atomic connectivity of the molecular reactant(s), the rules describing the types of chemical transformations the species can undergo, and the small set of structure/reactivity parameters that allow quantitative specification of reactivity. The output from this module includes the list of reaction intermediates and products, the assembly of these species as a reaction mechanism, and the reactivity indices for each reaction in the mechanism. The combination of the structure/reactivity parameters (model input) and the reactivity indices obtained as output from the model allows specification of the quantitative rate parameters.

With the reaction mechanism and rate constants in hand, the next module generated is the system of governing species balance equations. The ordinary differential equation generator interprets the output from the mechanism generator in the framework of its defined grammar and outputs the FORTRAN code used to describe the reaction mechanism. The rate constants are generally embedded directly in the differential equations as quantitative values, but care was taken to allow them to be coded as parameters suitable for specification or optimization.

The final module realizes the complete integration of computer-generated reaction predictions. The FORTRAN mechanism code is combined with reaction con-
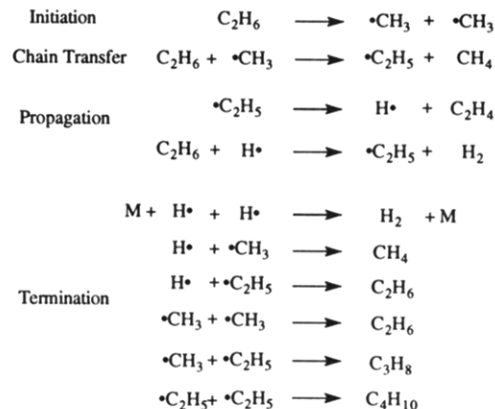

Figure 2. Representative elementary steps of free-radical chain pyrolysis of ethane.

ditions and rate parameters to solve the ordinary differential equations, obtaining predictions of the time dependence of species yields. The code can also be linked with a global optimization algorithm to obtain best-fit parameters describing the experimental data, if desired.

The present report focuses on the Reaction Mechanism module of Figure 1 and is intended to develop the chemical concepts more than the algorithm strategies, which will appear separately. These chemical concepts are developed using the familiar examples of (1) ethane pyrolysis and (2) cyclohexane pyrolysis as case studies. Ethane pyrolysis is reasonably well understood and thus provides a convenient vehicle for the discussion of graphs, reaction matrices, isomorphism, and mechanism building. Cyclohexane pyrolysis includes the added dimension of ring structures and the need for extended isomorphism algorithms.

Case Study 1. Ethane Pyrolysis

The specific problem of ethane pyrolysis was used to develop the logic and code of a more general computer-generated pyrolysis model. Ethane pyrolysis to ethylene and hydrogen proceeds via a free-radical chain process such as that shown in Figure 2. Homolytic fission of the C–C bond creates two methyl radicals that participate in a hydrogen-abstraction chain-transfer step to afford ethyl radical, whose β-scission generates ethylene and the chain carrier H$\cdot$. Bimolecular hydrogen abstraction by H$\cdot$ generates H$_2$ and the ethyl radical. Radical recombination and disproportionation steps consume two free radicals and create numerous and often minor termination products.

The discussion to follow describes the computer generation of schemes such as Figure 2 by the graph representation of molecules and radicals and the use of the reaction matrix operator to generate their reactions. Special care is taken to describe the isomorphism algorithms and the computer grammar that transforms the reaction operations into a mechanism and, ultimately, the set of ordinary differential equations. The following also highlights the "on-the-fly" specification of quantitative reaction parameters that significantly broadens the concept of computer generation of reaction mechanisms.

Representation of Chemical Species. The molecular graph of ethane is illustrated in Figure 3a. Note that it is comparable to the traditional notation used to denote molecules in chemistry. The nodes of the graph are assigned unique numbers initially in the input file and as the result of various graph-searching algorithms.

The connectivity of the ethane molecule is stored in its adjacency structure, depicted in Figure 3b. This is a set
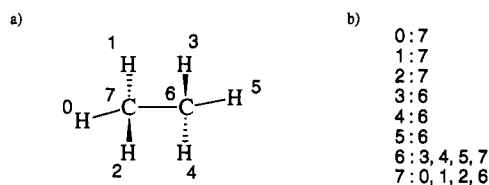
a)   b)


```
                                0:7
                                1:7
                                2:7
                                3:6
                                4:6
                                5:6
                                6:3,4,5,7
                                7:0,1,2,6
```

Figure 3. Equivalent representations of ethane: (a) molecular graph; (b) adjacency structure.

a)   b)


```
                                0:6
                                1:3
                                2:6
                                3:6,1,4
                                4:3
                                5:6
                                6:2,0,5,3
```
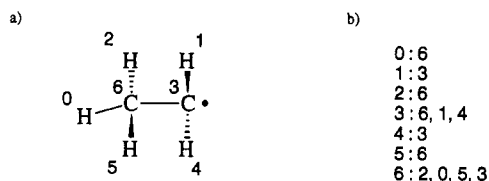
Figure 4. Radical species can also be represented in terms of (a) the molecular graph and (b) the adjacency structure.

```
C  0 1 1 1 0 0 0 1
C  1 0 0 0 1 1 1 0   C• 1 1 1 1 0 0 0
H  1 0 0 0 0 0 0 0   C  1 0 0 0 1 1 1   C  0 2 1 0 0 1
H  1 0 0 0 0 0 0 0   H  1 0 0 0 0 0 0   C  2 0 0 1 1 0
H  0 1 0 0 0 0 0 0   H  1 0 0 0 0 0 0   H  1 0 0 0 0 0
H  0 1 0 0 0 0 0 0   H  0 1 0 0 0 0 0   H  0 1 0 0 0 0
H  0 1 0 0 0 0 0 0   H  0 1 0 0 0 0 0   H  0 1 0 0 0 0
H  1 0 0 0 0 0 0 0   H  0 1 0 0 0 0 0   H  1 0 0 0 0 0
```

Figure 5. Bond–electron matrix representations that specify atomic connectivity and electronic environment for (a, left) ethane, (b, center) ethyl radical, and (c, right) ethene.

of lists that contain the identity of all of the vertices adjacent to a given vertex of interest. The adjacency list for ethane atom number 7, one of the carbon atoms, contains the values of its adjacent atoms, 0, 1, 2, and 6. The graph of a species is depicted for the user as an ordered string composed of the letters that represent the atoms contained in the molecule or radical.

Additional details of species' chemical environment were incorporated by using edge and vertex labels and graph instance variables. This allowed representation of radical centers. The graph and adjacency structure of the ethyl radical are depicted in Figure 4a and b, respectively. The adjacency structure reveals a distinction between radicals and molecules. Vertex 6, the carbon with its valency fully satisfied, has four adjacent atoms: 2, 0, 5, and 3. Vertex 3, the radical center, has only three adjacent atoms, 6, 1, and 4, and therefore indicates the presence of an unpaired electron. The radical graph has associated with it the radical identification number that is equal to the number of the atom to which the unpaired electron belongs. In this example, the radical identification number is 3. Radicals and molecules are also distinguished by the label, or atom type, of a graph vertex. For example, a carbon with a single unpaired electron has a string representation of "C." and a carbon atom with its valency satisfied is designated as simply "C".

The BE matrices of ethane, ethyl radical, and ethylene are shown in Figure 5a–c. The off-diagonal elements m_{ij} represent the bond order between the two atoms i and j, and the diagonal elements contain the number of valence electrons not involved in bonding. The sum of the entries in a given row is equal to the number of valence electrons for that atom. The diagonal elements in the ethane BE matrix are all zero, as all of the valence electrons for carbon and hydrogen are involved in bonding. The BE matrix for the ethyl radical contains a nonzero entry in the diagonal to denote the unpaired electron defining the radical center. Nonunity entries for multiple bonds are illustrated in the BE matrix of ethylene of Figure 5c. Each of these BE matrices is inherently sparse and, therefore, inefficient in terms of storage. However, a minimal
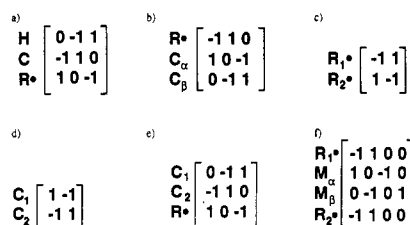
a)   b)   c)

d)   e)   f)

Figure 6. Reaction matrix representation for (a) H-abstraction, (b) β-scission, (c) recombination, (d) bond fission, (e) radical addition, and (f) disproportionation.

representation of the species BE matrix provides an efficient description of chemical transformations.

Implementation of Chemical Reaction. The reaction matrices for hydrocarbon pyrolysis are shown in Figure 6. These molecule- or radical-invariant operations were defined to involve only the subset of a species' atoms that are actually involved in the chemical reaction. This efficiency required sorting the m reactive centers that participate in the chemical transformation into an $m \times m$ BE submatrix. Note that the sorting of the reaction matrix through row and column permutation is specific to the type of reactive site involved in the transformation. For example, the application of the β-scission reaction matrix operator requires that the radical center be in row 1. Simple addition of the $m \times m$ reaction matrix could then effect reaction.

The original matrix from which the $m \times m$ BE submatrix is extracted is dependent upon the molecularity of the reaction. The full $n \times n$ BE matrix for unimolecular reactions comprises the environment of the n atoms of the reactant molecule. When the reactive atoms come from two different molecules, as in bimolecular reactions, the $i \times i$ BE matrix of reactant 1 and the $j \times j$ BE matrix of reactant 2 are merged into a single $(i + j) \times (i + j)$ BE matrix from which the minimal representation can be obtained.

User-defined reaction "rules" define the species' atoms that are involved in the chemical reaction. Bond fission requires a carbon–carbon single bond. The abstraction of every hydrogen atom in a molecule is allowed for H-abstraction. β-scission requires an atom that is in the β position to the radical center and is singly bonded to the α atom. Radical addition involves the reaction of a radical center with the atoms of a double or triple bond. Radical recombination involves the reaction of any two radical centers, and disproportionation requires one of the reacting radicals to have an atom in the β position to the radical center. All of the possible sets of atoms within a molecule or a radical that pass these rule-based criteria are extracted for participation in chemical reactions.

The conversion of a reactant(s) into product(s) was accomplished by addition of a reaction matrix to a reactant BE matrix to form a product matrix. The off-diagonal entries in the reaction matrix defined the formation or breakage of bonds, and the diagonal elements specified the loss or gain of free valence electrons. H-abstraction is shown in Figure 7a as an example of a bimolecular reaction where the reactant species' BE matrices must be merged prior to the extraction of the minimal BE matrix representation for implementation of reaction. Fission of ethane is illustrated in Figure 7b. This results in the cleavage of the C–C bond and distributes a free valence electron to each carbon. In the product matrix, shown in Figure 8, it is clear that the eight atoms in ethane have a different connectivity and the two carbon atoms have a new electronic environment.

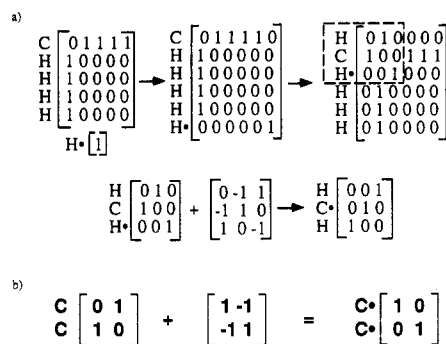**Connectivity Check: Determination of the Number**

Figure 7. (a) Hydrogen abstraction as a bimolecular reaction example. (b) Application of the fission reaction matrix to a minimal BE representation of ethane.
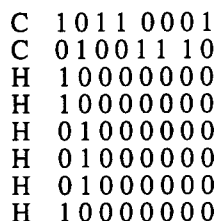
$$
\begin{array}{c}
C \\
C \\
H \\
H \\
H \\
H \\
H \\
H
\end{array}
\begin{bmatrix}
1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
$$

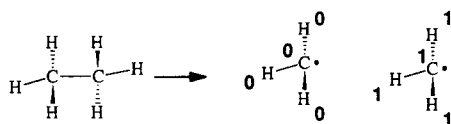**Figure 8.** Product matrix for fission of ethane.

Figure 9. Identification of the connected components of the product graph from fission of ethane.

of Products. These reaction operations altered the molecular graphs such that the number of molecules and their atomic connectivities were not, in general, preserved. This requires a systematic connectivity check of the graph resulting from reaction that determines the number and identity of the product species.

Product species were determined using a depth-first search (DFS) (Tarjan, 1972) that labeled the atoms belonging to the same molecule with the same identification number. An arbitrary first atom was chosen as the starting point. It and the atoms connected to it via any continuous path were labeled with a 0. When all connected atoms were exhausted, a new starting atom was chosen and labeled with a new identification number. The process was repeated until all of the atoms in the graph were visited. The bond fission example of Figure 9 reveals the presence of two distinct products by the two different species identification numbers. With the connectivity of these species in hand, their uniqueness as chemical structures remained to be determined before the formal reaction mechanism could be delineated.

Isomorphism: Determination of Unique Species. The determination of product graph isomorphism with any reactant or previously generated species graphs revealed the uniqueness of generated components. The simplest implementation of the isomorphism algorithm is for tree structures, i.e., connected graphs with no cycles. In chemical terms, molecules with no cycles or rings are represented as trees. The isomorphism algorithm consists of constructing a unique code for each tree and performing a lexicographical comparison with all other graphs that have the same empirical formula. The algorithm treats the graph iteratively until a unique code representation of the graph is determined. The efficiency of this algorithm is $O(N)$.
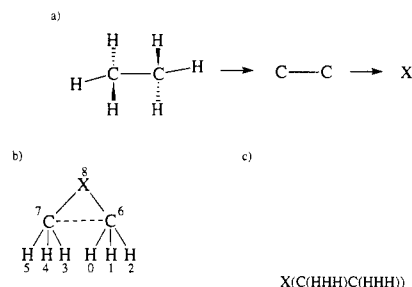
Figure 10. (a) Determination of canonical form of graph of ethane. (b) Canonical representation of graph of ethane. (c) Unique string code for ethane.

The application of this algorithm to ethane is illustrated in Figure 10. The vertices of degree 1, i.e., those with a single adjacent vertex, are successively removed from the tree until there is either one vertex, or two adjacent vertices, remaining. In the first case, the remaining vertex is designated the root node and the lexicographical ordering begins with its vertex label. In the second case, a dummy root node is fashioned, and the two remaining vertices are considered the children of this dummy root node. The descendants of the true root node or the descendants of the children of the dummy root node are then successively ordered lexicographically to construct the unique string code. The parentheses delineate generations of descendants. Isomorphism is established by a direct comparison of these unique string codes. The root labels are first compared. If these are identical, the subtrees are defined by the children of the root(s) and compared as lexicographical, ordered sequences.

These operations effect the transformation of a reactant graph(s) to a product graph(s), generating new species. Successive iteration that considers all possible reactions among starting reactants, intermediates, and products builds a list of possible reactions given the reaction rules that are provided as input. It is thus timely to turn to the integration of chemical species into a comprehensive list of their reactions.

Integrated Code for Reaction Generation

The logic of the algorithm for the generation of reaction mechanisms and rate constants is shown in Figure 11. One input is the reactant structure in the form of its graph. Complex multicomponent mixtures of reactants or mixtures of isotopically labeled reactants can be equally handled by providing each of the reactant graphs as input. The capability also exists to generate the molecular components of the mixture on-the-fly by using suitable probability density functions (Neurock et al., 1990). The second input is the set of parameters of the appropriate structure/reactivity correlation. These two form the basis for the generation of mechanism species and the calculation of reaction rate constants. Reaction generation is considered first.

The mechanism generator contains three species lists: unreacted components (molecules and/or radicals), reacted molecules, and reacted intermediates (radicals). These lists are visited through an iterative algorithm. Reaction generation begins by placing the reactants in the unreacted components array, which is a list of species that are as yet untested for reaction. The first species is then extracted from the unreacted component array. For pyrolysis of pure ethane, the initial unreacted component list contained only the graph of ethane itself. Ethane was first tested to determine its state as a radical or a molecule, since these species undergo different reactions. This simple
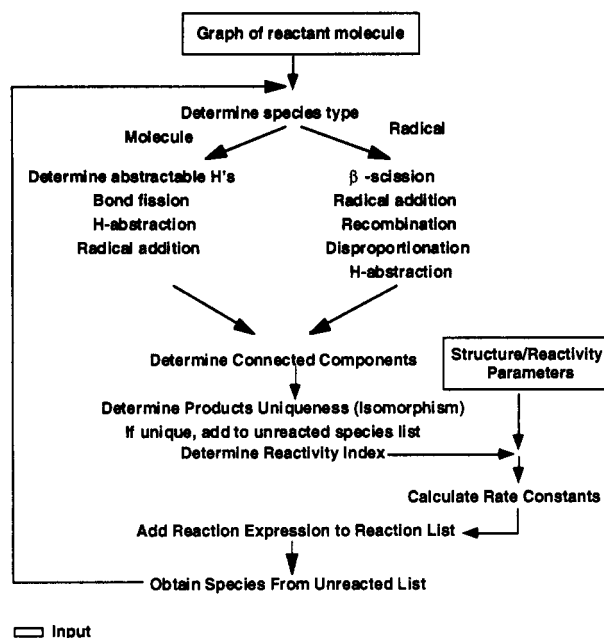
Figure 11. Logic of the algorithm for the generation of species and the assembly of the reaction mechanism.

species classification "filter" therefore reduced the number of reactions for which a given species was tested. Ethane was thus subjected to the rules for reaction of molecules.

The species reaction subroutine is a sequential test for the appropriate reaction sites for the set of reactions. A positive test requires that the species undergo reaction. In the pyrolysis mechanism generator, molecules are tested for bond fission, H-abstraction, and radical addition. For example, ethane was tested for bond fission by determining if there was a single carbon–carbon bond, the current rule or criterion for alkane homolysis. After this single carbon–carbon bond was located, the reactive sites were specified, and their intrasite connectivity information and chemical environment were placed into a BE matrix for implementation of reaction. As the initial and sole species thus far into the mechanism generation, ethane would fail the test for H-abstraction and radical addition, since there are no coreactant radicals in the species lists at this time. These reactions are developed further into the reaction generation process, ultimately through the application of the first bond fission reaction.

The fission reaction was carried out by addition of the reactant (ethane) submatrix and the reaction (fission) matrix as detailed above. A new BE submatrix describing the connectivity and electron configuration of the reactive sites of the product molecule was obtained. This is shown in Figure 7 for bond homolysis of ethane to two methyl radicals. The connectivity of these reactive sites to atoms unaltered by reaction has been maintained. The product molecule(s) is(are) reassembled by combining the adjacency information of the reacted and unreacted atoms. This entire array of atoms was then subjected to a connectivity check to determine the number of products formed. The isomorphism algorithm was then used to determine the uniqueness of the product molecules with respect to all of the previously encountered species. When a product was unique, it was placed at the end of the unreacted components list.

The transformation of the reactant graph of ethane into the product graphs of two methyl radicals defines a single reaction. The reaction was completely realized when the appropriate parameter(s) for the type of reaction carried out is associated with the reactant and product species.

The implied reaction is added to the unique reaction list. The division of species according to reacted and unreacted components guarantees that the reaction is unique. Reaction path degeneracy is accounted for by the repetition of reactions that involve the same species but differ by the specific reacting atoms.

The current molecule or radical subjected to the reaction tests (ethane, in this example) is now finished as a reactive component. It is removed from the unreacted component list and placed in the appropriate molecule or radical list so that subsequently generated species can participate with it in bimolecular reactions. This is why the initial failure of ethane to participate in H-abstraction reactions, for example, is not a flaw. All combinations of species for a given reaction type will ultimately be tested.

Subsequent passes through the generation algorithm follow the same logic and treat all species in a systematic manner to ensure all possible combinations are generated. The algorithm continues so long as it finds a new species in the unreacted components list. Thus, the methyl radical was tested for the pyrolysis radical reactions of recombination, radical addition, disproportionation, β-scission, and hydrogen abstraction. Note that radical addition and hydrogen abstraction are in both the molecule and radical reaction test lists since they are bimolecular reactions involving a molecule and a radical.

The methyl radical abstracts hydrogen from every species in the current molecule list that has abstractable hydrogen. At this stage of the mechanism generation only ethane contained abstractable hydrogen. Methane and ethyl radical were generated as new products. Thus, they were both placed in the unreacted components list. The methyl radical could not undergo radical addition at this point, since there were no species in the molecule list that possess unsaturated bonds. The subsequent generation of an unsaturated species will allow for methyl radical addition, however. This will occur as a reaction of a generated olefin with a methyl radical in the stored radical list. Methyl radical cannot undergo β-scission, since it does not possess an atom in a β position to the radical center. It also does not have the necessary reactive sites to undergo disproportionation. Methyl recombines with every species in the current radical list as well as itself. The current radical list is empty, and therefore, the only recombination reaction carried out is self-termination to form ethane. Ethane is a previously detected species, and as a result, is not appended to the unreacted components list. Note that methyl radical does not recombine with the newly formed ethyl radical, an entry in the unreacted components list. This reaction will be carried out when ethyl radical is the current species and finds methyl radical as an element in the radical list.

The generation of the mechanism is halted when the unreacted components list is empty. For pyrolysis chemistry where radical addition reactions are important, this will require the user to set the criterion for placing a unique component in the unreacted species list. The current condition for allowing a component to be placed in the unreacted component list is a maximum carbon count. In this way, large molecules generated as higher rank products are not allowed to undergo fission and generate additional radicals, and the algorithm terminates. The expected reaction conditions, where variables such as degree of conversion and temperature determine the relative contribution of secondary and higher reactions, guide the user's selection of the appropriate stopping criterion.

Code Development: Specification of Quantitative Mechanism Kinetic Parameters

The advantage of computer-generated reactions is the enumeration of thousands of elementary reactions and the strict accounting of hundreds of species without tedious manual effort that is prone to errors. A set of reactions has no quantitative value, however, without the associated rate parameters. This motivated our interest in a reaction-generation algorithm that would assign the appropriate rate constant as the reaction is revealed. This would also allow uninterrupted interaction of the mechanism generator with the method used to solve the balance equations that describe the mechanism.

Several paths forward for the specification of the quantitative mechanism parameters were apparent. At the lowest level, it is possible to assign a symbolic rate constant to each reaction during mechanism generation and leave the burden of supplying quantitative values for each with the user. However, the richness of chemical structure in the reaction-generation code invites *in situ* assignment of chemically significant rate parameters. For example, Clymans and Froment (1984) lumped pyrolysis rate constants in terms of reaction type and carbon type or radical type. In this way, the rate constant for abstraction of hydrogen to form a primary radical was different from that for abstraction of hydrogen to form a secondary radical.

We sought an even finer-scale form of lumping that would provide rate constants *in situ*. Our approach was to categorize the elementary reactions in terms of reaction families, the reactivity within each reaction family being characterized in terms of a linear free energy relationship (Nigam and Klein, 1993). Each sterically similar reaction family is defined by a single Arrhenius A factor, and its linear free energy relationship (LFER) relates the change in the activation energy to a reactivity index. The reactivity index can be a property of one of the molecules involved in the reaction, e.g., ΔH°_f of a carbenium ion, or a property of the reaction itself, e.g., ΔH°_R. In this manner, the full range of reactions within a reaction family can be parametrized in terms of only a slope and intercept.

The semiempirical relationship for exothermic reactions attributed to Polanyi (Evans and Polanyi, 1938) was used to relate the activation energy for elementary radical reactions to the heat of reaction.

$$E^* = E^*_0 + \alpha(\Delta H_R) \tag{1}$$

Semenov (1958) applied a Polanyi relation to radical-molecule H-abstraction reactions and determined the slope to be 0.25 for exothermic reactions and 0.75 for endothermic reactions and the intercept to be 11.5. LaMarca (1992) determined LFER parameters of E_0 =15.8 and α = 0.58 for β-scission of simple (propyl–pentyl) alkyl radicals.

Equation 2 shows that the use of linear free energy relationships transforms the problem of estimating reaction rate constants into the problem of estimating the reactivity index for each reaction. The reactivity indices

$$\log k_i = a + b \, RI_i \tag{2}$$

cannot be input *a priori*, since the species and reactions encountered are as of yet unknown. This is the power of the graph representation of species. Species connectivity is also the first step in estimating a wide range of species' physical properties. This can be accomplished in a variety of ways. For thermodynamic (enthalpic) data, linkage

with a group additivity database (Benson, 1976; Ritter and Bozzelli, 1991) would suffice. However, the appropriate reactivity index might require estimation of other species properties, such as electron affinity, electron density, proton affinity, etc. (Neurock, 1992). This suggests the value of semiempirical quantum mechanical software packages such as MOPAC (Stewart, 1990). This would allow direct calculation of a myriad of properties, e.g., electron affinity, electron density, bond order, and heat of formation, given the graph of a species. We thus sought to interface MOPAC and our reaction generator to accomplish "on-the-fly" calculation of properties. Clearly this broadens the concept of computer generated reaction mechanisms considerably.

Graph Conditioning for Use in Computational Quantum Chemistry Programs. The input to the semiempirical quantum calculation software packages requires an initial estimate of the three-dimensional conformation of the species as an additional level of detail. The two-dimensional graph of the species and the associated connectivity map provided the starting point for our representation of the 3D geometry using internal coordinates. The internal coordinates were defined so that, for any one atom i, there was an interatomic distance from the already defined parent atom j; there was an interatomic angle among atoms i and j and an already defined angle atom k; and there was a torsional angle involving atoms i, j, k, and an already defined dihedral atom l.

Optimization of the geometry corresponding to the initial estimate refines the structure to a minimum-energy configuration. For tree structures, the initial internal coordinate specification of the molecule was accomplished in a systematic fashion. A once-through translation of the 2D molecule or radical graph into its internal coordinate representation of 3D molecular structure was possible after parent and child atom relationships were established using a depth-first search. The atoms were then visited according to a breadth-first search to guarantee that all atoms attached to a given parent atom had the same parent atom, angle atom, and dihedral atom. Numerical values for bond length were specified according to the identity of the two bonded atoms and the associated bond order. Values for the bond angle were specified by the hybridization of the parent. To achieve the minimum-energy configuration, less favorable, sterically hindered geometries that would result in a local minimum were avoided through specifying carbon children first. This permitted accounting for the differences in energy of various conformations which could be as great as 4–5 kcal/mol. For example, the anti configuration of n-butane is 4.5 kcal/mol lower in energy than the eclipsed conformation (Streitweiser and Heathcock, 1985).

Keywords dependent upon the species type and reactivity index desired guide the semiempirical quantum calculation and were written as the first line of a buffer file. As the atoms of the species were traversed in the breadth-first search, the line of the MOPAC input file corresponding to each atom was written to the same buffer file in the appropriate format. When the breadth-first search was complete and the entire species was specified in internal coordinates, the buffer file was renamed as the MOPAC input file, and the species it represented was submitted for calculation of the electronic configuration via a semiempirical molecular orbital calculation.

Upon achievement of self-consistency, the MOPAC-generated output file was parsed for the appropriate species' contribution to the reactivity index. The value was stored as an adjacency structure instance variable for

a)
```
%% --- Component Definitions
1 : "(C(H3)C(H3))" ;
4 : "C(H4)" ;
5 : "(C(H2)C(H2))" ;
11 : "(H2)" ;
7 : "C(C(H3)C(H3)H2)" ;
8 : "(C(C(H3)H2)C(C(H3)H2))" ;
9 : "C(C(H3)C.(H2)H2)" ;
10 : "(C(C(H3)H2)C(C.(H2)H2))" ;
2 : "C.(H3)" ;
3 : "(C(H3)C.(H2))" ;
6 : "H." ;
```

c)
```
%% --- Component Definitions
1 : "(C(H3)C(H3))" = -18.131790 ;
4 : "C(H4)" = -13.014540 ;
5 : "(C(H2)C(H2))" = 16.629660 ;
11 : "(H2)" = -13.391140 ;
7 : "C(C(H3)C(H3)H2)" = -23.622160 ;
8 : "(C(C(H3)H2)C(C(H3)H2))" = -29.051069 ;
9 : "C(C(H3)C.(H2)H2)" = 9.407100 ;
10 : "(C(C(H3)H2)C(C.(H2)H2))" = 3.962060 ;
2 : "C.(H3)" = 27.964399 ;
3 : "(C(H3)C.(H2))" = 14.584590 ;
6 : "H." = 52.102001 ;
```

b)
```
/* (C(H3)C(H3)) -> C.(H3) + C.(H3) */
species1 -> k_lfer( A_init,E0_init,alpha_init, 74.06 ) species2 + species2 ;
/* (C(H3)C(H3)) + C.(H3) -> (C(H3)C.(H2)) + C(H4) */
species1 + species2 -> k_lfer( A_habs,E0_habs,alpha_habs, -8.26 ) species3 + species4 ;
/* (C(H3)C(H3)) + C.(H3) -> (C(H3)C.(H2)) + C(H4) */
species1 + species2 -> k_lfer( A_habs,E0_habs,alpha_habs, -8.26 ) species3 + species4 ;
/* (C(H3)C(H3)) + C.(H3) -> (C(H3)C.(H2)) + C(H4) */
species1 + species2 -> k_lfer( A_habs,E0_habs,alpha_habs, -8.26 ) species3 + species4 ;
/* (C(H3)C(H3)) + C.(H3) -> (C(H3)C.(H2)) + C(H4) */
species1 + species2 -> k_lfer( A_habs,E0_habs,alpha_habs, -8.26 ) species3 + species4 ;
/* (C(H3)C(H3)) + C.(H3) -> (C(H3)C.(H2)) + C(H4) */
species1 + species2 -> k_lfer( A_habs,E0_habs,alpha_habs, -8.26 ) species3 + species4 ;
/* (C(H3)C(H3)) + C.(H3) -> (C(H3)C.(H2)) + C(H4) */
species1 + species2 -> k_lfer( A_habs,E0_habs,alpha_habs, -8.26 ) species3 + species4 ;
/* C.(H3) + C.(H3) -> (C(H3)C(H3)) */
species2 + species2 -> k_lfer( A_recomb,E0_recomb,alpha_recomb, -74.06 ) species1 ;
/* (C(H3)C.(H2)) -> (C(H2)C(H2)) + H. */
species3 -> k_lfer( A_bscission,E0_bscission,alpha_bscission, 54.15 ) species5 + species6 ;
/* (C(H3)C.(H2)) -> (C(H2)C(H2)) + H. */
species3 -> k_lfer( A_bscission,E0_bscission,alpha_bscission, 54.15 ) species5 + species6 ;
/* (C(H3)C.(H2)) -> (C(H2)C(H2)) + H. */
species3 -> k_lfer( A_bscission,E0_bscission,alpha_bscission, 54.15 ) species5 + species6 ;
/* (C(H3)C(H3)) + (C(H3)C.(H2)) -> (C(H3)C.(H2)) + (C(H3)C(H3)) */
species1 + species3 -> k_lfer( A_habs,E0_habs,alpha_habs, 0.00 ) species3 + species1 ;
/* (C(H3)C(H3)) + (C(H3)C.(H2)) -> (C(H3)C.(H2)) + (C(H3)C(H3)) */
species1 + species3 -> k_lfer( A_habs,E0_habs,alpha_habs, 0.00 ) species3 + species1 ;
/* (C(H3)C(H3)) + (C(H3)C.(H2)) -> (C(H3)C.(H2)) + (C(H3)C(H3)) */
species1 + species3 -> k_lfer( A_habs,E0_habs,alpha_habs, 0.00 ) species3 + species1 ;
/* (C(H3)C(H3)) + (C(H3)C.(H2)) -> (C(H3)C.(H2)) + (C(H3)C(H3)) */
species1 + species3 -> k_lfer( A_habs,E0_habs,alpha_habs, 0.00 ) species3 + species1 ;
/* (C(H3)C(H3)) + (C(H3)C.(H2)) -> (C(H3)C.(H2)) + (C(H3)C(H3)) */
species1 + species3 -> k_lfer( A_habs,E0_habs,alpha_habs, 0.00 ) species3 + species1 ;
/* (C(H3)C(H3)) + (C(H3)C.(H2)) -> (C(H3)C.(H2)) + (C(H3)C(H3)) */
species1 + species3 -> k_lfer( A_habs,E0_habs,alpha_habs, 0.00 ) species3 + species1 ;
```

Figure 12. (a) Generated unique reaction mechanism species. (b) Sample output for pyrolysis of ethane constrained by initiation of species with no more than two carbon atoms. (c) Component definitions with their associated heat of formation values.

future incorporation into the linear free energy relationships. For example, if the heat of reaction is the relevant reactivity index, each of the stored species' heat of formation values obtained from MOPAC is accessed to calculate the overall heat of reaction. The reactivity indices combined with the linear free energy parameters provided as input completely specified the quantitative mechanism parameters.

Generation of Reaction Pathways from Species and Reaction Elucidation: Assembly of the Reaction Mechanism

The generation of a reaction pathway is the concerted generation and identification of product species and the retention of reactant information. The reactant is stored in memory as its possible reactions are elucidated. Successful bimolecular reactions require that the coreactant be stored in memory as well. In the algorithm, the determination of the unique, unconnected product components is followed by addition of the reaction pathway to the growing reaction list array. The reaction is written in text format, such that the user can inspect the reactions generated. The reaction is also written in terms of the species and their unique numbers, assigned after determination of species isomorphism. A label for the appropriate rate constant and the reactivity index are recorded in the reaction list as well. The final reaction list file is the input for OdeGen, the ordinary differential equation generator.

A sample output file for pyrolysis of ethane is shown in Figure 12a-c. Reaction of only species with two or fewer carbon atoms was considered in this example. In Figure 12a, the unique components are printed using their unique, canonical string representation and their identifying

a)
```
%% --- Component Definitions
1 : 0 : "[C[CCH2]C[CCH2]C[CCH2]C[CCH2]C[CCH2]C[CCH2]]" ;
2 : 0 : "(C(C(C.(H2)H2)H2)C(C(C.(H2)H2)H2))" ;
7 : 0 : "[C[CC.H2]C[CCH2]C[CCH2]C[CCH2]C[CC.H2]C.[CCH]]" ;
8 : 0 : "(C(C(C(H3)H2)H2)C(C(C.(H2)H2)H2))" ;
```

b)
```
/*    [C[CCH2]C[CCH2]C[CCH2]C[CCH2]C[CCH2]C[CCH2]]    ->
(C(C(C.(H2)H2)H2)C(C(C.(H2)H2)H2)) */
species1 -> k(init) species2 ;

/*    [C[CCH2]C[CCH2]C[CCH2]C[CCH2]C[CCH2]C[CCH2]]    +
(C(C(C.(H2)H2)H2)C(C(C.(H2)H2)H2))    ->
{C[CC.H2]C[CCH2]C[CCH2]C[CCH2]C[CC.H2]C.[CCH]}    +
(C(C(C(H3)H2)H2)C(C(C.(H2)H2)H2)) */
species1 + species2 -> k(habs) species7 + species8 ;
```

Figure 13. (a) Representative species for cyclohexane pyrolysis. (b) Examples of bond fission and H-abstraction reactions for cyclohexane pyrolysis.

species number. The presence of radical species is apparent by the period after the letter designation for the atom on which the unpaired electron resides. The chemical species represented by the canonical string can be reconstructed by placing the first atom, in the case of a single root node, or the two atoms at the first level of parentheses, in the case of two adjacent root nodes, at the top of a tree. The atoms within the next level of parentheses following a given atom are that atom's children and are placed on the tree directly below that atom and connected to it. The process is continued until the innermost set of parentheses are reached and all the species' atoms are specified.

Figure 12b shows a truncated list of representative reactions for this example with the rate constants represented in a form that is suitable for using linear free energy relationships and the Arrhenius form of the rate coefficient. The LFER and Arrhenius parameters were reaction rule dependent. For example, the rate coefficient for H-abstraction was written as "k_lfer(A_habs, E0_habs, alpha_habs, $\Delta H°_R$)". The species that were involved in each reaction were written in terms of the unique species numbers. Comment lines were included that show the reaction in a more traditional chemical form. Figure 12c shows the revised component definitions where the heat of formation of each species was included.

Case Study 2. Cyclohexane Pyrolysis

With two exceptions, direct application of the otherwise unaltered algorithm described for ethane pyrolysis created the cyclohexane pyrolysis mechanism. The cyclic species in the cyclohexane pyrolysis mechanisms placed special demands on the algorithms for determining isomorphism and for the translation of the two-dimensional graph into a three-dimensional structure for MOPAC calculations. These developments are the subject of a forthcoming algorithms-oriented manuscript.

The reaction rules and the pyrolysis chemistry reaction operators are completely general because of the insensitivity of the adjacency structure to the presence of cycles and the invariance of the reaction matrices. Thus, cyclohexane pyrolysis was modeled using the reaction rules of fission (to a biradical), H-abstraction, β-scission, radical recombination, disproportionation, and radical addition shown in Figure 6. Figure 13a lists the products generated from the representative bond fission and H-abstraction reactions of Figure 13b using the graph of cyclohexane as input. Pyrolysis of cyclohexane revealed two types of species not encountered for ethane pyrolysis: biradicals and cyclic species. The fission of any of the carbon–carbon bonds in cyclohexane resulted in the biradical indicated by the two C.'s in Figure 13b. The algorithm provided a strict accounting of the number of electrons surrounding each atom and therefore identified the two carbons with

their unpaired electrons as radical centers. Hydrogen abstraction from cyclohexane afforded a cyclohexyl radical.

The encoding of cycles necessitated the definition of additional symbols to obtain unique representations. The braces of Figure 13 enclose the atoms of the cycle and the atoms to which they are attached. Square brackets enclose the atoms to which a preceding cycle member is bonded. For example, cyclohexane is compactly represented as {C-[CCH2]C[CCH2]C[CCH2]C[CCH2]C[CCH2]C[C-CH2]}, and cyclohexyl radical is designated by {C-[CC.H2]C[CCH2]C[CCH2]C[CC.H2]C.[C-CH]}, where the final ring carbon has one fewer atom bonded to it, consistent with its identification as the radical center. A cycle atom is prefixed with a hyphen if it is bonded to an atom that is not of degree 1. For the sake of illustration, the string (C(C(C(H3)H2)H2)C(C(C(H2-{-C[CCCH]C[CCH2]C[CCH2]C[CCH2]C[CCH2]C[C-CH2]})H2)H2)) would represent n-hexylcyclohexane.

Representative Results: The Influence of User Specifications on the Ethane Pyrolysis Mechanism

The user-specified carbon-count convergence criterion impacted the number of reactions included and therefore the number of species and reactions generated. A wide range of ouputs was found. Allowing reaction of only species with two or fewer carbon atoms generated 11 species and 55 equations. When this termination criterion was increased to three carbon atoms, 99 species and 611 equations were created.

The numbers of species and equations are plotted as a function of the carbon-number stopping criterion in Figure 14a. Their increases were both essentially exponential with increasing carbon number for carbon numbers up to four. However, these quantities were independent of the performance of MOPAC calculations.

The performance of the mechanism generation code is shown in Figure 14b as a plot of required CPU time versus the carbon-number criterion for convergence. These results are shown for parametric cases with and without the use of MOPAC for the calculation of the heat of formation of each generated species. Without calling MOPAC, the CPU time for pyrolysis including the initiation of species with up to two carbons was 1 s and increased to 162 s when species with four carbon atoms were allowed to initiate. The CPU requirement increased by essentially an order of magnitude with the addition of the MOPAC calculations. However, fractional resource allocation for MOPAC calculations should decrease as the carbon number increases because the algorithm will spend a significant amount of time generating redundant species.

The species and equation numbers and even CPU demand were a strong function of the stopping criterion but insensitive to the starting alkane reactant. This is shown in Figures 14, where the points for reactant alkanes of chain length up to four lie directly on top of one another for a given carbon-number stopping criterion. The mechanisms generated are entirely equivalent. This impacts the generation of reaction mechanisms for a homologous series of reactants and for mixtures of these reactants. For example, the reaction mechanism for a binary mixture of ethane and butane that does not allow initiation of species with carbon numbers greater than four is simply the same as the mechanism for butane alone, since ethane is formed as a primary product of butane pyrolysis.

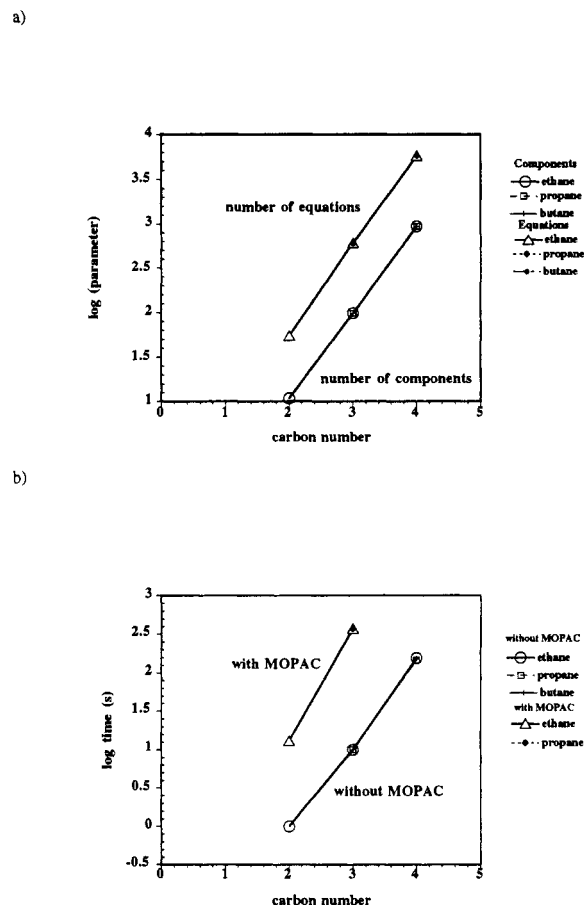Chemical and kinetic significance can also be used to

Figure 14. (a) Number of equations and species for ethane pyrolysis as a function of carbon number. (b) CPU time as a function of carbon-number stopping criterion for pyrolysis of ethane parametric in inclusion of MOPAC calculations.

guide the choice of stopping criterion. Frequently, the H-abstraction and β-scission reactions will be orders of magnitude faster than the bond fission and recombination reactions; i.e., the kinetic chain length is long. A high selectivity to ethylene and hydrogen and low yield of other minor products result during ethane pyrolysis, for example. It is therefore relevant to inspect the mechanism generation scheme using the allowance of only primary reactions as the stopping criterion. This was implemented by allowing radical species but not molecules to be placed in the unreacted components list and imposing a carbon-number stopping criterion of two. Radical addition reactions were stifled, for example. Using this criterion, nine species and 37 reactions were revealed. The resultant reaction mechanism is depicted in Figure 3.

In principle, there are no restrictions on the number of equations and species in the reaction mechanism that can be generated. However, hardware can impose practical limitations. The memory required by the computer to store all of the information generated as the mechanism grows also increases as the stopping criterion is increased. The memory in megabytes is $O(10^{n-1})$, where n is the carbon number of the stopping criterion. For large systems and incorporation of the full set of reaction types, a prohibitively large quantity of memory is required even when the fixed disk space is utilized as virtual memory. This emphasizes the importance of the careful choice of stopping criterion and user-defined reaction rules, to be examined in a following publication, so that unimportant reaction paths are not pursued.

Quantitative Computer-Generated Reaction Predictions

The reaction mechanism is neither useful nor subject to the ultimate confrontation with experimental data until its balance equations are reduced to solvable code. We have developed a grammar that translates the textual representation of the reaction mechanism of Figure 12b into the system of ordinary differential equations that describe it. The grammar recognizes individual components with numbered species names and interprets reactants as those species listed on the left hand side of the reaction arrow and products as those on the right hand side. The rate constant is directly after the reaction arrow. It can be defined in terms of a numeric or symbolic rate constant array index, or it can be implicit within the Arrhenius/LFER form through the use of grammar-defined macros. This allows for flexible rate constant specification. Thus, the output of the mechanism generator is in the format suitable for use as direct input for differential equation formulation.

Conclusions

The concept of computer-generated reaction modeling was broadened through the extension of work presented in the literature in both breadth and depth. The literature provided both guidance and specific modules. The concept of a molecular graph and the related matrix representations and operations that effect chemical reaction provided the foundation for automatic reaction mechanism generation. However, many of the modules were burdened by algorithm- and problem-specific limitations.

Special attention was paid to the systematic generation and identification of species, the specification of quantitative mechanism parameters by accessing "on-the-fly" quantum chemistry calculations, and the integration of species and reaction generation to yield quantitative reaction network predictions. The isomorphism algorithm allowed both tree structures and ring-containing compounds to be handled equally well. The translation of a two-dimensional species graph into a three-dimensional species enabled the calculation of species' properties as the species was generated in the reaction mechanism. A grammar was developed which allowed translation of reactions in text format to ordinary differential equations. All of these modules form an integrated system for computer-generated reaction predictions.

The integrated system of computer modules allows an iterative approach to kinetic analysis. The end-product model can be augmented, reduced, or altered to reflect additional or inconsistent chemical information obtained experimentally. The chemical and physical significance of the parameters estimated for the previously postulated mechanism also directs the alterations to the analysis. The facilitation of making reaction predictions by using the computer allows many more mechanistic scenarios to be explored and suggests ones that are currently unexplored.

Literature Cited

Benson, S. W. *Thermochemical Kinetics*; Wiley-Interscience: New York, 1976.

Biggs, N. L.; Lloyd, E. K.; Wilson, R. J. *Graph Theory 1736–1936*; Clarendon: Oxford, U.K., 1976.

Bonchev, D., Rouvray, D. H., Eds.; *Chemical Graph Theory Introduction and Fundamentals*; Gordon & Breach Science: New York, 1991.

Chevalier, C.; Warnatz, J.; Melenk, H. Automatic Generation of Reaction Mechanisms for Description of Oxidation of Higher Hydrocarbons. *Ber. Bunsen-Ges. Phys. Chem.* **1990**, *94* (11), 1362–1367

Clymans, P. J.; Froment, G. F. Computer-Generation of Reaction Paths and Rate Equations in the Thermal Cracking of Normal and Branched Paraffins. *Comput. Chem. Eng.* **1984**, *8* (2), 137–142.

Corneil, D. G.; Gotlieb, C. C. An Efficient Algorithm for Graph Isomorphism. *J. Assoc. Comput. Mach.* **1970**, *17* (1), 51–64.

DiMaio, F. P.; Lignola, P. G. KING, a KInetic Network Generator. *Chem. Eng. Sci.* **1992**, *47* (9–11), 2713–2718.

Evans, M. G.; Polanyi, M. Inertia and Driving Force of Chemical Reactions. *Trans. Faraday Soc.* **1938**, *34*, 11–29.

Foulds, L. R. *Graph Theory Applications*; Springer-Verlag: New York, 1992.

Greenberg, M. D. *Advanced Engineering Mathematics*; Prentice-Hall: Englewood Cliffs, NJ, 1988.

Hillewaert, L. P.; Dierickx, J. L.; Froment, G. F. Computer Generation of Reaction Schemes and Rate Equations for Thermal Cracking. *AIChE J.* **1988**, *34* (1), 17–24.

Hopcroft, J. E.; Tarjan, R. E. Isomorphism of Planar Graphs. In *Complexity of Computer Computations*; Miller, R. E., Thatcher, J. W., Bohlinger, J. D., Eds.; Plenum: New York, 1972.

LaMarca, C. Kinetic Coupling in Multicomponent Pyrolysis Systems. Ph.D. Dissertation, University of Delaware, 1992.

Neurock, M. A Computational Chemical Reaction Engineering Analysis of Complex Heavy Hydrocarbon Reaction Systems. Ph.D. Dissertation, University of Delaware, 1992.

Neurock, M. N.; Nigam, A.; Libanati, C.; Klein, M. T. Monte Carlo Simulation of Complex Reaction Systems: Molecular Structure and Reactivity in Modelling Heavy Oils. *Chem. Eng. Sci.* **1990**, *45* (8), 2083–2088.

Nigam, A.; Klein, M. T. A Mechanism-Oriented Lumping Strategy for Heavy Hydrocarbon Pyrolysis: Imposition of Quantitative Structure–Reactivity Relationships for Pure Components. *Ind. Eng. Chem. Res.* **1993**, *32* (7), 1297–1303.

Quann, R. J.; Jaffe, S. B. Structure-Oriented Lumping. Describing the Chemistry of Complex Hydrocarbon Mixtures. *Ind. Eng. Chem. Res.* **1992**, *31* (11), 2483–2497.

Ritter, E. R.; Bozzelli, J. W. THERM. Thermodynamic Property Estimation for Gas Phase Radicals and Molecules. *Int. J. Chem. Kinet.* **1991**, *23* (9), 767–778.

Semenov, N. N. *Some Problems in Chemical Kinetics and Reactivity*; Princeton University: Princeton, NJ, 1958; Vol. 2.

Stewart, J. J. P. *MOPAC Reference Manual and Release Notes*, 6th ed.; Frank J. Seiler Research Laboratory, United States Air Force Academy, CO, October 1990.

Streitweiser, A., Jr.; Heathcock, C. H. *Introduction to Organic Chemistry*, 3rd ed.; Macmillan: New York, 1985.

Tarjan, R. Depth-First Search and Linear Graph Algorithms. *SIAM J. Comput.* **1972**, *1* (2), 146–160.

Tarjan, R. E. Graph Algorithms in Chemical Computation. *Algorithms for Chemical Computations*; Christoffersen, R. E., Ed.; American Chemical Society: Washington, DC, 1977.

Ugi, I.; Bauer, J.; Brandt, J.; Freidrich, J.; Gasteiger, J.; Jochum, C.; Schubert, W. New Applications of Computers in Chemistry. *Angew. Chem., Int. Ed. Engl.* **1979**, *18*, 111–123.