J. Phys. Chem. A **1997,** *101,* 3731−3740

Rate-Based Construction of Kinetic Models for Complex Systems

Roberta G. Susnow, Anthony M. Dean, William H. Green,* and P. Peczak

Exxon Research & Engineering Company, Route 22 East, Annandale, New Jersey 08801

Linda J. Broadbelt*

Department of Chemical Engineering, Northwestern University, Evanston, Illinois 60208

Received: November 12, 1996; In Final Form: February 19, 1997⊗

A general-purpose rate-based algorithm for the construction of chemical kinetic models for systems with hundreds or thousands of reacting species is presented. The algorithm comprehensively works out the details of the chemistry implied by given reaction rate estimation rules, identifies the species and reactions that are numerically significant, and solves the resulting system of differential equations to compute the concentrations of the significant species as a function of time. A key innovation is a definition and numerical test for the "completeness" of the kinetic scheme. This approach obviates the need to arbitrarily neglect certain species and reactions in order to keep reaction schemes small enough to be manageable and allows chemical kinetic modelers to focus on the chemistry rather than on the computational details. Examples of hydrocarbon pyrolysis and combustion applications are presented, where the computer evaluates the importance of nearly 100 000 reactions in the process of identifying the few hundred species that are kinetically significant. The new algorithm, given reliable rate estimation rules, provides a framework for systematically constructing kinetic schemes including all of the numerically significant species, even for systems involving so many reactions that they could not be handled manually.

I. Introduction

The demand for detailed kinetic models is increasing as economic needs and environmental legislation require more explicit accounting of the composition of product streams. Kinetic models not only provide a basis for reactor design but also process improvement through facile exploration of diverse processing scenarios. They also provide a starting point for drawing inferences about the fundamental underlying, controlling chemistry. Kinetic models lie at the heart of chemistry, since they allow chemists to make testable predictions for systems of real interest based on measurements made in highly simplified laboratory conditions. The award of the 1995 Nobel Prize in Chemistry to Crutzen, Sherwood, and Molina for developing and using an accurate model of ozone layer depletion kinetics testifies to the scientific importance of these sorts of predictive models.

Models for important industrial processes are so valuable that companies often invest millions of dollars in their development. Governments routinely make major public policy decisions based on kinetic models, for example, of the ozone layer and urban airsheds.[1] Models of significant complexity represent substantial investments of time and manpower, since historically they have required tedious manual construction.[2,3] The advent of tools for computer construction of kinetic models has made them more flexible and dramatically reduced the time for their development. Computer-generated models become even more valuable as competitive and regulatory pressures increase while computing power is readily available.

A major problem in constructing a chemical kinetic model is the very large, sometimes infinite, number of possible reactions, products, and reaction intermediates involved. The number of species usually grows exponentially with the number of reaction steps one takes from the initial reactants. Several approaches are taken in an attempt to reduce the size of the mechanism of interest. One conventional approach involves the engineering technique of lumping product and reactant molecules into convenient pseudospecies and connecting them through a reaction scheme by parameter estimation.[4,5] Since these types of models are generated through data fitting, there is little transferrability or extrapolation to related systems. Another approach, often taken by the kineticist, is to make approximations as to which species and reactions are significant in an attempt to shorten the kinetic scheme. However, these approximations are not easily tested, and important steps in the mechanism may be left out. In fact, the literature is full of examples where important reactions and species were ignored.[6] Techniques are needed for including the important species and reactions, while excluding the unimportant ones. Complex reacting systems such as hydrocarbon pyrolysis and oxidation chemistry can consist of thousands of kinetically significant species. Manually assembling the chemistry into a kinetic model is formidable, and the only practical way to construct and use large models is on the computer.

Considerable activity in recent years has focused on the development of algorithms to computationally generate reaction mechanisms.[7−19] We are building on the computer-generated reaction mechanism NetGen program of Broadbelt, Stark, and Klein.[17−19] The original version of NetGen contains a species rank-based criterion for the rational halt of the generating mechanism so that an infinite reaction network does not result. As with other published termination approaches, NetGen's termination criteria have the potential for missing important species and may include many insignificant species. Because we would like to develop accurate models for complicated processes of technological importance, we need a near-optimal reaction scheme generator. This generator must include all the important species and should leave out the unimportant ones. In some cases, the number of chemically significant species (the number of differential equations required) may be $O(10^4)$, and solving additional unnecessary kinetic equations may overtax

---

⊗ Abstract published in *Advance ACS Abstracts,* May 1, 1997.

S1089-5639(96)03769-3 CCC: $14.00

our computer hardware. In all cases, it is desirable to keep the kinetic scheme as small as feasible to improve our chances of understanding the system. Also, a small scheme is desirable for coupling the kinetic equations with computationally demanding tasks such as fluid-flow modeling and process optimization. Therefore, the challenge is to incorporate strategies for keeping the reaction scheme as small as possible into an algorithm capable of generating a comprehensive mechanism.

Our solution to the mechanism size and termination problem is to modify NetGen to use rate and concentration information in its reaction scheme generation algorithm. In this algorithm we add the most important species and reactions first, guided by a user-specified precision level, which sets the boundary between "significant" and "insignificant" species. This precision criterion gives the user the flexibility to decide how detailed and, therefore, how large a mechanism is required. In some cases one is interested in only the major species, so a small reaction scheme will suffice, while in other situations one might need to predict all of the detectable products.

The purpose of this paper is to explain and demonstrate the new rate-based algorithm that has been incorporated into NetGen and to present various tests of its accuracy and efficiency. A second improvement to NetGen that is discussed is the incorporation of improved thermochemistry by accessing the Structure & Properties database program developed by NIST.[20] Of course, a kinetic model can't be more accurate than the rate constants it employs. This truth is brought into sharper focus in this work, where the size of the kinetic scheme and the choice of species included in the scheme depend on the rate estimates. As explained within, we have used reasonable rate constant estimates so the models are in qualitative accord with expectations. However, a large amount of work remains to be done to formulate the more detailed reaction rate estimation rules that will be required to quantitatively match experiment.

II. Overview of NetGen

The NetGen program has previously been described in the literature.[16−19] The key elements of the current modified program include a reaction scheme generator, rate estimation rules for various types of reactions, an interface to a stiff differential equation solver (DASSL),[21] an interface to NIST's Structures & Properties program[20] and MOPAC[22] for thermochemistry estimates, routines that manipulate various representations of molecules (including routines that generate new molecules from reactions of existing molecules), and a set of algorithms for testing species' uniqueness based on planar graph theory.[17,23]

The core of the program is the reaction scheme generator that generates all of the possible reactions of a given species. At any given point during the process of constructing the reaction scheme, the partial reaction scheme consists of all the possible reactions involving *N* species as reactants; see Figure 1. A number, *M*, of other species have been identified as products of those "reacted" species, but the computer has not yet generated any reactions where these "unreacted" species are reactants. To build the reaction mechanism, NetGen selects one of the *M* "unreacted" species, call it Y, to become a reactant and generates all possible reactions involving Y, either alone or reacting with one of the *N* "reacted" species. These new reactions generate *m* new product species, previously unknown to the program, which are added to the list of "unreacted" species. The selected species Y is then moved from the "unreacted" species list to become the (*N*+1)th "reacted" species. At the next iteration, one of the *M*+*m*−1 species on the "unreacted" list is selected, and the process is repeated. In this sequential way, the number of reactions grows, forming the reaction mechanism. This process is illustrated for ethane pyrolysis in Figure 1. On going from part a to part b of Figure 1, CH_4 has become a "reacted" species, generating a few additional reactions. In the next iteration, going from part b to part c of Figure 1, C_2H_4 becomes a "reacted" species, generating several new reactions and new "unreacted" species.

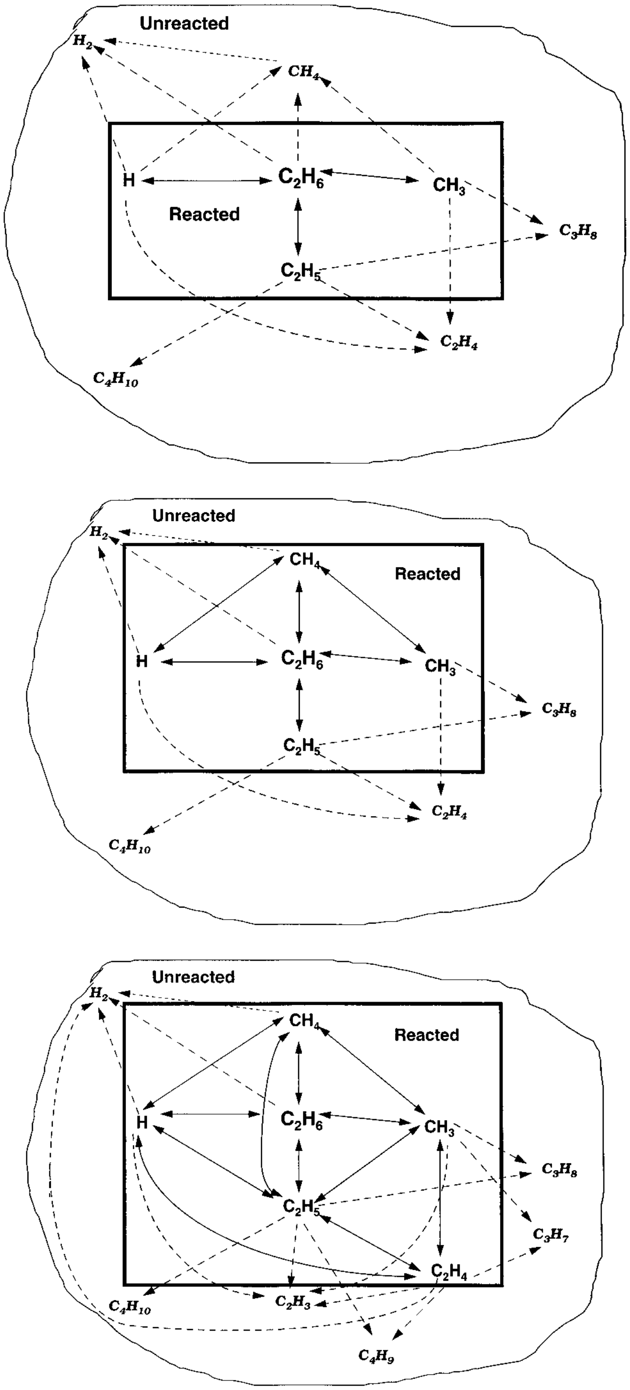

Figure 1. (a, top) Example of an initial pool for ethane pyrolysis, illustrating the set of "unreacted" and "reacted" species. (b, middle) Initially "unreacted" CH_4 from the initial pool of part a becomes "reacted", generating additional reactions and expanding the "reacted" species set. (c, bottom) Initially "unreacted" C_2H_4 from the initial pool of part a is the second species to become "reacted", generating a network of several new reactions and new "unreacted" species.

The set of "all possible reactions" in NetGen is determined by the reaction rate estimation rules provided by the user. For some reaction chemistries this iterative algorithm results in a manageable, compact reaction mechanism; that is, repetitive application of the possible reactions to all reacting species ultimately results in convergence. For example, if 2-methyl-1-hexene is the initial reactant, and the reaction rules include only bond migration, 1,2-methyl shifts, and hydrogenation, the reaction mechanism, consisting of all heptyl species, converges and results in a compact reaction mechanism. In contrast, the reaction rules appropriate for free radical chemistry usually allow unlimited molecular weight growth, resulting in reaction schemes that do not naturally converge, and extra termination criteria must be implemented to halt the generating mechanism so that an infinite network does not result.

In previous versions of NetGen, various restrictions were applied to limit the number of species and reactions generated. They included limitations on the rank of the products (the order in which they appear as products in the mechanism) and on the size of the radicals that could undergo reaction. Even with the implementation of these restrictive termination criteria, huge mechanisms resulted; in some cases more than 83% of the species were kinetically insignificant when compared with experimental data.[16] The limitations of these termination criteria motivated the present work aimed at developing a more natural ending point for the growing mechanism.

III. The New Rate-Based Scheme Generation Algorithm

Our new rate-based termination algorithm exploits the capability of NetGen to estimate rate constants as the mechanism is generated, allowing the partial mechanisms to be solved to estimate the concentration vs time profiles. These concentration vs time profiles are then used to evaluate those "unreacted" species, Figure 1, that are most significant. The process continues iteratively, as a growing reaction mechanism is alternatively generated and solved. Quantitative evaluation of the formation rates of a set of species during the mechanism-building process determines the next species allowed to undergo reaction. When the estimated production rate of all possible new byproducts has become small, relative to some criterion, R_{min}, we assume that the reaction scheme is complete.

The formation rates of the "unreacted" species are estimated by solving the differential equations corresponding to the current reaction scheme. Since the program does not include consumption reactions of the species it considers "unreacted", Figure 1, their rates of formation are all positive. The decision to allow an "unreacted" species to become "reacted" is based on its rate of formation. If the rate of formation of each of the "unreacted" species is less than a minimal rate, R_{min}, the current reaction scheme is considered adequately complete and the program terminates. Otherwise, the program selects the "unreacted" species with the largest formation rate, generates the reactions where the selected species is a reactant, and promotes it to be a "reacted" species in the scheme. The new products formed from this new reactant are added to the "unreacted" list and the program iterates.

The behavior of this algorithm is somewhat unpredictable in the early stages when the "reacted" species list may be missing kinetically important species, but in the later stages, when the scheme is substantially complete, each additional species does not change the overall kinetics much and the algorithm converges smoothly. The problem of finding all the kinetically significant species and reactions is in principle rather difficult, since there are an infinite number of species unknown to the computer that lie outside the ring of "unreacted" species in Figure 1. The reaction rates leading from the "unreacted" species to these "unknown" species are also unknown. The key reason why the rate-based algorithm can work is that all the possible products of the "reacted" species are included in the "unreacted" species list; the unknown species lying outside the ring of "unreacted" species can only be formed via one of the "unreacted"species. So if all the "unreacted" species are formed at negligible rates, we can be confident that all the "unknown" species are also formed at negligible rates. At convergence, all the reactions leading to "unreacted" species, the dashed arrows in Figure 1, are negligible, and we are justified in ignoring them and keeping only the reactions connecting the "reacted" species with each other in the final mechanism.

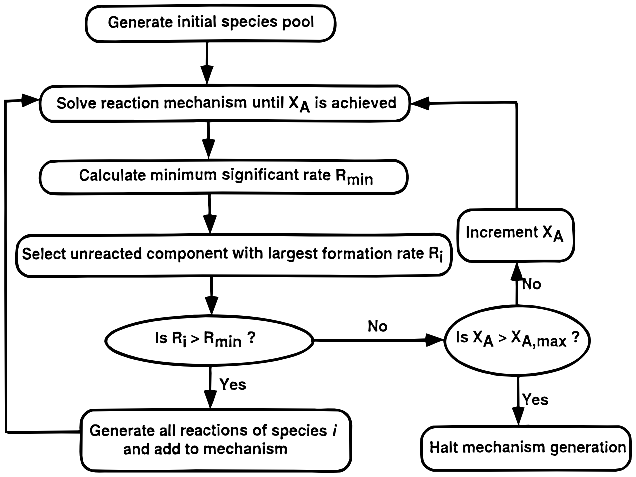

Figure 2. Flow diagram illustrating the iterative reaction scheme generation.

The externally defined rate criterion, R_{min}, is set to be a user-specified precision factor multiplied by a characteristic rate, R_{char}, of the system. Reactions very much slower than the characteristic rate are considered insignificant. In simple cases where the focus is on the uniform conversion of a single primary reactant, we take the characteristic rate

$$R_{\mathrm{char}} = \text{(amount of reactant converted)}/\text{(time it takes for conversion)}$$

The user must also specify the time scale of interest. In most cases this is better achieved by setting a desired conversion level than a particular time in seconds, since the time scale of the simulation may vary in unpredictable ways until the kinetically important species have been included in the reaction scheme.

IV. Details of the Rate-Based Algorithm

The rate-based strategy builds the mechanism iteratively, as a growing reaction mechanism is alternately generated and solved. The reaction scheme generator builds the reaction mechanism by appending reactions, and at any point in its generation, the mechanism can be solved. Rate constants are assigned to each of the reactions using the thermochemical or quantum mechanical data as described below, and differential equations are set up for each of the "reacted" species. These equations are solved using DASSL. After solution for the concentrations of the "reacted" species, algebraic evaluation of the formation rates for the set of "unreacted" species determines the next species allowed to undergo reaction.

The logic used to build the reaction mechanism iteratively is revealed in Figure 2. The first step is to generate an initial pool of species, the size of which is directed by the user. Either a maximum product rank, a maximum heavy atom count, or a maximum number of species in the initial pool is specified. At this point, the pool consists of "reacted" and "unreacted" species. An example of a minimal initial pool for ethane pyrolysis is shown in Figure 1a.

The mechanism describing the reactant/product relationships of the initial pool species is solved, incrementing time in regular intervals until a particular reactant conversion, X_A, is achieved at time τ. During solution, the maximum net rate of formation, $R_{i,\max}$, for each "unreacted" species i for the interval $0 < t < \tau$ is determined and stored. Note that the $R_{i,\max}$ values for the "unreacted" species are always positive since these species are not consumed by any reactions. The "unreacted" species with the largest $R_{i,\max}$ is chosen as the next species to become "reacted" if its value is greater or equal to a minimum significant rate $R_{\min}$. This species' reactions are added to the existing mechanism, generating a new mechanism that is solved starting from time zero.

The minimum rate $R_{\min}$ is derived from user-specified values and is governed by the kinetics of reactant conversion. The user specifies the desired percent conversion of the initial species, $X_{A,\max}$. Initially the conversion level for mechanism generation is set to a smaller value, X_A, and incremented as the mechanism gets larger. The logic behind this is that small mechanisms are expected to be accurate only for small conversions. The characteristic disappearance rate of the reactant is given by eq 1,

$$R_{\mathrm{char}} = [C_{A0} - C_A(\tau)]/\tau = X_A C_{A0}/\tau \qquad (1)$$

where C_{A0} is the initial concentration of the reactant, C_A is the concentration of the reactant at the particular conversion X_A, and τ is the time required to achieve this conversion. The user also specifies a precision level, related to the accuracy of the detection of minor species. The precision level is used to scale the reactant disappearance rate yielding $R_{\min}$, the minimum formation rate of a significant product:

$$R_{\min} = (\text{desired precision level}) \times R_{\mathrm{char}} \qquad (2)$$

$R_{\min}$ provides a relevant scale to which rates of formation of other species can be compared. Thus when $R_{i,\max} > R_{\min}$, the "unreacted" species i is added to the list of the significant, "reacted" species. Clearly, as the precision tightens, $R_{\min}$ decreases proportionally, and the number of species with formation rates large enough that they must be included in the mechanism increases. The mechanism generator iterates at a particular conversion level through the inner loop of the logic diagram in Figure 2 until no species have $R_{i,\max}$ values greater than the $R_{\min}$.

The user-specified maximum conversion, $X_{A,\max}$ ultimately halts mechanism generation. Once iteration through the inner loop at a current X_A value is complete, X_A is compared to the maximum conversion. If $X_A \leq X_{A,\max}$, X_A is incremented and the iteration based on the reactant rates is carried out. Reaction to higher conversions provides the opportunity for additional species to react, as their $R_{i,\max}$ values may now be greater than $R_{\min}$. When no "unreacted" species meet the minimum rate criterion and $X_{A,\max}$ has been reached, mechanism generation halts.

There are alternative definitions of R_{char} that would be more appropriate for more complicated situations where the kinetics dramatically change with time or there is more than one main reactant. We use the instantaneous $R_i(t)$ as our criteria rather than a time-averaged value to properly account for the kinetic significance of species that may only be important for a short period of time (e.g. during ignition). We don't believe there would be much advantage to adding more than one species to the "reacted" species list at each iteration, but we have not tested this possibility.

Another quantity that we calculate is the overall concentration of "unreacted" species produced using the current mechanism, eq 3, where the sums run over the n time steps and the i "unreacted" species:

$$\text{“unreacted” concentration} = \sum\sum R_i(t_n)\Delta t_n \qquad (3)$$

The quantity computed in eq 3 is analogous to a mass-balance discrepancy in analytical chemistry, since it is an approximate upper bound on the total production of molecules not included explicitly in the final model.

In the early stages of the mechanism generation, where the mechanism is quite incomplete, the rate of formation of "unreacteds" may be relatively large and the corresponding overall "unreacted" concentration relatively large as well. For example, in the rudimentary mechanism illustrated in Figure 1a, nearly 100% of the converted ethane ends up as "unreacted" species such as methane, ethylene, and hydrogen. As the iterations continue, Figures 1b,c, the most important of these species are no longer considered "unreacted" but are incorporated into the mechanism, the mechanism becomes more complete, and we expect that the overall "unreacted" yield will decrease. Eventually the "unreacted" rates and concentrations will be insignificant, and the mechanism will naturally terminate.

V. Species on the "Edge" of the Kinetic Model

In our approach, the "reacted" species in the center of Figure 1 are treated more or less exactly: all their reactions with each other are included, and inside this set of species all reactions are reversible. The double-headed solid arrows indicate that there are reversible reactions connecting the species. It is less clear how to deal with the "unreacted" species, for which only a very incomplete set of reactions are known (the dashed lines in Figure 1), none of which involve the "unreacted" species as reactants.

It is important to note that the classification of the species as "unreacted" or "reacted" dramatically reduces the computational burden required to solve the differential equations. By definition, the "unreacted" species are not reactants, so their concentrations do not appear on the right-hand side of the differential equations:

$$\mathrm{d}C_j/\mathrm{d}t = f(C_j) \qquad (4)$$

$$R_i = \mathrm{d}C_i/\mathrm{d}t = g(C_j) \qquad (5)$$

where all j are "reacted" species and all i are "unreacted" species.

Therefore, the differential equations of the "unreacted" species are decoupled from those of the "reacted" species. It is only necessary to solve the differential equations corresponding to the "reacted" species to yield the concentration profiles $C_j(t)$ of the "reacted" species. These concentration profiles are then used to calculate the rates of formation R_i of the "unreacted" species, which are needed to decide whether they are significant. In the examples presented, we observe that the number of "reacted" species is much smaller than the total number of species. In solving the differential equations only for the small subset of "reacted" species, the computational requirements are much less than for solving all the equations simultaneously.

To further reduce the computational effort, we only include reactions where all the species involved are on the "reacted" species list when constructing the differential equations for the "reacted" species' concentrations. This amounts to neglecting the effects of reactions that cross the boundary between the "reacted" and "unreacted" species, Figure 1, while estimating the “reacted” species’ concentration profiles $C_j(t)$. These $C_j(t)$ are then used to estimate the rate of formation of the “unreacted” species.

Any treatment of the reactions that “cross the boundary” is bound to be problematic, since the reaction scheme is always incomplete, and the subsequent behavior of the “unreacted” species formed is unknown at this stage. However, when the algorithm approaches convergence and the reaction scheme is substantially complete, the rates of all the reactions across the boundary are very small and these reactions are only a small perturbation on the overall reaction rates of the system. We have tried other methods for accounting for these reactions “across the boundary” (e.g. treating them reversibly or as irreversible sinks). All approaches give substantially identical results at tight precisions. The approximation we chose is significantly faster than the alternatives, since there are often 100 times as many reactions that cross the boundary as there are reactions involving only “reacted” species. Using this approximation, it is possible to generate reaction schemes to much higher precisions than would be feasible with other approaches.

VI. Rate Constant and Thermochemical Property Estimation

The rate constants assigned to each elementary reaction in the mechanism are calculated from linear free energy relationships that are often referred to as “Evans−Polanyi equations”. The linear free energy relationship is a semiempirical relationship that relates the activation energy to the heat of reaction:

$$E_a = E_0 + \alpha \Delta H_{rxn} \tag{6}$$

The rate constants are then calculated from the Arrhenius equation:

$$k(T) = A \exp(-E_a/RT) \tag{7}$$

Table 1 lists the linear free energy relationship parameters and the Arrhenius frequency factors used to compute the rate constants, with each reaction type having a unique set of parameters. These rate estimation rules are roughly correct, but omit certain reactions and details that would be important if one’s purpose were to develop a model that quantitatively matched an experiment rather than to develop a new algorithm as in the present work. For example, the rules presented in Table 1 neglect intramolecular addition reactions and the reverse intramolecular β-scissions, which are thought to be important for the formation of cyclic species and eventually coke. The form used assumes that the reactions are all in the high-pressure limit, and neglects complications such as heat-capacity corrections to ΔH_{rxn} and non-Arrhenius temperature dependence. These reaction rate estimation rules are not exactly consistent with the relation $\Delta G_{rxn} = -RT \ln K_{eq}$, since they do not include the dependence on ΔS_{rxn}. In the future, more complicated rate expressions $k(T,P)$ will be incorporated into NetGen.

TABLE 1: Rate Estimation Rules Used, See Eqs 6 and 7[a]

A	E_0	α	reaction family
10^{18} s^{-1}	0	1	$-C-C- \rightarrow -C^{\bullet} + {}^{\bullet}C-$ $-C-O- \rightarrow -C^{\bullet} + {}^{\bullet}O-$ $-O-O- \rightarrow -O^{\bullet} + {}^{\bullet}O-$
10^{11}	14	0.5	$R_1^{\bullet} + \mathrm{H}-\mathrm{R}_2 \rightarrow \mathrm{R}_1-\mathrm{H} + {}^{\bullet}\mathrm{R}_2$
3.2×10^{10}	12	0.5	$R^{\bullet} + C{=}C \rightarrow R-C-C^{\bullet}$ $R^{\bullet} + -C{\equiv}C- \rightarrow R-C{=}C.-$
10^{14} s^{-1}	12	0.5	$R-\mathrm{C}-\mathrm{C}^{\bullet} \rightarrow \mathrm{R}^{\bullet} + \mathrm{C}{=}\mathrm{C}$ $R-\mathrm{O}-\mathrm{C}^{\bullet} \rightarrow \mathrm{R}^{\bullet} + \mathrm{O}{=}\mathrm{C}$ $R-\mathrm{C}-\mathrm{O}^{\bullet} \rightarrow \mathrm{R}^{\bullet} + \mathrm{C}{=}\mathrm{O}$ $R-\mathrm{O}-\mathrm{O}^{\bullet} \rightarrow \mathrm{R}^{\bullet} + \mathrm{O}_2$
10^{12}	0	0	$R_1^{\bullet} + \mathrm{R}_2^{\bullet} \rightarrow \mathrm{R}_1-\mathrm{R}_2$ $R_1OO^{\bullet} + R_2OO^{\bullet} \nrightarrow R_1OOOOR_2$
3.2×10^{12}	0	0	$H-C-C^{\bullet} + R^{\bullet} \rightarrow RH + C{=}C$ $H-O-C^{\bullet} + R^{\bullet} \rightarrow RH + O{=}C$ $H-C-O^{\bullet} + R^{\bullet} \rightarrow RH + C{=}O$ $H-O-O^{\bullet} + R^{\bullet} \rightarrow RH + O_2$
3×10^{13}	5	1	$-C-H + O_2 \rightarrow -C^{\bullet} + HOO^{\bullet}$ $R-O-H + O_2 \rightarrow RO^{\bullet} + HOO^{\bullet}$
5×10^{12}	0	0	$-C^{\bullet} + O_2 \rightarrow -COO^{\bullet}$ $RO^{\bullet} + O_2 \rightarrow ROOO^{\bullet}$
10^{11} s^{-1}	5	1	$H-C.{=}O \rightarrow H^{\bullet} + CO$ $\sim C-C.{=}O \rightarrow \sim C^{\bullet} + CO$
3.2×10^{12}	0	0	$r_1OO^{\bullet} + r_2OO^{\bullet} \rightarrow r_1O^{\bullet} + r_2O^{\bullet} + O_2$
5×10^{12} s^{-1}	41	0.5	1,2 H-shift, e.g., $H-C-O^{\bullet} \rightarrow {}^{\bullet}C-O-H$
10^{12} s^{-1}	40	0.5	1,3 H-shift, e.g., $H-C-O-O^{\bullet} \rightarrow {}^{\bullet}C-O-O-H$
2×10^{11} s^{-1}	19	0.5	1,4 H-shift, e.g., $H-O-O-C-C^{\bullet} \rightarrow {}^{\bullet}O-O-C-C-H$
10^{11}	0	0	$H-C-O-O^{\bullet} + rOO^{\bullet} \rightarrow \sim C{=}O + rOH + O_2$
$1.3 \times 10^{7}\, T^{1.3}$	−0.8	0	${}^{\bullet}OH + CO \rightarrow H^{\bullet} + CO_2$
2×10^{14}	16.8	0	$H^{\bullet} + O_2 \rightarrow O + {}^{\bullet}OH$
10^{10}	0	0	$r_1OO^{\bullet} + r_2OO^{\bullet} \rightarrow r_1OOO^{\bullet} + r_2O^{\bullet}$

[a] The units of A are cm^3/(mol·s) unless otherwise stated; E_0 is in kcal/mol.

The heat of formation for a given species is retrieved from the NIST Structures and Properties thermochemical database program[20] or calculated using MOPAC.[22] As each new species is produced during the mechanism generation, the NIST database is first searched for the heat of formation value for the given species. This heat of formation is usually calculated using group additivity, but may also be stored as an experimentally derived value.[20] If the heat of formation for the species cannot be determined by the NIST program, a library containing heat of formation values calculated from MOPAC is searched. If the species data are not in the library, a MOPAC calculation is then performed to obtain the necessary heat of formation information. To perform the MOPAC calculation, an initial guess at the three-dimensional geometry is needed; this is provided by the Converter program distributed by Molecular Simulations.[24] The heat of formation values obtained from the MOPAC calculation are scaled for agreement with experimental data:

$$\Delta H_f = \Delta H_{MOPAC} - 3.7 \text{ kcal/mol} \quad \text{for closed-shell molecules} \tag{8}$$

$$\Delta H_f = \Delta H_{MOPAC} + 10.7 \text{ kcal/mol} \quad \text{for radicals} \tag{9}$$

For radicals, we used the “DOUBLET” keyword in MOPAC.

Since we are typically considering about 10 000 species, and the predicted kinetics are often quite sensitive to the heats of formation used, the speed and accuracy with which the program can estimate the thermochemical values for new molecules are quite important. To accelerate the calculations, we added several thermochemical groups to the NIST database, drawing heavily on results from THERM.[25] This greatly reduced the number of time-consuming MOPAC calculations required and probably increased the accuracy of the results as well, since the group additivity estimates are usually more accurate than heats of formation calculated using MOPAC.

VII. Examples

A. Ethane Pyrolysis. *1. Results with the New Rate-Based Algorithm.* Several runs for ethane pyrolysis at 1200 K at various precision levels were carried out for a rank zero initial species pool. “Rank” distinguishes primary products from secondary products; the exact definition of “rank” used is given in ref 18. Rank zero is the small, nine species, initial pool in

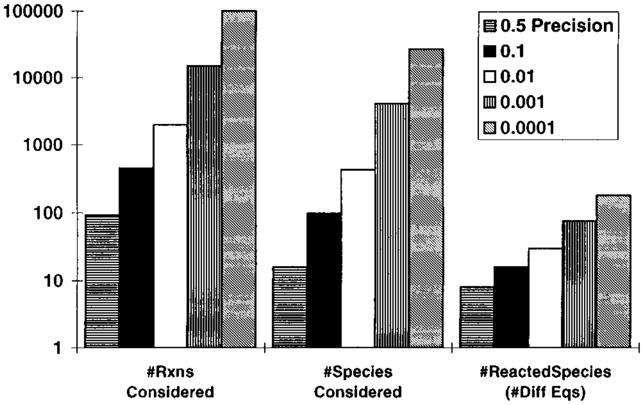

Figure 3. Ethane pyrolysis mechanism complexity, increasing dramatically with increasing precision.

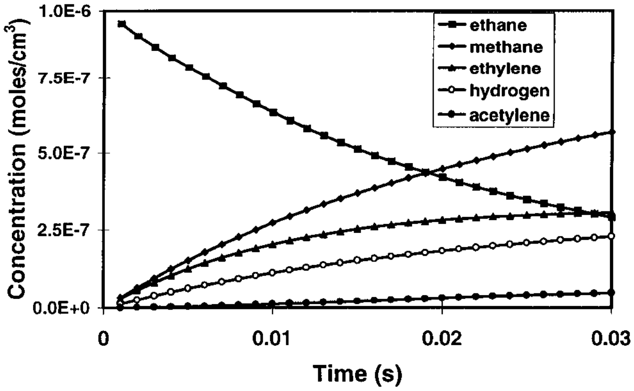

Figure 4. Predicted concentration profiles of major species of ethane pyrolysis at 0.0001 precision.

Figure 1a consisting of the initial reagent, ethane, and the radicals that can be formed from it; the primary products are placed in the "unreacted" species list. The initial ethane concentration was 10^{-6} mol/cm^3, and the conversion target was 70%. The precisions ranged from 0.5 to 0.0001. As illustrated in Figure 3, the ethane mechanism complexity increases rapidly with increasing precision. For this range of precisions, the number of reactions considered while generating the mechanism ranges from 92 to 99 644, and the total number of species considered ranges from 16 to 27 076. However, the increase in the number of "reacted" species (the resulting number of differential equations that must be solved) is not as dramatic; these range from 8 to 180.

The relatively slow rise in the number of "reacted" species makes it feasible to solve the decoupled differential equations for large mechanisms. However, our goal is to utilize as small a mechanism as possible, whereby we include important species and reactions while excluding unimportant ones. To evaluate how large a mechanism is necessary, we compare concentration profiles for the major species ethane, methane, ethylene, hydrogen, and acetylene, calculated at the various precision levels. We find that they are very close for mechanisms generated using 0.01 and 0.0001 precisions; the largest deviation between the concentrations calculated using these two mechanisms at 30 ms is less than 2% for the major species. These concentration profiles are shown in Figure 4 for a precision of 0.0001. The concentration profiles are considerably different from the profiles generated using a precision parameter of 0.5. We can conclude that the rate-based mechanism generation has produced a suitably complete mechanism at 0.1 precision, including all the necessary 16 "reacted" species and the 203 reactions involving them. The larger mechanisms generated using tighter precisions include reactions and species that are unnecessary unless one is concerned with minor species. As expected for substantially complete mechanisms, the total "unreacted" concentration, computed from eq 3, decreases smoothly as the precision is tightened, Figure 5. The rate-based algorithm naturally decided those reactions and species necessary for describing the 70% conversion of ethane.

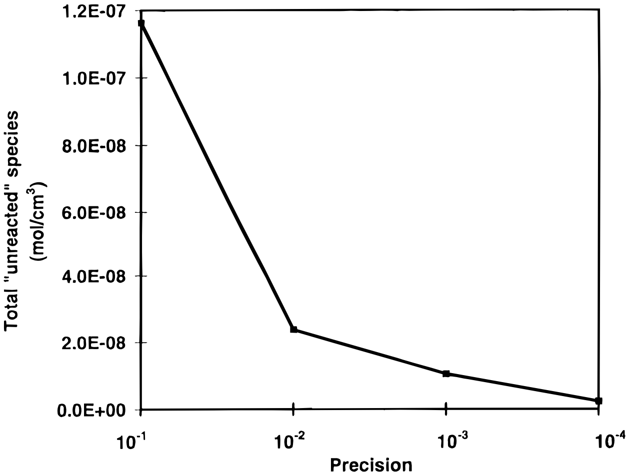

Figure 5. Total "unreacted" concentration, a measure of the truncation error (eq 3), for the ethane pyrolysis mechanism, which decreases as the precision is tightened.

2. Comparison with Other Approaches. The rate-based scheme generation algorithm of the present work was contrasted with the carbon count and rank-based termination criteria used in earlier versions of NetGen[16−19] by comparing the predictions of the various models generated. The carbon count- and rank-based models are designated by c_nr_m, where radicals with a number of carbon atoms greater than n and all species with a product rank greater than m are not allowed to react. Models were constructed allowing for the formation of primary (c_2r_0), secondary (c_3r_1, c_4r_1), tertiary (c_3r_2), and quaternary (c_3r_3) products. The same reaction conditions and thermochemical and rate estimation parameters were used in all cases.

The model results are compared at a reaction time of 30 ms in Table 2. The number of "reacted" species and therefore the number of differential equations solved are also reported. The model allowing for formation of only primary products is clearly inadequate. Acetylene, a product with a predicted selectivity as high as 7.5%, is not formed, and there is significant deviation of the major product yields from the predictions of the other models at longer reaction times. The predictions of the rate of ethane disappearance from the two models allowing for the formation of secondary products agree within 3% with the rate-based (0.0001 precision) and more comprehensive c_3r_3 and c_3r_2 models. However, a 12% deviation in major product yields is observed. As the plot of the temporal variations of ethylene formation shown in Figure 6 reveals, it is difficult to distinguish the results of the rate-based and c_3r_3 models by eye. The predictions differ by less than 2% for most important products and by less than 9% for the relatively minor product acetylene. However, the comprehensive rank and carbon count-based models include a significantly larger number of species in the model which are not required for a numerically precise description of the dominant chemistry. For example, the c_3r_3 model, which gives results similar to the rate-based model, involves 4 times as many differential equations and species (691 vs 169); that is, at least 75% of the species in the c_3r_3 model are apparently insignificant. The rate-based model is capable of including important higher rank species such as propadiene (allene), a tertiary product, without including a large number of insignificant species of the same rank.

TABLE 2: Comparison between Rank-Based and Rate-Based Models for Ethane Pyrolysis

model	ethane (10^{-7} mol/cm^3)	methane (10^{-7} mol/cm^3)	ethylene (10^{-7} mol/cm^3)	hydrogen (10^{-7} mol/cm^3)	acetylene (10^{-7} mol/cm^3)	no. of species in model
c2r0	3.23	5.03	4.24	1.72		9
c3r1	3.03	5.27	3.26	2.27	0.595	29
c4r1	3.03	5.27	3.26	2.28	0.593	49
c3r2	2.97	5.57	3.07	2.28	0.508	144
c3r3	2.95	5.61	3.06	2.26	0.502	691
rate-based	2.93	5.66	3.05	2.27	0.468	169

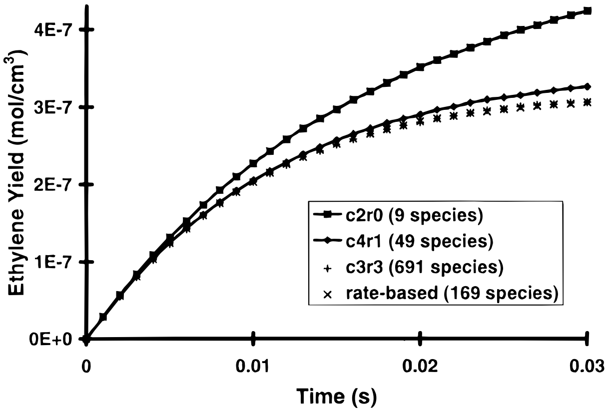

Figure 6. Concentration profiles for ethylene generated using rate-based and rank-based termination criteria, virtually identical for the c_3r_3 model.

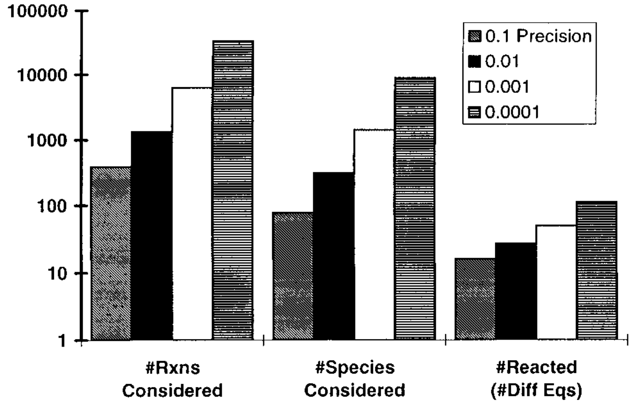

Figure 7. Butane pyrolysis mechanism complexity, increasing dramatically with increasing precision.

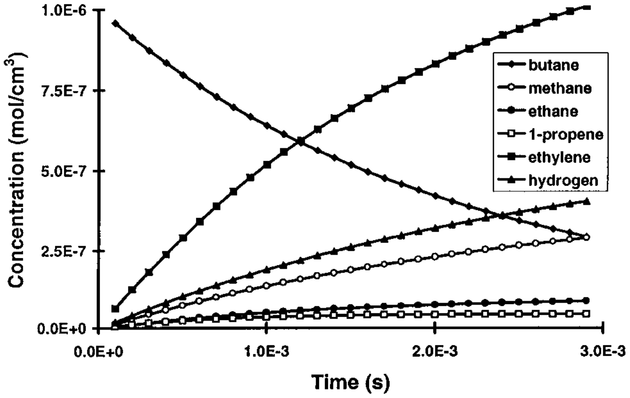

Figure 8. Predicted concentration profiles of major species of butane pyrolysis at 0.0001 precision.

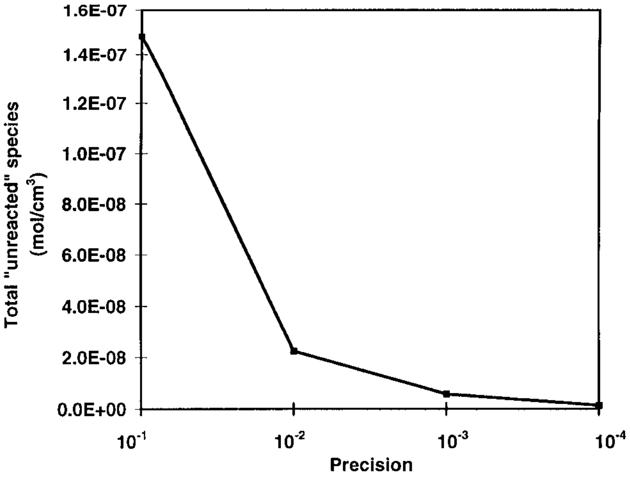

Figure 9. Total "unreacted" concentration, a measure of the truncation error (eq 3), for the butane pyrolysis mechanism, decreasing as the precision is tightened.

B. Butane Pyrolysis. Butane pyrolysis runs were carried out at 1200 K at precision levels of 0.1−0.0001 for a rank zero initial species pool consisting of 25 species. The initial butane concentration was 10^{-6} mol/cm^3, and the conversion level was 70%. From Figure 7, we see that the butane mechanism complexity increases with increasing precision as in the case of ethane.

Figure 8 shows the concentration profiles for the major species butane, methane, ethylene, hydrogen, ethane, and 1-propene calculated at the 0.0001 precision level. These profiles are very close for precisions 0.1−0.0001; the predicted concentrations at 3 ms change by less than 1% on going to the tighter precisions. The total "unreacted" concentration, computed from eq 3, decreases smoothly as the precision is tightened, Figure 9. These data suggest that the rate-based mechanism generation has produced a suitably complete mechanism at 0.1 precision, including all of the necessary 14 reactive species and the 260 reactions involving them.

1. Expected Precision Limits. As we approach the tight precision limit, the mechanism will be essentially correct, and there will typically be many thousands of very minor species in the "unreacted" species list. If we tighten the precision by another order of magnitude, moving the most important of these minor species into the mechanism, we expect that there will be a similar drop in the rate of formation of the remaining "unreacted" species, Figure 9, and that we will add another significant digit to our calculated major species concentrations. Alternatively, if we neglect reactions with rates that are a few percent of the characteristic reaction rate of the system, R_{char}, we cannot expect any of our predictions will be accurate to better than a few percent. We therefore expect that once the kinetic model is substantially complete, the predicted concentrations of the major species should lie within or close to the range (1 $\pm$ precision)C_{complete},where C_{complete} is the concentration prediction if the calculation were carried to infinitely tight precision. Figure 10a shows the concentrations of butane at 2 ms for butane pyrolysis, plotted along with the expected precision limits (presuming the 0.0001 precision level calculation is essentially exact). In this case, most of the major species concentrations actually converge more rapidly than we would expect, but the least important of the major products, acetylene, lies slightly outside the expected precision limits in some cases; see Figure

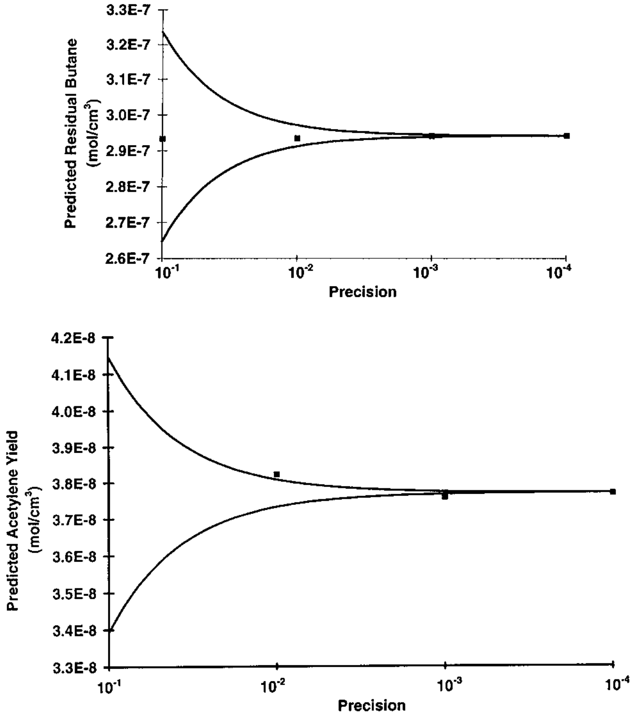

Figure 10. (a, top) Computed butane concentrations at 2 ms, lying within the expected precision limits. (b, bottom) Computed acetylene concentrations at 2 ms, lying just outside the expected precision limits.

10b. Note that acetylene is considered insignificant at the 0.1 precision level.

The minimum rate criteria we use is only indirectly related to the minor species concentrations. The models typically include only a few hundred species, so we are implicitly assuming the concentrations of all other species are zero. For minor species in the model, we expect that errors could be on the order of ±(precision)$R_{\mathrm{char}}\tau$, which may be larger than the predicted concentrations of these species. This is the situation for acetylene in the 0.1 precision model disussed above.

In some cases (e.g. catalysis with poisoning or long-chain free radical reactions) the overall reaction rate can be very sensitive to errors in certain minor species concentrations; in those cases higher precisions are required to achieve convergence for the major species. Systems with long induction periods or poisoning times, and where these effects involve several sequential reactions of minor species, may be difficult to converge, since they would require such high precisions that the computer resources could be exhausted during model construction. We are currently testing these limits on systems that are known to be highly nonlinear in certain minor byproduct concentrations.

2. Effects of Initial Pool Size. We created rank 1 generated pools of varying sizes and propagated them out to 0.01 precision to study the effects of the initial pool size on the mechanism generated, Table 3.

A plot of residual butane concentrations vs pool size for the 0.01 precision level, Figure 11, suggests that the concentration profiles of the major species at this level are only slightly sensitive to the initial pool size. The calculations become insensitive to pool size once they contain enough of the important molecules and radicals to efficiently produce a suitably complete mechanism. At high precisions the rate-based mechanism generation algorithm is not sensitive to the initial mechanism since it is able to determine the important reactions and species itself: it does not require a very good initial guess.

TABLE 3: Dependence of Butane Pyrolysis Mechanisms on Initial Pool Size

precision	pool size	no. of rxns considered	no. of species considered	no. of species in model	no. of rxns in model
0.01	25	1328	222	24	732
0.01	39	1328	222	24	741
0.01	55	1799	222	27	1240
0.01	73	2832	246	32	2120
0.01	104	3550	254	35	2765
0.01	177	4160	300	37	3815
0.01	255	5828	412	42	5379
0.0001	255	47710	8094	118	22653
0.0001	25	33405	7359	104	8059

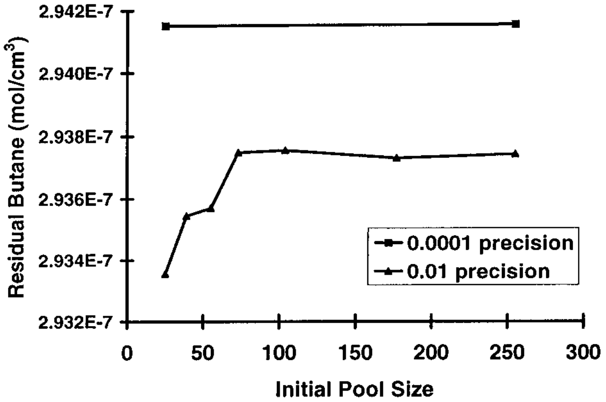

Figure 11. Residual butane concentrations at 2 ms vs the initial pool size at the 0.01 and 0.0001 precisions. Butane shows a small sensitivity to the smaller pools at 0.01 precision and is insensitive to the pool at 0.0001 precision.

C. Ethane Combustion. Isothermal "combustion" of ethane was studied in order to evaluate the rate-based mechanism generation algorithm for a more complex chemical system. An initial ethane concentration of 2 × 10^{-7} mol/cm^3, an initial oxygen concentration of 7 × 10^{-7} mol/cm^3, and a constant temperature of 1200 K were used. When the NIST program was unable to estimate a needed heat of formation, due to missing thermochemical groups, MOPAC was used to make that estimate.

The rate-based strategy was used to construct six mechanisms for ethane combustion using precision levels ranging from 1.0 to 0.0001. The initial species pool was constructed allowing for generation of primary products only and restricting the "reacted" species to the subset of the first 50 species with a rank less than 1. The maximum conversion with respect to the reactant ethane was set to 0.75.

The characteristics of the models generated are summarized in Table 4 as a function of the precision level. The residual reactant concentration and the concentrations of the major products, ethylene and acetylene, are reported at a reaction time of 10 ms. For precision levels poorer than 0.1, the models were almost identical. As discussed below, these poor-precision models are sensitive to the size of the initial pool. As the precision level was tightened from 0.1 to 0.0001, the total number of species considered increased dramatically from 191 to 10 400. However, as was true for the pyrolysis studies, the number of "reacted" species included in the final model increases more gradually, from 30 to 142.

It is clear from the results summarized in Table 4 that the mechanism is substantially complete at a precision level of 1.0: tightening the precision by 4 orders of magnitude changes the major product concentrations at 10 ms by less than 1% and the residual ethane concentration by less than 5%.

The effect of varying the initial species pool was also examined for ethane combustion. The parameter controlling the number of species permitted to react in the initial pool had a clear effect on the mechanisms generated with loose precisions

TABLE 4: Model Characteristics and Major Species Concentrations at 10 ms for Ethane Combustion Mechanisms as a Function of the User-Defined Precision Level

precision level	ethane (10^{-8} mol/cm^3)	ethylene (10^{-8} mol/cm^3)	acetylene (10^{-8} mol/cm^3)	no. of species considered	no. of species in model	"unreacted" species (10^{-8} mol/cm^3)
1.0	1.233	8.992	7.206	191	28	2.742
5.0×10^{-1}	1.233	8.992	7.205	191	29	0.829
1.0×10^{-1}	1.233	8.992	7.206	196	30	0.444
1.0×10^{-2}	1.192	8.966	7.196	629	43	0.085
1.0×10^{-3}	1.179	8.941	7.232	1597	65	0.024
1.0×10^{-4}	1.176	8.933	7.238	10 400	142	0.006

(1.0, 0.5, 0.1); for small initial pools, mechanisms generated to these precisions gave concentrations at 10 ms more than 10% different than those reported in Table 4. In each of the deficient mechanisms, ethyl peroxyl radical was formed, but not permitted to react. Evidently, this species is crucial to the overall chemistry, but is formed slowly enough that it is not picked up with a precision level of 0.1. Three strategies were used to successfully overcome this exclusion: (1) the precision was tightened, so ethyl peroxyl was considered a significant species and permitted to react; (2) ethyl peroxyl was input as a reactant; (3) a larger initial pool was used. The results reported in Table 4 were computed using the third approach.

Conclusions

Using the new algorithm, it is now possible to generate kinetically complete reaction schemes for cases much more complex than ever before. The computer will precisely follow out the implications of rate estimation rules of any complexity. Essentially, the user need only input the initial concentrations, the desired degree of conversion, and desired precision, and the computer will generate all the significant reactions, intermediates, and products, and their concentrations as a function of time. Kineticists can therefore focus on the rate estimation rules, rather than on the computational details and bookkeeping required to keep track of thousands of species and solve thousands of differential equations. In many cases, there is no longer a need to make untested approximations such as neglecting species or reactions *a priori*, using steady-state approximations, or lumping species together. The computer can handle the whole problem precisely. There are limits to the complexity that modern computers can handle, however, and we believe that it would be very difficult to handle a system with more than 10 000 significant reactive species with our approach on currently available hardware.

The program allows one to be certain that the predictions are not in error because particular species or reactions were arbitrarily left out. The estimate of the total concentration of "unreacted" species (i.e. all the species not included in the kinetic model) formed gives a quantitative measure of the error one is making by using a partial kinetic scheme instead of including all possible reactions and species. If this estimated concentration is very small, one can be confident that the kinetic scheme involving only "reacted" species adequately represents the full system. In the three cases studied, the algorithm converges smoothly as one tightens the precision. However, all the concentration vs time calculations rely on thermochemical and rate constant estimates, which are undoubtedly the largest sources of error. Our ability to quantitatively predict reactions is now limited by how well we know the chemistry (i.e. the rate constants) and not by computational issues, freeing the chemical kinetic modeler to focus solely on the chemistry rather than on the computational details. The computer keeps track of all species that are numerically significant and solves the system of differential equations to a user-specified level of precision.

In the course of this work, we performed calculations on several systems using slightly different thermochemistry and rate estimation parameters. Both the size of the required reaction mechanism and the calculated concentration vs time profiles were found to be sensitive to the values of some of the parameters; changes of only a few kcal/mol in some key parameters can drastically alter the predictions. These results will not be surprising to thermochemical kineticists, but they do highlight the need for accurate methods to estimate thermochemistry and rate constants, based on solid experimental data. It would appear to be relatively straightforward to compute the sensitivity of the predictions to uncertainties in the thermochemical and rate parameters using this program. A key point is that the number of parameters in the rate and thermoestimation routines is very much smaller than the number of species considered in constructing the model. Sensitivity calculations are, therefore, feasible even for systems containing large numbers of species.

A problem with virtually all chemical kinetic models is that they are only complete for a limited range of conditions (initial concentrations, temperatures, pressures, time scale). Usually the modeler selects which reactions to include and which to exclude by making assumptions about whether or not they will be significant in the range of interest. The range of applicability of the models is rarely known exactly. Sensitivity analysis can be used to get a feel for the range of applicability, but this can be misleading since the sensitivity analysis itself depends on the initial conditions, and it is only valid in the range where the kinetic model is adequately complete. Our rate-based algorithm explictly checks which reactions are numerically significant for particular initial conditions, so one can be quite confident that our models are complete for those initial conditions. To construct a model that is adequately complete over a range of initial conditions, the computer can generate models for several different initial conditions and then construct a grand model that is the union of these sets of chemical reactions. Since the computer can generate models much faster than a human, it is perfectly feasible to ask it to generate a dozen different models, each appropriate to different initial conditions, in order to construct a more general grand model. In some applications, the user really wants the predicted concentration profiles, not the kinetic model, as output; in these cases it may be appropriate to generate and immediately solve a new kinetic model for every different initial condition specified by the user.

A very significant limitation of the current program is that it does not consider inhomogeneity or transport issues. Current 3-D reactive fluid dynamic simulations can typically handle less than a dozen distinct species,[26] while 0-dimensional simulations are now feasible with more than 1000 reactive species, and in many cases more than 100 species are found to be significant by NetGen. *A posteriori* scheme reduction may allow some reduction in the number of species, but it is unlikely to completely bridge the gap between the need for detailed chemistry and the capabilities of computational fluid dynamics. A major challenge for the future is to couple chemistry and flow together without losing needed precision, in order to accurately simulate the often multiphase, reacting flows important in combustion, refining, and chemical processing.

The advent of general-purpose reaction scheme generation programs like NetGen places a renewed emphasis on the importance of developing accurate rate estimation techniques and provides an efficient way to use those rate estimation rules that are already available. It is hoped that in the near future, empirical and theoretical estimates for $k(T,P)$ will be developed for additional classes of reactions, allowing construction of more accurate chemical kinetic models for a broader range of systems.

References and Notes

(1) *Rethinking the Ozone Problem in Urban and Regional Air Pollution*; National Academy Press: Washington, DC, 1991.

(2) Ranzi, E.; Faravelli, T.; Gaffuri, P.; Sogaro, A. *Combust. Flame* **1995**, *102*, 179.

(3) Dente, M.; Pierucci, S.; Ranzi, E.; Bussant, G. *Chem. Eng. Sci*. **1992**, *47*, 2629−34.

(4) Weekman, V. R., Jr. *AIChE Monogr. Ser. 11*, **1979**, *75*.

(5) Jacob, S. M.; Gross, B.; Voltz, S. E.; Weekman, V. R., Jr. *AIChE J*. **1976**, *22*, 701.

(6) See for example: Dean, A. M. *J. Phys. Chem*. **1990**, *94*, 1432.

(7) Chinnick, S. J.; Baulch, D. L.; Ayscough, P. B. *Chemom. Intell. Lab. Syst*. **1988**, *5*, 39−52.

(8) Haux, L.; Cunin, P.-Y.; Griffiths, M.; Come, G.-M. *J. Chim. Phys*. **1988**, *85*, 739−43.

(9) Chevalier, C.; Warnatz, J.; Melenk, H. *Ber. Bunsen-Ges. Phys. Chem.* **1990**, *94*, 1362−7.

(10) Dente, M.; Pierucci, S.; Ranzi, E.; Bussant, G. *Chem. Eng. Sci*. **1992**, *47*, 2629−34.

(11) Chevalier, C.; Pitz, W. J.; Warnatz, J.; Westbrook, C. K.; Melenk, H. *Twenty-Fourth Symposium (International) on Combustion*; The Combustion Institute: Pittsburgh, PA, 1992; pp 93−101.

(12) Blurock, E. S. *J. Chem. Inf. Comput. Sci*. **1995**, *35*, 607−16.

(13) Nehse, M.; Warnatz, J.; Chevalier, C. *Twenty-Sixth Symposium (International) on Combustion*; The Combustion Institute: Pittsburgh, PA, in press.

(14) Ranzi, E.; Faravelli, T.; Gaffuri, P.; Sogaro, A. *Combust. Flame* **1995**, *102*, 179−92.

(15) Ranzi, E.; Sogaro, A.; Gaffuri, P.; Pennati, G.; Westbrook, C. K.; Pitz, W. J. *Combust. Flame* **1994**, *99*, 201−11.

(16) Klinke, D. J., II; Broadbelt, L. J. Mechanism Reduction during Computer Generation of Compact Reaction Models. *AIChE J.*, in press.

(17) Broadbelt, L. J.; Stark, S. M.; Klein, M. T. *Comput. Chem. Eng.* **1996**, *20*, 113−129.

(18) Broadbelt, L. J.; Stark, S. M.; Klein, M. T. *Ind. Eng. Chem. Res.* **1995**, *34*, 2566−73.

(19) Broadbelt, L. J.; Stark, S. M.; Klein, M. T. *Ind. Eng. Chem. Res.* **1994**, *33*, 790−9.

(20) Stein, S. E. *NIST Structures & Properties version 2.0*; NIST Standard Reference Database 25, Gaithersburg, MD, 1994.

(21) Brenan, K. E.; Campbell, S. L.; Petzold, L. R. *Numerical Solutions of Initial-Value Problems in Differential-Algebraic Equations*; Elsevier: New York, 1989.

(22) (a) Stewart, J. J. P. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers, Inc.: New York, 1990. (b) Stewart, J. J. P. *MOPAC Reference Manual and Release Notes*, 6th ed.; Frank J. Seiler Research Laboratory: United States Air Force Academy, CO, 1990.

(23) Corneil, D. G.; Gotlieb, C. C. *J. Assoc. Comput. Mach.* **1970**, *17*, 51−64.

(24) *Converter*, version 2.0; Biosym Technologies: San Diego, CA, 1992.

(25) Ritter, E. R.; Bozzelli, J. W. *Int. J. Chem. Kinet*. **1991**, *23*, 767−78.

(26) Oran, E. S.; Boris, J. P. *Numerical Approaches to Combustion Modeling*; Progress in Astronautics and Aeronautics Series, Vol. 135; American Institute of Aeronautics and Astronautics: Washington, DC, 1991.