Descriptive Simulation: Combining Symbolic and Numerical Methods in the Analysis of Chemical Reaction Mechanisms

Michael Eisenberg

A.I. Memo No. 1171 September, 1989

Abstract

The Kineticist's Workbench is a computer program currently under development whose purpose is to help chemists understand, analyze, and simplify complex chemical reaction mechanisms. This paper discusses one module of the program that numerically simulates mechanisms and constructs qualitative descriptions of the simulation results. These descriptions are given in terms that are meaningful to the working chemist (e.g., steady states, stable oscillations, and so on); and the descriptions (as well as the data structures used to construct them) are accessible as input to other programs.

Keywords: Symbolic computation, chemical kinetics

This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the Laboratory's artificial intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-86-K-0180.

Descriptive Simulation: Combining Symbolic and Numerical Methods in the Analysis of Chemical Reaction Mechanisms

Michael Eisenberg

Abstract

The Kineticist's Workbench is a computer program currently under development whose purpose is to help chemists understand, analyze, and simplify complex chemical reaction mechanisms. This paper discusses one module of the program that numerically simulates mechanisms and constructs qualitative descriptions of the simulation results. These descriptions are given in terms that are meaningful to the working chemist (e.g., steady states, stable oscillations, and so on); and the descriptions (as well as the data structures used to construct them) are accessible as input to other programs.

I. Introduction

A. Chemical Reaction Mechanisms

Most chemical reactions are complicated events. A "simple" chemical reaction as presented in an undergraduate chemistry textbook may in actuality correspond to a (possibly large) collection of elementary chemical steps (typically, unimolecular decompositions and bimolecular collisions). This collection of elementary steps constitutes a *mechanism* for the overall reaction. To give one brief example, consider the decomposition of acetaldehyde:

```
(1)        CH3CHO --> CH4 + CO
```

A hypothetical mechanism for this reaction consists of the following six elementary reactions:

```
(2.1)      CH3CHO --> CH3 + CHO
(2.2)      CHO --> CO + H
(2.3)      CH3 + CH3CHO --> CH4 + CH3CO
(2.4)      CH3CO --> CH3 + CO
```

```
(2.5)       H + CH3CHO --> H2 + CH3CO
(2.6)       2CH3 -->  C2H6
```

A mechanism such as (2.1)-(2.6) is invariably a hypothesis; one of the major tasks of kineticists, then, is to verify a proposed mechanism by comparing its predicted behavior with laboratory observations.

Hidden within the previous sentence is a formidable challenge: namely, predicting (and understanding) the behavior of a reaction mechanism. Mechanisms such as (2.1)-(2.6) above give rise, in general, to systems of (strongly) coupled nonlinear ordinary differential equations. Equations of this kind rarely lend themselves to a direct analytical solution [4, 11]; and even in those few instances where an analytical solution is possible, it is often in a form that conveys little qualitative information about the behavior of the mechanism. In some cases (for open systems far from equilibrium) the rate equations generated by a given mechanism can give rise to a bewildering variety of behavior: reactions may exhibit stable oscillations, birhythmicity, and "chemical chaos"[9]. Finally, it should be noted that sheer size is a factor in determining a mechanism's complexity: it is not uncommon to see mechanisms containing dozens (or even hundreds) of elementary steps.* For the working chemist, then, understanding how a large collection of interacting reactions can give rise to overall behavior represents a task of immense complexity and subtlety.

B. Qualitative Approximations and Numerical Simulation

When a chemist is confronted with a complex mechanism such as (2.1)-(2.6), there are two broad approaches that he or she may take:

- The chemist can make certain plausible simplifying assumptions that can be used to rewrite the mechanism (and corresponding differential equations) in a simpler form.
- The chemist can simulate the mechanism numerically on a computer, and tabulate the results.

▷ Making Approximations to Simplify Mechanisms

Suppose we have the following mechanism for the overall reaction in which reactant A is converted to a mixture of products B and C:

---

* In fact, if the system in question is not homogeneous, the corresponding mathematical model will consist of partial rather than ordinary differential equations—making the task of understanding even more difficult.

```
(3.1)  A --> B
(3.2)  B --> C
(3.3)  C --> B
```

Depending on the assumptions that we make at this point, we can anticipate a range of different behaviors for this simple mechanism. For example if the rates of reactions (3.2) and (3.3) are extremely large compared to that of (3.1), then the mechanism might be described as a rapid (near-) equilibrium between B and C, with a slow increase in both species due to reaction (7.1). In this case we would expect that the ratio of the concentration of B to that of C would be close to constant throughout most of the reaction, and that the concentration of A would slowly decline while those of B and C slowly increase.

Techniques of this type, using notions such as "rapid equilibrium" and "quasi-steady-state" approximations, are standard among kineticists (and in kinetics textbooks). Nevertheless, these techniques have their drawbacks. First, and most obviously, it is often the case that the reaction mechanism under investigation does not lend itself to straightforward simplification; for example, the rates of the reactions (3.1)-(3.3) might all be comparable. Slightly more subtle is the fact that even when the simplifications are justified, they may not apply throughout the course of the reaction. For instance, even if the rates of reactions (3.2) and (3.3) are indeed much larger than that of (3.1), if we start out with only substance A present, it may take time to achieve the "near-equilibrium" ratio between B and C. It is in part due to such complications that chemists must often forgo the possibility of simplifying a mechanism directly, and must turn to numerical simulation.

▷ Numerical Simulation of Mechanisms

Although numerical simulation of chemical mechanisms might be viewed as a technique of last resort—something to be used only when analytical or simplifying methods fail—in point of fact, for large and complex mechanisms, numerical simulation is probably the only way to obtain meaningful information. By now, computational methods for integrating systems of differential equations such as those generated by (2.1)-(2.6) are well-developed, and a variety of powerful integration packages are available.[1, 10] Moreover, the continuing trends in computational hardware have been toward increased speed and decreased cost, and these trends can only make simulation increasingly attractive as an option for understanding mechanisms.

Despite its popularity and utility, however, there are important limitations

in numerical simulation. First, the end result of a simulation is typically a large body of numerical data; it is up to the chemist to look for patterns in these numbers, and to interpret the results in terms of the mechanism that generated them. For example, there may be periods during the simulation in which a few select elementary reactions in the mechanism are responsible for the most dramatic changes in the concentrations of individual substances; but if this is so, it will be up to the chemist to make that interpretation without additional assistance from the computer. Similarly, it may be unclear from looking at the numerical results how (and when) a particular elementary reaction contributed to the overall behavior of the mechanism.

Moreover, the computer that is used for straightforward numerical simulation provides no assistance in *summarizing* the results of the simulation in compact form. Typically, simulations produce only raw data; but for the chemist, it is often the case that certain qualitative features of the simulation (e.g., the presence of oscillations, or the fact that some substances appear in near-constant concentrations throughout the simulation) represent the key results of a numerical experiment. In this case, we would like our computer program to provide (and understand) a representation of the results that includes these qualitative features. Note that simply graphing the results is not sufficient for these purposes: a plotting routine does serve to summarize data for the user, but it fails to provide that summarized data in a form that may then be further examined by the computer itself. For example, there is no current program that can run a sequence of simulations of mechanism (2.1)-(2.6) with a range of values for the rate of reaction (2.3), and identify how the rate of that reaction affects the possibility of a steady-state assumption for the intermediate CH3.

It would be desirable, then, to have some form of data structure representing an overall qualitative description of the results of a simulation; ideally, this data structure would be "coarse-grained" enough to provide a substantial compaction of the numerical data while still retaining enough "fine-grained" detail to capture those features of the simulation that are of interest to the researcher. Additionally, this data structure should be in a form that can be examined, classified, and manipulated in interesting ways by a computer program.

II. The Episode Structure as a Descriptive Primitive for Chemical Simulations

As the discussion of the previous section suggests, understanding the behavior of a reaction mechanism is a difficult task. The chemist may analyze the mechanism before simulating it, in the hopes of finding useful simplifying approximations; he may then resort to direct numerical simulation of the mechanism; he may use the results of his simulation to suggest further approximations; and throughout this process, he must keep in mind the relation between the various mechanisms that he tries out and the behavior of the systems governed by those mechanisms.

To date, computers have been employed primarily to assist with the numerical side of the chemist's task; but there is also a tremendous potential value in having computers assist chemists in the task of simplifying mechanisms and relating mechanism structure to behavior. This is the goal of the Kineticist's Workbench program, described below.

A. The Kineticist's Workbench

The Kineticist's Workbench is a program currently under development. Its purpose is to expand the role of the computer by integrating numerical simulation routines with a variety of other algorithms, both numeric and symbolic, for analyzing, describing, and simplifying reaction mechanisms. In the remainder of this paper, only one portion of the Workbench will be discussed at any length—namely, that portion which is dedicated to performing numerical simulations of mechanisms and generating qualitative descriptions of the simulation results. The final (discussion) section of this paper will briefly describe some other modules within the Workbench program. It should be mentioned before proceeding, however, that much of the eventual strength of the Workbench program will be derived by integrating the various subparts so that each can make use of the results and suggestions of the others in the course of analyzing a given mechanism.

B. Qualitative Descriptions of Simulations

Chemists, in describing the behavior of a particular reaction mechanism, often fall into a kind of narrative discourse. Consider the following description of a certain mechanism giving rise to an oscillating reaction (in this passage,

α and β refer to small collections of elementary steps in the mechanism, and symbols such as F3 refer to individual steps):

"The rate of (α) is proportional to Y, and when this process is dominant X attains a steady state approximated by X_{min}... At such a time, Y will be depleted by the dominance of (α). If (F3) is rate-determining for (F3-4), then when (β) is dominant the rate is proportional to X and X attains a steady state concentration approximated by X_{max}... Transition between dominance by (α) and (β) is strongly dependent on Y and takes place whenever that concentration passes through $Y_{critical}$." [7]

It is clear from the quote above that a significant part of what it means to "understand a mechanism" is embodied in this sort of narrative understanding. The chemist can observe the result of a simulation (or laboratory experiment) and discern a sequence of significant events — e.g., rapid jumps in concentrations, or long periods of monotonic growth or decline of certain concentrations. These significant events may be seen as part of a larger repertoire of qualitative behavior in chemical mechanisms. Moreover, the chemist's understanding is not limited to the observation of significant events; he is also able to relate those events to the activity of individual reactions of the hypothesized mechanism.

The Kineticist's Workbench contains modules whose purpose is to support this kind of reasoning and interpretation by numerically simulating a given mechanism and then constructing a qualitative history of the reaction. The elements of this qualitative history are precisely the kinds of significant events mentioned above — rapid jumps or decreases in concentrations, periods of nearly-constant concentrations, and so forth. Additionally, the program relates these events to the relative numerical contributions of the individual steps in the given mechanism.

Before going into the way in which the Workbench program constructs qualitative histories, it is worth presenting a simple example of the program in operation. Suppose the chemist wishes to simulate the sample mechanism described earlier, and reproduced below:*

---

* The symbols k1, k2, and k3 in this mechanism denote *rate constants* for the elementary steps (4.1)-(4.3). Briefly, a rate constant is a proportionality constant for the differential equation terms generated by a given elementary step. For instance, the rate of change of species A due to reaction (4.1) can be written `d[A]/dt = -k1[A]`. For more explanation, see Laidler[5].

```
(4.1) A --> B   k1 = 20
(4.2) B --> C   k2 = 2
(4.3) C --> B   k3 = 1
```

The chemist enters this mechanism into the program as follows: *

```
Number of elementary reactions: 3
Step 1:  (A   --> B    20)
Step 2:  (B   --> C     2)
Step 3:  (C   --> B     1)

Initial concentrations: ((A 5) (B 0) (C 0))
Starting and ending times for simulation: (0 10)
```

Once this information has been entered (as well as some parameters for graphing routines), the program then simulates the mechanism and produces the graph of concentrations for species A, B, and C shown in Figure 1. The program then constructs and prints out the following qualitative history of the simulation:

```
Qualitative History of the Simulation:

No oscillations observed.

Episode number: 1  (short)
   Starting time: .01
   Species A: Large rapid monotonic decrease.
              Most important step(s): (1)
   Species B: Large rapid increase.
              Most important step(s): (1)
   Species C: Large rapid monotonic increase.
              Most important step(s): (2)

Episode number: 2
   Starting time: .14
   Species A: Large rapid monotonic decrease.
              Most important step(s): (1)
```

* Note that units are not included here; as long as the choices of units for concentrations, times, and rate constants are mutually consistent, there is no need to specify them for the Workbench. For example, we might stipulate that concentrations are in moles/liter, time in seconds, and all three (first order) rate constants in seconds^{-1}.

```
   Species B: Large rapid monotonic decrease.
              Most important step(s): (2)
   Species C: Large monotonic increase.
              Most important step(s): (2)

Episode number: 3  (long; final)
   Starting time: 1.73
   Species A: Large rapid monotonic decrease.
              Most important step(s): (1)
   Species B: Slow monotonic decrease. Steady state.
              Most important step(s): (2 3)
   Species C: Slow monotonic increase. Steady state.
              Most important step(s): (2 3)
```

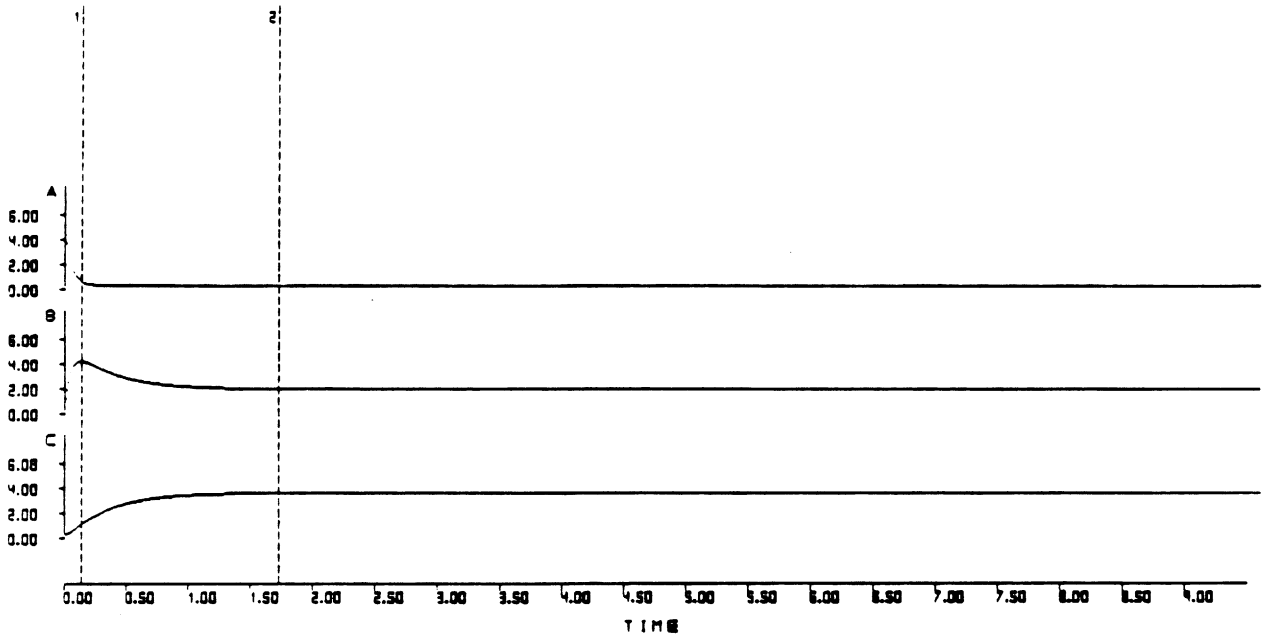

Figure 1: [A], [B], and [C] versus time. The dashed vertical lines are episode boundaries (see next section).

The Workbench's response is worth elaborating upon. First, note that it has classified the events of the simulation into "episodes"; the meaning of this term will be explained in the following section, but for now suffice it to say that the Workbench has divided the just-performed simulation into three major

periods of time. During the first of these periods, the concentrations of B and C rise quickly (C's concentration rises monotonically), and the concentration of A declines quickly and monotonically; we are also told that this first period is brief. The second period is similar, but the rate of growth of C is smaller than before, and this "episode" is longer. Finally, the last period is lengthy; here, the concentrations of B and C are both near-constant, as suggested by the program's notation that both seem to be at a steady state.* The "important steps" information provided by the program will be explained in the next section.

The reader might compare the Workbench's summary of the simulation results with his or her own assessment of the graphs in Figure 1. In essence, the program is attempting to capture those aspects of the graph that would be included in, say, a lab assistant's report.

C. Simulation Episodes

How does the Workbench construct a qualitative explanation of its results like the one shown above? The basic data structure used by the Workbench in this process is called an *episode*. An episode is a period of time during which the relative importance of each term in the differential equations governing the rate of change of the various species remains constant. To take an example, consider the differential equation governing the concentration of species B in the system above:

$$\frac{d[B]}{dt} = 20[A] - 2[B] + [C] \qquad (5)$$

There are three terms on the right side of this differential equation. Suppose that at a given moment in some simulation we have concentrations of 1, 0.2, and 0.1 for species A, B, and C respectively (as before, the particular units chosen are irrelevant as long as they are mutually consistent). Then the three terms

* This is not inconsistent with the program's note that both B and C are changing monotonically in concentration; it just happens that the changes are so small that they do not affect the judgment that both species are at steady-state concentrations. Similarly, A's concentration, although varying little in absolute terms during the last two episodes, is not deemed to be at a steady state, since during each episode period A declines by a large amount relative to its concentration at the beginning of the episode.

on the right have absolute values of 20, 0.4, and 0.1 in this case. Thus, the term having the largest local effect on the rate of change of [B] is the first term, 20[A]; the next most important term is the second, and the least important is the third. We can construct similar orderings for [A] and [C]; and this complete set of orderings is the defining characteristic for the current episode at this moment of the simulation. If any or all of these orderings should change, we regard this as the signal that a new episode in the simulation has begun. In Figure 1 above, episode boundaries are depicted as vertical lines in the graph of concentrations of species A, B, and C.

As the Workbench program performs a simulation, it simultaneously constructs ongoing orderings of the type shown above for each species of interest in the system. If any of these orderings should change during the simulation, the Workbench notes this fact as the beginning of a new episode. Episode structures thus constitute a way of demarcating regions of interest "on the fly" during a simulation. In the example shown in the previous section, the three episodes corresponded to changes in the contributions of the three terms in (5) above. During the first episode, the 20[A] term was dominant; during the second, the 2[B] term was dominant; and during the third the two terms 2[B] and [C] were so close in value that they were deemed to be of approximately equal importance in determining the derivative of [B].

For each episode, the Workbench program maintains additional information about events during that episode. For instance, the Workbench records initial and final concentrations of each species during an episode; it also records maximum and minimum concentrations, maximum and minimum derivatives, and a few additional features. Thus, the sequence of episodes constructed during a simulation constitutes a basis for a qualitative reconstruction of the simulation history: by examining the data retained in the episode structures, the Workbench can reconstruct—approximately—the events of the original simulation. It is just this kind of reconstruction that is shown in the example above, where the program prints out a "qualitative history" of the simulation: here, the program scans the episode structures created during the course of the simulation and identifies features such as the "large decrease in A" during the first stage of the reaction. The printed history also indicates, for each species, which differential equation term had the largest absolute value during a particular episode: this is shown in the "important reaction" field of the output.

There are of course other ways in which we could attempt to demarcate regions of interest in a reaction simulation: we might, for example, look for

instances of near-zero first or second derivatives for certain species. The major reason for using "episodes" as the natural units of simulation histories is that this notion seems to capture some of the chemist's natural intuition about the relationship between the mechanism and the simulation. That is to say, the chemist is not merely interested in where large or small derivatives occur; she is concerned with how the events she has just witnessed in the course of a simulation are determined by the varying contributions of the steps within the mechanism.* (Again, see the quote from Noyes [7] above for an instance of this type of intuition.) The episode data structure provides other advantages that will be touched upon in the following section.

D. The Episode Data Structure

The episodes created by the Workbench are composite structures containing the following information:

- The time (since the simulation's beginning) that the episode starts.
- The duration of the episode.
- Concentrations of species at the beginning of the episode.
- Ordering of differential equation terms defining the episode.
- Maximum/minimum species concentrations during the episode.
- Maximum/minimum species derivatives during the episode.

In addition, episodes may contain some special information; for example, the final episode constructed during a simulation will also contain the final concentrations of all species.

This information is used to construct the type of qualitative history shown earlier. For example, an episode in which the minimum derivative for a given species is positive is known to be monotonic in that species. Similarly, an episode in which the concentration of a species increases by more than fifty percent over its starting value is said to have a large increase in that species. The duration of an episode is judged according to a time scale created by comparing the longest episode during a simulation to the briefest; a logarithmic "time ruler" is constructed according to the following formula:

---

* Although the complete term orderings that demarcate episode boundaries are not shown in the qualitative history output by the program, the chemist may examine them through other procedures if she wishes.

```
Time-ruler
  = log(longest episode duration) - log(shortest episode duration)
```

A long episode is one whose duration meets the following criterion:

```
log(duration) > log(shortest episode duration) + 0.85(Time-ruler)
```

Steady states and rapid increases/decreases are determined according to slightly more elaborate formulas that involve both the rates of change of species and the relative duration of an episode.

Obviously, these criteria for qualitative judgments (for length of episodes, or the relative sizes of species changes) are approximations only. Experience with the program to date indicates that these criteria work well with the examples tried so far. The user may of course examine the episode structures themselves to see how the qualitative judgments are arrived at; ultimately the interface will provide the user with the opportunity to create custom procedures for making qualitative judgments.

As a final point regarding the episode data structure, it should be noted that this structure provides a natural way in which to perform finer- or coarser-grained qualitative analyses. To wit: the episode boundaries are, at the finest grain, determined by changes in the relative orderings of the contributions of *all* terms within the differential equations for each species. It is possible, however, to define "coarser-grained" episode boundaries according to only the most important term in the various differential equations: in other words, we need not distinguish two episodes in which the most important term in each differential equation remains the same (the changes might have occurred in the ordering of terms with smaller absolute values). Similarly, we might define "medium-grain" episode boundaries by using only the identities of the two (or, in general n) most important terms in each differential equation. Thus the episode structure provides a meaningful dimension along which to look for more (or less) detail in a qualitative analysis of a given simulation.

III. An Extended Example: The Brusselator Mechanism

As a more challenging example of constructing qualitative histories, we can consider the so-called Brusselator mechanism [6, 11]:

```
(6.1)     A --> X
(6.2)     B + X --> Y + D
```

(6.3) $2X + Y \rightarrow 3X$

(6.4) $X \rightarrow E$

This four-step mechanism is a simple mathematical model illustrating the possibility of sustained oscillations in an open chemical system. The concentrations of A and B in the mechanism are assumed to be constant (implying that there are external sources of these species), and the species D and E may be treated as "driven off" and hence constant at zero. (These latter two species do not enter as reactants in any reaction step so that their concentrations do not affect those of the other species.) Thus, the state of the system determined by reactions (6.1) - (6.4) is determined entirely by the concentrations of X and Y, and the differential equations for these species may be written out as follows:*

$$\frac{dX}{dt} = k1[A] - k2[B][X] - 2k3[X]^2[Y] + 3k3[X]^2[Y] - k4[X] \tag{7.1}$$

$$\frac{dY}{dt} = k2[B][X] - k3[X]^2[Y] \tag{7.2}$$

By finding a fixed point for these two equations and linearizing about that point, one can find conditions on [A], [B], and the various rate constants such that the fixed point is unstable and the system supports sustained oscillations (corresponding to a stable limit cycle in the X,Y-phase plane).

For the examples discussed here, we will let the concentrations of A and B be fixed at 1 and 3 respectively, and we will let the values of k1, k2, and k4 all be equal to 1. Thus, the differential equations above may be written:

$$\frac{dX}{dt} = 1 - 3[X] - 2k3[X]^2[Y] + 3k3[X]^2[Y] - [X] \tag{8.1}$$

---

* It is worth noting that the first term in (7.1) is a constant, since [A] is constant. Similarly, the third and fourth terms are commonly combined, as they arise from the same reaction; but since the Workbench leaves these two terms separate (as a "product" term and a "reactant" term for the third step), I have written out (7.1) in an "expanded" form.

$$(8.2)\quad \frac{dY}{dt} = 3[X] - k3[X]^2[Y]$$

Using the mechanism (6.1)-(6.4) with the constraints expressed in (8.1)-(8.2), we can now use the Workbench to explore how the qualitative behavior of the Brusselator mechanism varies with the choice of the constant k3.*

With a value of 1 for k3, the Brusselator should reach a stable limit cycle. We simulate this mechanism with the Workbench (for a total of 50 seconds), and Figure 2 depicts the resulting graphs of concentration of X and Y over time. The Workbench produces the following qualitative history for this simulation (only the first six episodes are shown):

```
Qualitative History of the Simulation:

Apparent oscillations located.


Episode number: 1  (long)
   Starting time: .05
   Species X: Large monotonic increase.
              Most important step(s): (1)
   Species Y: Large monotonic increase.
              Most important step(s): (2)

Episode number: 2
   Starting time: 4.8
   Species X: Large monotonic increase.
              Most important step(s): (3)
   Species Y: Monotonic increase.
              Most important step(s): (2)


   *** Repeating Group ***

Episode number: 3  (short)
```

* For these examples, episode boundaries are determined only by the single most important term in each differential equation, corresponding to the "coarse-grained" analysis discussed in the previous section. This coarse graining appears to retain enough information to analyze the numerical results.

```
   Starting time: 8.1
   Species X: Large rapid increase.
              Most important step(s): (3)
   Species Y: Large rapid monotonic decrease.
              Most important step(s): (3)

Episode number: 4  (long)
   Starting time: 9.3
   Species X: Large rapid decrease.
              Most important step(s): (3)
   Species Y: Large monotonic increase.
              Most important step(s): (2)



   *** Repeating Group ***

Episode number: 5  (short)
   Starting time: 15.25
   Species X: Large rapid increase.
              Most important step(s): (3)
   Species Y: Large rapid monotonic decrease.
              Most important step(s): (3)

Episode number: 6  (long)
   Starting time: 16.45
   Species X: Large rapid decrease.
              Most important step(s): (3)
   Species Y: Large monotonic increase.
              Most important step(s): (2)
```

etc.

The Workbench has identified a series of repeating episode structures, each consisting of two episodes apiece, and beginning at time 8.1. Each repeating unit may be described roughly as a rapid increase in X and decrease in Y (over a brief timespan), followed by an increase in Y and rapid decrease in X (over a longer timespan). These periods continue until the end of the history, so we can conclude that the mechanism has achieved a stable limit cycle. As for the first two episodes in the history, these might be said to correspond to an "induction period" before the beginning of the oscillations proper.

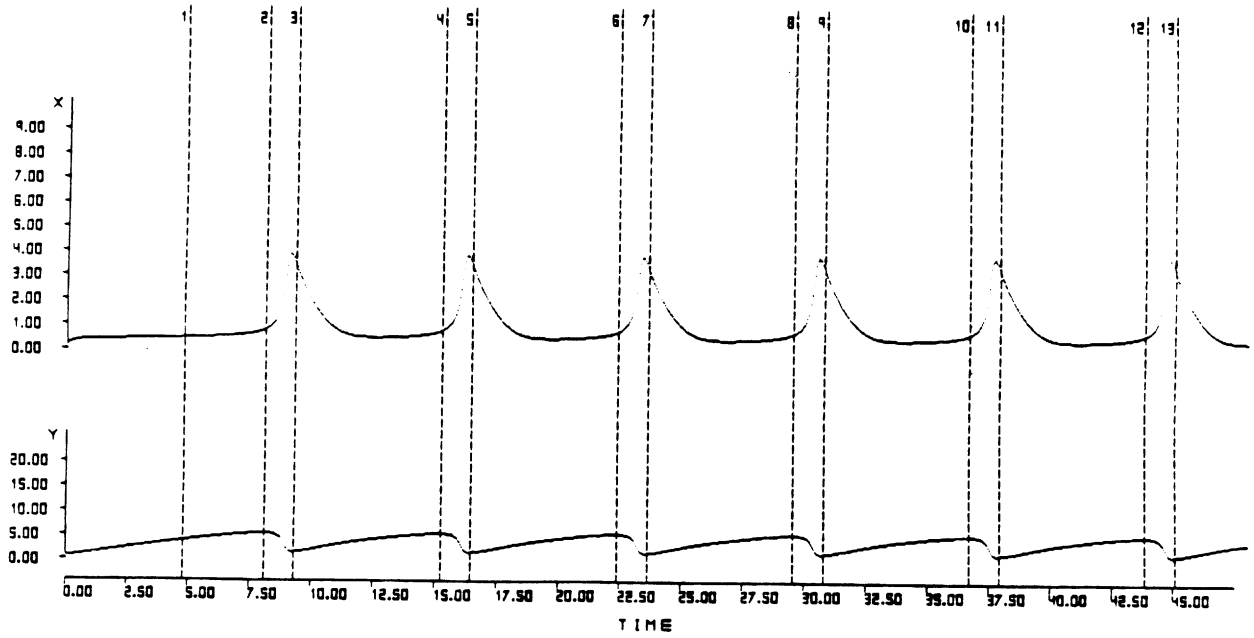

Figure 2: [X] and [Y] versus time; k3 = 1.

When we run the same mechanism with a value of 10 for k3, the Workbench produces the concentration graph shown in Figure 3, and the following qualitative history:

```
Qualitative History of the Simulation:

No oscillations observed.

Episode number: 1  (short)
   Starting time: .05
   Species X: Large rapid monotonic increase.
              Most important step(s): (1)
   Species Y: Large rapid monotonic increase.
              Most important step(s): (2)

Episode number: 2  (short)
   Starting time: .9
   Species X: Rapid monotonic increase.
              Most important step(s): (3)
   Species Y: Rapid increase.
              Most important step(s): (2)
```

```
Episode number: 3  (short)
   Starting time: 1.6
   Species X: Rapid monotonic increase.
              Most important step(s): (3)
   Species Y: Large rapid monotonic decrease.
              Most important step(s): (3)

Episode number: 4  (long; final)
   Starting time: 2.5
   Species X: Slow increase. Steady state.
              Most important step(s): (3)
   Species Y: Slow decrease. Steady state.
              Most important step(s): (3 2)
```

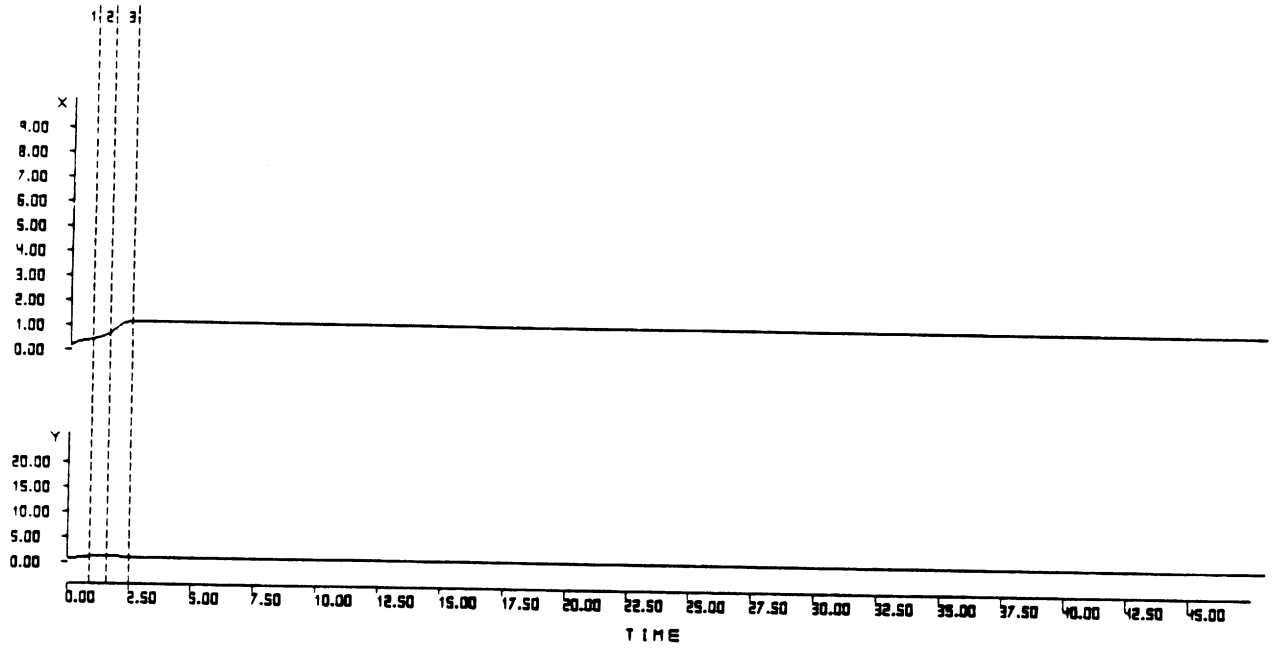

Figure 3: [X] and [Y] versus time; k3 = 10.

Here, the program finds only three brief episodes followed by a long period (starting at time 2.5) after which both X and Y seem to reach steady state concentrations. In point of fact, the equilibrium point is stable in the Brusselator for this value of k3, so the Workbench's analysis is qualitatively correct.

For a value of 2.1 for k3, the Brusselator's equilibrium point is stable, but is approached via damped oscillations.* The Workbench has more difficulty with this situation: the graph produced is shown in Figure 4. The qualitative history can be summarized as follows: the Workbench finds a brief two-episode oscillation beginning at time 4.5, and a second set of four-episode oscillations beginning at time 15.05 and persisting until nearly the end of the simulation. The final episode is deemed to be long, and to have steady-state concentrations of both X and Y; this suggests that the Workbench has correctly identified the arrival of the system at a stable equilibrium.

Below we reproduce the first two-episode oscillation found by the Workbench; the first four-episode oscillation; and the final episode, as taken from the qualitative history produced by the program.

Two-episode oscillation:

```
   *** Repeating Group ***

Episode number: 3
   Starting time: 4.5
   Species X: Large increase.
              Most important step(s): (3)
   Species Y: Large rapid monotonic decrease.
              Most important step(s): (3)

Episode number: 4
   Starting time: 5.8
   Species X: Large rapid decrease.
              Most important step(s): (3)
   Species Y: Large monotonic increase.
              Most important step(s): (2)
```

Four-episode oscillation:

```
   *** Repeating Group ***
```

* This is in contrast to the k3 = 10 situation, in which the equilibrium point is approached "directly"; the distinction has to do with the fact that at k3 = 2.1, the eigenvalues of the linearized system around the equilibrium point have a nonzero imaginary component.

```
Episode number: 8
   Starting time: 15.05
   Species X:
              Most important step(s): (3)
   Species Y: Monotonic increase.
              Most important step(s): (2)

Episode number: 9  (short)
   Starting time: 17.25
   Species X: Monotonic increase.
              Most important step(s): (3)
   Species Y: Slow decrease. Steady state.
              Most important step(s): (2 3)

Episode number: 10
   Starting time: 17.55
   Species X:
              Most important step(s): (3)
   Species Y: Monotonic decrease.
              Most important step(s): (3)

Episode number: 11  (short)
   Starting time: 19.15
   Species X: Monotonic decrease.
              Most important step(s): (3)
   Species Y: Slow increase. Steady state.
              Most important step(s): (3 2)
```

Final episode:

```
Episode number: 35  (long; final)
   Starting time: 44.65
   Species X: Slow decrease. Steady state.
              Most important step(s): (3)
   Species Y: Slow decrease. Steady state.
              Most important step(s): (3 2)
```

Part of the problem for the program here is that the Workbench does not as yet have any representation for trends within oscillations—such as "damped oscillations" or "unstable oscillations." (Examination of the episode history shows indeed that the peak concentrations of X and Y are both declining over the

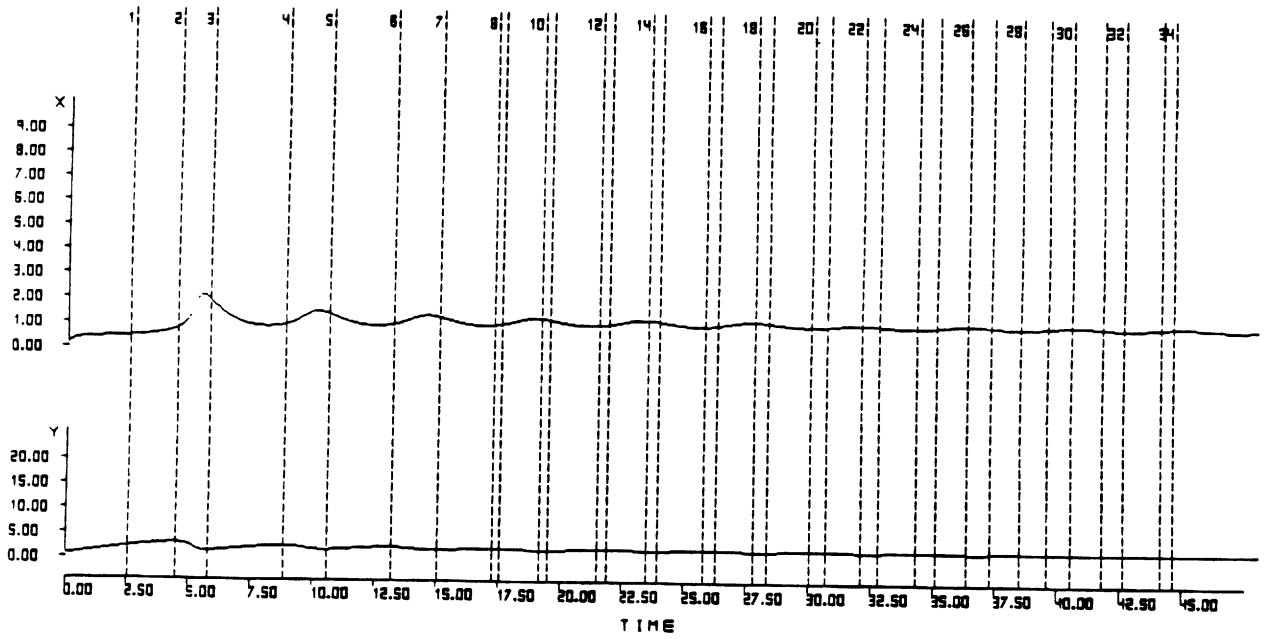

Figure 4: [X] and [Y] versus time; k3 = 2.1.

course of successive oscillations.) We will return to this problem in the discussion section below. It should however be noted that the Workbench has correctly located the presence of oscillations during the course of the simulation, and a steady state for both X and Y concluding the simulation.

IV. Discussion

At the moment, the Workbench program is still at a fairly early stage, and has only been tested with a handful of examples. The portion of the program that maintains and analyzes episode histories of simulations will no doubt undergo changes and elaborations as experience with the program grows. The results thus far, however, are encouraging in that they indicate that the feasibility of automating some of the routine qualitative interpretation usually left to the chemist.

To reiterate an earlier point, the major benefits of having the computer take over some of the duty of qualitative analysis is not merely so that these results may be presented directly to the user, but rather so that they may be used as inputs to other computer programs. For example we could imagine writing a

program that tells the Workbench to search a parameter space to find those regions in which a particular system exhibits bistability or oscillations. Similarly, the information maintained by the Workbench includes not only those reactions that are most *important* in determining the local changes in species concentrations, but also those which are most *unimportant.* Thus the Workbench could use the episode information to suggest which reactions in a given mechanism appear to contribute only little to the qualitative history of the mechanism; subsequently, the program could remove these reactions from the mechanism, resimulate the mechanism, and compare the qualitative behavior of the original and newly-simplified systems. In this way, the computer could be given the job of attempting (and assessing) some immediate candidates for simplification of mechanisms. Experiments along these lines in using the program have just begun.

More generally, we would like to integrate the data structures provided by qualitative analysis with the other portions of the Workbench currently under development. As currently envisioned, the Workbench will include a separate module that examines symbolic representations for mechanisms directly, and looks for ways in which these mechanisms may be decomposed or simplified directly. This module includes algorithms that perform graphical analysis of mechanisms (based on the work of M. Feinberg and others [2]); the module also includes heuristics for isolating poritons of mechanisms that appear to be likely candidates for simplification (e.g., pairs of opposing reactions that may constitute a "rapid equilibrium"). Thus, we would hope that eventually the Workbench might be given an initial mechanism, and propose (prior to numerical simulation) a candidate for a simplified mechanism as a result of graphical analysis; the original and simplified mechanisms could then be compared automatically to see if their qualitative behaviors are similar.

There are also certain weaknesses in qualitative analysis as shown here. Many of the terms employed ("large decrease," "rapid increase," and so forth) are by nature imprecise and context-dependent. Thus, the distinction between a concentration increase described as "rapid" and one not so described may hinge on small numerical differences. One possibility for augmenting the current program would be to introduce some finer qualitative distinctions (e.g., "very large increase") that are still coarse-grained enough to summarize large portions of numerical data; another approach to be pursued in parallel is to provide the user with the opportunity to examine the default criteria for qualitative judgments and change or expand these criteria at will. Beyond this, there are

many remaining difficulties to be overcome in having the Workbench recognize somewhat higher-level behavioral concepts such as "quasiperiodicity" (likewise, recall the problem with "damped oscillations" in the earlier example). Enriching both the vocabulary of a program like the Workbench, as well as its ability to analyze qualitative histories, represents a major challenge.

A more philosophical problem with the Workbench approach involves understanding the differences between the sort of qualitative analyses provided (at present) by this program and the kinds of qualitative classifications often treated as standard in dynamical systems theory. Typically, the important formal distinction between dynamical systems lies in notions such as topological conjugacy and stability of fixed points. These are important ideas but their relationship to other definitions of "qualitative behavior" is problematic in some respects. For instance, a topological classification of systems may be too strong for some purposes: it may distinguish between two systems that, for the chemist's purposes, are more interesting for their similarities than for their differences. A topological treatment would distinguish, say, between a system with a stable limit cycle and one with a stable focus whose eigenvalues have a very small (negative) real component; to the chemist, these systems are similar in that both exhibit qualitative oscillations over certain periods of time. Conversely, the topological treatment may fail to distinguish systems that the chemist regards as different: imagine a system in which the concentration of some species declines exponentially (and monotonically) to a zero value, and one in which the concentration first jumps to a maximum and then approaches a zero value. In both cases, the system has a simple, stable equilibrium, and both may in fact have the same equilibrium concentration values; but to the chemist, the difference in the path to equilibrium may be an important clue to the underlying mechanism.

The upshot of these considerations is that a useful tool should provide chemists with information about both the global characteristics of a chemical system's phase space trajectory, as well as information about the stability and uniqueness of equilibria. A particularly exciting prospect would be ultimately to augment the analyses of a program such as the Workbench with the kind of stability analyses made possible by symbolic algebra programs [8].

Acknowledgments

This work has benefited from the guidance and support of many people. Special thanks go to Harold Abelson, Gerald Sussman, Ken Yip, and Jacob Katzenelson.

Bibliography

[1] Byrne, George D.
Software for Differential Systems and Applications Involving Macroscopic Kinetics.
Computers and Chemistry, 5:4, pp. 151-158, 1981

[2] Feinberg, M.
Chemical Oscillations, Multiple Equilibria, and Reaction Network Structure.
In *Dynamics and Modelling of Reactive Systems*,
W. E. Stewart, W. Harmon Ray, and C. C. Conley, eds.
Academic Press, 1980

[3] Francisco, J. F.; Hase, W. L.; and Steinfeld, J. I.
Chemical Kinetics and Dynamics
Prentice-Hall, 1989

[4] Guckenheimer, J. and Holmes, P.
Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields
Springer-Verlag, 1983

[5] Laidler, K. J.
Chemical Kinetics
Harper and Row, 1987

[6] Nicolis, G. and Prigogine, I.
Self-Organization in Nonequilibrium Systems
John Wiley, 1977

[7] Noyes, R. M.
Mechanisms of Chemical Oscillators.
In *Synergetics: Far from Equilibrium*, A. Pacault and C. Vidal, eds.
Springer-Verlag, 1979

[8] Pavelle, R.
MACSYMA: Capabilities and Applications to Problems in Engineering and the

Sciences.
In *Applications of Computer Algebra*, R. Pavelle, ed.
Kluwer, 1985

[9] Rössler, O. E.
Chaos and Strange Attractors in Chemical Kinetics.
In *Synergetics: Far from Equilibrium*, A. Pacault and C. Vidal, eds.
Springer-Verlag, 1979

[10] Shacham, M.
Comparing Software for the Solution of Systems of Nonlinear Algebraic Equations Arising in Chemical Engineering.
Computers and Chemical Engineering, 9:2, pp. 103-112, 1985

[11] Thompson, J. M. T. and Stewart, H. B.
Nonlinear Dynamics and Chaos
John Wiley, 1986

This blank page was inserted to preserve pagination.

Scanning Agent Identification Target

Scanning of this document was supported in part by the **Corporation for National Research Initiatives,** using funds from the **Advanced Research Projects Agency** of the **United states Government** under Grant: **MDA972-92-J1029.**

The scanning agent for this project was the **Document Services** department of the **M.I.T Libraries.** Technical support for this project was also provided by the **M.I.T. Laboratory for Computer Sciences.**

Scanned

Date: 1/9/1995

M.I.T. Libraries
Document Services

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AD-A214678

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AIM 1171	2. GOVT ACCESSION NO.: AD-A214678
3. RECIPIENT'S CATALOG NUMBER:	
4. TITLE (and Subtitle): Descriptive Simulation: Combining Symbolic and Numerical Methods in the Analysis of Chemical Reaction Mechanisms	5. TYPE OF REPORT & PERIOD COVERED: memorandum
	6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): Michael Eisenberg	8. CONTRACT OR GRANT NUMBER(s): N00014-86-K-0180
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Artificial Intelligence Laboratory, 545 Technology Square, Cambridge, MA 02139	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:
11. CONTROLLING OFFICE NAME AND ADDRESS: Advanced Research Projects Agency, 1400 Wilson Blvd., Arlington, VA 22209	12. REPORT DATE: September 1989 13. NUMBER OF PAGES: 25
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office): Office of Naval Research, Information Systems, Arlington, VA 22217	15. SECURITY CLASS. (of this report): UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report)

Distribution is unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

None

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

symbolic computation
chemical kinetics

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

see reverse

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Block 20 cont'd

The Kineticist's Workbench is a computer program currently under development whose purpose is to help chemists understand, analyze, and simplify complex chemical reaction mechanisms. This paper discusses one module of the program that numerically simulates mechanisms and constructs qualitative descriptions of the simulation results. These descriptions are given in terms that are meaningful to the working chemist (e.g., steady states, stable oscillations, and so on); and the descriptions (as well as the data structures used to construct them) are accessible as input to other programs.