# Constructing useful statistical master equations for clusters

R. Stephen Berry *, Jun Lu, Chi Zhang

*Department of Chemistry, The University of Chicago, James Franck Institute, 5735 South Ellis Avenue, Chicago, IL 60637, USA*

## Abstract

Models and computational experiments provide insights into ways to construct and use master equations to describe kinetics of very complex systems. The goal of the work is establishing reliable ways to create such equation systems for small statistical samples of very large complex systems, that are good mimics of their full but unmanageable counterparts. Two converging approaches are described: one is based on any of a set of very simplified models of a complex potential surface; the other is based on simplifying the representation of the realistic surface itself. The results from systems of moderate size indicate that optimism is justified, regarding attainment of the goal for very large systems.
© 2005 Published by Elsevier B.V.

Describing the time evolution of atomic systems presents a set of challenges that probably cannot be met with one unified approach or method. It is now straightforward to describe the detailed dynamics of such systems by solving explicit equations of motion, whether classical or quantum-mechanical. However this is done at the cost of describing only rather brief intervals of time, of exploring only as much of the system's available phase space as can be reached in that interval, and under the constraint that the system cannot consist of very many constituent particles. (This last is not necessarily a severe limitation; astronomers do follow the evolution of large clusters by solving their coupled equations of motion.) Monte Carlo methods get around the time limitation by surrendering all constraints related to how a system gets from one place at one time to the next place at the next time; they do allow the system to explore its full phase space. This is particularly the case with variations of Monte Carlo search procedures that generate steps across long intervals in the system's configuration or phase space. But this is done at the cost of getting

no information about how real processes occur. Another way to approach the problem, the one we address here, is to use the formalism of *kinetics*, an approach more coarse-grained than explicit dynamics methods but one that makes possible the description of the system's behavior over arbitrarily long time intervals, and retains some but not all of the information about how the system actually evolves. In that sense, kinetics is an approach less detailed than explicit dynamics and less efficient at searching a phase space than Monte Carlo methods, but provides an extremely useful kind of information.

We think of a kinetic description as one that tells us the time history of a system or an ensemble undergoing some kind of reaction or relaxation, starting from an initial condition or population distribution that we can take as given. The description itself is expressed in terms of the populations of regions of local stability, that is, of the regions around local minima on the system's multidimensional effective potential energy surface. Each *state* of the system corresponds to a population around a specific local minimum. It is usually assumed, and can frequently be shown to be so, that the vibrational modes of the system attain thermal equilibrium on a time scale short relative to that of state-to-state passages. The

---

kinetic description begins with the rate equations for the time evolution of the population $P_i(t)$ (or probability) at time $t$, of each local state $i$, when an initial population distribution $P_i(0)$ is given. With the assumption of thermal equilibration within each state, one can use transition state theory to estimate the well-to-well or state-to-state rate coefficients. For clusters, this has been shown to be plausible [1].

The set of coupled rate equations for the populations is the well-studied "Master Equation" [2]. This set of equations can be expressed in terms of a matrix of rate coefficients. The eigenvalues $\lambda_i$ of this matrix are (negative) rate coefficients for simple exponential relaxation of the populations to their equilibrium values; one eigenvalue is zero, and its eigenvector is the population distribution at equilibrium. (This assumes that the elements of the matrix are not separable into disconnected blocks.) All the other eigenvalues are nonzero; those closest to zero correspond, of course, to the slowest processes. Explicitly, the populations, in terms of the eigenvalues and corresponding eigenvectors $\boldsymbol{u}_i$ are

$$P_i(t) = \sum_{j=1}^{N} u_i^{(j)} e^{\lambda_j t} \left[ \sum_{m=1}^{N} u_m^{(j)} \frac{P_m(0)}{P_m^{eq}} \right] \quad (1)$$

Master equations can be constructed readily for systems with tens or even hundreds of local minima. However the number of minima on a potential surface increases rapidly with the number $n$ of particles in the system. The number of minima for systems of atoms bound by Lennard–Jones pairwise interactions grows roughly as $n! \exp(n)$, probably even a little faster. This means that even a cluster of 20 atoms has so many minima that it is impractical to write a master equation for the full potential surface. Instead, one must find a way to sample the potential surface and use the master equation based on such sampling to understand the behavior of the full system. Of course the slowest processes are typically the most important, and the very fast processes, of very little chemical interest. This enables us to define a problem: How can we construct a statistical sample of a complex, many-dimensional potential surface so that the master equation based on that sample has eigenvalues, especially those nearest zero, that mimic the eigenvalues of the full system? One would also like the eigenvectors of the sample system to be similar to those of the full system, but that is a problem of a different magnitude and will not be addressed here.

One step toward finding a solution to this problem was taken with the clusters $Ar_{11}$ and $Ar_{13}$, systems small enough that one can work with their full potential surfaces but large enough that one can construct meaningful statistical samples of their topographies. This work [3] established sufficient conditions for a statistical sample to reproduce the main features of the kinetics based on the full surface and also the kinetics exhibited in molecular dynamics simulations. However these systems are small enough that one cannot readily generalize and extrapolate the conditions to systems significantly larger.

Here, we use two methods to learn how to use sample distributions to construct master equations whose eigenvalues, especially those of the slowest processes, are representative of the full system. In so doing, we gain useful insights into the kinetics of complex systems. The first method is based on variations of a "toy" net of linked states; the second is based on the realistic minima and saddles of Lennard–Jones clusters. The simplest version of the "toy" model consists of layers of local states, each layer with a fixed energy; the energy spacings between levels and all the saddle energies are equal. More complex variants change these assumptions, varying the numbers of states from level to level, the magnitudes of the level spacings and the saddle energies, and the ways the levels are linked. Fig. 1 is a schematic representation of a bit of the simplest version; it shows only 12 states in three tiers; in the calculations, the basic model has 50 levels, each with 50 states.

This model enables one to study the stability of the highest eigenvalues (those nearest zero) as a function of the size and nature of the subsystem chosen as the statistical sample. The results are encouraging; while eigenvalue spectra based on smaller and smaller subsystems give poorer and poorer representations of the behavior of the fast processes, the sparse eigenvalue spectrum in the slow-process region persists rather well even when quite small subsystems are used. Fig. 2 shows the forms of the eigenvalue spectra for the full model of 50 columns × 50 rows and for smaller subsets. In (panel a), the number of columns is reduced; in (panel b), the number of layers is reduced. The important information revealed here is the persistence of the long, flat tail at the right ends of the curves. This suggests that small subsets
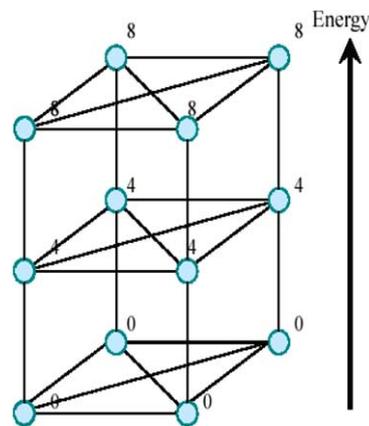


Fig. 1. A schematic picture of a part of the simplest model for the statistical master equation: equal energy intervals between layers, simple links as shown between states. All the saddle or barrier energies are also assumed identical in the simplest version.
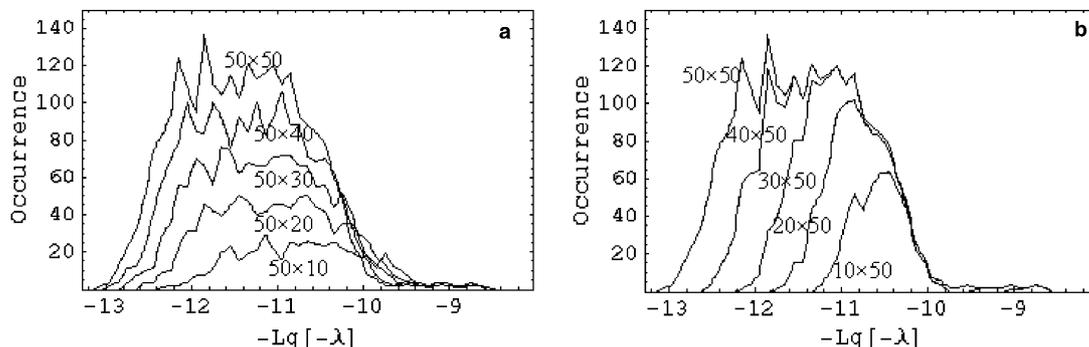
Fig. 2. Distributions of the eigenvalues for the full $50 \times 50$ model and for smaller subsets of the states thereof. Note the similarity of all the curves in the regions of their largest eigenvalues, at right, in each panel. (In panel (a) (left), the number of columns is reduced; in panel (b) (right), the number of layers is reduced.).

of sequences of minima and saddles for realistic models have a reasonable likelihood of yielding master equations whose high eigenvalues are similar to those of the full systems.

This method has also been used in other variant forms. One has the number of states increasing with the energy of each tier, making a sort of tree-like array. Another set have somewhat different energies or barriers for the states within an approximate layer; the energies of the saddles between levels in a given layer need not be the same as that from level to level. All these variations have led to approximately the same optimistic conclusion, that the slow eigenvalues are the most robust toward reductions in sample size, by far, to variations in the model. This all points toward the utility of the master equation approach for describing kinetics of quite complex systems.

The next step we report here is a jump to a realistic system, the $Ar_{13}$ cluster held by pairwise Morse forces, with a range to make the system as similar as possible to a Lennard–Jones counterpart. The surface has (at least) 1408 minima and 21,843 saddles connecting them,

mostly of rank 1. We construct sample subsets by choosing a maximum barrier height, lowest for the smallest subset, and find the minima that can be reached from a saddle with that barrier height. Hence a minimum may be picked in this way on one side of a saddle but not on the other. However we then include all minima that are linked to the saddles in this set. Clearly the higher the maximum allowable barrier, the more saddles and minima in the set. The subset sizes ranged from 128 minima up to the full set. Eigenvalues and eigenvectors were found for the master equations of all the sample systems, at several temperatures. For each sample system, we studied the relaxation from three initial conditions: a uniform population of the highest 10 minima, likewise of the highest 20 minima, and a uniform population of all the minima. The relaxation times for each of these initial conditions and for various temperatures $T$ were obtained from the corresponding eigenvalues and eigenvectors of the master equations. Master equations were obtained in two ways: in one, only the saddles were included that were no higher than the chosen barrier maximum; in the other, all saddles connected to the
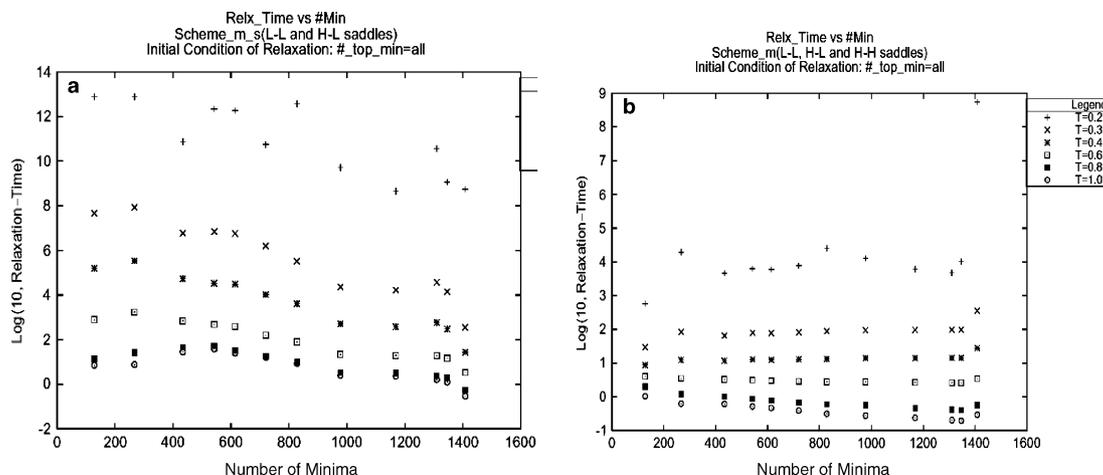


Fig. 3. Relaxation times for $Ar_{13}$ at various temperatures, $T$; (a) based only on the saddles presenting at least one barrier at or below the threshold defining the sample size and (b) with all barriers included that connect to the minima in the subset.
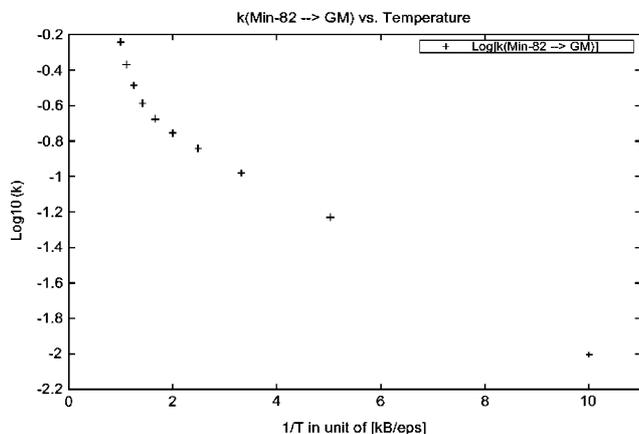
Fig. 4. Arrhenius plot, $\ln k(T)$ vs. $1/T$, for passage from a typical higher minimum to the global minimum of the potential surface of $Ar_{13}$. The curvature indicates that there are multiple pathways from the upper minimum to reach the global minimum.

admitted minima were included. Typical results of these calculations are shown in Fig. 3a and b. The first shows relaxation times based only on saddles with one low barrier, and the second, on all saddles. The results show clearly that high saddles do play a significant role, as we shall see below.

These results give strong support to the supposition that fairly reliable models of relaxation of complex systems can be obtained from master equations based on rather small subsets of the stationary points of full potential surfaces.

Finally, we use this system to demonstrate how important multiple pathways are, even for systems as small and apparently simple as this $Ar_{13}$ cluster. The simplest, clearest demonstration of multiple pathways is curvature in the dependence of the logarithm of the rate coefficient, $\ln k(T)$, on $1/T$ [4]. Classical Arrhenius theory, based on the concept that the lowest saddle must dominate because of the exponential dependence of rates of crossing on the activation energy, holds well for simple systems, but complex systems may, and frequently do, have so many high-energy pathways that these dominate, as an aggregate, over the lowest single path. The calculations for passage from one of the higher minima to the global minimum shown in Fig. 4, that even a system as small as this cluster can show such behavior.

We conclude that there are good reasons to be optimistic about the use of master equations based on small statistical samples of potential surfaces to give fairly reliable representations of the relaxation processes of complex systems. The hope is that the approach can be used to study such phenomena as the annealing of nanoparticles and the folding, unfolding and reactions of proteins.

## References

[1] B. Vekhter, K.D. Ball, J. Rose, R.S. Berry, J. Chem. Phys. 106 (1997) 4644.
[2] N.G. van Kampen, Stochastic Processes in Physics and Chemistry, North-Holland, Amsterdam, 1981.
[3] K.D. Ball, R.S. Berry, J. Chem. Phys. 111 (1999) 2060.
[4] M. Zaman, T.R. Sosnick, R.S. Berry, Phys. Chem. Chem. Phys. 5 (2003) 2589.