

# Calculating rate constants and committor probabilities for transition networks by graph transformation

David J. Wales<sup>a)</sup>

University Chemical Laboratories, Lensfield Road, Cambridge CB2 1EW, United Kingdom

(Received 13 February 2009; accepted 22 April 2009; published online 29 May 2009)

A graph transformation procedure is described that enables waiting times, rate constants, and committor probabilities to be calculated within a single scheme for finite-state discrete-time Markov processes. The scheme is applicable to any transition network where the states, equilibrium occupation probabilities, and transition probabilities are specified. For networks involving many states or slow overall kinetics, the deterministic graph transformation approach is faster and more accurate than direct diagonalization of the transition matrix, kinetic Monte Carlo, or iterative procedures. © 2009 American Institute of Physics. [DOI: 10.1063/1.3133782]

## I. INTRODUCTION

Transition networks are widely used to model the kinetics of complex systems in the physical sciences.<sup>1-3</sup> In this framework, we specify a set of states and the transition rates between them. Global dynamical properties are then usually analyzed using a Markov assumption, where transition probabilities only depend on the current state and not on how this state was reached. A pathway defined in terms of such transitions for a discrete, finite, or countably infinite-state space is called a Markov chain.<sup>4-6</sup> Transition networks of this type can be specified using data from molecular dynamics simulations to produce Markov state models.<sup>7-19</sup> Alternatively, transition states on the potential energy surface can be located using geometry optimization techniques, with rate constants between local minima defined using unimolecular rate theory. For example, in the discrete path sampling (DPS) approach,<sup>20-22</sup> a database of connected stationary points is systematically expanded<sup>23-25</sup> in an effort to represent the global kinetics by a finite set of states. A similar philosophy has been used by other groups in constructing connected stationary point databases for a variety of systems.<sup>1,26-30</sup>

Extracting the overall kinetic information from large transition networks can become computationally expensive or intractable, especially for slow dynamics corresponding to “rare events,”<sup>31-37</sup> which is the usual motivation for DPS studies. In the previous work, we have described how a graph transformation (GT) approach<sup>2,38,39</sup> can be used to overcome the slow convergence and numerical problems associated with iterative and stochastic schemes. In the present contribution, we define an alternative transformation, which enables us to calculate phenomenological two-state rate constants within a hierarchy of approximations in a single framework. This formulation therefore provides an efficient means to evaluate the accuracy of a steady-state (SS) approximation for states in the intervening region between the product and the reactant, and the extent to which local equilibration within the latter regions is achieved on the time scale of interest for the given network.

## II. RATE CONSTANT FORMULATIONS

Once we have adopted the Markov assumption for transitions between the states of the system, the resulting network can be represented as a graph<sup>4,10,40,41</sup> where nodes correspond to states and edges correspond to direct connections with nonzero transition probabilities. We define  $P_{\alpha\beta}$  as the transition probability from state  $\beta$  to state  $\alpha$ , where  $\sum_{\alpha} P_{\alpha\beta} = 1$  for a sum over the adjacent nodes that are directly connected to  $\beta$  (or all nodes, since the remaining direct transition probabilities are zero). This convention for the ordering of subscripts is adopted for transition probabilities and rate constants throughout the present contribution, following van Kampen.<sup>43</sup> Within this framework, we have previously derived a hierarchy of expressions for phenomenological two-state rate constants.<sup>2,20,22</sup>

In the DPS scheme, we calculate transition probabilities as  $P_{\beta\alpha} = k_{\beta\alpha} / \sum_{\delta} k_{\delta\alpha}$  using rate constants for directly connected minima,  $k_{\delta\alpha}$ . We normally estimate the  $k_{\delta\alpha}$ , partition functions for individual minima, and equilibrium occupation probabilities,  $p_{\alpha}^{\text{eq}}$ , using harmonic densities of states and transition state theory, but the present analysis is independent of how these parameters or the network itself is obtained. Treating the possible transitions from a given minimum,  $\alpha$ , as independent Poisson processes, the waiting time for any transition from that state follows a Poisson distribution<sup>42</sup> with escape rate  $\sum_{\delta} k_{\delta\alpha}$  and mean waiting time  $\tau_{\alpha} = 1 / \sum_{\delta} k_{\delta\alpha}$ , so that  $P_{\beta\alpha} = k_{\beta\alpha} \tau_{\alpha}$ .

The assumption of Markovian transitions between adjacent local minima enables the global kinetics to be written in terms of a set of linear master equations,<sup>43,44</sup>

$$\frac{dp_{\alpha}(t)}{dt} = \sum_{\beta} k_{\alpha\beta} p_{\beta}(t) - p_{\alpha}(t) \sum_{\beta} k_{\beta\alpha}, \quad (1)$$

where  $p_{\alpha}(t)$  is the occupation probability of minimum  $\alpha$  at time  $t$ . To define phenomenological two-state rate constants, we now classify the states as reactant,  $A$ , and product,  $B$ , with all other states assigned as intervening and belonging to a third set,  $I$ . If local equilibration is assumed to be fast in the  $A$  and  $B$  regions and the steady-state (SS) approximation is

<sup>a)</sup>Electronic mail: dw34@cam.ac.uk.

applied to the  $I$  set, we obtain the following expression for the two-state rate constant from  $A$  to  $B$ :<sup>20</sup>

$$k_{BA}^{SS} = \frac{1}{P_A^{\text{eq}}} \sum_{b \leftarrow a} \frac{k_{b i_1} k_{i_1 i_2} \cdots k_{i_n a} p_a^{\text{eq}}}{\sum_{\alpha_1} k_{\alpha_1 i_1} \sum_{\alpha_2} k_{\alpha_2 i_2} \cdots \sum_{\alpha_n} k_{\alpha_n i_n}},$$

$$= \frac{1}{P_A^{\text{eq}}} \sum_{b \leftarrow a} P_{b i_1} P_{i_1 i_2} \cdots P_{i_{n-1} i_n} P_{i_n a} p_a^{\text{eq}} \tau_a^{-1}, \quad (2)$$

where  $P_A^{\text{eq}} = \sum_{a \in A} P_a^{\text{eq}}$  is the equilibrium occupation probability for the  $A$  set. The corresponding expression for  $k_{AB}^{SS}$  is obtained by exchanging the labels corresponding to the  $A$  and  $B$  sets. The sum is over all paths that start and finish on the  $A$  and  $B$  boundaries, respectively, including arbitrary revisits to states in the  $I$  set.<sup>20</sup> The products of branching probabilities therefore provide a statistical weight for each path between states, and the sum defines a committor probability,<sup>45</sup>  $C_a^B$ , the probability that a random walk starting from  $a$  will encounter a  $B$  minimum before it visits an  $A$  state.<sup>38</sup>

Hence

$$k_{AB}^{SS} = \frac{1}{P_B^{\text{eq}}} \sum_{b \in B} \frac{C_b^A p_b^{\text{eq}}}{\tau_b} \quad \text{and} \quad k_{BA}^{SS} = \frac{1}{P_A^{\text{eq}}} \sum_{a \in A} \frac{C_a^B p_a^{\text{eq}}}{\tau_a}. \quad (3)$$

An alternative to the SS approximation is obtained by considering a state space restricted to the  $A$  and  $B$  sets to give<sup>38</sup>

$$k_{AB}^{\text{NSS}} = \frac{1}{P_B^{\text{eq}}} \sum_{b \in B} \frac{C_b^A p_b^{\text{eq}}}{t_b} \quad \text{and} \quad k_{BA}^{\text{NSS}} = \frac{1}{P_A^{\text{eq}}} \sum_{a \in A} \frac{C_a^B p_a^{\text{eq}}}{t_a}, \quad (4)$$

where the superscript ‘‘NSS’’ stands for ‘‘non-SS.’’ Here  $t_b$  is the average waiting time for transitions between minimum  $b$  and any minimum in the  $A$  or  $B$  sets, so that  $\tau_b \leq t_b$ . The SS limit is defined by  $\dot{p}_i(t) \rightarrow 0$  for all  $i \in I$ . In this limit  $t_b \rightarrow \tau_b$  because the waiting times in  $I$  minima are negligible, and we recover the  $k^{SS}$  results in Eq. (3).<sup>2</sup> Hence a comparison of these rate constants provides a quantitative assessment of the SS approximation for the static transition network in question, which is required for overall two-state kinetics to apply.

In fact, the SS formulation may be useful even when the SS approximation is not accurate because it enables us to break down the overall rate constants into a sum over paths directly. Paths that make the largest contributions to  $k_{AB}^{SS}$  or  $k_{BA}^{SS}$  can be extracted from the network using Dijkstra’s shortest-path algorithm<sup>46</sup> or the recursive enumeration algorithm.<sup>47,48</sup> To do this, the edge weights are chosen to match the product of branching probabilities in Eq. (2), and the remaining factor of  $p_a^{\text{eq}}/\tau_a$  or  $p_b^{\text{eq}}/\tau_b$  is included when comparing paths from different starting points.<sup>20,21</sup> Each path can therefore be written as a conditional occupation probability times a flux, as expected.<sup>49–52</sup> We have used such paths to choose pairs of local minima for subsequent connection attempts within the DPS framework,<sup>24,25</sup> identifying structures separated by the largest barriers, or corresponding to short minimized distances but lacking a direct connection.

We now introduce  $\mathcal{W}_\xi$ , the product of branching probabilities for a path  $\xi$ ,<sup>39</sup> which is useful in analyzing the conservation of the mean first-passage time (MFPT) in Sec. III B. If detailed balance is obeyed for every direct connection, then it is straightforward to show that

$\mathcal{W}_\xi p_a^{\text{eq}}/\tau_a = \mathcal{W}_{-\xi} p_b^{\text{eq}}/\tau_b$  for a path from state  $a$  to state  $b$ , where  $\mathcal{W}_{-\xi}$  is the product of branching probabilities for the reverse path.<sup>20</sup> Summing over paths between the two states then gives  $C_b^a p_a^{\text{eq}}/\tau_a = C_b^a p_b^{\text{eq}}/\tau_b$ , where  $C_b^a$  is the probability that a path starting from  $a$  encounters a minimum  $b$  before any other states in  $A \cup B$ , etc. These results lead to an overall detailed balance condition for the SS rate constants,  $k_{AB}^{SS} p_B^{\text{eq}} = k_{BA}^{SS} p_A^{\text{eq}}$ , as noted in earlier work.<sup>20</sup> This condition does not necessarily hold for  $k_{AB}^{\text{NSS}}$  and  $k_{BA}^{\text{NSS}}$ , or for the rate constants defined in Eq. (6) below. Hence deviation from the overall detailed balance condition provides a quantitative assessment of the SS approximation for the  $I$  minima.

The committor probabilities in Eqs. (3) and (4) can be obtained iteratively using a first-step analysis<sup>53</sup> with

$$C_\alpha^B = \sum_{\beta \in A} P_{\beta\alpha} C_\beta^B, \quad (5)$$

where transitions into reactant minima (here the  $A$  set) are specifically excluded so that nonzero values can be defined and calculated for  $C_a^B$  with  $a \in A$ .  $C_\alpha^A$  values are obtained analogously, and the same framework has been used in studies of proteins to calculate the folding probability,<sup>10</sup>  $P_\alpha^{\text{fold}}$ . We have previously implemented successive over-relaxation<sup>54</sup> (SOR) to obtain the committor probabilities using a compressed row storage scheme<sup>55</sup> for the sparse branching probability matrix. This approach is also used in the present work to check the results of the new GT (NGT) procedure.

The mean waiting time for a transition from state  $\alpha$  to a directly connected state,  $\tau_\alpha$ , is calculated directly, so  $k_{AB}^{SS}$  and  $k_{BA}^{SS}$  are available once the committor probabilities are known. To calculate the waiting times  $t_a$  and  $t_b$  that appear in  $k_{AB}^{\text{NSS}}$  and  $k_{BA}^{\text{NSS}}$ , we can average over multiple kinetic Monte Carlo (KMC)<sup>56–60</sup> runs using the rejection-free scheme of Bortz, Kalos, and Lebowitz (BKL),<sup>56</sup> as in previous work.<sup>38</sup> The simulation time is incremented by  $\tau_\alpha$  for each state encountered before an  $A$  or  $B$  minimum is reached, so that an average over different random number sequences provides the MFPT between the starting state and  $A$  or  $B$ .

The final formulation of the rate constants that we consider here is obtained by averaging the MFPT,  $T_{Ba}$ , for a product state in  $B$  to be reached over KMC trajectories initiated from each starting state,  $a$ , and similarly for interchange of  $A$  and  $B$ . The rate constants obtained by inverting these MFPT values represent reference values for the given network that depend only on the validity of the Markov assumption, assuming that all the transition rates and occupation probabilities are exact,

$$k_{AB} = \frac{1}{P_B^{\text{eq}}} \sum_{b \in B} \frac{p_b^{\text{eq}}}{T_{Ab}} \quad \text{and} \quad k_{BA} = \frac{1}{P_A^{\text{eq}}} \sum_{a \in A} \frac{p_a^{\text{eq}}}{T_{Ba}}. \quad (6)$$

We have previously referred to these rate constants using a superscript ‘‘KMC,’’ but this superscript is omitted here to emphasize that the values calculated from KMC, GT, and NGT are formally equivalent. Obtaining these rate constants using KMC is usually more computationally expensive than the corresponding SS and NSS calculations because the corresponding MFPT includes an arbitrary number of revisits to minima in the starting region, which increases exponentially

as  $k_B T$  decreases.<sup>38,61</sup> The KMC approach used to calculate  $t_a$  and  $t_b$  generally converges more rapidly because these runs terminate as soon as they reach either reactants or products. However, for sufficiently large databases, or overall barriers that are large compared to  $k_B T$ , we find that both the iterative calculation of the committor probabilities and the KMC calculation of  $\mathcal{T}_{Ba}$  or  $\mathcal{T}_{Ab}$  become unfeasible. In such situations we find that solution of the master equation by diagonalization of the transition matrix or weighted subspace projection methods<sup>62</sup> also encounters numerical problems, and attempts to evaluate  $k_{AB}^{SS}$  and  $k_{BA}^{SS}$  by matrix multiplication<sup>20,63,64</sup> converge too slowly. These problems motivated our previous development of the GT approach, as described below.<sup>2,38,39</sup> This framework provides  $k_{AB}$  and  $k_{BA}$  defined in Eq. (6) via a formally exact, deterministic algorithm, with a computational cost that is independent of the temperature and the time scale of the slowest relaxation. The new formulation, described in Sec. III B, provides the SS and NSS rate constants as a by-product, including the committor probabilities.

### III. GRAPH TRANSFORMATION

The GT approach removes states from the network one at a time, renormalizing branching probabilities and waiting times to preserve the overall kinetics. This theory extends the “leapfrog” moves to second neighbors described for KMC calculations.<sup>22</sup> It is also related to the absorbing Markov chains approach of Novotny,<sup>65</sup> the  $\tau$ -leap algorithm,<sup>66–68</sup> and to methods that eliminate “flickering” associated with low barriers in KMC trajectories.<sup>69–71</sup> The principal difference is that the GT procedure is applied successively to eliminate complete sets of minima while conserving the MFPT’s of interest.<sup>2,38,39</sup> There are also connections to the mean value analysis and aggregation/disaggregation techniques used in the performance and reliability evaluation of queueing networks,<sup>4,72–74</sup> and to dynamic graph algorithms.<sup>75–78</sup>

#### A. Previous formulation

In our previous implementation of the GT approach, we derived renormalized branching probabilities,  $P'_{\gamma\beta}$ , and waiting times,  $\tau'_\beta$ , for state  $\beta$  when one of its neighbors, state  $x$ , is removed.<sup>2,38,39</sup> The renormalized branching probabilities and waiting time for a transition from  $\beta$  to any state adjacent to  $x$  or  $\beta$ , excluding  $\beta$  and  $x$  themselves, were

$$P'_{\gamma\beta} = \frac{P_{\gamma x} P_{x\beta} + P_{\gamma\beta}}{1 - P_{\beta x} P_{x\beta}} \quad \text{and} \quad \tau'_\beta = \frac{\tau_\beta + P_{x\beta} \tau_x}{1 - P_{\beta x} P_{x\beta}}. \quad (7)$$

We proved that successively removing states with repeated renormalizations conserves the probability associated with  $a \leftrightarrow b$  paths. It does not conserve the individual MFPT’s because all steps out of a given state are associated with the same average time increment,  $\tau'_\beta$ . However, the MFPT averaged over product states is conserved,<sup>2,38,39</sup> as demonstrated in detail for the new scheme in Sec. III B.

Since the renormalization excludes transitions back to  $\beta$  itself, as for the BKL procedure,<sup>56</sup> all the diagonal branching probabilities are zero in this approach:  $P_{\beta\beta} = 0$ . In the first phase of the calculation, we remove all the  $I$  states. The corresponding branching probabilities and waiting times for

minimum  $\alpha \in A \cup B$  at this point correspond to transitions from  $\alpha$  to any other minima in  $A$  or  $B$ , aside from  $\alpha$  itself. Hence they do not correspond precisely to the committor probabilities and waiting times in the expressions for the NSS rate constants in Eq. (4), as noted before.<sup>2,38</sup> However, if there is only one reactant state, say  $a$ , then  $\sum_{b \in B} P'_{ba} = 1$ ,  $\tau'_a = \mathcal{T}_{Ba}$ , and  $k_{BA}^{NSS} = k_{BA}$ , once all the  $I$  states are removed. If there is more than one reactant state, then for each of them we remove all the other reactants using the same renormalization procedure, and this gives the appropriate  $\tau'_a \equiv \mathcal{T}_{Ba}$ .  $k_{BA}$  is then evaluated by summing over reactant states weighted according to the equilibrium occupation probabilities, as in Eq. (6).

#### B. New formulation

The new formulation, NGT, considered here differs from the previous version in that steps from the starting state back to itself, corresponding to  $P_{\beta\beta}$ , are allowed in the renormalization. It is this feature that allows us to identify the branching probabilities and waiting times after removal of the  $I$  states with the committor probabilities and waiting times in Eq. (4).

We again consider the effect of removing a state  $x$  on a pathway corresponding to a Markov chain of states, which arrives at state  $\beta$  adjacent to  $x$ . Let  $\Gamma$  be the set of states adjacent to either  $\beta$  or  $x$ , but excluding  $x$ . In contrast to the previous scheme,<sup>2,38,39</sup> here we include  $\beta \in \Gamma$  so that steps from  $\beta$  to  $\beta$  are included explicitly rather than renormalized away. We now define renormalized branching probabilities,  $P'_{\gamma\beta}$ , for minima  $\gamma \in \Gamma$  so that the probability of stepping from  $\beta$  to  $\gamma$  is conserved. Before  $x$  is removed the set of possible paths includes an arbitrary number of transitions from  $x$  to itself before the step to  $\gamma$ , so we need

$$P'_{\gamma\beta} = P_{\gamma\beta} + P_{\gamma x} P_{x\beta} \sum_{m=0}^{\infty} P_{xx}^m = P_{\gamma\beta} + \frac{P_{\gamma x} P_{x\beta}}{1 - P_{xx}}. \quad (8)$$

This renormalized transition probability from  $\beta$  to  $\gamma$  therefore preserves the overall probability of this step in the absence of state  $x$ . The sum of the branching probabilities from  $\beta$  is also conserved:

$$\begin{aligned} \sum_{\gamma \in \Gamma} P'_{\gamma\beta} &= \sum_{\gamma \in \Gamma} P_{\gamma\beta} + \sum_{\gamma \in \Gamma} \frac{P_{\gamma x} P_{x\beta}}{1 - P_{xx}} \\ &= 1 - P_{x\beta} + \frac{(1 - P_{xx}) P_{x\beta}}{1 - P_{xx}} = 1. \end{aligned} \quad (9)$$

The renormalized waiting time,  $\tau'_\beta$ , must allow for all the possible  $x \leftrightarrow x$  transitions.  $\tau'_\beta$  can be calculated directly by summing a geometric progression, but a more general approach is to replace each branching probability  $P_{\alpha\beta}$  by  $\tilde{P}_{\alpha\beta} = P_{\alpha\beta} \exp(\zeta \tau_\beta)$  and employ the result

$$\left[ \frac{d}{d\zeta} \tilde{P}_{\alpha_1 \alpha_2} \tilde{P}_{\alpha_2 \alpha_3} \tilde{P}_{\alpha_3 \alpha_4} \cdots \tilde{P}_{\alpha_{n-1} \alpha_n} \right]_{\zeta=0} \\ = P_{\alpha_1 \alpha_2} P_{\alpha_2 \alpha_3} P_{\alpha_3 \alpha_4} \cdots P_{\alpha_{n-1} \alpha_n} (\tau_{\alpha_2} + \tau_{\alpha_3} + \cdots + \tau_{\alpha_n}), \quad (10)$$

giving the total waiting time multiplied by the probability associated with the given chain. For any function of the branching probabilities collected in the matrix  $\mathbf{P}$ , we now define  $\tilde{f}(\mathbf{P}) = f(\tilde{\mathbf{P}})$ . The function  $\tilde{f}(\mathbf{P})$  is therefore obtained from  $f(\mathbf{P})$  by replacing every branching probability,  $P_{\beta\alpha}$  by  $\tilde{P}_{\beta\alpha}$ . Hence we obtain

$$\tau'_\beta = \sum_{\gamma \in \Gamma} \left[ \frac{d}{d\zeta} \tilde{P}'_{\gamma\beta} \right]_{\zeta=0} = \sum_{\gamma \in \Gamma} \left[ \frac{d}{d\zeta} \left( \tilde{P}_{\gamma\beta} + \frac{\tilde{P}_{\gamma x} \tilde{P}_{x\beta}}{1 - \tilde{P}_{xx}} \right) \right]_{\zeta=0} \\ = \tau_\beta + \frac{P_{x\beta} \tau_x}{1 - P_{xx}}. \quad (11)$$

This result can be verified directly by considering the series

$$P_{\gamma\beta} \tau_\beta + P_{\gamma x} P_{x\beta} [(\tau_\beta + \tau_x) + P_{xx}(\tau_\beta + 2\tau_x) + P_{xx}^2(\tau_\beta + 3\tau_x) \\ + \cdots] = P_{\gamma\beta} \tau_\beta + \frac{P_{\gamma x} P_{x\beta} (\tau_x + (1 - P_{xx}) \tau_\beta)}{(1 - P_{xx})^2}. \quad (12)$$

Summing over  $\gamma$  now gives

$$\sum_{\gamma \in \Gamma} \left( P_{\gamma\beta} \tau_\beta + \frac{P_{\gamma x} P_{x\beta} (\tau_x + (1 - P_{xx}) \tau_\beta)}{(1 - P_{xx})^2} \right) \\ = (1 - P_{x\beta}) \tau_\beta + \frac{(1 - P_{xx}) P_{x\beta} (\tau_x + (1 - P_{xx}) \tau_\beta)}{(1 - P_{xx})^2} \\ = \tau_\beta + \frac{P_{x\beta} \tau_x}{1 - P_{xx}} = \tau'_\beta. \quad (13)$$

We note here that direct transitions back into the same state that occur before any renormalization make no difference to any of the calculated rate constants.<sup>22</sup> When the states in the network are local minima of a potential energy surface, such transitions correspond to degenerate rearrangements.<sup>21,79</sup> Including such processes for a state  $\alpha$  simply rescales the initial branching probabilities,  $P_{\beta\alpha}$  for  $\beta \neq \alpha$ , and the waiting time for any transition,  $\tau_\alpha$ , by the same factor of  $1 - P_{\alpha\alpha}$ . The rate constant expressions only involve ratios of such quantities and are therefore unaffected.

The above transformations clearly conserve the probability and waiting time associated with the  $\gamma \leftarrow \beta$  transition locally in the network. We now consider the overall conservation of transition probabilities between particular reactant and product states and the MFPT between the reactant and the product set. Consider the family of paths that start at  $a \in A$ , arrive at minimum  $\beta \in \Gamma$ , step to  $\gamma \in \Gamma$  either directly or with an arbitrary number of revisits to  $x$ , and ultimately reach  $b \in B$ , where they terminate. Every path in this family is associated with a product of branching probabilities, and we wish to conserve the sum of these products by subsuming the visits to  $x$  into renormalized branching probabilities. The probability that  $\gamma \in \Gamma$  is the next member of the set  $\Gamma$  visited after  $\beta$  for the set of paths is  $P'_{\gamma\beta}$ , as calculated above. We can therefore remove state  $x$  and conserve the probability

associated with every  $a \leftrightarrow b$  set of paths by renormalizing the branching probabilities and waiting times of the states in  $\Gamma$ . Further states can be removed progressively by applying the same transformation to the current renormalized values of the branching probabilities.

Using  $\mathcal{W}_\xi$ , the product of branching probabilities for path  $\xi$ , the total probability that the  $b \leftarrow a$  transition occurs via one of the family of paths considered above can be written as

$$\sum_{\xi_3 \in b \leftarrow \gamma} \mathcal{W}_{\xi_3} \sum_{\xi_2 \in \gamma \leftarrow \beta} \mathcal{W}_{\xi_2} \sum_{\xi_1 \in \beta \leftarrow a} \mathcal{W}_{\xi_1} \\ = \sum_{\xi_3 \in b \leftarrow \gamma} \mathcal{W}_{\xi_3} \sum_{\gamma} (P_{\gamma\beta} + P_{\gamma x} P_{x\beta} + P_{\gamma x} P_{xx} P_{x\beta} \\ + P_{\gamma x} P_{xx}^2 P_{x\beta} + P_{\gamma x} P_{xx}^3 P_{x\beta} + \cdots) \sum_{\xi_1 \in \beta \leftarrow a} \mathcal{W}_{\xi_1} \\ = \sum_{\xi_3 \in b \leftarrow \gamma} \mathcal{W}_{\xi_3} \sum_{\gamma} P'_{\gamma\beta} \sum_{\xi_1 \in \beta \leftarrow a} \mathcal{W}_{\xi_1} \quad (14)$$

with  $\gamma \in \Gamma$ . The contribution to the MFPT from these paths is obtained by introducing  $\tilde{\mathcal{W}}_\xi$ , corresponding to  $\mathcal{W}_\xi$  where every branching probability  $P_{\alpha\delta}$  is replaced by  $\tilde{P}_{\alpha\delta} = P_{\alpha\delta} \exp(\zeta \tau_\delta)$ , as above. We therefore consider

$$\left[ \frac{d}{d\zeta} \sum_{\xi_3 \in b \leftarrow \gamma} \tilde{\mathcal{W}}_{\xi_3} \sum_{\gamma \in \Gamma} \tilde{P}'_{\gamma\beta} \sum_{\xi_1 \in \beta \leftarrow a} \tilde{\mathcal{W}}_{\xi_1} \right]_{\zeta=0} \\ = \sum_{\xi_3 \in b \leftarrow \gamma} \mathcal{W}_{\xi_3} \sum_{\gamma \in \Gamma} P'_{\gamma\beta} \sum_{\xi_1 \in \beta \leftarrow a} \left[ \frac{d\tilde{\mathcal{W}}_{\xi_1}}{d\zeta} \right]_{\zeta=0} \\ + \sum_{\xi_3 \in b \leftarrow \gamma} \mathcal{W}_{\xi_3} \sum_{\gamma \in \Gamma} \left[ \frac{d\tilde{P}'_{\gamma\beta}}{d\zeta} \right]_{\zeta=0} \sum_{\xi_1 \in \beta \leftarrow a} \mathcal{W}_{\xi_1} \\ + \sum_{\xi_3 \in b \leftarrow \gamma} \left[ \frac{d\tilde{\mathcal{W}}_{\xi_3}}{d\zeta} \right]_{\zeta=0} \sum_{\gamma \in \Gamma} P'_{\gamma\beta} \sum_{\xi_1 \in \beta \leftarrow a} \mathcal{W}_{\xi_1}.$$

Only the middle term changes when we renormalize after removing state  $x$  from this part of the family of paths,

$$\sum_{\xi_3 \in b \leftarrow \gamma} \mathcal{W}_{\xi_3} \sum_{\gamma \in \Gamma} \left[ \frac{d\tilde{P}'_{\gamma\beta}}{d\zeta} \right]_{\zeta=0} \sum_{\xi_1 \in \beta \leftarrow a} \mathcal{W}_{\xi_1} \quad (15) \\ = \sum_{\xi_3 \in b \leftarrow \gamma} \mathcal{W}_{\xi_3} \sum_{\gamma \in \Gamma} \left( P_{\gamma\beta} \tau_\beta + \frac{P_{\gamma x} P_{x\beta} (\tau_x + (1 - P_{xx}) \tau_\beta)}{(1 - P_{xx})^2} \right) \\ \times \sum_{\xi_1 \in \beta \leftarrow a} \mathcal{W}_{\xi_1}, \quad (16)$$

as for Eq. (12). If we now sum over product states, then  $\sum_{b \in B} \sum_{\xi_3 \in b \leftarrow \gamma} \mathcal{W}_{\xi_3} = 1$  for any  $\gamma$ , and the sum over  $\gamma$  becomes  $\tau'_\beta = \sum_{\gamma} P'_{\gamma\beta} \tau'_\beta$ , as in Eq. (13). The MFPT from reactant state  $a$  to product states  $B$  is therefore conserved if we associate the renormalized waiting time  $\tau'_\beta$  with the renormalized branching probability  $P'_{\gamma\beta}$  for all  $\gamma \in \Gamma$ , and we can apply this transformation to any section of a family of trajectories that passes through the set  $\Gamma$ . The individual MFPT's to particular product states are not conserved (unless there is only one



product state) because every step  $\gamma \leftarrow \beta$  is associated with the same averaged  $\tau_\beta^I$ .

The first part of an NGT calculation involves removing all the  $i \in I$  states one at a time, renormalizing at each step. We identify the renormalized branching probabilities and waiting times at the end of this phase by a superscript I. These branching probabilities can be summed to obtain committor probabilities

$$\sum_{b \in B} P_{ba}^I = P_{Ba}^I \equiv C_a^B \quad \text{and} \quad \sum_{a \in A} P_{ab}^I = P_{Ab}^I \equiv C_b^A. \quad (17)$$

In the previous GT scheme, the branching probabilities did not allow return to the same state, and so the above connection did not exist. In fact, the renormalized branching probabilities and waiting times after all  $I$  states have been removed in the previous scheme correspond to the NGT values divided by  $1 - P_{\alpha\alpha}^I$  for the state in question. The ratios required for the calculation of  $k_{AB}^{\text{NSS}}$  and  $k_{BA}^{\text{NSS}}$  are therefore available in both schemes.

The waiting times  $t_a$  and  $t_b$  required in Eq. (4) correspond to the average waiting time for a transition out of the state in question to any member of the  $A$  and  $B$  sets. This is simply the renormalized waiting time for the  $a$  or  $b$  state in question once all the  $I$  states have been removed, so that  $t_a = \tau_a^I$  and  $t_b = \tau_b^I$  at this point. We therefore obtain

$$k_{AB}^{\text{SS}} = \frac{1}{P_B^{\text{eq}}} \sum_{b \in B} \frac{P_{Ab}^I P_b^{\text{eq}}}{\tau_b} \quad \text{and} \quad k_{BA}^{\text{SS}} = \frac{1}{P_A^{\text{eq}}} \sum_{a \in A} \frac{P_{Ba}^I P_a^{\text{eq}}}{\tau_a}, \quad (18)$$

$$k_{AB}^{\text{NSS}} = \frac{1}{P_B^{\text{eq}}} \sum_{b \in B} \frac{P_{Ab}^I P_b^{\text{eq}}}{\tau_b^I} \quad \text{and} \quad k_{BA}^{\text{NSS}} = \frac{1}{P_A^{\text{eq}}} \sum_{a \in A} \frac{P_{Ba}^I P_a^{\text{eq}}}{\tau_a^I}.$$

The waiting times required for  $k_{AB}^{\text{SS}}$  and  $k_{BA}^{\text{SS}}$  are just the expected waiting times for a transition from the given state to any directly connected state in the original database, i.e.,  $\tau_a = 1 / \sum_{\delta} k_{\delta a}$  and  $\tau_b = 1 / \sum_{\delta} k_{\delta b}$ .  $\tau_a$  and  $\tau_b$  therefore correspond to the initial values of the waiting time for these states before any renormalization.

To complete the calculation of  $k_{AB}$  and  $k_{BA}$ , we save the renormalized branching probabilities and waiting times after removing all the  $I$  states. Then for each member of the  $A$  and  $B$  sets, we remove all other members of the same set to produce final renormalized branching probabilities and waiting times that we will label with a superscript F. Hence, for a given state  $a \in A$ , we obtain  $P_{ba}^F$  for every  $b \in B$ , together with  $\tau_a^F$ . At this point,  $P_{aa}^F = 1 - \sum_{b \in B} P_{ba}^F = 1 - P_{Ba}^F$  and  $\tau_a^F$  corresponds to the mean waiting time for a transition to any  $b \in B$  or back to  $a$ . The MFPT for a transition from  $a$  to  $B$  is therefore

$$\tau_a^F P_{Ba}^F (1 + 2P_{aa}^F + 3(P_{aa}^F)^2 + 4(P_{aa}^F)^3 + \dots) = \tau_a^F P_{Ba}^F / (1 - P_{aa}^F)^2 = \tau_a^F / P_{Ba}^F. \quad (19)$$

Hence for Eq. (6) we identify  $\mathcal{T}_{Ba} = \tau_a^F / P_{Ba}^F$  and  $\mathcal{T}_{Ab} = \tau_b^F / P_{Ab}^F$  and obtain

$$k_{AB} = \frac{1}{P_B^{\text{eq}}} \sum_{b \in B} \frac{P_{Ab}^I P_b^{\text{eq}}}{\tau_b^F} \quad \text{and} \quad k_{BA} = \frac{1}{P_A^{\text{eq}}} \sum_{a \in A} \frac{P_{Ba}^I P_a^{\text{eq}}}{\tau_a^F}, \quad (20)$$

A calculation of  $k_{AB}$  within the NGT scheme produces both  $k_{AB}^{\text{SS}}$  and  $k_{AB}^{\text{NSS}}$  directly, providing a quantitative assessment of an overall two-state description for the kinetics, including the definition of the  $A$  and  $B$  states. The results described in the following section were obtained after improving the efficiency and numerical precision in several ways. For example, to retain numerical precision in the calculation of  $1 - P_{xx}$  for values of  $P_{xx}$  that approach unity, we can write  $1 - P_{xx} = \sum_{j \neq x} P_{jx}$  and we evaluated the latter sum directly when  $P_{xx} > 0.99$ . During the renormalization procedure the number of states is reduced by one at every step, but the number of nonzero branching probabilities can increase or decrease. At the beginning of the calculation, the branching matrix is generally sparse, and a compressed row storage scheme<sup>55</sup> provides an appropriate data structure. However, as the size of the matrix decreases, the density of nonzero elements increases, and eventually it becomes more efficient to store the full matrix. The results described in Sec. IV were obtained by switching to the dense storage scheme when more than 2% of the elements became nonzero, so long as fewer than 11 000 states remained. Although the rate constants do not depend on the order in which the states are removed, the execution time can vary significantly for a sparse transition probability matrix. As for the previous GT implementation, we therefore removed the  $I$  states with fewest connections first.<sup>39</sup> We also summed the transition probabilities over all product states during the second phase of each NGT calculation, where branches to the other reactant states are removed for each reactant in turn. All the tests were conducted with the PATHSAMPLE program, available for download under the Gnu General Public License.<sup>80</sup>

## IV. TESTS

A number of tests were performed to check both the accuracy and efficiency of the NGT scheme. The committor probabilities and MFPT's can be calculated using successive over-relaxation (SOR)<sup>54</sup> and a compressed row storage scheme<sup>55</sup> for the branching probability matrix. For example, the MFPT to the  $B$  region for state  $\alpha$  satisfies  $\mathcal{T}_{B\alpha} = \tau_\alpha + \sum_{\beta} P_{\beta\alpha} \mathcal{T}_{B\beta}$ , with  $\mathcal{T}_{Bb} = 0$  for all  $b \in B$ . The waiting times  $t_a$  and  $t_b$  that occur in Eq. (4) can also be found from KMC simulations, as in previous work.<sup>38</sup> We therefore compared the NGT results for  $k_{AB}^{\text{SS}}$  and  $k_{BA}^{\text{SS}}$  with committor probabilities obtained from SOR;  $k_{AB}^{\text{NSS}}$  and  $k_{BA}^{\text{NSS}}$  with SOR and KMC results for  $t_a$  and  $t_b$ , and with GT;  $k_{AB}$  and  $k_{BA}$  with separate KMC, SOR, and GT results for  $\mathcal{T}_{Ab}$  and  $\mathcal{T}_{Ba}$ . Direct diagonalization of the master equation was compared to GT in previous work.<sup>38</sup> Accurate eigenvalues could not be obtained for any of the examples in Table I by diagonalization of the transition matrix or weighted subspace projection methods.<sup>62</sup> Precise agreement between the other methods was obtained in every case where the calculations could be converged. A variety of small networks were checked first, including cases where the SS, NSS, and KMC rate constants are significantly different. Details are omitted for brevity. A more practical test was provided by the stationary point database previously obtained in DPS studies<sup>25</sup> of the three-stranded  $\beta$ -sheet peptide Beta3s.<sup>81</sup> In fact, it was necessary

TABLE I. Total execution time in seconds for a single AMD 248 processor required to calculate  $k_{AB}^{SS}$ ,  $k_{BA}^{SS}$ ,  $k_{AB}^{NSS}$ ,  $k_{BA}^{NSS}$ ,  $k_{AB}$ , and  $k_{BA}$  from NGT, and  $k_{AB}$  and  $k_{BA}$  from SOR results for the MFPT's. The relaxation parameter used in the SOR calculations was 1.999 in each case, and the timings correspond to convergence of the MFPT to three significant figures. Results of KMC runs are also provided for the cases that could be converged (to two or three significant figures).  $N_{\min}$  and  $N_{ts}$  are the total number of states and connections corresponding to nonzero branching probabilities in each case. The number of states varies in the first batch of tests according to the threshold barrier height  $\Delta F_{\text{barrier}}$ , which is used in a recursive regrouping scheme (Refs. 23 and 25). Blank entries correspond to calculations that were unconverged.

$k_B T$ (kcal/mol)	$\Delta F_{\text{barrier}}$	$N_{\min}$	$N_{ts}$	NGT (s)	SOR (s)	KMC (s)
0.592	10.0	64	63	8	8	18
0.592	5.0	272	287	8	13	85 138
0.592	4.9	630	650	8	550	1 437 600 <sup>a</sup>
0.592	4.8	1 018	1 049	8	930	
0.592	4.7	1 457	1 507	8	2 337	
0.592	4.6	1 883	1 957	8	60 347	
0.592	4.5	2 344	2 462	8	217 830	
1.0	...	40 000	58 410	34.7		
2.0	...	40 000	58 410	35.3	281	1 020 540 <sup>a</sup>
3.0	...	40 000	58 410	39.2	122 242	
5.0	...	40 000	58 410	41.1		

<sup>a</sup>Converged to only two significant figures.

to consider a relatively small subset of this database containing only 40 000 minima and 58 410 transition states in order to converge the SOR calculations for the MFPT. To further accelerate the SOR calculations, calculations were run at higher temperatures or using a free energy regrouping scheme to reduce the number of states.<sup>23,25</sup> As expected, the NGT timings are almost temperature independent, and change by less than 10% for the different regrouping thresholds in this range (Table I). Timings for KMC calculations of the MFPT using the BKL algorithm<sup>56</sup> are also given where convergence was possible. These KMC runs used the leap-frog scheme described elsewhere<sup>22</sup> to speed up the simulations for low barriers where the product of forward and reverse branching probabilities exceeded 0.9. More sophisticated KMC schemes would be needed to treat most of these test cases.<sup>61</sup>

The relaxation parameter,  $\omega$ , used in the SOR calculations was 1.999 in each case, which was generally the largest value for which the iterations exhibited stable behavior for both  $\mathcal{T}_{Ab}$  and  $\mathcal{T}_{Ba}$ . The optimal rate of convergence is actually determined by the spectral radius of the branching probability matrix.<sup>82</sup> The SOR timings should therefore be regarded as upper bounds since specific optimization of  $\omega$  might be possible.<sup>82</sup> Nevertheless, the efficiency of the NGT calculations seems clear from these comparisons. The SOR calculations for the database without any regrouping run fastest at the temperature where the smallest of the two rate constants,  $k_{AB}$  and  $k_{BA}$ , has its largest value.

## V. CONCLUSIONS

An alternative renormalization scheme has been described that enables us to extract global kinetic properties from a transition network modeled as a graph theory representation of a Markov chain. As for the previous GT procedure,<sup>2,38,39</sup> the calculation is noniterative and nonstochastic, and the operation count is independent of tempera-

ture (or total energy for microcanonical kinetics). The new formulation (NGT) enables committor probabilities for transitions from each state to product or reactant to be calculated directly, along with the corresponding waiting times. Two-state rate constants can therefore be compared automatically under the additional assumptions of equilibrium within the product and reactant regions, and a steady-state approximation for states classified as intervening. Comparing these results as a function of regrouping thresholds<sup>21,23,25,63,64,83</sup> for the states in the transition network can then provide useful insight into whether a two-state description is appropriate. The GT and NGT schemes are the only methods that we have been able to apply successfully to large transition networks and systems with slow overall kinetics. The computational complexity of the NGT algorithm is essentially the same as for GT, as described in previous work.<sup>39</sup>

## ACKNOWLEDGMENTS

I am grateful to Dr. J. M. Carr and Dr. M. Khalili for their comments on the original manuscript.

- 1 F. Noé and S. Fischer, *Curr. Opin. Struct. Biol.* **18**, 154 (2008).
- 2 D. J. Wales, *Int. Rev. Phys. Chem.* **25**, 237 (2006).
- 3 B. Strodel and D. J. Wales, *Chem. Phys. Lett.* **466**, 105 (2008).
- 4 G. Bolch, S. Greiner, H. de Meer, and K. S. Trivedi, *Queueing Networks and Markov Chains* (Wiley, New York, 1998).
- 5 G. R. Grimmett and D. R. Stirzaker, *Probability and Random Processes* (Oxford University Press, Oxford, 2005).
- 6 G. R. Grimmett and D. R. Stirzaker, *One Thousand Exercises in Probability* (Oxford University Press, Oxford, 2005).
- 7 W. C. Swope, J. W. Pitera, F. Suits, M. Pitman, M. Eleftheriou, B. G. Fitch, R. S. Germain, A. Rayshubski, T. J. C. Ward, and Y. Zhestkov, *J. Phys. Chem. B* **108**, 6582 (2004).
- 8 W. C. Swope, J. W. Pitera, and F. Suits, *J. Phys. Chem. B* **108**, 6571 (2004).
- 9 S. V. Krivov and M. Karplus, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 14766 (2004).
- 10 N. Singhal, C. D. Snow, and V. S. Pande, *J. Chem. Phys.* **121**, 415 (2004).
- 11 S. P. Elmer, S. Park, and V. S. Pande, *J. Chem. Phys.* **123**, 114902

- (2005).
- <sup>12</sup> G. Jayachandran, V. Vishal, and V. S. Pande, *J. Chem. Phys.* **124**, 164902 (2006).
- <sup>13</sup> S. V. Krivov and M. Karplus, *J. Phys. Chem. B* **110**, 12689 (2006).
- <sup>14</sup> I. Horenko, E. Dittmer, A. Fischer, and C. Schütte, *Multiscale Model. Simul.* **5**, 802 (2006).
- <sup>15</sup> S. Park, T. E. Klein, and V. S. Pande, *Biophys. J.* **93**, 4108 (2007).
- <sup>16</sup> J. D. Chodera, K. A. Dill, N. Singhal, V. S. Pande, W. C. Swope, and J. W. Pitera, *J. Chem. Phys.* **126**, 155101 (2007).
- <sup>17</sup> F. Noé, I. Horenko, C. Schütte, and J. C. Smith, *J. Chem. Phys.* **126**, 155102 (2007).
- <sup>18</sup> C. Schultheis, T. Hirschberger, H. Carstens, and P. Tavan, *J. Chem. Theory Comput.* **1**, 515 (2005).
- <sup>19</sup> P. Metzner, C. Schütte, and E. Vanden-Eijnden, *Multiscale Model. Simul.* **7**, 1192 (2009).
- <sup>20</sup> D. J. Wales, *Mol. Phys.* **100**, 3285 (2002).
- <sup>21</sup> D. J. Wales, *Energy Landscapes* (Cambridge University Press, Cambridge, 2003).
- <sup>22</sup> D. J. Wales, *Mol. Phys.* **102**, 891 (2004).
- <sup>23</sup> J. M. Carr and D. J. Wales, *J. Chem. Phys.* **123**, 234901 (2005).
- <sup>24</sup> B. Strodel, C. S. Whittleston, and D. J. Wales, *J. Am. Chem. Soc.* **129**, 16005 (2007).
- <sup>25</sup> J. M. Carr and D. J. Wales, *J. Phys. Chem. B* **112**, 8760 (2008).
- <sup>26</sup> S. V. Krivov and M. Karplus, *J. Chem. Phys.* **117**, 10894 (2002).
- <sup>27</sup> S. F. Chekmarev, S. V. Krivov, and M. Karplus, *J. Phys. Chem. B* **109**, 5312 (2005).
- <sup>28</sup> F. Noé, D. Krachtus, J. C. Smith, and S. Fischer, *J. Chem. Theory Comput.* **2**, 840 (2006).
- <sup>29</sup> F. Noé, M. Oswald, G. Reinelt, S. Fischer, and J. C. Smith, *Multiscale Model. Simul.* **5**, 393 (2006).
- <sup>30</sup> G. C. Boulougous and D. N. Theodorou, *J. Chem. Phys.* **127**, 084903 (2007).
- <sup>31</sup> A. F. Voter, *Phys. Rev. B* **57**, R13985 (1998).
- <sup>32</sup> C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler, *J. Chem. Phys.* **108**, 1964 (1998).
- <sup>33</sup> C. Dellago, P. G. Bolhuis, and D. Chandler, *J. Chem. Phys.* **110**, 6617 (1999).
- <sup>34</sup> P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, *Annu. Rev. Phys. Chem.* **53**, 291 (2002).
- <sup>35</sup> D. Hamelberg, J. Mongan, and J. A. McCammon, *J. Chem. Phys.* **120**, 11919 (2004).
- <sup>36</sup> A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12562 (2002).
- <sup>37</sup> G. Bussi, F. L. Gervasio, A. Laio, and M. Parrinello, *J. Am. Ceram. Soc.* **128**, 13435 (2006).
- <sup>38</sup> S. A. Trygubenko and D. J. Wales, *Mol. Phys.* **104**, 1497 (2006).
- <sup>39</sup> S. A. Trygubenko and D. J. Wales, *J. Chem. Phys.* **124**, 234110 (2006).
- <sup>40</sup> M. S. Apaydin, D. L. Brutlag, C. Guestrin, D. Hsu, and J.-C. Latombe, *RECOMB '02: Proceedings of the Sixth Annual International Conference on Computational Biology* (ACM, New York, 2002), pp. 12–21.
- <sup>41</sup> M. S. Apaydin, C. E. Guestrin, C. Varma, D. L. Brutlag, and J.-C. Latombe, *Bioinformatics* **18**, S18 (2002).
- <sup>42</sup> M. G. Bulmer, *Principles of Statistics* (Dover, New York, 1979).
- <sup>43</sup> N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1981).
- <sup>44</sup> R. E. Kunz, *Dynamics of First-Order Phase Transitions* (Deutsch, Thun, 1995).
- <sup>45</sup> L. Onsager, *Phys. Rev.* **54**, 554 (1938).
- <sup>46</sup> E. W. Dijkstra, *Numer. Math.* **1**, 269 (1959).
- <sup>47</sup> V. M. Jiménez and A. Marzal, in *Algorithm Engineering: Third International Workshop, WAE'99*, edited by J. S. Vitter and C. D. Zaroliagis (Springer, Berlin, 1999), Vol. 1668, pp. 15–29.
- <sup>48</sup> J. M. Carr and D. J. Wales, in *Latest Advances in Atomic Cluster Collisions: Structure and Dynamics from the Nuclear to the Biological Scale* (Imperial College Press, London, 2008).
- <sup>49</sup> J. B. Anderson, *J. Chem. Phys.* **58**, 4684 (1973).
- <sup>50</sup> C. H. Bennett, in *Algorithms for Chemical Computations*, edited by R. E. Christofferson (American Chemical Society, Washington, D.C., 1977), pp. 63–97.
- <sup>51</sup> D. Chandler, *J. Chem. Phys.* **68**, 2959 (1978).
- <sup>52</sup> E. A. Carter, G. Ciccotti, J. T. Hynes, and R. Kapral, *Chem. Phys. Lett.* **156**, 472 (1989).
- <sup>53</sup> H. M. Taylor and S. Karlin, *An Introduction to Stochastic Modeling* (Academic, Orlando, 1984).
- <sup>54</sup> W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in FORTRAN*, 2nd ed. (Cambridge University Press, Cambridge, 1992).
- <sup>55</sup> *Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide*, edited by Z. Bai, J. Demmel, J. Dongarra, A. Ruhe, and H. van der Vorst (SIAM, Philadelphia, 2000).
- <sup>56</sup> A. B. Bortz, M. H. Kalos, and J. L. Lebowitz, *J. Comput. Phys.* **17**, 10 (1975).
- <sup>57</sup> D. A. Reed and G. Ehrlich, *Surf. Sci.* **105**, 603 (1981).
- <sup>58</sup> A. F. Voter, *Phys. Rev. B* **34**, 6819 (1986).
- <sup>59</sup> K. A. Fichtorn and W. H. Weinberg, *J. Chem. Phys.* **95**, 1090 (1991).
- <sup>60</sup> A. F. Voter, *Radiation Effects in Solids* (Springer-Verlag, New York, 2005), pp. 1–22.
- <sup>61</sup> W. Cai, M. H. Kalos, M. de Koning, and V. V. Bulatov, *Phys. Rev. E* **66**, 046703 (2002).
- <sup>62</sup> T. J. Frankcombe and S. C. Smith, *J. Comput. Chem.* **21**, 592 (2000).
- <sup>63</sup> D. A. Evans and D. J. Wales, *J. Chem. Phys.* **121**, 1080 (2004).
- <sup>64</sup> D. A. Evans and D. J. Wales, *J. Chem. Phys.* **118**, 3891 (2003).
- <sup>65</sup> M. A. Novotny, *Phys. Rev. Lett.* **74**, 1 (1995).
- <sup>66</sup> D. T. Gillespie, *J. Comput. Phys.* **22**, 403 (1976).
- <sup>67</sup> D. T. Gillespie, *J. Phys. Chem.* **81**, 2340 (1977).
- <sup>68</sup> D. G. Vlachos, *Phys. Rev. E* **78**, 046713 (2008).
- <sup>69</sup> M. Athenes, P. Bellon, and G. Martin, *Philos. Mag. A* **76**, 565 (1997).
- <sup>70</sup> L. Xu and G. Henkelman, *J. Chem. Phys.* **129**, 114104 (2008).
- <sup>71</sup> F. El-Mellouhi, N. Mousseau, and L. J. Lewis, *Phys. Rev. B* **78**, 153202 (2008).
- <sup>72</sup> N. M. van Dijk, *Queueing Networks and Product Forms* (Wiley, New York, 1993).
- <sup>73</sup> E. Gelenbe and G. Pujolle, *Introduction to Queueing Networks* (Wiley, New York, 1998).
- <sup>74</sup> A. E. Conway and N. D. Georganas, *Queueing Networks—Exact Computational Algorithms* (The MIT Press, Cambridge, 1989).
- <sup>75</sup> D. Eppstein, Z. Galil, and G. F. Italiano, in *Algorithms and Theory of Computation Handbook*, edited by M. J. Atallah (CRC, Boca Raton, 1999), Chap. 8.
- <sup>76</sup> B. V. Cherkassky, A. V. Goldberg, and T. Radzik, in “SODA '94,” Proceedings of the Fifth Annual ACM-SIAM Symposium on Discrete Algorithms (Society for Industrial and Applied Mathematics, Philadelphia, 1994), pp. 516–525.
- <sup>77</sup> G. Ramalingam and T. Reps, *J. Algorithms* **21**, 267 (1996).
- <sup>78</sup> G. Ramalingam and T. Reps, *Theor. Comput. Sci.* **158**, 233 (1996).
- <sup>79</sup> R. E. Leone and P. V. R. Schleyer, *Angew. Chem., Int. Ed. Engl.* **9**, 860 (1970).
- <sup>80</sup> D. J. Wales, PATHSAMPLE: A program for generating connected stationary point databases and extracting global kinetics (<http://www-wales.ch.cam.ac.uk/software.html>).
- <sup>81</sup> E. De Alba, J. Santoro, M. Rico, and M. A. Jiménez, *Protein Sci.* **8**, 854 (1999).
- <sup>82</sup> A. Nicholls and B. Honig, *J. Comput. Chem.* **12**, 435 (1991).
- <sup>83</sup> D. A. Evans and D. J. Wales, *J. Chem. Phys.* **119**, 9947 (2003).