Introduction to the Kinetic Monte Carlo Method

Arthur F. Voter

Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545 USA afv@lanl.gov

1 Introduction

Monte Carlo refers to a broad class of algorithms that solve problems through the use of random numbers. They first emerged in the late 1940's and 1950's as electronic computers came into use [1], and the name means just what it sounds like, whimsically referring to the random nature of the gambling at Monte Carlo, Monaco. The most famous of the Monte Carlo methods is the Metropolis algorithm [2], invented just over 50 years ago at Los Alamos National Laboratory. Metropolis Monte Carlo (which is *not* the subject of this chapter) offers an elegant and powerful way to generate a sampling of geometries appropriate for a desired physical ensemble, such as a thermal ensemble. This is accomplished through surprisingly simple rules, involving almost nothing more than moving one atom at a time by a small random displacement. The Metropolis algorithm and the numerous methods built on it are at the heart of many, if not most, of the simulations studies of equilibrium properties of physical systems.

In the 1960's researchers began to develop a different kind of Monte Carlo algorithm for evolving systems *dynamically* from state to state. The earliest application of this approach for an atomistic system may have been Beeler's 1966 simulation of radiation damage annealing [3]. Over the next 20 years, there were developments and applications in this area (e.g., see [3, 4, 5, 6, 7]), as well as in surface adsorption, diffusion and growth (e.g., see [8, 9, 10, 11, 12, 13, 14, 15, 16, 17]), in statistical physics (e.g., see [18, 19, 20]), and likely other areas, too. In the 1990's the terminology for this approach settled in as *kinetic Monte Carlo*, though the early papers typically don't use this term [21]. The popularity and range of applications of kinetic Monte Carlo (KMC) has continued to grow and KMC is now a common tool for studying materials subject to irradiation, the topic of this book. The purpose of this chapter is to provide an introduction to this KMC method, by taking the reader through the basic concepts underpinning KMC and how it is typically implemented, assuming no prior knowledge of these kinds of simulations. An appealing property of KMC is that it can, in principle, give the *exact* dynamical evolution of a system. Although this ideal is virtually never achieved, and usually not even attempted, the KMC method is presented here from this point of view because it makes a good framework for

A.F. Voter, in Radiation Effects in Solids, edited by K. E. Sickafus and E. A. Kotomin (Springer, NATO Publishing Unit, Dordrecht, The Netherlands, 2005) in press.

(perhaps check with me for final citation after it has appeared)

understanding what is possible with KMC, what the approximations are in a typical implementation, and how they might be improved. Near the end, we discuss a recently developed approach that comes close to this ideal. No attempt is made to fully survey the literature of KMC or applications to radiation damage modeling, although some of the key papers are noted to give a sense of the historical development and some references are given for the reader who wants a deeper understanding of the concepts involved. The hope is that this introductory chapter will put the reader in a position to understand (and assess) papers using KMC, whether for simulations of radiation damage evolution or any other application, and allow him/her to write a basic KMC program of their own if they so desire.

2 Motivation: the time-scale problem

Our focus is on simulating the dynamical evolution of systems of atoms. The premiere tool in this class of atomistic simulation methods is molecular dynamics (MD), in which one propagates the classical equations of motion forward in time. This requires first choosing an interatomic potential for the atoms and a set of boundary conditions. For example, for a cascade simulation, the system might consist of a few thousand or million atoms in a periodic box and a high velocity for one atom at time zero. Integrating the classical equations of motion forward in time, the behavior of the system emerges naturally, requiring no intuition or further input from the user. Complicated and surprising events may occur, but this is the correct dynamical evolution of the system for this potential and these boundary conditions. If the potential gives an accurate description of the atomic forces for the material being modeled, and assuming both that quantum dynamical effects are not important (which they can be, but typically only for light elements such as hydrogen at temperatures below T=300K) and that electron-phonon coupling (non-Born-Oppenheimer) effects are negligible (which they will be unless atoms are moving extremely fast), then the dynamical evolution will be a very accurate representation of the real physical system. This is extremely appealing, and explains the popularity of the MD method. A serious limitation, however, is that accurate integration requires time steps short enough ($\sim 10^{-15}$ s) to resolve the atomic vibrations. Consequently, the total simulation time is typically limited to less than one microsecond, while processes we wish to study (e.g., diffusion and annihilation of defects after a cascade event) often take place on much longer time scales. This is the "time-scale problem."

Kinetic Monte Carlo attempts to overcome this limitation by exploiting the fact that the long-time dynamics of this kind of system typically consists of diffusive jumps from state to state. Rather than following the trajectory through every vibrational period, these state-to-state transitions are treated directly, as we explain in the following sections. The result is that KMC can reach vastly longer time scales, typically seconds and often well beyond.

3 Infrequent-event systems, state-to-state dynamics, and the KMC concept

An infrequent-event system is one in which the dynamics is characterized by occasional transitions from one state to another, with long periods of relative inactivity between these transitions. Although the infrequent-event designation is fairly general (and hence also the possible applications of KMC), for simplicity we will restrict our discussion to the case where each state corresponds to a single energy basin, and the long time between transitions arises because the system must surmount an energy barrier to get from one basin to another, as indicated schematically in Fig. 1. This is an appropriate description for most solid-state atomistic systems. For a system that has just experienced a knock-on event causing a cascade, this infrequent-event designation does not apply until the excess initial energy has dissipated and the system has thermally equilibrated. This usually takes a few ps or a few tens of ps.

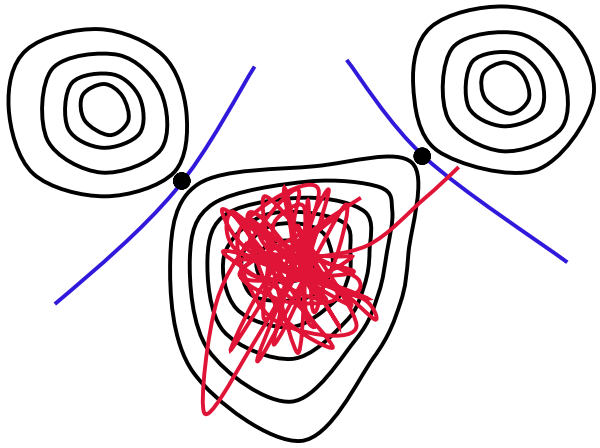

Fig. 1. Contour plot of the potential energy surface for an energy-barrier-limited infrequent-event system. After many vibrational periods, the trajectory finds a way out of the initial basin, passing a ridgetop into a new state. The dots indicate saddle points.

To be a bit more concrete about the definition of a state, consider a 256-atom system in a perfect fcc crystal geometry with periodic boundary conditions. Remove one of the atoms and put it back into the crystal somewhere else to create an interstitial. Now, using a steepest descent or conjugate gradient algorithm, we can "relax" the system: we minimize the energy to obtain the geometry at the bottom of the energy basin where the forces on every atom are zero. This defines a particular state i of the system and the

geometry at the minimum is R_i. If we heat the system up a bit, e.g., by giving each atom some momentum in a random direction and then performing MD, the system will vibrate about this minimum. As it vibrates, we still say it is in state i (assuming it has not escaped over a barrier yet) because if we stop the MD and minimize the energy again, the system will fall back to the exact same geometry R_i. Adjacent to state i there are other potential basins, each separated from state i by an energy barrier. The lowest barriers will correspond to moving the interstitial (perhaps through an interstitialcy mechanism) or moving an atom into the vacancy. Even though only one or a few atoms move in these cases, the entire system has been taken to a new state. This is an important point – we don't move atoms to new states, we move the entire system from state to state.

The key property of an infrequent-event system caught in a particular basin is that because it stays there for a long time (relative to the time of one vibrational period), it forgets how it got there. Then, for each possible escape pathway to an adjacent basin, there is a *rate constant* k_{ij} that characterizes the probability, per unit time, that it escapes to that state j, and these rate constants are independent of what state preceded state i. As we will discuss below, each rate constant k_{ij} is purely a property of the shape of the potential basin i, the shape of the ridge-top connecting i and j, and (usually to a much lesser extent) the shape of the potential basin j. This characteristic, that the transition probabilities for exiting state i have nothing to do with the history prior to entering state i, is the defining property of a Markov chain [22, 23]. The state-to-state dynamics in this type of system correspond to a Markov walk. The study of Markov walks is a rich field in itself, but for our present purposes we care only about the following property: because the transition out of state i depends only on the rate constants $\{k_{ij}\}$, we can design a simple stochastic procedure to propagate the system correctly from state to state. If we know these rate constants exactly for every state we enter, this state-to-state trajectory will be indistinguishable from a (much more expensive) trajectory generated from a full molecular dynamics simulation, in the sense that the probability that we see a given sequence of states and transition times in the KMC simulation is the same as the probability for seeing that same sequence and transition times in the MD. (Note that here we are assuming that at the beginning of an MD simulation, we assign a random momentum to each atom using a fresh random number seed, so that each time we perform the MD simulation again, the state-to-state trajectory will in general be different.)

4 The rate constant and first-order processes

Because the system loses its memory of how it entered state i on a time scale that is short compared to the time it takes to escape, as it wanders around vibrationally in the state it will similarly lose its memory repeatedly about

just where it has wandered before. Thus, during each short increment of time, it has the same probability of finding an escape path as it had in the previous increment of time. This gives rise to a first-order process with exponential decay statistics (i.e., analagous to nuclear decay). The probability the system has not yet escaped from state i is given by

$$p_{survival}(t) = \exp(-k_{tot}t) \tag{1}$$

where k_{tot} is the total escape rate for escape from the state. We are particularly interested in the the probability distribution function $p(t)$ for the time of first escape from the state, which we can obtain from this survival probability function. The integral of $p(t)$ to some time t' gives the probability that the system has escaped by time t', which must equate to $1 - p_{survival}(t')$. Thus, taking the negative of the time derivative of $p_{survival}$ gives the probablity distribution function for the time of first escape,

$$p(t) = k_{tot}\exp(-k_{tot}t). \tag{2}$$

We will use this first-passage-time distribution in the KMC procedure. The average time for escape τ is just the first moment of this distribution,

$$\tau = \int_0^\infty t\ p(t)dt = \frac{1}{k_{tot}}. \tag{3}$$

Because escape can occur along any of a number of pathways, we can make the same statement as above about each of these pathways – the system has a fixed probability per unit time of finding it. Each of these pathways thus has its own rate constant k_{ij}, and the total escape rate must be the sum of these rates:

$$k_{tot} = \sum_j k_{ij}. \tag{4}$$

Moreover, for each pathway there is again an exponential first-escape time distribution,

$$p_{ij}(t) = k_{ij}\exp(-k_{ij}t), \tag{5}$$

although only one event can be the first to happen. For more discussion on the theory of rate processes in the context of stochastic simulations, see [24, 25].

We are now almost ready to present the KMC algorithm. Not surprisingly, given the above equations, we will need to be able to generate exponentially distributed random numbers, which we quickly describe.

4.1 Drawing an exponentially distributed random number

Generating an exponentially distributed random number, i.e., a time t_{draw} drawn from the distribution $p(t) = k\ \exp(-kt)$, is straightforward. We first draw a random number r on the interval (0,1), and then form

$$t_{draw} = -(1/k)\ln(r). \tag{6}$$

A time drawn in this way is an appropriate realization for the time of first escape for a first-order process with rate constant k. Note that the usual definition of the uniform deviate r is either $0 < r < 1$ or $0 < r \leq 1$; a random number generator implemented for either of these ranges will give indistinguishable results in practice. However, some random number generators also include $r = 0$ in the bottom of the range, which is problematic (causing an ill-defined $\ln(0)$ operation), so zero values for r must be avoided.

5 The KMC procedure

Having laid the conceptual foundation, it is now straightforward to design a stochastic algorithm that will propagate the system from state to state correctly. For now, we are assuming all the rate constants are known for each state; in later sections we will discuss how they are calculated and tabulated.

Before discussing the procedure usually used by KMC practitioners, it is perhaps instructive to first present a more transparent approach, one that is less efficient, though perfectly valid. Our system is currently in state i, and we have a set of pathways and associated rate constants $\{k_{ij}\}$. For each of these pathways, we know the probability distribution for the first escape time is given by Eq. 5. Using the procedure in Sect. 4.1, we can draw an exponentially distributed time t_j from that distribution for each pathway j. Of course, the actual escape can only take place along one of these pathways, so we find the pathway j_{min} which has the lowest value of t_j, discard the drawn time for all the other pathways, and advance our overall system clock by t_{jmin}. We then move the system to state j_{min}, and begin again from this new state. That is all there is to it. This is less than ideally efficient because we are drawing a random number for each possible escape path, whereas it will turn out that we can advance the system to the next state with just two random numbers.

We now describe the KMC algorithm that is in common use. The pathway selection procedure is indicated schematically in Fig. 2a. We imagine that for each of the M escape pathways we have an object with a length equal to the rate constant k_{ij} for that pathway. We put these objects end to end, giving a total length k_{tot}. We then choose a single random position along the length of this stack of objects. This random position will "touch" one of the objects, and this is the pathway that we choose for the system to follow. This procedure gives a probability of choosing a particular pathway that is proportional to the rate constant for that pathway, as it should. To advance the clock, we draw a random time from the exponential distribution for the rate constant k_{tot} (see Sect. 4.1). Note that the time advance has nothing to do with *which* event is chosen. The time to escape depends only on the total

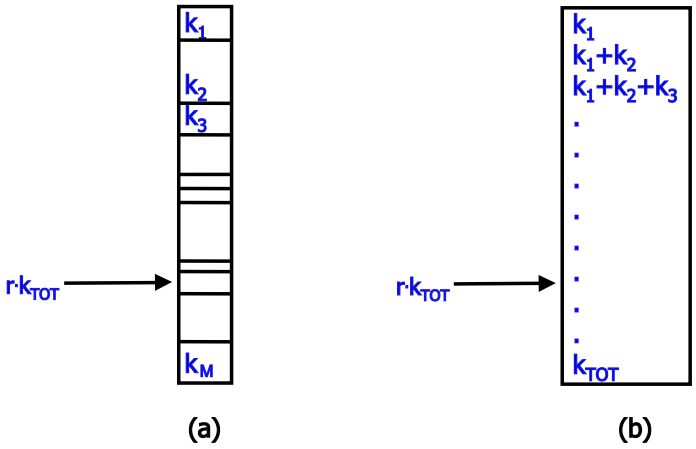

Fig. 2. Schematic illustration of the procedure for picking the reaction pathway to advance the system to the next state in the standard KMC algorithm. (a) Objects (boxes for this illustration), each with a length proportional to the rate constant for its pathway, are placed end to end. A random number r on (0,1), multiplied by k_{tot}, points to one box with the correct probability. (b) In a computer code, this is achieved by comparing rk_{tot} to elements in an array of partial sums.

escape rate. Once the system is in the new state, the list of pathways and rates is updated (more on this below), and the procedure is repeated.

In a computer implementation of this procedure, we make an array of partial sums. Let array element $s(j)$ represents the length of all the objects up to and including object j,

$$s(j) = \sum_{q}^{j} k_{iq}, \tag{7}$$

as shown in Fig. 2b. One then draws a random number r, distributed on (0,1), multiplies it by k_{tot}, and steps through the array s, stopping at the first element for which $s(j) > rk_{tot}$. This is the selected pathway.

This rejection-free "residence-time" procedure is often referred to as the BKL algorithm (or the "n-fold way" algorithm), due to the 1975 paper by Bortz, Kalos and Lebowitz [18], in which it was proposed for Monte Carlo simulation of Ising spin systems. However, the idea goes back further. For example, rejection-free KMC algorithms were used by others [8, 9, 10] earlier to study surface adsorption and dynamics, and the residence-time algorithm is described in the 1965 textbook by Cox and Miller [26].

6 Determining the rates

Assuming we know about the possible pathways, we can use transition state theory (TST) [27, 28, 29], to compute the rate constant for each pathway. Although TST is approximate, it tends to be a very good approximation for solid-state diffusive events. Moreover, if desired, the rate computed from TST can be corrected for recrossing effects to give the exact rate. By underpinning the KMC in this way, using high-quality TST rates that can be extended to exact rates if desired, the state-to-state dynamics of the KMC simulations can, in principle, be made as accurate as real molecular dynamics on the underlying potential. This concept was first proposed in [17].

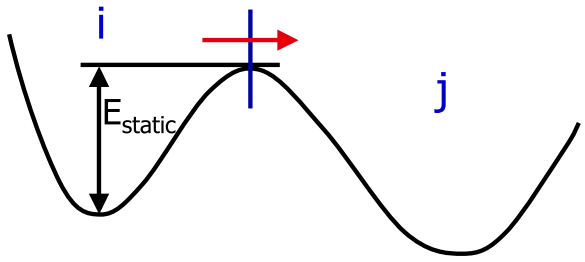

Fig. 3. Illustration of the transition state theory rate constant. The unimolecular rate constant for escape from state i to state j, k_{ij}, is given by the equilibrium outgoing flux through the dividing surface separating the states.

6.1 Transition state theory

Transition state theory (TST), first proposed in 1915 [27], offers a conceptually straightforward approximation to a rate constant. The rate constant for escape from state i to state j is taken to be the equilibrium flux through a dividing surface separating the two states, as indicated in Fig. 3. We can imagine having a large number of two-state systems, each allowed to evolve long enough that many transitions between these states have occurred, so that they represent an equilibrium ensemble. Then, looking in detail at each of the trajectories in this ensemble, if we count the number of forward crossings of the dividing surface per unit time, and divide this by the number of trajectories, on average, that are in state i at any time, we obtain the TST rate constant, k_{ij}^{TST}. The beauty of TST is that, because it is an equilibrium property of the system, we can also calculate k_{ij}^{TST} without ever looking at dynamical trajectories. For a thermal ensemble (the only kind we are considering in this chapter), k_{ij}^{TST} is simply proportional to the Boltzmann probability of being at the dividing surface relative to the probability of being anywhere

in state i. Specifically, for a one-dimensional system with a dividing surface at $x = 0$,

$$k_{ij}^{\mathrm{TST}} = \langle |dx/dt| \, \delta(x) \rangle_i \,, \tag{8}$$

where the angular brackets indicate a canonical ensemble average over the position coordinate x and momentum p, the subscript i indicates evaluation over the phase space belonging to state i ($x \leq 0$ in this case), and $\delta(x)$ is the Dirac delta function. Extension to many dimensions is straightforward [30], but the point is that once the dividing surface has been specified, k_{ij}^{TST} can be evaluated using, for example, Metropolis Monte Carlo methods [30, 31].

The implicit assumption in TST is that successive crossings of the dividing surface are uncorrelated; i.e., each forward crossing of the dividing surface corresponds to a full reactive event that takes the system from residing in state i to residing in state j. However, in reality, there is the possibility that the trajectory may recross the dividing surface one or more times before either falling into state j or falling back into state i. If this happens, the TST rate constant overestimates the exact rate, because some reactive events use up more than a single outgoing crossing. As stated above, the exact rate can be recovered using a dynamical corrections formalism [32, 33, 34], in which trajectories are initiated at the dividing surface and integrated for a short time to allow the recrossing events to occur. While the best choice of dividing surface is the one that minimizes the equilibrium flux passing through it (the best surface usually follows the ridgetop), this dynamical corrections algorithm recovers the exact rate constant even for a poor choice of dividing surface. This dynamical corrections formalism can also be extended to correctly account for the possibility of multiple-jump events, in which case there can be nonzero rate constants k_{ij} between states i and j that are not adjacent in configuration space [35].

In principle, then, classically exact rates can be computed for each of the pathways in the system. In practice, however, this is never done, in part because the TST approximation is fairly good for solid-state diffusive processes. In fact, most KMC studies are performed using a further approximation to TST, which we describe next.

6.2 Harmonic transition state theory

The harmonic approximation to TST, and further simplifications to it, are often used to calculate KMC rate constants. Harmonic TST (HTST) is often referred to as Vineyard theory [36], although equivalent or very similar expressions were derived earlier by others [37]. In HTST, we require that the transition pathway is characterized by a saddle point on the potential energy surface (e.g., the dots in Fig. 1). One assumes that the potential energy near the basin minimum is well described (out to displacements sampled thermally) with a second-order energy expansion – i.e., that the vibrational

modes are harmonic – and that the same is true for the modes perpendicular to the reaction coordinate at the saddle point. The dividing surface is taken to be the saddle plane (the hyperplane perpendicular to the reaction coordinate at the saddle point), and evaluation of the ensemble average in Eq. 8 for a system with N moving atoms gives the simple form

$$k^{HTST} = \frac{\prod_i^{3N} \nu_i^{min}}{\prod_i^{3N-1} \nu_i^{sad}} \exp(-E_{static}/k_B T). \tag{9}$$

Here E_{static} is the static barrier height (energy difference between the saddle point and the minimum) and k_B is the Boltzmann constant. In the preexponential factor (or prefactor), $\{\nu_i^{min}\}$ are the $3N$ normal mode frequencies at the minimum and $\{\nu_i^{sad}\}$ are the $3N-1$ nonimaginary normal mode frequencies at the saddle [38]. The computation of k^{HTST} thus requires information only about the minimum and the saddle point for a given pathway. The HTST rate tends to be a very good approximation to the exact rate (e.g., within 10-20%) up to at least half the melting point for diffusion events in most solid materials (e.g., see [39, 40]), although there can be exceptions [41]. Further, since prefactors are often in the range of 10^{12} s^{-1} - 10^{13} s^{-1} (though they can be higher; e.g., see Fig. 4 in [42]), a common approximation is to choose a fixed value in this range to save the computational work of computing the normal modes for every saddle point.

The form of the Vineyard approximation merits further comment. Note that the only temperature dependence is in the exponential, and depends only on the static (i.e., T=0) barrier height [43]. No correction is needed, say, to account for the extra potential energy that the system has as it passes over the saddle region at a finite temperature. This, and all the entropy effects, cancel out in the integration over the normal modes, leaving the simple form of Eq. (9). Also note that Planck's constant h does not appear in Eq. (9). The kT/h preexponential sometimes found in TST expressions is an artifact of incomplete evaluation of the partition functions involved, or a dubious approximation made along the way. TST is a classical theory, so h cannot remain when the integrals are all evaluated properly. Confusion about this expression, which introduces the wrong temperature dependence and an inappropriate physical constant, has persisted because $k_B T/h$ at T=300K (6.2×10^{12} Hz) is coincidentally similar to a typical preexponential factor.

7 The lattice assumption and the rate catalog

Typically in KMC simulations, the atoms in the system are mapped onto a lattice. An event may move one atom or many atoms, perhaps in a complicated way, but in the final state, each atom will again map onto a unique

lattice point. Note that if, for example, harmonic TST is used to compute the rates, it requires that the system be relaxed to find the energy and frequencies at the minimum. After relaxation, the atoms will in general no longer be positioned on the lattice points, especially for atoms near defects. However, if each atom is much closer to one lattice point than any other lattice point, and if the mapping of the atoms onto the lattice points does not change during the relaxation, then it is safe to map the system onto a lattice in this way to simplify the KMC and the generation of the rate constants.

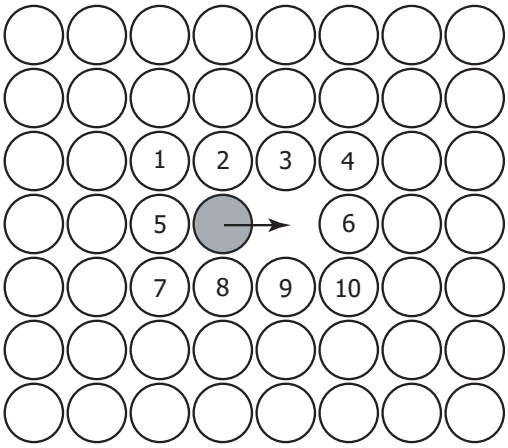

Fig. 4. Schematic illustration of the rate catalog concept for the diffusional jump of a vacancy. Atoms in the lattice sites labelled 1-10 can affect the rate constant significantly, so TST rate constants are computed for all possible occupations (differing atom types or vacancies) of these sites. This list of rates makes up a *rate catalog*, which can be accessed during the KMC simulation to determine the rate constant for the jump of a vacancy in any direction for any environment.

Lattice mapping also makes it easy to exploit locality in determining rates. We assume that only the atoms near a defect affect the rate constant for any change or migration of that defect. An example of this is shown in Fig. 4 for a pathway schematically corresponding to the jump of a lattice vacancy. For each of the local environments of the jumping atom (i.e., in which each of the numbered sites in Fig. 4 is either vacant or filled with one atom type or another), we can compute a TST rate constant [17]. The number of possible rates, ignoring symmetry, is

$$n_{rate} = (n_{type} + 1)^{n_{site}}, \tag{10}$$

where n_{site} is the number of sites explicitly considered ($n_{site} = 10$ in Fig. 4) and n_{type} is the number of possible atom types that can be at each of those sites. Equation 10 results from the fact that each site, independently, may

either be vacant, or have an atom of one of the n_{type} types. (For the purposes of this formal counting, we are overlooking the fact that some of the pathways involving multiple vacancies in the environment may be ill defined if the adatom has no neighbors.)

The set of rates computed in this way comprise a "rate catalog" [17], which we can then use to look up the rates we need for every state the system visits. By making the local environment larger, we can make the rates in it more accurate, and in principle we can make the environment as large as we need to achieve the accuracy we desire. In practice, however, the fact that the number of rates that will have to be computed grows as a strong power law in n_{site} means that we may settle for less than ideal accuracy. For example, for vacancy moves in an fcc metal, including just nearest-neighbor sites of the jumping atom, n_{site}=18 and n_{type}=1, giving 2^{18} = 262,144 rates to be computed (many equivalent by symmetry). For a classical interatomic potential, this is feasible, using an automated procedure in which a nudged elastic band calculation [44, 45] or some other saddle-finding algorithm (e.g., Newton-Raphson) is applied to each configuration. However, just increasing this to include second nearest neighbors ($2^{28} = 2.7\text{x}10^8$ rates, ignoring symmetry) or to consider a binary alloy ($3^{18} = 3.8\text{x}10^8$ rates, ignoring symmetry) increases the computational work enormously.

This work can be reduced somewhat by splitting the neighborhood into two sets of sites [46], one set of sites that most influence the active atom in the initial state (e.g., sites 1,2,3,5,7,8, and 9 in Fig. 4) and another set of sites that most influence the active atom at the saddle point (e.g., sites 2,3,8, and 9 in Fig. 4). Two catalogs are then generated, one for the minima and one for the saddle points. Each catalog entry gives the energy required to remove the active atom (the one involved in the jump) from the system. Subtracting these special vacancy formation energies for a given minimum-saddle pair gives the energy barrier for that process.

Another way to reduce the work is to create the rate catalog as the KMC simulation proceeds, so that rate constants are computed only for those environments encountered during the KMC.

While achieving convergence with respect to the size of the local environment is formally appealing, we will see in Sect. 9 that it is usually more important to make sure that all *types* of pathway are considered, as missing pathways often cause larger errors in the final KMC dynamics.

Finally, we note that the locality imposed by this rate-catalog approach has the benefit that in the residence-time procedure described in Sect. 5, updating the list of rates after a move has been accepted requires only fixed amount of work, rather than work scaling as the number of atoms in the entire system.

7.1 Assuming additive interactions

As discussed above (see Eq. 10), computing every rate necessary to fill the rate catalog may be undesirably expensive if n_{site} and/or n_{type} are large, or if a computationally expensive electronic structure calculation is employed to describe the system. Within the HTST framework, where the rate is specified by a barrier height and a preexponential factor, an easy simplification is to assume that the barrier height can be approximated by additive interactions. For example, beginning from the example shown in Fig. 4, the neighboring atoms can be categorized as class $m1$ (nearest neighbors to the jumping atom when the system is at the minimum, sites 2, 5 and 8), class $m2$ (second nearest neighbors to the jumping atom when the system is at the minimum, sites 1, 3, 7 and 9), class $s1$ (first neighbors to the jumping atom when the system is at the saddle point, sites 2, 3, 8, and 9), and so forth. This example is shown in Fig. 5. The barrier energy is then approximated by

$$E_{static} = E_{sad} - E_{min}, \tag{11}$$

where the energy of the minimum (E_{min}) and the energy of the saddle (E_{sad}) are given by

$$E_{min} = E^0_{min} + n_{m1}E_{m1} + n_{m2}E_{m2} \tag{12}$$

$$E_{sad} = E^0_{sad} + n_{s1}E_{s2} + n_{s2}E_{s2}. \tag{13}$$

Here, n_{m1} is the number of atoms in $m1$ positions, and similarly for n_{m2}, n_{s1}, and n_{s2}. In this way, the rate catalog is replaced by a small number of additive interaction energies. The energies $E^0_{min}E_{m1}, E_{m2}, E^0_{sad}, E_{s1}$, and E_{s2} can be simply specified ad hoc or adjusted to give simulation results that match experiment (this is the way almost all KMC simulations were done until the mid 1980's, and many still are), or they can be obtained from a best fit to accurately calculated rate constants (e.g., see [47]). For the prefactor, $10^{12} - 10^{13}s^{-1}$ is a good estimate for many systems.

7.2 Obeying detailed balance

In any chemical system, we can make general statements about the behavior of the system when it is in equilibrium that are useful for understanding the dynamical evolution when the system is out of equilibrium (as it typically is). Formally, exact equilibrium properties can be obtained by gathering statistics on a very large number of systems, each of which has run for an extremely long time before the measurements are made. At equilibrium, the fractional population of state i, χ_i, is proportional to $\exp(-G_i/k_BT)$, where G_i is the free energy of state i. For every pair of connected states i and j, the number of transitions per unit time (on average) from i to j must equal the number of transitions per unit time from j to i. Because the number of escapes per

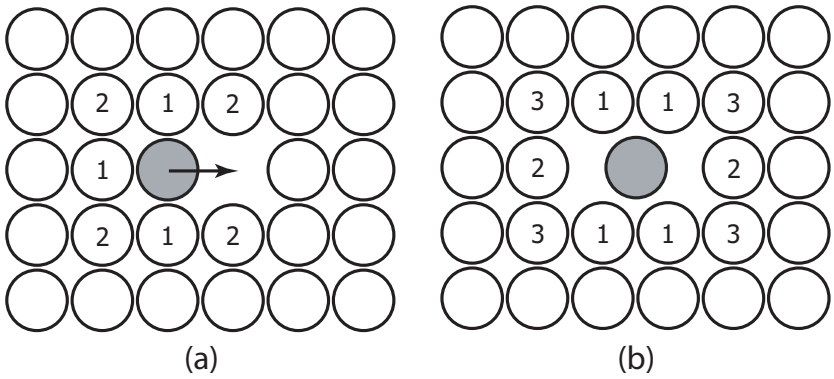

Fig. 5. Schematic illustration of the additive rate catalog for the diffusional jump of a vacancy. Sites are labeled by class for (a) the minimum and (b) the saddle point.

time from i to j is proportional to the population of state i times the rate constant for escape from i to j, we have

$$\chi_i k_{ij} = \chi_j k_{ji}, \tag{14}$$

and the system is said to "obey detailed balance." Because the equilibrium populations and the rate constants are constants for the system, this detailed balance equation, which must hold even when the system is not in equilibrium, places requirements on the rate constants. If a rate catalog is constructed that violates detailed balance, then the dynamical evolution will not correspond to a physical system. This ill-advised situation can occur, for example, if a rate constant is set to zero, but the reverse rate is not, as might happen in a sensitivity analysis or a model study. It can also arise if there is an asymmetry in the procedure for calculating the rates (e.g., in the way that saddle points are found) that gives forward and reverse rates for a connected pair of states that are not compatible.

8 Computational scaling with system size

For a system with M escape pathways, the residence-time algorithm described in Sect. 5, in its simplest implementation, would require searching through a list of M rates to find the pathway that is selected by the random number. The computational work to choose each KMC step would thus scale as M. Over the years, papers [48, 49, 50] have appeared discussing how to implement the residence-time procedure with improved efficiency. Blue, Beichl and Sullivan [49] pointed out that by subdividing the list of rates into hierarchical sublists, the work can be reduced to that of searching a binary tree, scaling as $\log(M)$.

Recently, Schulze [50] demonstrated that for a system in which there are equivalent rates that can be grouped (e.g., the rate for a vacancy hop in one part of the system is equivalent to the rate for a vacancy hop in another part of the system), the work can be reduced further, becoming independent of M.

After the pathway is selected and the system is moved to the new state, the rate list must be updated. In general, for this step the locality of the rate constants can be exploited, as discussed in Sect. 7, so that only a fixed amount of work is required, independent of M.

The overall computational scaling of KMC also depends on how far the system advances in time with each KMC step. In general, for a system with N atoms, the number of pathways M will be proportional to the number of atoms N [51]. If we increase the size of a system in a self-similar way, e.g., doubling N by placing two equivalent systems side by side, then the total escape rate k_{tot} will be proportional to N (see Eq. 4). Since the average time the system advances is inversely proportional to k_{tot} (see Eq. 3), this means that the overall work required to propagate a system of N atoms forward for a certain amount of time is proportional to N (within the Schulze assumption that there is a fixed number of unique rate constants) or at worst $N\log N$.

9 Surprises – the real reason KMC is not exact

As claimed in the introduction, KMC can, in principle, give the exact state-to-state dynamics for a system. This assumes that a complete rate catalog has been generated, containing an accurate rate constant for every escape pathway for every state that will be encountered in the dynamics. We have discussed above the fact that the TST rate is not exact (unless augmented with dynamical corrections) and the difficulty in fully converging the environment size (see Eq. 10). However, for a typical system, neither of these effects are the major limitation in the accuracy of the KMC dynamics. Rather, it is the fact that the real dynamical evolution of a system will often surprise us with unexpected and complex reaction pathways. Because these pathways (before we have seen them) are outside our intuition, they will typically not be included in the rate catalog, and hence cannot occur during the KMC simulation.

The field of surface diffusion provides a classic example of such a surprising pathway. Until 15 years ago, diffusion of an adatom on the simple fcc(100) surface was assumed to occur by the adatom hopping from one four-fold site to the next. Using density functional theory calculations, Feibelman discovered in 1990 that the primary diffusion pathway on Al(100) actually involves the exchange event shown in Fig. 6, in which the adatom plunges into the surface, pushing a substrate atom up into the second nearest neighbor binding site [52]. In field-ion microscope experiments, this exchange mechanism was shown to be the dominant pathway for Pt/Pt(100) [53] and Ir/Ir(100) [54].

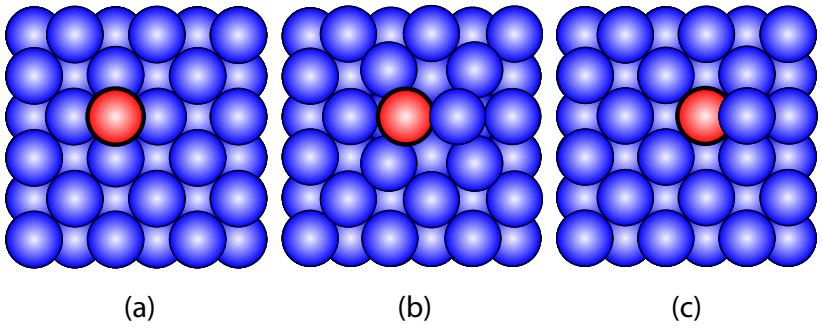

Fig. 6. Exchange mechanism for adatom on fcc(100) surface. (a) initial state; (b) saddle point; (c) final state. This mechanism, unknown until 1990 [52], is the dominant diffusion pathway for some fcc metals, including Al, Pt, and Ir.

For Pt/Pt(100), the barrier for the hop mechanism is roughly 0.5 eV higher than for the exchange mechanism. Thus, a KMC simulation of Pt adatoms on a Pt(100) surface, using a rate catalog built assuming hop events only (which was standard practice for KMC on the fcc(100) surface until recently), would give a seriously flawed description of the diffusion dynamics.

More recently, there have been many examples of unexpected surface and bulk diffusion mechanisms [55, 56, 57, 58, 59]. In some cases, the discovered mechanisms are so complex that it would not be easy to incorporate them into a KMC rate catalog, even afte the existence of the pathway is known. This issue was the primary motivation for the development of the accelerated molecular dynamics methods [60] described in the next chapter.

10 Simulation time achievable with KMC

The total simulation time that can be achieved in a KMC simulation is strongly system dependent. Each KMC step advances the system by a time (on average) no greater than the inverse of the fastest rate for escape from the current state. This rate depends exponentially on the barrier height divided by the temperature, and the size of the lowest barrier can change, perhaps dramatically, as the system evolves. However, to get some sense of what is possible, we observe that on present-day computers, one can take roughly 10^{10} steps in a few hours of computer time (the exact value, of course, depends on the type and size of the system). If we assume that for every state there is one fast escape pathway with a fixed lowest barrier E_a and a prefactor of 10^{13}, then we can achieve a simulation time of $10^{10}/(10^{13}\exp(-E_a/k_BT))$. For $E_a = 0.5$ eV, this gives a total simulation time of 2.5x10^5 s at T=300K, 16 s at T=600K and 0.33 s at T=1000K. For a very low barrier, times are even shorter but the temperature dependence is much weaker. For exam-

ple, E_a = 0.1 eV gives 50 ms at T=300K, 10 ms at T=600K, and 3 ms at T=1000K. These times are all significantly longer than one can achieve with direct molecular dynamics simulation (typically between 1 ns and 1 μs).

11 The low-barrier problem

It is interesting to note how important the lowest barrier is. A persistent low barrier can significantly decrease the total accessible simulation time, and many systems exhibit persistent low barriers. For example, in metallic surface diffusion, adatoms that diffuse along the edge of a two-dimensional cluster or a step edge usually do so with a much lower barrier than for diffusion on an open terrace [17]. In bulk fcc materials, interstitials typically diffuse with very low barriers in the range of 0.1 eV or less. In glassy materials, low-barriers abound. This is a common and long-standing problem with KMC simulations.

One approximate approach to the problem is to raise the lowest barriers artificially to slow down the fastest rates. This will give accurate dynamics if the fast processes are reasonably well equilibrated under the conditions of interest, and if they are still able to reach equilibration when they are slowed down. In general, though, it may be hard to know for sure if this is corrupting the dynamics.

Often the structure of the underlying potential energy surface is such that the system repeatedly visits a subset of states. Among the states in this *superbasin*, an equilibrium may be achieved on a much shorter time scale than the time it takes the system to escape from the superbasin. In this situation, if all the substates are known and a list of all processes that take the system out of the superbasin can be enumerated, then one of these processes can be selected with an appropriate Boltzmann probability. Two difficulties typically arise: 1) efficiently identifying and recognizing all the substates, the number of which may be very large; and 2) being sure the system is truly equilibrated in the superbasin.

Regarding the first problem, Mason et al [61] have recently pointed out that a hashing procedure based on the Zobrist key [62], developed for recognizing previously stored configurations in chess, can be used to efficiently identify revisited states in a KMC simulation. In lattice-based KMC, as in chess, the number of possible states is typically astronomical, vastly exceeding the number of indices that can be stored in computer memory. The Zobrist approach maps a lattice-based configuration onto a non-unique index (key) that has a low probability of colliding with other configurations. Mason et al showed that retrieving previously visited states based on their Zobrist key saved substantial time for a KMC simulation in which rates were calculated from scratch for each new configuration. This type of indexing could also be powerful for enumerating and recognizing states in a superbasin.

Novotny has presented a general method [63] that circumvents the second problem (that of establishing equilibration in the superbasin). The implementation requires setting up and diagonalizing a transition matrix over the revisited states. While this probably becomes too costly for superbasins with large numbers of substates, the generality of the method is very appealing. It yields the time of the first transition out of the set of revisited states and the state the system goes to, while making no requirement of a local equilibrium or even that the complete set of states in the superbasin be known.

Finally, we note that the low-barrier problem is also an issue for the accelerated molecular dynamics methods discussed in the next chapter, as well as for the on-the-fly KMC discussed in Section 13.

12 Object kinetic Monte Carlo

A higher level of simulation, which is still in the KMC class, can be created by constructing state definitions and appropriate rate constants for multi-atom entities such as interstitial clusters, vacancy clusters, etc. This type of simulation is becoming more common in radiation damage annealing studies. As in basic atomic-level KMC, this *object KMC* approach is usually performed on a lattice using the residence time algorithm, perhaps augmented by additional rules. By treating the diffusive motion of a cluster, for example, as proceeding by simple KMC steps of the center of mass, rather than as a cumulative result of many individual basin-to-basin moves that move atoms around in complex (and often unproductive) ways, object KMC can reach much longer time and length scales than pure atom-based KMC. A good example of this approach, with references to earlier work, can be found in [64].

The tradeoff in this approach, however, is that important pathways may go missing from the rate catalog as atomistic details are eliminated. For example, the diffusion rate as a function of cluster size must be specifed directly, as well as rules and rates for coalescense or annihilation two clusters that encounter one another. When the real dynamics are explored, these dependencies sometimes turn out to be surprisingly complicated. For example, in the case of intersitial clusters in MgO, the diffusion constants are strongly non-monotonic with cluster size, and, worse, the cluster resulting from coalescence of two smaller clusters can sometimes form in a long-lived metastable state with dramatically different diffusion properties [59]. So, while object KMC is an effective way to reach even greater time and size scales than standard KMC, it is perhaps even more important to keep in mind the dangers of missing pathways.

13 On-the-fly Kinetic Monte Carlo

As discussed above, while ideally KMC simulations can be carried out in a way that is faithful to the real dynamics for the underlying interatomic potential, this is virtually never the case in real applications due to the fact that reaction pathways are invariably missing from the rate catalog. In part, this deficiency arises from the fact that keeping the system on lattice precludes certain types of diffusive events, but the far more dominant reason is simply that we usually make up rate catalogs based on our intuition about how the system will behave, and the real dynamics is almost always more complicated. This situation can be improved by gaining more experience on the system to be simulated, e.g., by observing the types of events that occur during extensive direct MD simulations. However, even this approach is usually inadequate for finding all the reactive events that could occur during the evolution of the system.

It is this situation that has motivated research in recent years to develop alternative methods that can reach long time scales while maintaining (or coming close to) the accuracy of direct MD. The next chapter in this book describes one powerful approach to this problem, accelerated molecular dynamics, in which the classical trajectory is retained (rather than collapsing the description to a set of states, as in KMC), and this classical trajectory is coaxed into finding each escape path more quickly. We finish the present chapter with a brief description of another approach, one which retains the flavor of KMC.

Recently, Henkelman and Jónsson [67] have proposed a variation on the KMC method, in which one builds a rate catalog on the fly for each state. The key to this approach is having an efficient way to search for saddle points that are connected to the current state of the system. For this, they use the "dimer" method [58]. Given a random starting position within the energy basin, the dimer algorithm climbs uphill along the lowest eigenvector of the Hessian matrix, reaching the saddle point at the top. Because it only requires first derivatives of the potential [68], it is computationally efficient. In principle, if *all* the bounding saddle points can be found (and hence all the pathways for escape from the state), the rate for each of these pathways can be supplied to the KMC procedure described in Section 5, propagating the system in a dynamically correct way to the next state, where the procedure is begun again. In practice, it is hard (probably impossible) to demonstrate that all saddles have been found, especially considering that the number of saddle points bounding a state grows exponentially with the dimensionality of the system. However, with a large number of randomly initiated searches, most of the low-lying barriers can be found, and this approach looks very promising [69]. Examining the pathways that these systems follow from state to state, often involving complicated multiple-atom moves, it is immediately obvious that the quality of the predicted dynamical evolution is substantially better than one could hope to obtain with pre-cataloged KMC. This approach

can be parallelized efficiently, as each dimer search can be performed on a separate processor. Also, for large systems, each dimer search can be localized to a subset of the system (if appropriate). On the other hand, this type of on-the-fly KMC is substantially more expensive than standard KMC, so the user must decide whether the increased quality is worth the cost.

14 Conclusions

Kinetic Monte Carlo is a very powerful and general method. Given a set of rate constants connecting states of a system, KMC offers a way to propagate dynamically correct trajectories through the state space. The type of system, as well as the definition of a state, is fairly arbitrary, provided it is appropriate to assume that the system will make first-order transitions among these states. In this chapter, we have focused on atomistic systems, due to their relevance to radiation damage problems, which are the subject of this book. In this case, the states correspond to basins in the potential energy surface.

We have emphasized that, if the rate catalog is constructed properly, the easily implemented KMC dynamics can give exact state-to-state evolution of the system, in the sense that it will be statistically indistinguishable from a long molecular dynamics simulation. We have also pointed out, however, that this ideal is virtually never realizable, due primarily to the fact that there are usually reaction pathways in the system that we don't expect in advance. Thus, if the goal of a KMC study is to obtain accurate, predictive dynamics, it is advisable to perform companion investigations of the system using molecular dynamics, on-the-fly kinetic Monte Carlo (see Section 13), or accelerated molecular dynamics (see next chapter).

Despite these limitations, however, KMC remains the most powerful approach available for making dynamical predictions at the meso scale without resorting to more dubious model assumptions. It can also be used to provide input to and/or verification for higher-level treatments such as rate theory models or finite-element simulations. Moreover, even in situations where a more accurate simulation would be feasible (e.g., using accelerated molecular dynamics or on-the-fly kinetic Monte Carlo), the extreme efficiency of KMC makes it ideal for rapid scans over different conditions, for example, and for model studies.

15 Acknowledgments

The author thanks P.A. Rikvold and H.L. Heinisch for supplying helpful information about early KMC work, and B.P. Uberuaga for a critical reading of the manuscript. This work was supported by the United States Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences, Division of Materials Science.

References

1. N. Metropolis, *Los Alamos Science*, **12**, 125 (1987).
2. N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
3. J.R. Beeler, Jr., Phys. Rev. **150**, 470 (1966).
4. D.G. Doran, Radiat. Eff. **2**, 249 (1970).
5. J.-M. Lanore, Rad. Eff. **22** 153 (1974).
6. H.L. Heinisch, D.G. Doran, and D.M. Schwartz, ASTM Special Technical Publication **725**, 191 (1981).
7. H.L. Heinisch, J. Nucl. Mater. **117** 46 (1983).
8. R. Gordon, J. Chem. Phys. **48**, 1408 (1968).
9. F.F. Abraham and G.W. White, J. Appl. Phys. **41**, 1841 (1970).
10. C.S. Kohli and M.B. Ives, J. Crystal Growth **16**, 123 (1972).
11. G.H. Gilmer, J. Crystal Growth **35**, 15 (1976).
12. M. Bowker and D.A. King, Surf. Sci. **71**,583 (1978).
13. D.A. Reed and G. Ehrlich, Surf. Sci. **105**, 603 (1981).
14. P.A. Rikvold, Phys. Rev. A **26**, 647 (1982).
15. E.S. Hood, B.H. Toby, and W.H. Weinberg, Phys. Rev. Lett. **55**, 2437 (1985).
16. S.V. Ghaisas and A. Madhukar, J. Vac. Sci. Technol. B **3**, 540 (1985).
17. A.F. Voter, Phys. Rev. B **34**, 6819 (1986).
18. A.B. Bortz, M.H. Kalos, and J.L. Lebowitz, J. Comp. Phys. **17**, 10 (1975).
19. K. Binder, in *Monte Carlo Methods in Statistical Physics* (Springer Topics in Current Physics, Vol. 7) edited by K Binder (Springer, Berlin 1979) p. 1.
20. K. Binder and M.H. Kalos, in *Monte Carlo Methods in Statistical Physics* (Springer Topics in Current Physics, Vol. 7) edited by K Binder (Springer, Berlin 1979) p. 225.
21. Early (and even some recent) KMC work can be found under various names, including "dynamic Monte Carlo," "time-dependent Monte Carlo," and simply "Monte Carlo."
22. R. Norris, *Markov Chains* (Cambridge University Press, Cambridge, UK, 1997).
23. W. Feller, *An Introduction to Probability Theory and its Applications, Vol. 1*, Wiley, New York (1966).
24. D.T. Gillespie, J. Comp. Phys. **22**, 403 (1976).
25. K.A. Fichthorn and W.H. Weinberg, J. Chem. Phys. **95**, 1090 (1991).
26. D.R. Cox and H.D. Miller, *The Theory of Stochastic Processes* (Methuen, London, 1965), pp 6-7.
27. R. Marcelin, Ann. Physique **3**, 120 (1915).
28. E. Wigner, Z. Phys. Chem. **B 19** 203 (1932).
29. H. Eyring, J. Chem. Phys. **3**, 107 (1935).
30. A.F. Voter and J.D. Doll, J. Chem. Phys. **80**, 5832 (1984).
31. A.F. Voter, J. Chem. Phys. 82, 1890 (1985).
32. J.C. Keck, Discuss. Faraday Soc. **33**, 173 (1962).
33. C.H. Bennett, in *Algorithms for Chemical Computation*, edited by R.E. Christofferson (American Chemical Society, Washington, DC, 1977), p. 63.
34. D. Chandler, J. Chem. Phys. **68**, 2959 (1978).
35. A.F. Voter and J.D. Doll, J. Chem. Phys. **82**, 80 (1985).
36. G.H. Vineyard, J. Phys. Chem. Solids **3**, 121 (1957).
37. P. Hanggi, P. Talkner, and M. Borkovec, Rev. Mod. Phys. **62**, 251 (1990).

38. For a three-dimensional periodic system with all atoms moving, discarding the translational modes leaves $3N$-3 and $3N$-4 real normal mode frequencies at the minimum and saddle, respectively. For a system that is free to rotate, there are $3N$-6 and $3N$-7 relevant modes.
39. G. DeLorenzi, C.P. Flynn, and G. Jacucci, Phys. Rev. B **30**, 5430 (1984).
40. M.R. Sørensen and A.F. Voter, J. Chem. Phys. **112**, 9599 (2000).
41. G. Boisvert and L.J. Lewis, Phys. Rev. B **54**, 2880 (1996).
42. F. Montalenti and A.F. Voter, Phys. Stat. Sol. (b) **226**, 21 (2001).
43. An additional T dependence is introduced if the quasiharmonic method is employed to give a different lattice constant (and hence different barrier height and preexponential) at each temperature, but here we are assuming a fixed lattice constant.
44. H. Jónsson, G. Mills, and K.W. Jacobsen, in *Classical and Quantum Dynamics in Condensed Phase Simulations*, edited by B.J. Berne, G. Ciccotti and D.F. Coker (World Scientific, 1998), chapter 16.
45. G. Henkelman, B.P. Uberuaga, and H. Jónsson, J. Chem. Phys. **113** 9901 (2000).
46. A.F. Voter, in *Modeling of Optical Thin Films*, M.R. Jacobson, Ed., Proc. SPIE **821**, 214 (1987).
47. H. Mehl, O. Biham, K. Furman, and M. Karimi, Phys. Rev. B **60**, 2106 (1999).
48. P.A. Maksym, Semicond. Sci. Technol. **3**, 594 (1988).
49. J.L. Blue, I. Beichl and F. Sullivan, Phys. Rev. E **51**, R867, (1994).
50. T.P. Schulze, Phys. Rev. E **65**, 036704 (2002).
51. In fact, the number of saddle points accessible to a state may scale more strongly than linear in N if complicated, high-barrier mechanisms are considered, but in almost all KMC implementations it will be proportional to N.
52. P.J. Feibelman, Phys. Rev. Lett. **65**, 729 (1990).
53. G.L. Kellogg and P.J. Feibelman, Phys. Rev. Lett. **64**, 3143 (1990).
54. C. Chen and T.T. Tsong, Phys. Rev. Lett. **64**, 3147 (1990).
55. C.L. Liu and J.B. Adams, Surf. Sci. **268**, 73 (1992).
56. R. Wang and K.A. Fichthorn, Molec. Sim. **11**, 105 (1993).
57. J.C. Hamilton, M.S. Daw, and S.M. Foiles, Phys. Rev. Lett. **74**, 2760 (1995).
58. G. Henkelman and H. Jónsson, J. Chem. Phys. **111**, 7010 (1999).
59. B.P. Uberuaga, R. Smith, A.R. Cleave, F. Montalenti, G. Henkelman, R.W. Grimes, A.F. Voter, and K.E. Sickafus, Phys. Rev. Lett. **92**, 115505 (2004).
60. A.F. Voter, F. Montalenti and T.C. Germann, Annu. Rev. Mater. Res., **32**, 321 (2002).
61. D.R. Mason, T.S. Hudson, and A.P. Sutton, Comp. Phys. Comm. **165**, 37 (2005).
62. A.L. Zobrist, Technical report 88, Computer Science Department, University of Wisconsin, Madison, 1970; Reprinted in: ICCA J. **13**, 69 (1990).
63. M.A. Novotny, Phys. Rev. Lett. **74**, 1 (1994); Erratum **75**, 1424 (1995).
64. C. Domain, C.S. Becquart, and L. Malerba, J. Nucl. Mater. **335**, 121 (2004).
65. S. Liu, Z. Zhang, J. Norskov, and H. Metiu, Surf. Sci. **321**, 161 (1994).
66. Z-P. Shi, Z. Zhang, A.K. Swan, and J.F. Wendelken, Phys. Rev. Lett. **76**, 4927 (1996).
67. G. Henkelman and H. Jónsson, J. Chem. Phys. **115**, 9657 (2001).
68. A.F. Voter, Phys. Rev. Lett. **78**, 3908 (1997).
69. G. Henkelman and H. Jonsson, Phys. Rev. Lett. **90**, 116101-1 (2003).

Index