Annu. Rev. Mater. Res. 2002. 32:321–46
doi: 10.1146/annurev.matsci.32.112601.141541

EXTENDING THE TIME SCALE IN ATOMISTIC SIMULATION OF MATERIALS

Arthur F. Voter, Francesco Montalenti, and Timothy C. Germann
Los Alamos National Laboratory, Los Alamos, New Mexico 87545; e-mail: afv@lanl.gov; franz@t12.lanl.gov; tcg@lanl.gov

Key Words infrequent events, transition state theory, accelerated dynamics, hyperdynamics, parallel replica dynamics, temperature-accelerated dynamics, kinetic Monte Carlo

■ **Abstract** Obtaining a good atomistic description of diffusion dynamics in materials has been a daunting task owing to the time-scale limitations of the molecular dynamics method. We discuss promising new methods, derived from transition state theory, for accelerating molecular dynamics simulations of these infrequent-event processes. These methods, hyperdynamics, parallel replica dynamics, temperature-accelerated dynamics, and on-the-fly kinetic Monte Carlo, can reach simulation times several orders of magnitude longer than direct molecular dynamics while retaining full atomistic detail. Most applications so far have involved surface diffusion and growth, but it is clear that these methods can address a wide range of materials problems.

INTRODUCTION

The past few decades have shown a dramatic increase in the use of atomistic simulation as a partner with experiment in addressing problems in materials science. The most direct of these simulation techniques is the molecular dynamics (MD) method, in which one chooses an appropriate interatomic potential to describe the forces between atoms and then integrates the classical equations of motion with suitable boundary conditions. The increasing power and popularity of this type of simulation arise from the fact that the quality of interatomic potentials has become quite good for some systems and because computer speeds continue to increase rapidly.

An appealing feature of MD [e.g., compared with Metropolis Monte Carlo (1) or molecular statics] is that it follows the actual dynamical evolution of the system. However, the limitation in the accessible simulation time represents a substantial obstacle in making useful predictions with MD. Resolving individual atomic vibrations requires a time step of approximately femtoseconds in the integration of the equations of motion, so that on today's fastest processors, reaching even one microsecond is very difficult. Because this integration is inherently sequential in

nature, direct parallelization does not help much; it just makes it possible to reach nanoseconds on much larger systems.

In the past five years, new methods have been developed that look very promising for circumventing this time scale problem. For systems in which the long-time dynamical evolution is characterized by a sequence of activated events (i.e., diffusive events), these methods can extend the accessible time scale by orders of magnitude relative to direct MD, while retaining full atomistic detail. So far these new methods have been applied primarily to metallic surface diffusion and surface growth. With a little more development, they will be useful for a much broader range of materials problems because this activated-event description applies to such varied processes as vacancy diffusion, pipe diffusion along a dislocation core, dislocation climb and kink nucleation, impurity clustering, void growth, grain growth, and surface morphology evolution. With this potential in mind, we present here an introduction to these methods, discuss their current strengths and limitations, and predict how their capabilities may develop in the next few years.

The main emphasis in this review is on "accelerated dynamics methods," a class including hyperdynamics, parallel replica dynamics, and temperature accelerated dynamics. As discussed below, the conceptual link between these methods is that the system trajectory, caught in its current state, is stimulated to find an appropriate path for escape more quickly than it would with direct MD. As in MD, no a priori information about what this escape path might look like is imposed on the procedure; the trajectory simply finds its own way out of the state. Although each method accomplishes this acceleration in a different way, transition state theory provides the theoretical foundation in each case. We also discuss another promising approach for reaching longer time scales that we term "on-the-fly kinetic Monte Carlo." As our focus is on methods that can extend the MD simulation time in an accurate way for materials problems, we make no attempt to discuss the large body of related work involving enhanced sampling methods and approximate dynamical approaches (e.g., for macromolecule systems).

BACKGROUND

Infrequent Event Systems

We begin by defining an infrequent-event system, the type of system we discuss throughout this article. The dynamical evolution of such a system consists of vibrational excursions within a potential basin, punctuated by occasional transitions between basins; these transition events are infrequent in the sense that the average time between events is many (perhaps very, very many) vibrational periods. A simple example of such a system is an adatom on a metal surface at a temperature that is low relative to the energy barrier for a diffusive jump. We exclusively consider thermal systems characterized by a temperature T, a fixed number of

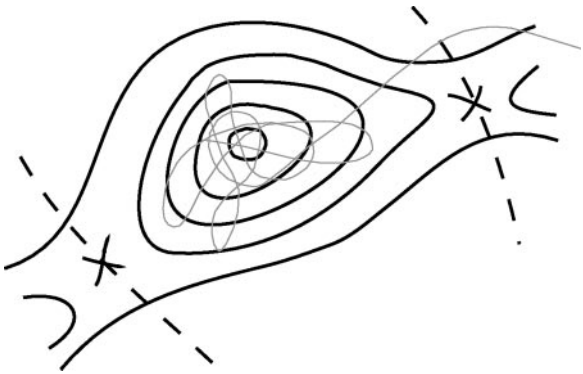

Figure 1 Schematic illustration of an infrequent-event system. The trajectory wanders around in the basin over a time scale of many (perhaps a huge number of) vibrational periods. At some point in time, when enough energy has been localized into a reactive mode, the trajectory passes through a dividing surface, entering another state. In essence, it "accidentally" exits the state. During this brief period of excitation, it may recross the dividing surface, but shortly thereafter it settles into the new state (or the original state if it recrosses), beginning a new session of vibrational wandering, with no memory of how it arrived in that state. Although it may never again visit this state, and "sees" only a single dividing surface as it exits, it nonetheless chooses an escape direction (relative to other possible escape directions) with the correct probability. This property of any infrequent-event system, that a trajectory will automatically choose an appropriate escape path with no prior information, is the basic concept exploited in the accelerated dynamics methods. The key is to coax the trajectory into making this choice more quickly without corrupting the relative escape probabilities (i.e., the rate constants for escape to various states).

moving atoms N, and a fixed volume V; i.e., the canonical ensemble (2). Although for simplicity we often draw systems in one or two dimensions, as in Figure 1, it is important to remember that they are actually $3N$-dimensional. Forgetting this complexity can sometimes lead to errors in our intuition. There may be a large number of possible paths for escape from any given basin to an adjacent basin. As a trajectory in this $3N$-dimensional coordinate space passes from one basin to another, it crosses a $(3N\text{-}1)$-dimensional "dividing surface" at the ridge top separating the two basins. Although on average these crossings are infrequent, successive crossings can sometimes occur within just a vibrational period or two; these are termed correlated dynamical events. An example would be a double jump of the adatom on the surface. For this presentation, it is sufficient, but important, to know that such events can occur—that successive crossings of dividing surfaces can be correlated in this way. The interested reader is encouraged to consult the rich literature on this topic (3–9). In most of the methods presented below, we

assume that these correlated events do not occur (this is the primary assumption of transition state theory), which is actually a very good approximation for many solid-state diffusive processes. We define the correlation time (τ_{corr}) of the system as the duration of the system memory. A trajectory that has resided in a particular basin for more than τ_{corr} has completely forgotten how it got there, in the sense that when it later escapes from this basin, the probability for escape along a given path is independent of how it entered the state. The relative probability for escape to a given adjacent state is proportional to the rate constant for that escape path; this rate is defined below.

To summarize and restate: An infrequent-event system is one in which the trajectory resides in a basin for many vibrational periods before finding a way to escape from the state after a time $\tau_{rxn} \gg \tau_{corr}$. In essence, the trajectory wanders around until it accidentally crosses a dividing surface. This reactive crossing event may (or may not) be followed quickly by subsequent crossings, leading to yet another state or perhaps recrossing back into the initial state. After a time τ_{corr} has passed since the initial crossing, the trajectory will have settled into some state, and the sequence begins again. The wandering within a basin between these bursts of excitement may consist of a huge number of vibrational excursions. The key to the accelerated dynamics methods presented below is recognizing that to obtain the right sequence of state-to-state transitions, we need not evolve the vibrational dynamics perfectly, as long as the relative probability of finding each of the possible escape paths is preserved.

It is easy to imagine systems that are not in this infrequent-event class. For example, if many of the barriers are extremely low, or if the system is at a temperature that is high relative to the typical barrier, then the correlated crossing events will be intermingled with the reactive events, violating the infrequent-event assumption. In the extreme case of this intermingling, even the rate constants themselves are ill-defined. Another example is a system that makes rapid transitions because it is driven by a high stress or strain rate. For many solid-state materials systems, however, the long-time dynamical evolution does fall into this infrequent-event class, and is thus, in principle, amenable to treatment with the accelerated dynamics methods described below.

Transition State Theory

Transition state theory (TST) (10–14) is the formalism that underpins all of the accelerated dynamics methods, directly or indirectly. In the TST approximation, the classical rate constant for escape from state *A* to some adjacent state *B* is taken to be the equilibrium flux through the dividing surface between *A* and *B* (Figure 2). If there are no correlated dynamical events, the TST rate is the exact rate constant. To expand on this equilibrium flux concept, imagine that for a two-state system we run a long classical trajectory, weakly coupled to a heat bath in some way to guarantee overall canonical (thermal) behavior. We run the trajectory for so long that it establishes equilibrium, visiting both states an extremely large number of

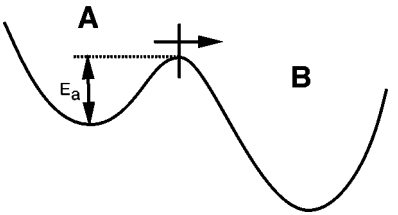

Figure 2 A two-state system that illustrates the definition of the transition state theory rate constant as the outgoing flux through the dividing surface bounding state *A*.

times. By examining this trajectory, we could accurately determine the fraction, χ_A, of the time it spends in state *A* and the number of crossings, per unit time, of the dividing surface. The TST rate constant for escape from *A*, $k_{A\rightarrow}^{\mathrm{TST}}$, would then be half of this crossing rate (i.e., counting only those crossings that are exiting state *A*) divided by χ_A.

The real beauty of TST, though, is that because this flux is an equilibrium property of the system, we can compute the TST rate without ever propagating a trajectory. The appropriate ensemble average is

$$k_{A\rightarrow}^{\mathrm{TST}} = \langle |dx_1/dt|\, \delta(x_1 - q) \rangle_A. \qquad 1.$$

Here the angular brackets indicate the ratio of Boltzmann-weighted integrals over $6N$-dimensional phase space (configuration space $\mathbf{r}$ and momentum space $\mathbf{p}$); i.e., for some property $P(\mathbf{r}, \mathbf{p})$,

$$\langle P \rangle = \frac{\int\int P(\mathbf{r}, \mathbf{p}) \exp[-H(\mathbf{r}, \mathbf{p})/k_B T]\, d\mathbf{r}\, d\mathbf{p}}{\int\int \exp[-H(\mathbf{r}, \mathbf{p})/k_B T]\, d\mathbf{r}\, d\mathbf{p}}, \qquad 2.$$

where k_B is the Boltzmann constant. The subscript *A* in Equation 1 indicates the configuration space integrals are restricted to the space belonging to state *A* (eliminating the need to divide by χ_A), and the dividing surface, for simplicity here, is at $x_1 = q$, involving only the reaction coordinate x_1 ($x_1 \in \mathbf{r}$). If the effective mass *m* of the reaction coordinate is constant over the dividing surface, Equation 1 reduces to a simpler ensemble average over configuration space only (15),

$$k_{A\rightarrow}^{\mathrm{TST}} = \sqrt{2k_B T/\pi m}\ \langle \delta(x_1 - q) \rangle_A. \qquad 3.$$

The essence of this expression, and of TST, is that the Dirac delta function picks out the probability of the system being at the dividing surface, relative to everywhere it can be in state *A*. Note that there is no dependence on the nature of the final state *B*. Evaluating Equation 3 to find the TST rate for a given temperature is relatively straightforward using, e.g., Metropolis Monte Carlo (1), once the dividing surface has been specified (15). At low *T*, the Metropolis walk will only rarely reach the dividing surface region, so importance sampling techniques become crucial (16–19).

In a system with correlated events, not every dividing surface crossing corresponds to a reactive event, so that, in general, the TST rate is an upper bound on the exact rate. [Adjusting the position of the dividing surface to minimize $k_{A\rightarrow}^{\mathrm{TST}}$,

thereby obtaining the best approximation to the true rate, is the basis of variational TST (20).] If desired, the exact rate can be recovered by running short trajectories from the dividing surface to compute a dynamical correction factor (9), a concept that goes back at least to Keck (3) and to Bennett (4, 21) for condensed phases. For two-state condensed phase systems, Chandler developed an elegant dynamical corrections formalism (5, 22) that has been extended to many-state systems (7). For diffusive events in materials at moderate temperatures, these correlated dynamical events typically do not cause a large change in the rate constants, so TST is often an excellent approximation. This is a key point; this behavior is markedly different from that chemical systems (e.g., molecular reactions in solution or the gas phase), where TST is just a starting point and dynamical corrections often lower the rate significantly (23, 24).

Whereas in the traditional use of TST, rate constants are computed after the dividing surface is specified; in the accelerated dynamics methods, we exploit the TST formalism to design approaches that do not require knowing in advance where the dividing surfaces will be, or even what product states might exist.

Harmonic Transition State Theory

A common and useful approximation to TST can be applied if we have identified a saddle point on the potential energy surface for the reaction pathway. We assume 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. Then the TST rate constant (in this case, the flux through the saddle plane), becomes simply

$$k^{HTST} = \nu_0 \exp(-E_a/k_B T), \tag{4.}$$

where

$$\nu_0 = \frac{\prod_i^{3N} \nu_i^{min}}{\prod_i^{3N-1} \nu_i^{sad}}. \tag{5.}$$

Here E_a is the static barrier height, or activation energy (energy difference between the saddle point and the minimum) (Figure 2), $\{\nu_i^{min}\}$ are the normal mode frequencies at the minimum, and $\{\nu_i^{sad}\}$ are the nonimaginary normal mode frequencies at the saddle (25). This is often referred to as the Vineyard (26) equation, although equivalent or very similar expressions were derived by others earlier (27). The analytic integration over the whole phase space thus leaves a very simple Arrhenius temperature dependence. (Although the exponent depends only on the static barrier height, there is no assumption that the trajectory passes exactly through the saddle point.) To the extent that there are no recrossings and the modes are truly harmonic, this is an exact expression for the rate. This harmonic TST

expression is employed in the temperature-accelerated dynamics (without need for the prefactor), and in the on-the-fly kinetic Monte Carlo method. It is interesting to note that the frequency of the imaginary mode (the reaction coordinate) does not enter, although it would come into play if we considered a high-friction system (27–29) or quantum effects (27).

Complex Infrequent Event Systems

The motivation for developing accelerated molecular dynamics methods becomes particularly clear when we try to understand the dynamical evolution of what we term complex infrequent event systems. In these systems, we simply cannot guess where the state-to-state evolution might lead. The underlying mechanisms may be too numerous, too complicated, and/or have an interplay whose consequences are unpredictable. While in very simple systems we can raise the temperature to make diffusive transitions occur on an MD-accessible time scale, in more complex systems this strategy will cause the system to travel down a different path in state space. Ultimately, this will lead to a completely different kind of system, making it impossible to address the questions that the simulation was attempting to answer.

Often, even systems that seem very simple can turn out to be in this complex class. For example, until 1990, we did not know, nor did we expect, that an adatom on the fcc(100) surface diffused in any way other than by a simple hop mechanism. The exchange mechanism (30–32), involving the adatom and a substrate atom (see Figure 3) is now known to be the preferred mechanism for diffusion on fcc(100) surfaces for many metals (e.g., Al, Pd, Pt, and Au), and the surface science community has since discovered a large variety of multiple-atom concerted diffusion mechanisms (33–40). In addition, the interplay of surface diffusion mechanisms during vapor-deposited growth can lead to surprising effects. For example, both fcc(100) (41, 42) and fcc(111) (43, 44) metals grow smoothly at high temperature, rougher as T is lowered, and then smooth again as the T is lowered further, and these are the simplest low-index surfaces! Many, if not most, materials problems

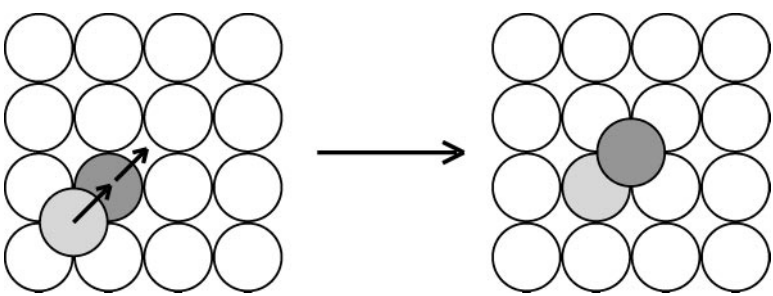

Figure 3 Adatom exchange mechanism on an fcc(100) surface. This, as opposed to a jump, is the preferred adatom diffusion mechanism on a number of fcc metals, demonstrating the complexity of even simple atomistic systems.

fall into this complex infrequent-event system category. We may want to know what happens on the time scale of milliseconds, seconds or longer, whereas with MD we cannot even reach one microsecond. Running at higher T or trying to guess what the unit processes are can mislead us about how the system really behaves. Often for these systems, if we could get a glimpse of what happens at these longer times, even if we could only afford to do a single trajectory that reached these times, our understanding of the system would improve substantially. This, in a sentence, is the chief motivation for the development of the methods described in this review.

Dividing Surfaces and Transition Detection

We have implied that the ridge tops between basins are the appropriate dividing surfaces in these systems. For a system that obeys TST, these ridge tops are the optimal dividing surfaces; for any other choice of dividing surface, recrossings will occur. A ridge top can be defined in terms of steepest descent paths—it is the $3N$-1-dimensional boundary surface that separates those $\mathbf{r}$ points whose steepest descent paths fall into one basin from those that fall into an adjacent basin. This definition also leads to a simple way to detect transitions as a simulation proceeds, as required in parallel replica dynamics and temperature-accelerated dynamics. Intermittently, the trajectory is paused, and a steepest descent minimization is initiated from its current $\mathbf{r}$ position. If this minimization leads to a basin minimum that is distinguishable from the minimum of the basin in which the system previously resided, a transition has occurred. An appealing feature of this approach is that it requires virtually no knowledge of the nature of the transition. Often only a few steepest descent steps are required to determine that there has been no transition. Although this is a fairly robust detection algorithm, and the one used for the simulations presented below, more efficient approaches can be designed. Richie et al. (45) have proposed a wavelet-based method that avoids the minimization requirement. It may also be possible to design very efficient system-specific algorithms; e.g., for covalently bonded systems, changes in bond connectivity can signal a transition.

PARALLEL REPLICA DYNAMICS

The parallel replica method (46) is the simplest and most accurate of the accelerated dynamics techniques, with the only assumption being that of infrequent events obeying first-order kinetics (exponential decay); i.e., for any time greater than τ_{corr} after entering a state, the probability distribution function for the time of the next escape is given by

$$p(t) = k \exp(-kt), \qquad 6.$$

where k is the rate constant for escape. For example, Equation 6 arises naturally for ergodic, chaotic exploration of an energy basin. The general approach is shown in Figure 4. Starting with an N-atom system in a particular state (basin), the entire system is replicated on each of M available parallel or

distributed processors. After a short dephasing stage ($\Delta t_{deph} \geq \tau_{corr}$), during which momenta are periodically randomized to eliminate correlations between replicas, each processor carries out an independent constant-temperature MD trajectory for the entire N-atom system, thus exploring phase space within the particular basin M times faster than a single trajectory would. Whenever a transition is detected on any processor, all processors are alerted to stop. The simulation clock is advanced by the accumulated trajectory time summed over all replicas, i.e., the total time spent exploring phase space within the basin until an escape pathway is found. It is readily shown (46) that this procedure gives an escape time that is correctly drawn from the distribution in Equation 6, even if the processor speeds are inequivalent, allowing the use of heterogeneous clusters or widely distributed machines running parallel replica dynamics as low-level background processes (47).

The parallel replica method also correctly accounts for correlated dynamical events (i.e., there is no requirement that the system obeys TST), unlike the other three methods presented here. This is accomplished by allowing the trajectory that made the transition to continue on its processor for a further amount of time $\Delta t_{corr} \geq \tau_{corr}$, during which recrossings or follow-on events may occur. The simulation clock is then advanced by Δt_{corr}, the final state is replicated on all processors, and the whole process is restarted. This overall procedure then gives exact state-to-state dynamical evolution because the escape times obey the correct probability distribution; nothing about the procedure corrupts the relative probabilities of the possible escape paths, and the correlated dynamical events are properly accounted for.

The efficiency of the method is limited by both the dephasing stage, which does not advance the system clock, and the correlated event stage, during which only one processor accumulates time. (This is illustrated schematically in Figure 4, where dashed line trajectories advance the simulation clock but dotted line trajectories do not.) Thus, the overall efficiency will be high when

$$\tau_{rxn}/M \gg \Delta t_{deph} + \Delta t_{corr}. \qquad 7.$$

Some tricks can further reduce this requirement. For example, whenever the system revisits a state, on all but one processor the interrupted trajectory from the previous visit can be immediately restarted, eliminating the dephasing stage. Also, the correlation stage (which involves only one processor) can be overlapped with the subsequent dephasing stage for the new state on the other processors, in the hope that there are no correlated crossings that lead to a different state.

Figure 5 shows an example of a parallel replica simulation; an Ag(111) island-on-island structure decays over a period of 1 μs at $T = 400$ K (F. Montalenti, T.C. Germann & A.F. Voter, in preparation). Many of the transitions involve concerted mechanisms. It is interesting to note that the boost on just 32 processors exceeds what we could currently obtain from either hyperdynamics or temperature-accelerated dynamics because of the relatively high temperature (400 K) of this simulation. Although these other two methods offer dramatically increasing boosts as the temperature is reduced (see below), parallel replica dynamics provides a consistently predictable boost for any situation where Equation 7 holds.

Birner et al. (49) have used parallel replica dynamics with up to 32 processors to study the growth of silicon interstitial clusters, scanning for trapping and diffusion mechanisms much more quickly than they could have with conventional MD. J.W. Halley & Y. Duan (in preparation) have used parallel replica to study the diffusion of Li ions in polyethylene oxide. To simplify and speed up the simulation, transition detection was based solely on the displacements of the Li ions, ignoring motion of the polymer chains. They verified that the distribution of first escape times was well approximated by an exponential, an important test in this type of case (see below).

This method also has a generality beyond the atomistic solid-state systems discussed here; the only requirements are that the first-escape-time distribution is exponential and that transitions can be identified. Pande and coworkers (47, 51) have used parallel replica dynamics to accelerate the folding of small polypeptides. In their systems, each state is primarily a free-energy basin (ignoring the low barriers corresponding to dihedral rotations of the backbone chain), entropically trapped by the difficulty the chain has in finding a better (or different) way to fold. They developed a novel transition detection scheme for these systems based on the potential energy fluctuations. Using widely distributed machines running parallel replica dynamics as low-level background processes (47), they have been able to follow the complete folding dynamics of some fast-folding proteins on the microsecond time scale.

Shirts & Pande (52) have analyzed the more general case when the escape-time distribution is not exponential. Parallelizing accelerates the evolution, although knowing how to advance the simulation clock becomes more complicated because the distribution shape must be known to make an estimate of the speed-up. For atomistic systems (or any energy-basin system), caution must be exercised if a nonexponential first-escape distribution is observed, as it indicates that either τ_{corr} has not been reached or the definition of a "state" is omitting some important substate transitions. For either case, there is a danger that the parallel replica procedure will give an incorrect sequence of transitions, corrupting the dynamics.

Parallel replica dynamics has the advantage of being fairly simple to program, with very few knobs to adjust, Δt_{deph} and Δt_{corr}, which can be conservatively set at a few picoseconds for most systems. The replica trajectories can themselves be accelerated trajectories, using, for instance, hyperdynamics on each processor, thus giving multiplicative boost factors (53, 54). As multiprocessing environments become more ubiquitous, with more processors within a node or even on a chip, and with loosely linked Beowulf clusters of such nodes, parallel replica dynamics will become an increasingly important simulation tool.

HYPERDYNAMICS

Method

In the hyperdynamics approach (55), the potential surface $V(\mathbf{r})$ of the system is modified by adding to it a non-negative bias potential $\Delta V(\mathbf{r})$, and a canonical

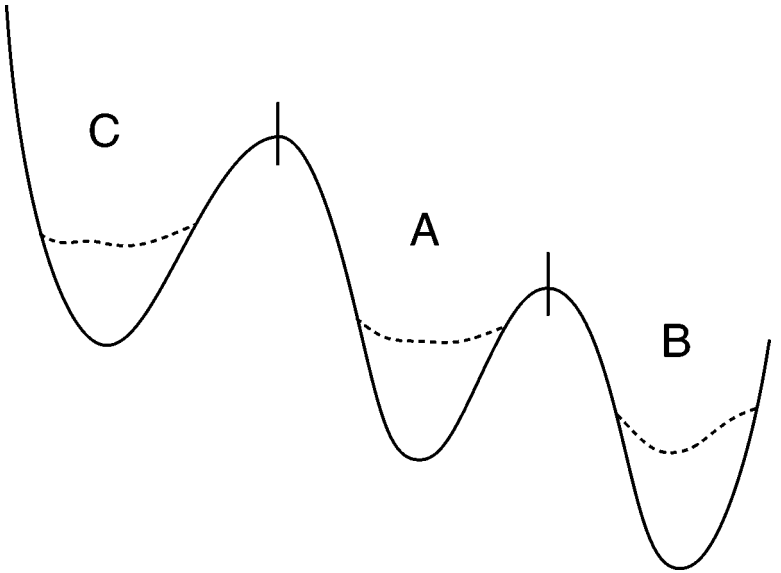

Figure 6 Schematic illustration of the hyperdynamics method. A bias potential ($\Delta V(\mathbf{r})$), is added to the original potential ($V(\mathbf{r})$, solid line). Provided that $\Delta V(\mathbf{r})$ meets certain conditions, primarily that it be zero at the dividing surfaces between states, a trajectory on the biased potential surface ($V(\mathbf{r}) + \Delta V(\mathbf{r})$, dashed line) escapes more rapidly from each state without corrupting the relative escape probabilities. The accelerated time is estimated as the simulation proceeds.

[constant-temperature (56)] classical trajectory is then propagated on this surface. A schematic illustration is shown in Figure 6. The derivation of the method requires that the system obeys TST; i.e, it assumes there are no correlated events. There are also important requirements on the form of the bias potential. It must be zero at all the dividing surfaces, and the system must still obey TST for dynamics on the biased potential. Assuming that we can construct such a bias potential, a challenging task discussed in some depth below, we obtain a very appealing result: A trajectory on this modified surface, while relatively meaningless on vibrational time scales, evolves correctly from state to state at an accelerated pace. (The evolution is correct in the sense that the probability of observing any particular sequence of states is the same for a simulation on the biased potential as for the unbiased potential.) Moreover, the accelerated time is easily estimated as the simulation proceeds. For a regular MD trajectory, the time advances at each integration step by Δt_{MD}, the MD time step (e.g., $\sim$1 fs). In hyperdynamics, the time advance at each step is Δt_{MD} multiplied by an instantaneous boost factor, the inverse Boltzmann factor for the bias potential at that point, so that the total time after n integration steps is

$$t_{hyper} = \sum_{j=1}^{n} \Delta t_{MD} \exp[\Delta V(\mathbf{r}(t_j))/k_B T]. \qquad 8.$$

Time thus takes on a statistical nature, advancing monotonically but nonlinearly. In the long-time limit, it converges on the correct value for the accelerated time with vanishing relative error. The overall computational speedup is then given by the average boost factor,

$$\text{boost(hyperdynamics)} = t_{hyper}/t_{MD} = \langle \exp[\Delta V(\mathbf{r})/k_B T] \rangle_b, \qquad 9.$$

divided by the extra computational cost of calculating the bias potential and its forces. In Equation 9, the $\langle\rangle_b$ indicates an average over the trajectory on the biased potential. If all the visited states are equivalent (e.g., this is common in calculations to test or demonstrate a particular bias potential), Equation 9 takes on the meaning of a true ensemble average.

The rate at which the trajectory escapes from a state is enhanced because the positive bias potential within the well lowers the effective barrier. Note, however, that the shape of the bottom of the well after biasing is irrelevant; no assumption of harmonicity is being made. The evolution from state to state is correct because the bias potential does not change the relative TST rates for different escape paths from a given state. Because of the delta function in Equation 1, only the denominator (see Equation 2) is affected by ΔV (which by construction is zero at the dividing surface), so the ratio of two TST rates is independent of ΔV.

The derivation of the hyperdynamics method builds on the basic concept of importance sampling, a well known technique in statistical mechanics (16, 23). That importance sampling (or umbrella sampling) can be used to fill in a potential basin to increase the probability for the system to be found near the transition state has been known for many years. Grimmelmann et al. (17) presented an excellent example of the power of this approach for a surface desorption process. More recently, it was suggested that it may be possible to construct a general bias potential without having prior knowledge of the nature of the transition state(s) or possible escape paths. Grubmüller (57) proposed a "conformational flooding" approach for macromolecular systems, in which the system is evolved in its current basin (actually a superbasin of basins for a macromolecular system) with regular MD to determine a representative configuration-space density. A coarse-grained representation of this population density (a multivariate Gaussian) is then obtained from the lowest eigenvectors of the covariance matrix, and this Gaussian in turn defines a bias potential of a harmonic form. In the hyperdynamics approach (55), independently developed later, a Hessian-based bias potential was proposed (see below), and the direct connection to a boosted time scale was derived, giving the exact long time dynamics for any bias potential meeting the requirements stated above.

Bias Potentials and Applications

The ideal bias potential should give a large boost factor, should have low computational overhead (although more overhead is acceptable if the boost factor is very high), and should to a good approximation meet the requirements given above. This is challenging because we want, as much as possible, to avoid utilizing any prior knowledge of the dividing surfaces or the available escape paths. The bias

potentials in the first hyperdynamics paper (55) were based on the lowest eigenvalue (ϵ_1) of the Hessian (the matrix $\partial^2 V/\partial x_i \partial x_j$). The bias potential was made positive for regions where $\epsilon_1 > 0$, and zero elsewhere, exploiting the fact that ϵ_1 is positive near the bottom of a basin and negative at saddle points. For a periodic two-dimensional example system (58), this gave substantial boosts (in the thousands when $k_B T$ was $\sim 1/20$ of the barrier height) and excellent accuracy, even when some recrossings were present. This approach was also tested on a more realistic atomistic system by simulating motion of a Ni adatom on a very small patch of Ni(100) using an embedded atom method (59) interatomic potential (60). The system was trimmed to just nine moving atoms to reduce the cost of diagonalizing the Hessian matrix. At $T = 500$ K, a boost factor of 40 was obtained (not counting the computational overhead of ~ 6), and this boost increased to 434 if the bias was turned on even when ϵ_1 was negative (but not as negative as the known value of ϵ_1 at the saddle point). This latter simulation reached 20 μs. These boost factors would have been larger at lower temperatures because a general property of hyperdynamics is that the boost increases roughly exponentially with decreasing temperature, as is evident from Equation 9. However, the overall rate of transitions (i.e., measured in computer time or human time), decreases even faster with temperature, and $T = 500$ K was the lowest temperature at which enough transitions could be observed (on computers in the year 1996) to test whether the rate constant was correctly predicted. As discussed below, this temperature dependence leads to the interesting observation that for many systems the power of hyperdynamics grows superlinearly with increasing computer speed.

A bias potential requiring a diagonalization of the full $3N$-dimensional Hessian at every time step becomes prohibitively expensive as N is increased beyond a few tens of atoms. Moreover, except for certain very simple systems, the fraction of configuration space with $\epsilon_1 > 0$ decreases as N increases, so that the boost vanishes for large systems. In a later paper (61), the bias potential and its calculation were modified in a few ways to address these issues. First, a more sophisticated form for the bias potential was developed that goes smoothly to zero in the proximity of any ridge top. Following the proposal of Sevick et al. (62), the system is assumed to be at a ridge top when the following conditions hold:

$$\epsilon_1 < 0 \quad \text{and} \quad g_{1p} = 0, \qquad 10.$$

where g_{1p} is the projection of the gradient onto the lowest eigenvector. This definition is not guaranteed to coincide with the exact ridge top between basins (63), but it is exact at a saddle point and a good approximation nearby (64).

A novel iterative method was also presented to obtain the lowest eigenvalue using only first derivatives of the potential, eliminating the need to diagonalize the Hessian, or even to construct it (61). The basic concept is to define the numerical approximation to the second derivative along an arbitrary direction $\mathbf{S}$ in $3N$-space {i.e., $d^2V/d\mathbf{S}^2 = [V(r+\eta\mathbf{S}) + V(r-\eta\mathbf{S}) - 2V(\mathbf{r})]/\eta^2$}, and then rotate $\mathbf{S}$ to minimize the value of this derivative. Within this numerical approximation, the optimized vector, $\mathbf{S}_{opt}$, is the direction of the lowest eigenvector, and the numerical

second derivative along $\mathbf{S}_{opt}$ is the lowest eigenvalue, ϵ_1. [Both Munro & Wales (65) and Henkelman & Jónsson (66) have used this approach to do mode-following (67) searches for saddle points, and the latter authors (66) have developed it into a method (the dimer method) efficient enough to use as the basis of an on-the-fly kinetic Monte Carlo procedure, as discussed below.] An iterative method (the lambda method) was also presented (61) for finding g_{1p} and its derivatives, as needed for a definition of ΔV based on Equation 10.

This bias potential gives larger boost factors than an expression based on the eigenvalue alone. A boost factor of 8310 was obtained (with a computational overhead of $\sim$30) in a 221.2 μs simulation of the $T = 300$ K diffusion of a ten-atom cluster of silver on Ag(111), a system with 70 moving atoms. However, the derivatives needed for the hyperdynamics forces in this bias form can be somewhat noisy because they are based on iteratively minimized quantities. Although in principle this noise can be reduced to give the correct derivatives to arbitrary accuracy by insisting on a high level of convergence, the derivatives of the projected gradient (g_{1p}) are especially sensitive to incomplete convergence in the lambda method. Stabilizing this procedure is the subject of ongoing study by our group and others (C.F. Sanz-Navarro, personal communication). Sanz-Navarro & Smith (69, 70) have tested an approximation to the g_{1p} derivative that is computationally less expensive and may in some cases be accurate enough for iterative Hessian-based hyperdynamics.

Alternative bias potentials have also been explored, the simplest of which is the "flat" bias potential proposed by Steiner et al. (71). They choose a fixed energy V_{flat}, which is higher than the basin minimum but lower than the energy of the lowest saddle point. As the trajectory propagates, if the potential energy $V(\mathbf{r})$ is greater than V_{flat}, the bias potential is zero. If $V(\mathbf{r})$ is below V_{flat}, the biased surface becomes flat ($V + \Delta V = V_{flat}$), and the trajectory skates across the ice on the pond. Interestingly, this type of bias potential actually has negative computational overhead because when the bias is turned on, the energy derivatives need not be calculated—the total force is zero. They demonstrated this approach for a two-dimensional model system and for a Lennard-Jones realization of the small fcc(100) surface diffusion system discussed above.

Because the average energy (relative to the basin minimum) of a system grows linearly with the number of degrees of freedom, a flat bias potential set below the lowest barrier will give vanishing boost for large systems [e.g., see (72)]; the system will only rarely have a potential energy lower than V_{flat}. Indeed, to obtain a substantial boost for their small fcc(100) terrace diffusion system with only nine moving atoms, Steiner et al. (71) had to put V_{flat} well above the lowest saddle point, requiring the calculation of a "numerator correction" (involving prior knowledge of the dividing surface) to obtain accurate hopping rates. This is an example of how our intuition can mislead us if we visualize a low-dimensional system. The thermal contribution to the total potential energy in a large system ($\sim 3Nk_BT/2$) is almost always greater than the lowest barrier, but the system only goes "over" the barrier when it has enough energy in the particular mode corresponding to the reaction

coordinate. For similar reasoning, however, for a system with many dimensions, it becomes safe to set V_{flat} somewhat above the lowest barrier (71) because the transition state subsystem has an average potential energy of $(3N-1)k_BT/2$. This approach has not been explored in detail, although D. Choudhary & P. Clancy (in preparation) have recently shown in annealing an amorphous silicon system that such a flat bias can be effective for gaining a small but useful boost factor with little programming investment, even for a system with hundreds of atoms.

In their paper, Steiner et al. (71) also proposed and tested the concept of a local bias potential—one based on the potential energy of only one or a few atoms. The idea is to focus attention on a known defect region, for instance around an adatom or vacancy, where the next activated event is most likely to occur. This can be a powerful approach in principle: By decoupling the size of the biased region (or regions) from the total system size, a large system can be studied with a simple and inexpensive bias potential.

Gong & Wilkins (74) subsequently extended the local bias potential method to a more sophisticated Hessian-based form. They achieved boost factors of 10^2 to 10^5 for diffusion of a dimer on a Lennard-Jones fcc(111) surface, diagonalizing the Hessian for the dimer adatoms and their neighbors (the active region). Similar boost factors have been obtained for monovacancy diffusion on an Al(100) surface, where the atoms immediately adjacent to the vacancy are taken as the active region (75). A single-atom, Hessian-based bias potential was also used by Fang & Wang (76) to study the diffusion of a vacancy in the core of an edge dislocation in Al, obtaining a boost factor of 10^4.

The local bias concept has also been explored by Fichthorn's group (77, 78). As in the Steiner et al. approach, they construct the bias energy based purely on the energy of an atom. Rather than a flat bias, they use a form such that from the point of view of that atom, the basin is softer and shallower. A significant feature of their approach, introduced in the second paper (78), is that the bias potential at any instant is defined based on the atom in the system that would give the lowest value for the bias potential. In this way, if any atom is about to jump, the bias potential will be defined based on that atom, turning the bias potential off. Substantial boost factors were obtained (78) for adatom diffusion on the Lennard-Jones fcc(100) surface; a lesser boost was observed for fcc(111) due to the lower jump barrier.

The local bias approaches should be applied very carefully, however, because mechanisms involving many atoms, or mechanisms in a seemingly unimportant region, will be suppressed if the bias potential is defined such that it does not recognize the mechanism (i.e., if $\Delta V > 0$ when the system is at the dividing surface for that mechanism). A good example of the caution required in such an approach is offered by the fcc(100) exchange mechanism discussed above, which would likely be suppressed by a local bias potential based purely on the energy of the adatom. Indeed, it has been demonstrated (54) that the local Hessian-based approaches artificially block almost all exchange events unless at least two neighbor shells around an adatom are included in the active region.

Nonetheless, the impressive boost factors demonstrated thus far using the various flavors of bias potentials are quite tantalizing and motivate the ongoing research into more powerful bias potentials.

TEMPERATURE-ACCELERATED DYNAMICS

In the temperature-accelerated dynamics (TAD) method (79), the idea is to speed up the transitions by increasing the temperature, while filtering out the transitions that should not have occurred at the original temperature. This filtering is critical, since without it the state-to-state dynamics will be inappropriately guided by entropically favored higher-barrier transitions. The TAD method is more approximate than the previous two methods in that it relies on the harmonic TST approximation (see above), but for many applications this additional approximation is acceptable, and the TAD method often gives substantially more boost than hyperdynamics or parallel replica dynamics. Consistent with the accelerated dynamics concept, the trajectory in TAD is allowed to wander on its own to find each escape path, so that no prior information is required about the nature of the reaction mechanisms.

In each basin, the system is evolved at a high temperature T_{high} (while the temperature of interest is some lower temperature T_{low}). Whenever a transition out of the basin is detected, the saddle point for the transition is found, e.g., using the nudged elastic band method (80–82). The trajectory is then reflected back into the basin and continued. This basin-constrained molecular dynamics (BCMD) procedure generates a list of escape paths and attempted escape times for the high-temperature system. [Chekmarev & Krivov have proposed a basin-confined MD approach for other applications (83).] Assuming that TST holds and that the system is chaotic and ergodic, the probability distribution for the first-escape time for each mechanism is an exponential (Equation 6). Because harmonic TST gives an Arrhenius dependence of the rate on temperature (Equation 4), depending only on the static barrier height, we can then extrapolate each escape time observed at T_{high} to obtain a corresponding escape time at T_{low} that is drawn correctly from the exponential distribution at T_{low}. This extrapolation, which requires knowledge of the saddle point energy, but not the preexponential factor, can be illustrated graphically in an Arrhenius-style plot ($\ln(1/t)$ versus $1/T$), as shown in Figure 7. The event with the shortest time at low temperature is the correct transition for escape from this basin. Because the extrapolation can in general cause a reordering of the escape times, a new shorter-time event may be discovered as the BCMD is continued at T_{high}. If we make the additional approximation that there is a minimum preexponential factor, ν_{min}, which bounds from below all the preexponential factors in the system, we can define a time at which the BCMD trajectory can be stopped, knowing that the probability that any transition observed after that time would replace the first transition at T_{low} is less than δ. This stop time is given by

$$t_{high,stop} \equiv \frac{\ln(1/\delta)}{\nu_{min}} \left(\frac{\nu_{min} t_{low,short}}{\ln(1/\delta)} \right)^{T_{low}/T_{high}}, \qquad 11.$$

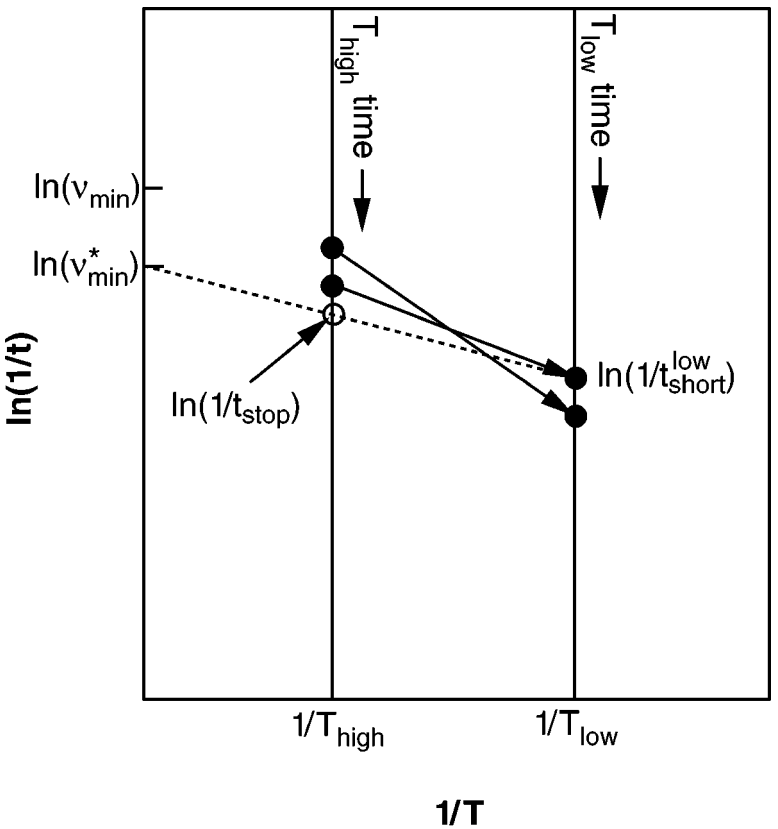

Figure 7 Schematic illustration of the temperature-accelerated dynamics method. Progress of the high-temperature trajectory can be thought of as moving down the vertical time line at $1/T_{high}$. For each transition detected during the run, the trajectory is reflected back into the basin, the saddle point is found, and the time of the transition (solid dot on left time line) is transformed (arrow) into a time on the low-temperature time line. Plotted in this Arrhenius-like form, this transformation is a simple extrapolation along a line whose slope is the negative of the barrier height for the event. The dashed termination line connects the shortest-time transition recorded so far on the low temperature time line (solid dot) with the confidence-modified minimum pre-exponential ($\nu^*_{min} = \nu_{min}/\ln(1/\delta)$) on the y axis. The intersection of this line with the high-T time line gives the time (t_{stop}, open circle) at which the trajectory can be terminated. With confidence 1-δ, we can say that any transition observed after t_{stop} could only extrapolate to a shorter time on the low-T time line if it had a pre-exponential lower than ν_{min}.

where $t_{low,short}$ is the shortest transition time at T_{low}. Once this stop time is reached, the system clock is advanced by $t_{low,short}$, the transition corresponding to $t_{low,short}$ is accepted, and the TAD procedure is started again in the new basin. The average boost in TAD can be dramatic when T_{high}/T_{low} is large. Any anharmonicity error at T_{high} transfers to T_{low}; a rate that is twice the Vineyard harmonic rate owing to anharmonicity at T_{high} will cause the transition times at T_{high} for that pathway to be 50% shorter, which in turn extrapolate to transition times that are 50% shorter at T_{low}. If the Vineyard approximation is perfect at T_{low}, these events will occur at twice the rate they should. This anharmonicity error can be controlled by choosing a T_{high} that is not too high.

As in hyperdynamics, the boost is limited by the lowest barrier, although this effect can be mitigated somewhat by treating repeated transitions in a "synthetic" mode (79). This is in essence a kinetic Monte Carlo treatment of the low-barrier transitions, in which the rate is estimated accurately from the observed transitions at T_{high}, and the subsequent low-barrier escapes observed during BCMD are excluded from the extrapolation analysis.

One of the fields where accelerated MD is desperately needed is vapor-deposited crystal growth, where the typical time scale can exceed minutes (44, 84). The application of accelerated MD to vapor-deposited crystal growth is discussed in Reference (85), and Figure 8 shows an example of the power of the TAD method for this problem (T.C. Germann, F. Montalenti & A.F. Voter, in preparation). We simulate vapor deposited growth of a Cu(100) surface at a deposition rate of one monolayer per 15 s and a temperature $T = 77$ K, exactly matching (except for the system size) the experimental conditions of Egelhoff & Jacob (41). Each deposition event is simulated using direct molecular dynamics for 2 ps, long enough for the atom to collide with the surface and settle into a binding site. A TAD simulation with $T_{high} = 550$ K then propagates the system for the remaining time (0.3 s on average for this system size and deposition rate) until the next deposition event is required. The overall boost factor is $\sim 10^7$, although it is not as large at higher temperatures. Even at this low temperature, many events accepted during the growth process involve concerted mechanisms, a few examples of which, observed during growth at $T = 77$ K and $T = 100$ K, are shown in Figure 9.

Especially interesting is the concerted sliding of an eight-atom cluster on a fcc(111) facet that formed during the growth. The embedded atom method potential used for this simulation (46) predicts a diffusion barrier for a Cu adatom on Cu(111) (87) that is probably too low, but this particular sliding event would have occurred before the next deposition even if the barrier (0.046 eV) was four times higher.

This MD/TAD procedure for simulating film growth has also been applied to Ag/Ag(100) at low temperatures (88), where a balance between normal-incidence steering (89) and thermally activated events was found to give a roughness that decreased with both temperature and the kinetic energy of the deposited atoms. Heteroepitaxial systems can also be considered, and experimental time scales have been reached in studying Cu/Ag(100) growth (J.A. Sprague, in preparation).

Heteroepitaxial systems are especially hard to treat with kinetic Monte Carlo (discussed below) because of the increased tendency for the system to go off lattice

owing to mismatch strain and because the rate table needs to be considerably larger when neighboring atoms can have multiple types.

Very recently, an enhancement to TAD, beyond the synthetic mode mentioned above, has been developed for systems that revisit states (91). The time required to accept an event can be reduced at each repeat visit by taking advantage of the time accumulated in previous visits. This procedure is exact; no assumptions beyond the ones required by the original TAD method are needed. After many visits, the procedure converges. The minimum barrier for escape from that state (E_{min}) is then known to within uncertainty δ. In this converged mode (ETAD), the average time at T_{high} required to accept an event no longer depends on δ, and the average boost factor becomes simply

$$\text{boost(ETAD)} = \frac{\overline{t}_{low,short}}{\overline{t}_{high,stop}} = \exp\left[E_{min}\left(\frac{1}{k_B T_{low}} - \frac{1}{k_B T_{high}} \right) \right] \qquad 12.$$

for that state. The additional boost (when converged) compared with the original TAD is an order of magnitude or more (91, 92).

For systems that seldom (or never) revisit the same state, it may still be possible to exploit this extra boost by running in ETAD mode with E_{min} supplied externally, a concept we explore below.

ON-THE-FLY KINETIC MONTE CARLO

Here we describe an alternative to the accelerated dynamics concept that has an identical goal: to make the system pass from state to state in a valid way as quickly as possible. In the standard kinetic Monte Carlo (KMC) approach, which has existed for many years (93–96), there is no classical trajectory at all. Instead, from a list of possible transition events, one escape path is chosen randomly, weighted by the rate constant, and the system is advanced to that new state. The clock is then incremented in a way that is consistent with the average time for escape from that state, which can be determined easily from the rate constants for the possible escape paths in the list. The rates for each of the events can be computed accurately from TST, making a "rate catalog" (95), using classical interatomic potentials, or even first-principles calculations (97). If the rate catalog is complete, the state-to-state dynamical evolution is exact.

The long-standing problem with this type of approach is that to generate a list of possible escape paths, approximations are usually made. Usually, only those mechanisms in which atoms stay on lattice are considered, a restriction that is sometimes unrealistic. The other, probably more serious problem, is that one assumes that the possible escape mechanisms can be guessed in advance. As discussed above, we now know that highly concerted events can be common and can be important to the dynamical evolution of the system. The number of possible events of this type can be very large, and sometimes the atomic motion is counterintuitive. As a result, it may be extremely difficult if not impossible to compile an adequate KMC catalog.

Recently, Henkelman & Jónsson (72) proposed a variation on the KMC method, in which one builds a state-specific catalog on the fly. The key to this new approach is the dimer method, which, as discussed above, is essentially a way to follow the lowest eigenvector up a trough to find the saddle at the top. Preliminary studies (66) indicate that this type of saddle search can be made so efficient [e.g., ∼400 force calls per saddle for adatom diffusion on Al(100)] that one can afford to start searches from many different randomly placed configurations, with the goal being to find, with high probability, all of the low-lying saddles surrounding the state, thus building a KMC catalog on the fly. If all saddles are found, then one has an exact KMC, and even if only the relevant low-lying saddles are found, one has a very accurate KMC; thus this on-the-fly KMC (OFKMC) is a very appealing approach. However, there is no guarantee that all relevant saddles can be found, and in fact this type of approach may systematically miss some obscured saddles, thus corrupting the dynamics. Also, if a state is revisited many times, it becomes increasingly important to find not just the low-lying saddles but also the higher-lying saddles. Otherwise, one could be incorrectly trapped in a superbasin of states, without finding the important escape path that is operative on longer time scales.

Despite these caveats, initial results with this method (which we will refer to as dimer-OFKMC) look promising; it has been applied to island ripening on Al(100) at $T = 300$ K, and vapor deposition (at 1 ML/ms) of Al on Al(100) at $T = 100$ K (72). In each case, a rich variety of mechanisms was observed, including many concerted mechanisms. 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).

This OFKMC concept has been applied in a more approximate way by Barkema & Mousseau (98, 99). With the goal of annealing amorphous systems, they designed the "activation-relaxation technique". Beginning from a random displacement, a first-derivative-only procedure pushes the system up the valley (the activation step), after which the system is relaxed into the basin on the other side of the saddle point. Accepting or rejecting the resulting basin-to-basin move according to the Metropolis probability (1) gives a pseudodynamics with thermally appropriate behavior.

COMPUTATIONAL SCALING WITH SYSTEM SIZE

As the above dynamical methods are developed further and come into wider use, their computational scaling with system size will become an important issue. Although the exact scaling depends on the type of system and many aspects of the implementation, a few general points can be made.

It can be shown that the computational work in both ETAD (91) and dimer-OFKMC (72) scales as O(N) for a finite-range classical interatomic potential (i.e., the same scaling as regular MD, resulting from the cost of each force call) for the trivial case of a system that has been enlarged by a simple replication. For the original TAD, the scaling for this case goes as $N^{2-T_{low}/T_{high}}$ (91). These scalings assume that all the work of dimer searches, transition detection, and saddle optimization can

be localized. However, real systems are more complicated. For example, states with low-barrier escape paths (which lower the boost) are more likely to show up somewhere in a large system than in a small system. To our knowledge, this type of effect has not been studied systematically.

The scaling in hyperdynamics depends both on the nature of the system and the form of the bias potential. Even for the ideal case of a bias potential requiring no computational overhead, there could still be an overall scaling worse than N for the reasons just given.

DISCUSSION AND PREDICTIONS

Applications of these methods to problems in materials research will become more prevalent as we gain experience with the methods and as computers get faster. In addition, development of the methods will remain a rich research area for a while. We conclude this article by discussing some of the their common properties and distinguishing features. We also point out some issues of importance for the near future and make a few predictions about how the methods may evolve.

Parallel replica dynamics is the most accurate of the methods, making no assumptions beyond that of infrequent events and exponential decay. Because it properly accounts for correlated dynamical events, it may be especially powerful for processes such as surface diffusion at a solid-liquid interface, where the whole concept of a saddle point is ill defined due to the myriad of possible orientations of the liquid molecules during the transition. It may even have important applications beyond atomistic systems. The maximum boost in parallel replica is the number of processors, although it is possible to combine parallel replica with hyperdynamics (53, 54) or TAD for multiplicative boost (until Equation 7 is no longer satisfied).

Hyperdynamics offers a general framework that will grow in power as better bias potentials are developed. It makes no harmonic approximation and even appears to be applicable in some situations where TST is violated (55; A.F. Voter, unpublished). In favorable cases, the boost factor can be dramatic, and it increases exponentially with decreasing temperature.

TAD is the most approximate of the three accelerated dynamics methods but is typically easy to implement and, in our experience so far, generally gives the largest boost factor. It requires that the system is well described by the harmonic TST approximation at both T_{low} and T_{high}.

Dimer-OFKMC is a powerful approach, and a significant improvement in quality over usual KMC, releasing the lattice restriction and eliminating the rate table (which is usually woefully incomplete). We have not yet performed a direct comparison, but for a typical case it appears that it is probably faster than TAD, the most closely related accelerated MD approach. This disparity will probably diminish as each of the methods is developed further (see below). Dimer-OFKMC relies on the harmonic approximation (Equation 4). Compared with TAD, it should have less anharmonicity error because it is only the anharmonicity at the simulation temperature that matters (however, depending on the temperature dependence of the anharmonicity, even assuming it increases with T, a cancellation of errors can

result in TAD having less error). Vineyard prefactors must be calculated (or a fixed prefactor assumed, a sometimes dangerous approximation). This work scales as N^3, but for large systems, this dependence can be eliminated by restricting the Hessian diagonalization to a subspace (100). The major approximation in dimer-OFKMC arises from the fact that there is no guarantee, given a finite number of dimer searches, that all the relevant saddle points will be found. This becomes a more serious issue as a system is evolved for many transitions, since there is an increasing cumulative probability that a high-barrier event (which may have been missed) should be selected in the KMC step.

The temperature dependence of the boost in hyperdynamics and TAD gives rise to an interesting prediction about their power and utility in the future. Sometimes, even accelerating the dynamics may not make the activated processes occur frequently enough to study a particular process. A common trick is to raise the temperature just enough that at least some events will occur in the available computer time, hoping, of course, that the behavior of interest is still representative of the lower-T system. When a faster computer becomes available, the same system can be studied at a lower, more desirable, temperature. This in turn increases the boost factor (e.g., see Equations 9 and 11), so that, effectively, there is a superlinear increase in the power of accelerated dynamics with increasing computer speed. (For hyperdynamics, this is an especially important effect, because if the boost does not exceed the overhead cost of the bias potential, there is no point in doing the hyperdynamics simulation at all—direct MD is faster.) Thus the accelerated dynamics approaches will become increasingly more powerful in future years simply because computers keep getting faster.

Using forces from first-principles calculations (e.g., density functional theory or quantum chemistry) for atomistic simulations is an increasingly popular approach (101). Because this is much more expensive than using a classical interatomic potential, the converse of the argument given above indicates that first-principles accelerated dynamics simulations will not give much useful boost on current computers (i.e., using first principles to calculate the forces is like having a very slow computer). First-principles hyperdynamics may be a powerful tool 5–10 years down the road, when breakeven (boost = overhead) is reached, and this could happen sooner with development of less-expensive bias potentials. For small systems, the flat bias potential of Steiner et al. (71) could be effective now. TAD is probably close to being viable for first-principles dynamics, while parallel replica dynamics, dimer-OFKMC, and the dimer-based ETAD discussed below could probably be used on today's computers for first-principles studies on some systems.

An interesting and important area for research is the low-barrier problem. Low barriers degrade the boost of hyperdynamics and TAD (and even parallel replica, if they cause Equation 7 to be no longer satisfied), and also reduce the efficiency in standard KMC, OFKMC, and TAD in KMC (synthetic) mode. For example, in hyperdynamics, which is the most susceptible to low barriers, it may be possible to develop advanced bias potentials that can safely block low barriers to give the boost necessary to escape from a superbasin (a basin of basins), although the possibility that detailed balance is disrupted then becomes an issue.

As mentioned above, the TAD method can be particularly efficient in boosting the dynamics when the minimum barrier E_{min} to escape from each state is known (ETAD). One potentially powerful future development would be to combine the dimer approach with ETAD. Upon entering a new state, a small number of dimer searches would be used to find the minimum barrier for escape, after which ETAD would be employed to quickly find a dynamically appropriate escape path. This would exploit the power of the dimer method to quickly find low-barrier pathways, while eliminating the danger associated with the possibility that it might miss important escape paths. Although the dimer method could fail to find the lowest barrier correctly, this is a much weaker demand on the dimer method than trying to find all relevant barriers (including all of the lowest ones). This approach would also eliminate the need to compute prefactors (Equation 5).

TAD itself can also be cast in an OFKMC form (A.F. Voter et al., in preparation). For each mechanism discovered at high temperature, the rate constant at T_{low} would be computed directly, using the full Vineyard expression (Equations 4 and 5). An expression similar to Equation 11 (and augmented by an additional run time to allow for anharmonicity) indicates when it is safe to stop the BCMD at T_{high}, knowing that the OFKMC catalog for this state is complete enough (with δ confidence) to accept the next KMC event. Each revisit of a state requires successively less BCMD time to establish this. In this approach, the rate at T_{low} can even be computed with a more accurate treatment, e.g., including quantum effects (23) or using first-principles.

As we hope we have demonstrated, the future of these methods looks quite bright. They are powerful enough now to study a wide range of materials problems, and there is plenty of room for further development.

ACKNOWLEDGMENTS

We gratefully acknowledge vital collaborations with Mads Sørensen, James Sprague, Blas Uberuaga, Sriram Swaminarayan, Stephane Mazevet, Walter Rudd, Jeongnim Kim, and Wolfgang Windl. Discussions with Hannes Jónsson, Graeme Henkelman, Tony Redondo, Joel Kress, John Hamilton, Jacques Amar, Carlos Sanz-Navarro, Vijay Pande, and Steve Stuart have also been very helpful. This work was supported by the United States Department of Energy (DOE), Office of Basic Energy Sciences, under DOE Contract No. W-7405-ENG-36.

The *Annual Review of Materials Research* is online at http://matsci.annualreviews.org

LITERATURE CITED

1. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. 1953. *J. Chem. Phys.* 21:1087–92
2. Frenkel D, Smit B. 1996. *Understanding Molecular Simulation*. San Diego: Academic
3. Keck JC. 1962. *Discuss. Faraday Soc.* 33: 173

4. Bennett CH. 1977. In *Algorithms for Chemical Computation*, ed. RE Christofferson, pp. 63–97. Washington, DC: Am. Chem. Soc.
5. Chandler D. 1978. *J. Chem. Phys.* 68: 2959–70
6. Toller M, Jacucci G, DeLorenzi G, Flynn CP. 1985. *Phys. Rev. B* 32:2082–95
7. Voter AF, Doll JD. 1985. *J. Chem. Phys.* 82:80–92
8. Voter AF. 1989. *Phys. Rev. Lett.* 63:167–70
9. Anderson JB. 1995. *Adv. Chem. Phys.* 91: 381–431
10. Marcelin R. 1915. *Ann. Phys.* 3:120–231
11. Wigner E. 1932. Z. *Phys. Chem. B* 19:203
12. Eyring H. 1935. *J. Chem. Phys.* 3:107–15
13. Truhlar DG, Garrett BC, Klippenstein SJ. 1996. *J. Phys. Chem.* 100:12771–800
14. Zwanzig R. 2001. *Nonequilibrium Statistical Mechanics*. Oxford, UK: Oxford Univ. Press
15. Voter AF, Doll JD. 1984. *J. Chem. Phys.* 80:5832–38
16. Valleau JP, Whittington SG. 1977. In *Modern Theoretical Chemistry*, ed. BJ Berne, 5:137–68. New York: Plenum
17. Grimmelmann EK, Tully JC, Helfand E. 1981. *J. Chem. Phys.* 74:5300–10
18. Voter AF, Doll JD. 1984. *J. Chem. Phys.* 80:5814–17
19. Voter AF. 1985. *J. Chem. Phys.* 82:1890–99
20. Truhlar DG, Garrett BC. 1980. *Acc. Chem. Res.* 13:440–48
21. Bennett CH. 1975. In *Diffusion in Solids: Recent Developments*, ed. AS Nowick, JJ Burton, pp. 73–113. New York: Academic
22. Montgomery JA Jr, Chandler D, Berne BJ. 1979. *J. Chem. Phys.* 70:4056–66
23. Berne BJ, Ciccotti G, Coker DF, eds. 1998. *Classical and Quantum Dynamics in Condensed Phase Simulations*. Singapore: World Scientific
24. Berne BJ, Borkovec M, Straub JE. 1988. *J. Phys. Chem.* 92:3711–25
25. Discarding translational modes, there are 3*N*-3 and 3*N*-4 real normal mode frequencies at the minimum and saddle, respectively, for a three-dimensional periodic solid system. For a system that is free to rotate, there are 3*N*-6 and 3*N*-7.
26. Vineyard GH. 1957. *J. Phys. Chem. Solids* 3:121–27
27. Hanggi P, Talkner P, Borkovec M. 1990. *Rev. Mod. Phys.* 62:251–341
28. Kramers HA. 1940. *Physica* 7:284–304
29. Chandrasekhar S. 1943. *Rev. Mod. Phys.* 15:1–89
30. Feibelman PJ. 1990. *Phys. Rev. Lett.* 65:729–32
31. Kellogg GL, Feibelman PJ. 1990. *Phys. Rev. Lett.* 64:3143–46
32. Chen C, Tsong TT. 1990. *Phys. Rev. Lett.* 64:3147–50
33. Stoltze P, Norskov JK. 1993. *Phys. Rev. B* 48:5607–11
34. Villarba M, Jónsson H. 1994. *Phys. Rev. B* 49:2208–11
35. Wang R, Fichthorn KA. 1994. *Surf. Sci.* 301:253–59
36. Hamilton JC, Daw MS, Foiles SM. 1995. *Phys. Rev. Lett.* 74:2760–63
37. Shi Z-P, Zhang Z, Swan AK, Wendelken JF. 1996. *Phys. Rev. Lett.* 76:4927–30
38. Montalenti F, Ferrando F. 1999. *Phys. Rev. Lett.* 82:1498–501
39. Linderoth TR, Horch S, Petersen L, Helveg S, Laegsgaard E, et al. 1999. *Phys. Rev. Lett.* 82:1494–97
40. Trushin OS, Ala Nissila A. 2000. *Phys. Rev. B* 62:1611–14
41. Egelhoff WF Jr, Jacob I. 1989. *Phys. Rev. Lett.* 62:921–24
42. Stoldt CR, Caspersen KJ, Bartelt MC, Jenks CJ, Evans JW, Thiel PA. 2000. *Phys. Rev. Lett.* 85:800–3
43. Kunkel R, Poelsema B, Verheig LK, Comsa G. 1990. *Phys. Rev. Lett.* 65:733–36
44. Jónsson H. 2000. *Annu. Rev. Phys. Chem.* 51:623–53
45. Richie DA, Kim J, Wilkins JW. 2001. *Mater. Res. Soc. Symp. Proc.* 677:AA5.1
46. Voter AF. 1998. *Phys. Rev. B* 57:13985–88

47. Shirts M, Pande VS. 2000. *Science* 290: 1903–4
48. Deleted in proof
49. Birner S, Kim J, Richie DA, Wilkins JW, Voter AF, Lenosky T. 2001. *Solid State Commun.* 120:279–82
50. Deleted in proof
51. Zagrovic B, Sorin EJ, Pande V. 2001. *J. Mol. Biol.* 313:151–69
52. Shirts MR, Pande VS. 2001. *Phys. Rev. Lett.* 86:4983–87
53. Voter AF, Germann TC. 1998. *Mater. Res. Soc. Symp. Proc.* 528:221–36
54. Rudd WG, Voter AF. 1999. *Mater. Res. Soc. Symp. Proc.* 538:485–90
55. Voter AF. 1997. *J. Chem. Phys.* 106:4665–77
56. Coupling to a thermostat guarantees that the hyperdynamics trajectory is canonical on long-time scales, but by keeping the coupling relatively weak, the trajectory gives nearly perfect microconanical (energy conserving) behavior on short-time scales.
57. Grubmüller H. 1995. *Phys. Rev. E* 52: 2893–906
58. Note that there is a typographical error in the definition of the 2-D potential in that paper (Reference 55). The second term in Equation 20, involving d_2 should be multiplied by 2π
59. Daw MS, Foiles SM, Baskes MI. 1993. *Mater. Sci. Rep.* 9:251–310
60. Voter AF, Chen SP. 1987. *Mater. Res. Soc. Symp. Proc.* 82:175–80
61. Voter AF. 1997. *Phys. Rev. Lett.* 78:3908–11
62. Sevick EM, Bell AT, Theodorou DN. 1993. *J. Chem. Phys.* 98:3196–212
63. Hoffman DK, Nord RS, Ruedenberg K. 1986. *Theor. Chim. Acta* 69:265–79
64. Equation 10 can also be satisfied in regions away from the ridge tops; one situation is when the lowest two eigenvalues (ε_1 and ε_2) cross, causing g_{1p} to pass through zero. In Reference 61, to reduce the probability of this occurrence, a repulsive term was added to ΔV, turning the bias potential on when ε_2–ε_1 was less than a threshold value; this threshold should be kept small to reduce or prevent blocking of the dividing surface.
65. Munro LJ, Wales DJ. 1999. *Phys. Rev. B* 59:3969–80
66. Henkelman G, Jónsson H. 1999. *J. Chem. Phys.* 111:7010–22
67. Cerjan CJ, Miller WH. 1981. *J. Chem. Phys.* 75:2800–6
68. Deleted in proof
69. Sanz-Navarro CF, Smith R. 2001. *Nucl. Instr. Methods Phys. Res. B* 180:130–38
70. Sanz-Navarro CF, Smith R. 2001. *Comp. Phys. Commun.* 137:206–21
71. Steiner MM, Genilloud P-A, Wilkins JW. 1998. *Phys. Rev. B* 57:10236–39
72. Henkelman G, Jónsson H. 2001. *J. Chem. Phys.* 115:9657–66
73. Deleted in proof
74. Gong XG, Wilkins JW. 1999. *Phys. Rev. B* 59:54–57
75. Duan XM, Sun DY, Gong XG. 2001. *Comput. Mater. Sci.* 20:151–56
76. Fang QF, Wang R. 2000. *Phys. Rev. B* 62:9317–24
77. Pal S, Fichthorn KA. 1999. *Chem. Eng. J.* 74:77–83
78. Wang J-C, Pal S, Fichthorn KA. 2001. *Phys. Rev. B* 63:085403:1–9
79. Sørensen MR, Voter AF. 2000. *J. Chem. Phys.* 112:9599–606
80. Jónsson H, Mills G, Jacobsen KW. 1998. In *Classical and Quantum Dynamics in Condensed Phase Simulations,* ed. BJ Berne, G Ciccotti, DF Coker, Chpt. 16. Singapore: World Scientific
81. Henkelman G, Jónsson H. 2000. *J. Chem. Phys.* 113:9978–85
82. Henkelman G, Uberuaga BP, Jónsson H. 2000. *J. Chem. Phys.* 113:9901–4
83. Chekmarev SF, Krivov SV. 1998. *Chem. Phys. Lett.* 287:719–24
84. Brune H. 1998. *Surf. Sci. Rep.* 31:125–229
85. Montalenti F, Voter AF. 2001. *Phys. Stat. Solidi B* 226:21–27

86. Deleted in proof
87. Liu CL, Cohen JM, Adams JB, Voter AF. 1991. *Surf. Sci.* 253:334–44
88. Montalenti F, Sørensen MR, Voter AF. 2001. *Phys. Rev. Lett.* 87:126101:1–4
89. Montalenti F, Voter AF. 2001. *Phys. Rev. B* 64:081401(R):1–4
90. Deleted in proof
91. Montalenti F, Voter AF. 2002. *J. Chem. Phys.* 116:4819–28
92. Deleted in proof
93. Bortz AB, Kalos MH, Lebowitz JL. 1975. *J. Comp. Phys.* 17:10–18
94. Gillespie DT. 1976. *J. Comp. Phys.* 22: 403–34
95. Voter AF. 1986. *Phys. Rev. B* 34:6819–29
96. Fichthorn KA, Weinberg WH. 1991. *J. Chem. Phys.* 95:1090–96
97. Jacobsen J, Jacobsen KW, Stoltze P, Norskov JK. 1995. *Phys. Rev. Lett.* 74:2295–98
98. Barkema GT, Mousseau N. 1996. *Phys. Rev. Lett.* 77:4358–61
99. Barkema GT, Mousseau N. 2001. *Comput. Mater. Sci.* 20:285–92
100. Mishin Y, Sørensen MR, Voter AF. 2001. *Phil. Mag. A* 81:2591–612
101. Payne MC, Teter MP, Allan DC, Arias TA, Joannopoulos JD. 1992. *Rev. Mod. Phys.* 64:1045–95

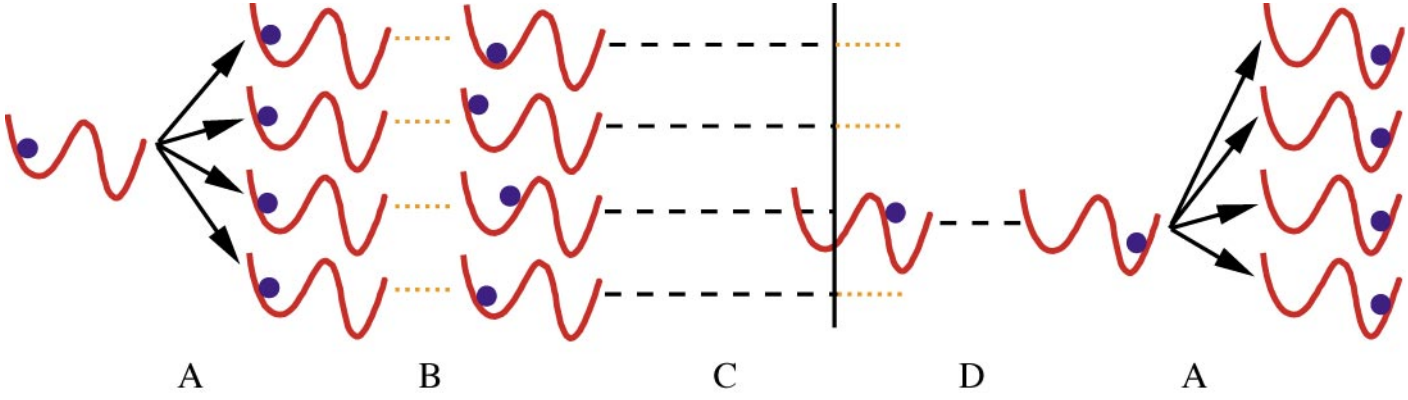

Figure 4 Schematic illustration of the parallel replica method. The four steps, described in the text are (*a*) replication of the system into *M* copies, (*b*) dephasing of the replicas, (*c*) independent trajectories until a transition is detected in any of the replicas, and (*d*) brief continuation of the transitioning trajectory to allow for correlated events such as recrossings or follow-on transitions to other states. The resulting configuration is the replicated, beginning the process again.

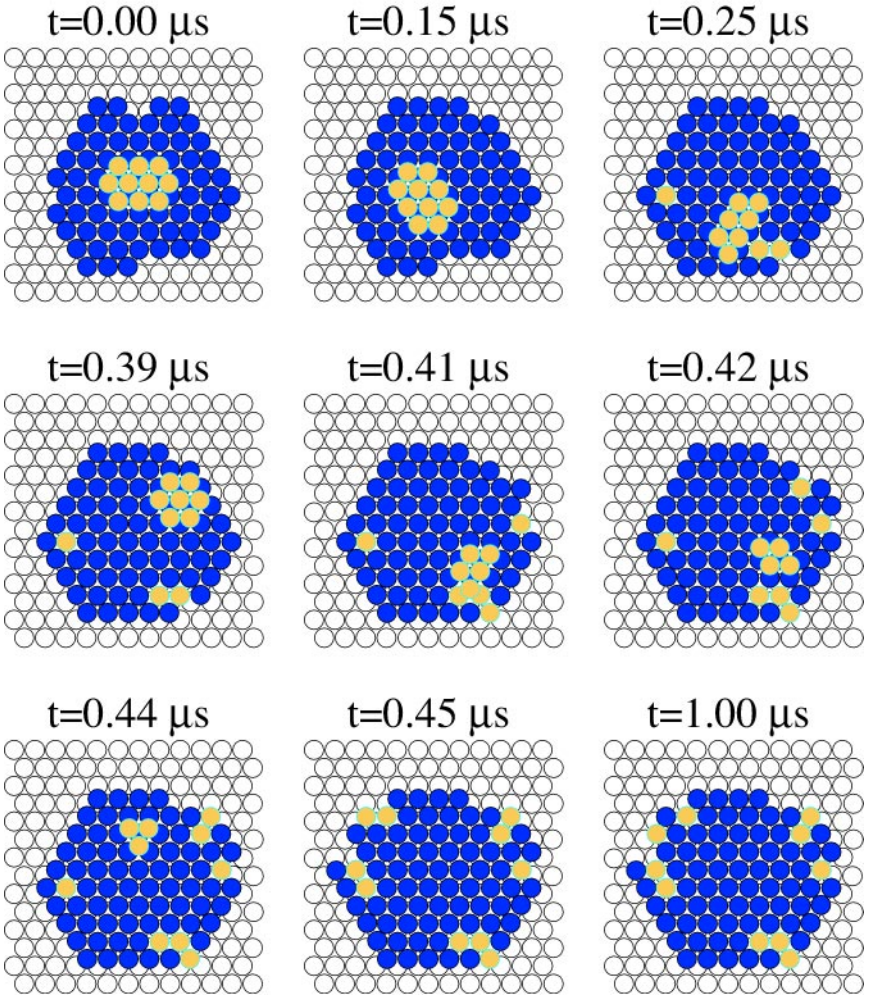

Figure 5 Nine snapshots from a parallel replica simulation of an island on top of an island on the Ag(111) surface at $T = 400$ K. On a microsecond time scale, the upper island gives up all its atoms to the lower island, filling vacancies and kink sites as it does so. This simulation took five days to reach one microsecond on 32 1-GHz Pentium III processors.

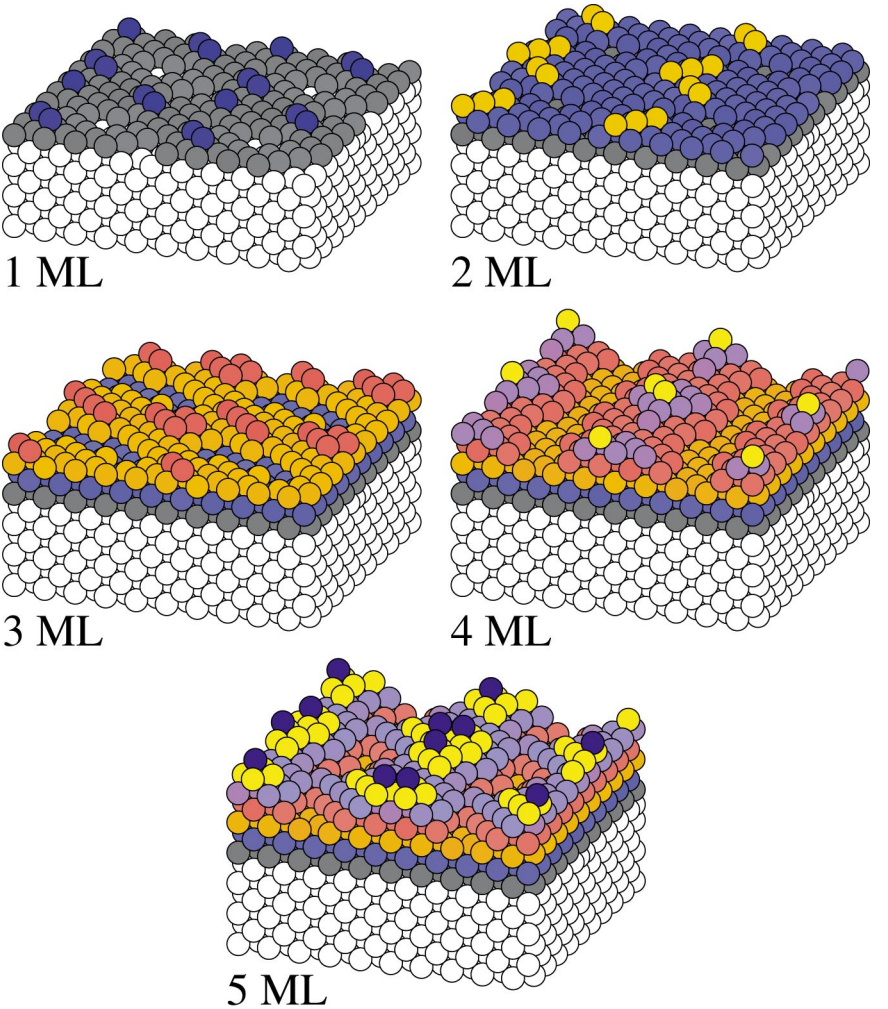

Figure 8 Snapshots from a TAD simulation of the deposition of five monolayers (ML) of Cu onto Cu(100) at 0.067 ML/s and $T = 77$ K, matching the experimental conditions of Egelhoff & Jacob (41). Deposition of each new atom was performed using direct molecular dynamics for 2 ps, while the intervening time (0.3 s on average for this 50 atom/layer simulation cell) was simulated using the TAD method. This simulation would take 10^7 years using direct MD.

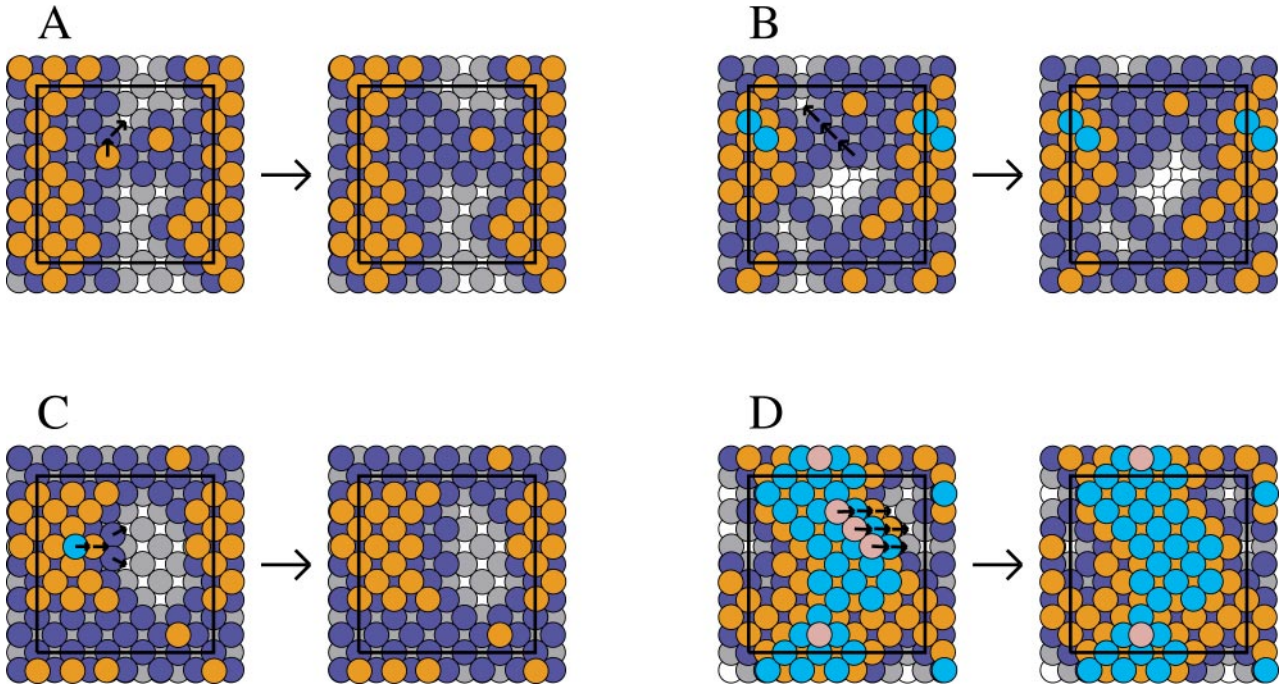

Figure 9 Examples of concerted mechanisms observed during TAD simulations of Cu/Cu(100) epitaxial growth, described in the text. Atoms are deposited at an average flux of one monolayer per 15 s, which corresponds to the experimental rate of Egelhoff & Jacob (41). (*a*) A 2-atom downward exchange, with $E_a = 0.255$ eV, which occurred after 3.00 ms at T = 100 K; (*b*) a 3-atom row sliding, $E_a = 0.241$ eV (28.3 ms at 100 K); (*c*) downward exchange pushing two additional atoms in a metastable off-lattice site, $E_a = 0.056$ eV (88.9 ps at 100 K); (*d*) sliding of an 8-atom cluster down a {111} facet, $E_a = 0.046$ eV (45.2 ns at 77 K).