2.11

ACCELERATED MOLECULAR DYNAMICS METHODS

Blas P. Uberuaga[1], Francesco Montalenti[2], Timothy C. Germann[3], and Arthur F. Voter[4]

[1] *Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA*
[2] *INFM, L-NESS, and Dipartimento di Scienza dei Materiali, Università degli Studi di Milano-Bicocca, Via Cozzi 53, I-20125 Milan, Italy*
[3] *Applied Physics Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA*
[4] *Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA*

Molecular dynamics (MD) simulation, in which atom positions are evolved by integrating the classical equations of motion in time, is now a well established and powerful method in materials research. An appealing feature of MD is that it follows the actual dynamical evolution of the system, making no assumptions beyond those in the interatomic potential, which can, in principle, be made as accurate as desired. However, the limitation in the accessible simulation time represents a substantial obstacle in making useful predictions with MD. Resolving individual atomic vibrations – a necessity for maintaining accuracy in the integration – requires time steps on the order of femtoseconds, so that reaching even one microsecond is very difficult on today's fastest processors. Because this integration is inherently sequential in nature, direct, spatial parallelization does not help significantly; it just allows simulations of nanoseconds on much larger systems.

Beginning in the late 1990s, methods based on a new concept have been developed for circumventing this time scale problem. For systems in which the long-time dynamical evolution is characterized by a sequence of activated events, these "accelerated molecular dynamics" methods [1] can extend the accessible time scale by orders of magnitude relative to direct MD, while retaining full atomistic detail. These methods – hyperdynamics, parallel-replica dynamics, and temperature accelerated dynamics (TAD) – have already been demonstrated on problems in surface and bulk diffusion and surface growth. With more development they will become useful for a broad range of key materials problems, such as pipe diffusion along a dislocation core, impurity clustering, grain

S. Yip (ed.),
Handbook of Materials Modeling, 629–648.

growth, dislocation climb and dislocation kink nucleation. Here we give an introduction to these methods, discuss their current strengths and limitations, and predict how their capabilities may develop in the next few years.

1. Background

1.1. Infrequent Event Systems

We begin by defining an "infrequent-event" system, as this is the type of system we will focus on in this article. The dynamical evolution of such a system is characterized by the occasional activated event that takes the system from basin to basin, events that are separated by possibly millions of thermal vibrations within one basin. A simple example of an infrequent-event system is an adatom on a metal surface at a temperature that is low relative to the diffusive jump barrier. We will exclusively consider thermal systems, characterized by a temperature T, a fixed number of atoms N, and a fixed volume V; i.e., the canonical ensemble. Typically, there is a large number of possible paths for escape from any given basin. As a trajectory in the $3N$-dimensional coordinate space in which the system resides passes from one basin to another, it crosses a $(3N-1)$-dimensional "dividing surface" at the ridgetop separating the two basins. While on average these crossings are infrequent, successive crossings can sometimes occur within just a few vibrational periods; these are termed "correlated dynamical events" [2–4]. An example would be a double jump of the adatom on the surface. For this discussion it is sufficient, but important, to realize that such events can occur. In most of the methods presented below, we will 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 longer than τ_{corr} has no memory of its history and, consequently, how it got to that basin, in the sense that when it later escapes from the basin, the probability for escape 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, which we will define below.

An infrequent event system, then, is one in which the residence time in a state (τ_{rxn}) is much longer than the correlation time (τ_{corr}). We will focus here on systems with energetic barriers to escape, but the infrequent-event concept applies equally well to entropic bottlenecks.[1] The key to the accelerated

[1]For systems with entropic bottlenecks, the parallel-replica dynamics method can be applied very effectively [1].

dynamics methods described here 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.

1.2. Transition State Theory

Transition state theory (TST) [5–9] is the formalism underpinning 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 (Fig. 1). If there are no correlated dynamical events, the TST rate is the exact rate constant for the system to move from state A to state B.

The power of TST comes from the fact that this flux is an *equilibrium* property of the system. Thus, we can compute the TST rate without ever propagating a trajectory. The appropriate ensemble average for the rate constant for escape from A, $k_{A\rightarrow}^{\mathrm{TST}}$, is

$$k_{A\rightarrow}^{\mathrm{TST}} = \langle |\mathrm{d}x/\mathrm{d}t|\, \delta(x-q)\rangle_A, \tag{1}$$

where $x \in \mathbf{r}$ is the reaction coordinate and $x = q$ the dividing surface bounding state A. The angular brackets indicate the ratio of Boltzmann-weighted integrals over $6N$-dimensional phase space (configuration space $\mathbf{r}$ and momentum space $\mathbf{p}$). That is, for some property $P(\mathbf{r}, \mathbf{p})$,

$$\langle P\rangle = \frac{\int\int P(\mathbf{r},\mathbf{p})\exp[-H(\mathbf{r},\mathbf{p})/k_{\mathrm{B}}T]\,\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{p}}{\int\int \exp[-H(\mathbf{r},\mathbf{p})/k_{\mathrm{B}}T]\,\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{p}}, \tag{2}$$

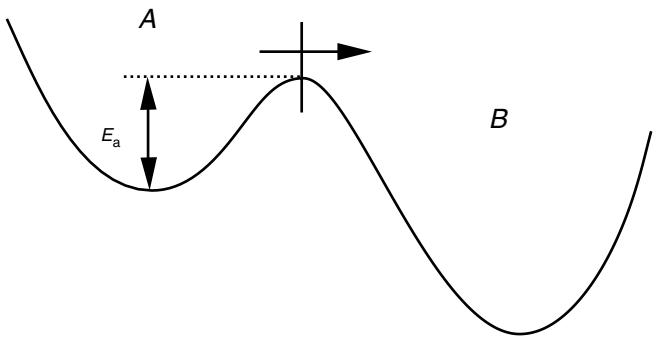

Figure 1. A two-state system illustrating the definition of the transition state theory rate constant as the outgoing flux through the dividing surface bounding state A.

where k_B is the Boltzmann constant and $H(\mathbf{r}, \mathbf{p})$ is the total energy of the system, kinetic plus potential. The subscript A in Eq. (1) indicates the configuration space integrals are restricted to the space belonging to state A. If the effective mass (m) of the reaction coordinate is constant over the dividing surface, Eq. (1) reduces to a simpler ensemble average over configuration space only [10],

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

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 else it can be in state A. Note that there is no dependence on the nature of the final state B.

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. 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 than in some chemical systems, such as molecular reactions in solution or the gas phase, where TST is just a starting point and dynamical corrections can lower the rate significantly [11].

While 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.

1.3. Harmonic Transition State Theory

If we have identified a saddle point on the potential energy surface for the reaction pathway between A and B, we can use a further approximation to TST. 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. Under these conditions, the TST rate constant becomes simply

$$k_{A\rightarrow B}^{\mathrm{HTST}} = \nu_0 e^{-E_a/k_B T}, \tag{4}$$

where

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

Here E_a is the static barrier height, or activation energy (the difference in energy between the saddle point and the minimum of state A (Fig. 1)), $\{\nu_i^{\min}\}$ are the normal mode frequencies at the minimum of A, and $\{\nu_i^{\text{sad}}\}$ are the non-imaginary normal mode frequencies at the saddle separating A from B. This is often referred to as the Vineyard [12] equation. The analytic integration of Eq. (1) over the whole phase space thus leaves a very simple Arrhenius temperature dependence.[2] 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 method (without requiring calculation of the prefactor ν_0).

1.4. 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 will 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 cannot be predicted by considering them individually. In very simple systems we can raise the temperature to make diffusive transitions occur on an MD-accessible time scale. However, as systems become more complex, changing the temperature causes corresponding changes in the relative probability of competing mechanisms. Thus, this strategy will cause the system to select a different sequence of state-to-state dynamics, ultimately leading to a completely different evolution of the system, and making it impossible to address the questions that the simulation was attempting to answer.

Many, if not most, materials problems are characterized by such complex infrequent events. We may want to know what happens on the time scale of milliseconds, seconds or longer, while with MD we can barely reach one microsecond. Running at higher T or trying to guess what the underlying atomic 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 run a single trajectory for that long, our understanding of the system would improve substantially. This, in essence, is the primary motivation for the development of the methods described here.

[2]Note that although the exponent in Eq. (4) depends only on the static barrier height E_a, in this HTST approximation there is no assumption that trajectory passes exactly through the saddle point.

1.5. Dividing Surfaces and Transition Detection

We have implied that the ridgetops between basins are the appropriate dividing surfaces in these systems. For a system that obeys TST, these ridgetops are the optimal dividing surfaces; recrossings will occur for any other choice of dividing surface. A ridgetop can be defined in terms of steepest-descent paths – it is the $3N-1$-dimensional boundary surface that separates those points connected by steepest descent paths to the minimum of one basin from those that are connected to the minimum of an adjacent basin. This definition also leads to a simple way to detect transitions as a simulation proceeds, a requirement of parallel-replica dynamics and temperature accelerated dynamics. Intermittently, the trajectory is interrupted and minimized through steepest descent. If this minimization leads to a basin minimum that is distinguishable from the minimum of the previous basin, a transition has occurred. An appealing feature of this approach is that it requires virtually no knowledge of the type of transition that might occur. Often only a few steepest descent steps are required to determine that no transition has occurred. While this is a fairly robust detection algorithm, and the one used for the simulations presented below, more efficient approaches can be tailored to the system being studied.

2. Parallel-Replica Dynamics

The parallel-replica method [13] is the simplest and most accurate of the accelerated dynamics techniques, with the only assumption being that the infrequent events obey first-order kinetics (exponential decay); i.e., for any time $t > \tau_{\text{corr}}$ after entering a state, the probability distribution function for the time of the next escape is given by

$$p(t) = k_{\text{tot}}\, e^{-k_{\text{tot}} t}, \tag{6}$$

where k_{tot} is the rate constant for escape from the state. For example, Eq. (6) arises naturally for ergodic, chaotic exploration of an energy basin. Parallel-replica allows for the parallelization of the state-to-state dynamics of such a system on M processors. We sketch the derivation here for equal-speed processors. For a state in which the rate to escape is k_{tot}, on M processors the effective escape rate will be Mk_{tot}, as the state is being explored M times faster. Also, if the time accumulated on one processor is t_1, on the M processors a total time of $t_{\text{sum}} = Mt_1$ will be accumulated. Thus, we find that

$$p(t_1)\,\mathrm{d}t_1 = Mk_{\text{tot}}\, e^{-Mk_{\text{tot}}t_1}\,\mathrm{d}t_1 \tag{7a}$$

$$p(t_1)\,\mathrm{d}t_1 = k_{\text{tot}}\, e^{-k_{\text{tot}}\,t_{\text{sum}}}\,\mathrm{d}t_{\text{sum}} \tag{7b}$$

$$p(t_1)\,\mathrm{d}t_1 = p(t_{\text{sum}})\,\mathrm{d}t_{\text{sum}} \tag{7c}$$

and the probability to leave the state per unit time, expressed in t_{sum} units, is the same whether it is run on one or M processors. A variation on this derivation shows that the M processors need not run at the same speed, allowing the method to be used on a heterogeneous or distributed computer; see Ref. [13].

The algorithm is schematically shown in Fig. 2. 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 during which each replica is evolved forward with independent noise for a time $\Delta t_{\mathrm{deph}} \geq \tau_{\mathrm{corr}}$ 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 τ_{rxn} spent exploring phase space within the basin until the transition occurred.

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 accelerated dynamics methods. This is accomplished by allowing the trajectory that made the transition to continue on its processor for a further amount of time $\Delta t_{\mathrm{corr}} \geq \tau_{\mathrm{corr}}$, 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 repeated. Parallel-replica dynamics 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.

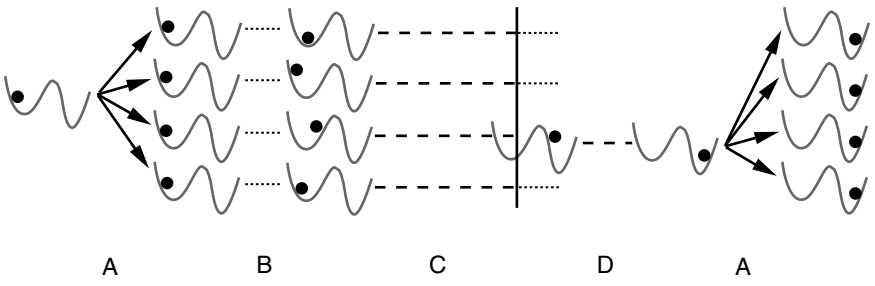

Figure 2. Schematic illustration of the parallel-replica method (after Ref. [1]). The four steps, described in the text, are (A) replication of the system into M copies, (B) dephasing of the replicas, (C) evolution of independent trajectories until a transition is detected in any of the replicas, and (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 then replicated, beginning the process again.

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 Fig. 2, where dashed line trajectories advance the simulation clock but dotted line trajectories do not.) Thus, the overall efficiency will be high when

$$\tau_{\mathrm{rxn}}/M \gg \Delta t_{\mathrm{deph}} + \Delta t_{\mathrm{corr}}. \tag{8}$$

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 only involves 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 3 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. Many of the transitions involve concerted mechanisms.

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 ps for most systems. As multiprocessing environments become more ubiquitous, with more processors within a node or even on a chip, and loosely linked Beowulf clusters of such nodes, parallel-replica dynamics will become an increasingly important simulation tool.

Recently, parallel-replica dynamics has been extended to driven systems, such as systems with some externally applied strain rate. The requirement here is that the drive rate is slow enough that at any given time the rates for the processes in the system depend only on the instantaneous configuration of the system.

3. Hyperdynamics

Hyperdynamics builds on the basic concept of importance sampling [14, 15], extending it into the time domain. In the hyperdynamics approach [16], the potential surface $V(\mathbf{r})$ of the system is modified by adding to it a non-negative *bias* potential $\Delta V_b(\mathbf{r})$. The dynamics of the system is then evolved on this biased potential surface, $V(\mathbf{r}) + \Delta V_b(\mathbf{r})$. A schematic illustration is shown in Fig. 4. The derivation of the method requires that the system obeys TST – that 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 modified potential surface. If such a bias potential can be constructed, a challenging

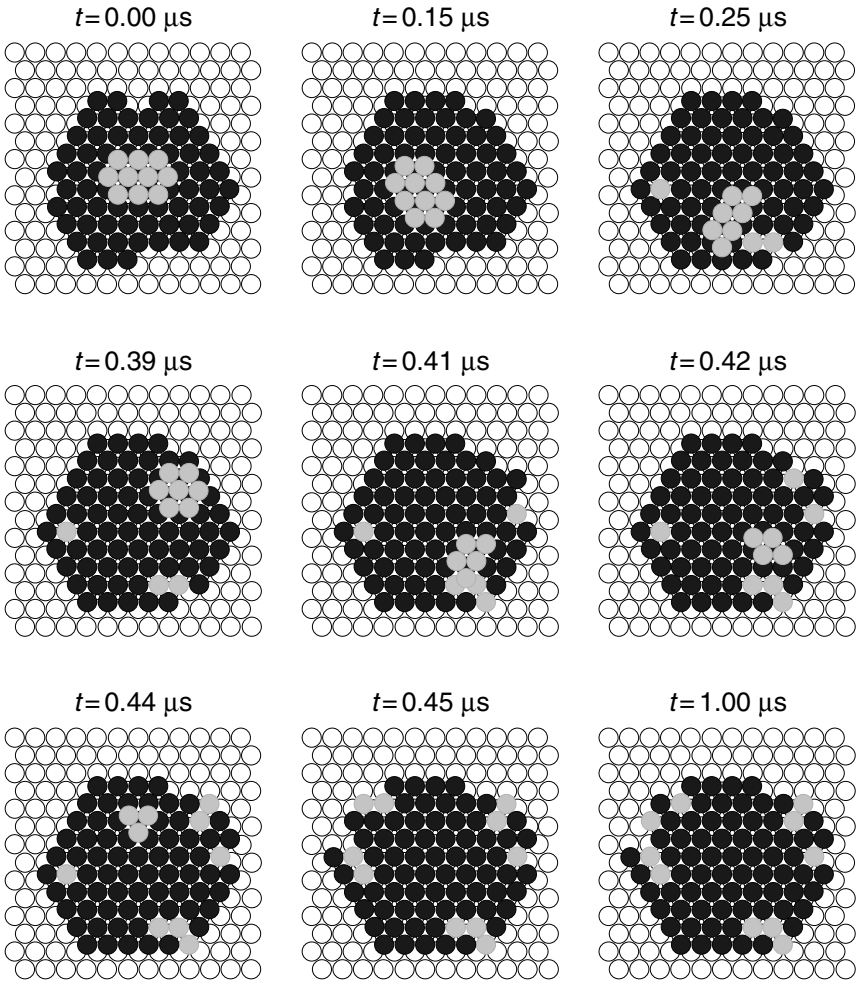

Figure 3. Snapshots from a parallel-replica simulation of an island on top of an island on the Ag(111) surface at $T = 400\,\mathrm{K}$ (after Ref. [1]). On a microsecond time scale, the upper island gives up all its atoms to the lower island, filling vacancies and kink sites as it does so. This simulation took 5 days to reach 1 μs on 32 1 GHz Pentium III processors.

task in itself, we can substitute the modified potential $V(\mathbf{r}) + \Delta V_b(\mathbf{r})$ into Eq. (1) to find

$$k_{A\rightarrow}^{\mathrm{TST}} = \frac{\langle |v_A|\, \delta(x - q)\rangle_{A_b}}{\left\langle e^{\beta \Delta V_b(\mathrm{r})}\right\rangle_{A_b}}, \tag{9}$$

where $\beta = 1/k_{\mathrm{B}}T$ and the state A_b is the same as state A but with the bias potential ΔV_b applied. This leads to a very appealing result: a trajectory on this modified surface, while relatively meaningless on vibrational time scales,

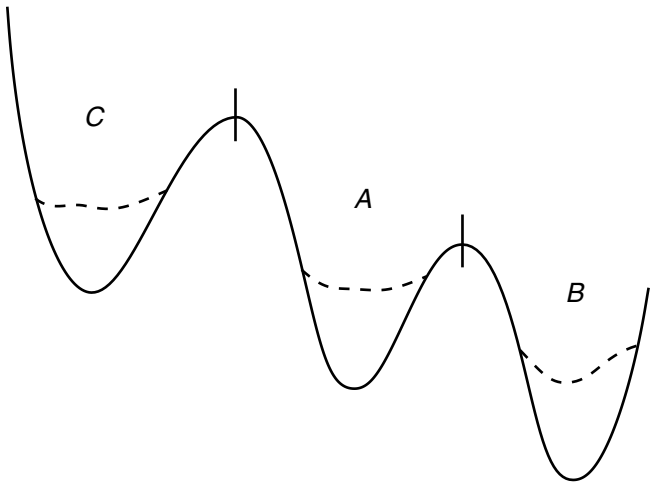

Figure 4. Schematic illustration of the hyperdynamics method. A bias potential ($\Delta V(\mathbf{r})$), is added to the original potential ($V(\mathbf{r})$, solid line). Provided that $\Delta V(\mathbf{r})$ meets certain conditions, primarily that it be zero at the dividing surfaces between states, 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.

evolves *correctly* from state to state at an accelerated pace. That is, the relative rates of events leaving A are preserved:

$$\frac{k^{\mathrm{TST}}_{A_b \to B}}{k^{\mathrm{TST}}_{A_b \to C}} = \frac{k^{\mathrm{TST}}_{A \to B}}{k^{\mathrm{TST}}_{A \to C}}. \tag{10}$$

This is because these relative probabilities depend only on the numerator of Eq. (9) which is unchanged by the introduction of ΔV_b since, by construction, $\Delta V_b = 0$ at the dividing surface.

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 (often on the order of 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

$$t_{\mathrm{hyper}} = \sum_{j=1}^{n} \Delta t_{\mathrm{MD}}\, e^{\Delta V(\mathbf{r}(t_j))/k_{\mathrm{B}}T}. \tag{11}$$

Time thus takes on a statistical nature, advancing monotonically but non-linearly. 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,

$$\text{boost(hyperdynamics)} = t_{\text{hyper}}/t_{\text{MD}} = \left\langle e^{\Delta V(\mathbf{r})/k_{\text{B}}T} \right\rangle_{A_b}, \tag{12}$$

divided by the extra computational cost of calculating the bias potential and its forces. If all the visited states are equivalent (e.g., this is common in calculations to test or demonstrate a particular bias potential), Eq. (12) 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 made.

Figure 5 illustrates an application of hyperdynamics for a two-dimensional, periodic model potential using a Hessian-based bias potential [16]. The hopping diffusion rate was compared against MD at high temperature, where the two calculations agreed very well. At lower temperatures where the MD calculations would be too costly, it is compared against the result computed

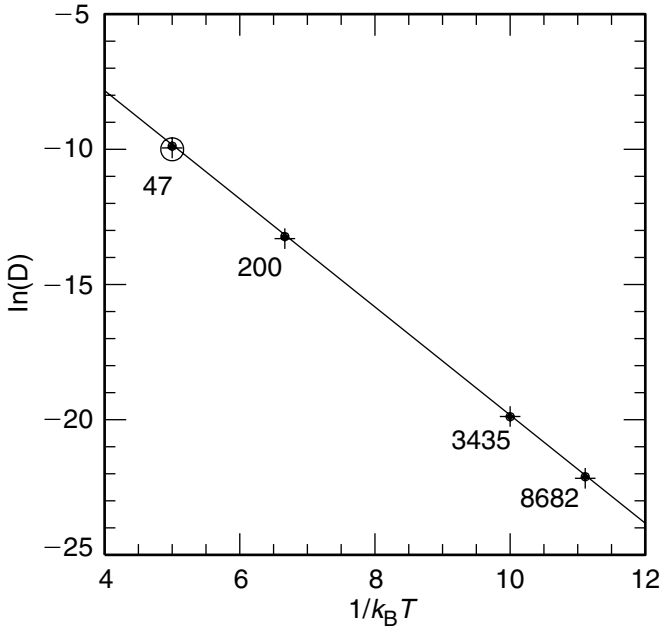

Figure 5. Arrhenius plot of the diffusion coefficients for a model potential, showing a comparison of direct MD (◯), hyperdynamics (•), and TST + dynamical corrections (+). The symbols are sized for clarity. The line is the full harmonic TST approximation, and is indistinguishable from a least-square line through the TST points (not shown). Also shown are the boost factors, relative to direct MD, for each hyperdynamics result. The boost increases dramatically as the temperature is lowered (after Ref. [16]).

using TST plus dynamical corrections. As the temperature is lowered, the effective boost gained by using hyperdynamics increased to the point that, at $k_B T = 0.09$, the boost factor was over 8500. See Ref. [16] for details.

The ideal bias potential should give a large boost factor, have low computational overhead (though more overhead is acceptable if the boost factor is very high), and, to a good approximation, meet the requirements stated above. This is very challenging, since we want, as much as possible, to avoid utilizing any prior knowledge of the dividing surfaces or the available escape paths. To date, proposed bias potentials typically have either been computationally intensive, have been tailored to very specific systems, have assumed localized transitions, or have been limited to low-dimensional systems. But the potential boost factor available from hyperdynamics is tantalizing, so developing bias potentials capable of treating realistic many-dimensional systems remains a subject of ongoing research by several groups. See Ref. [1] for a detailed discussion on bias potentials and results generated using various forms.

4. Temperature Accelerated Dynamics

In the temperature accelerated dynamics (TAD) method [17], 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, as it relies on harmonic TST, 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. 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. 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 (Eq. (6)). Because harmonic TST gives an Arrhenius dependence of the rate on temperature (Eq. (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 ($\ln(1/t)$ vs. $1/T$), as shown in Fig. 6. The time for each event seen at T_{high} extrapolated to T_{low} is then

$$t_{\text{low}} = t_{\text{high}}\, e^{E_{\text{a}}(\beta_{\text{low}} - \beta_{\text{high}})}, \tag{13}$$

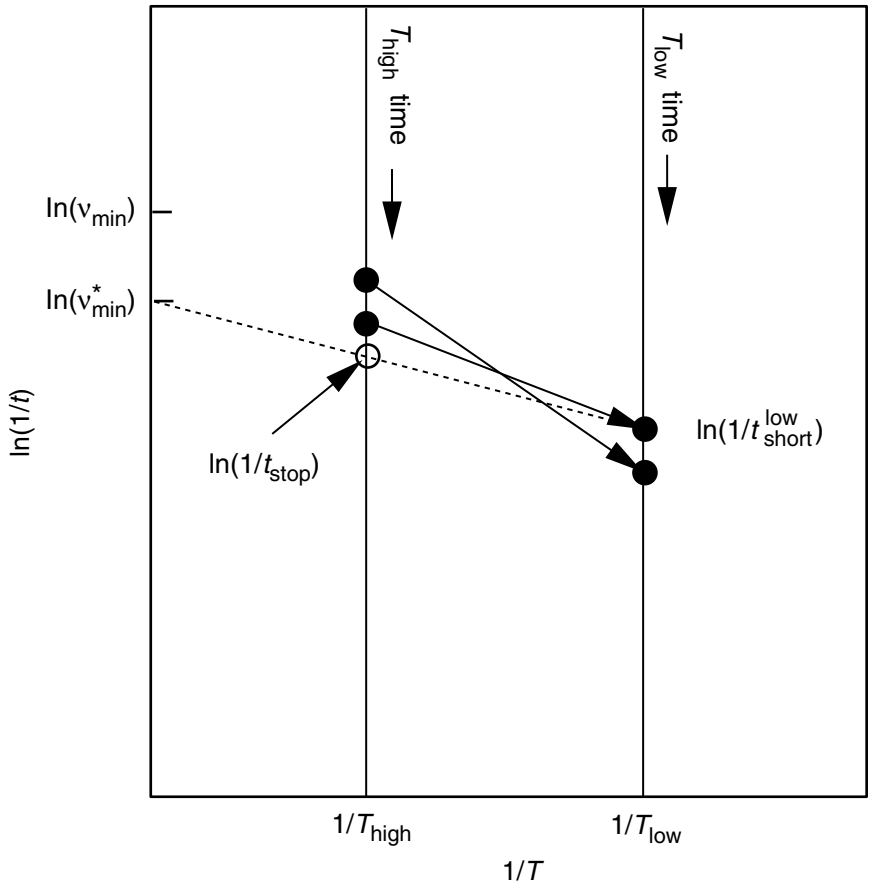

Figure 6. Schematic illustration of the temperature accelerated dynamics method. Progress of the high-temperature trajectory can be thought of as moving down the vertical time line at $1/T_{\text{high}}$. For each transition detected during the run, the trajectory is reflected back into the basin, the 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 with the confidence-modified minimum preexponential ($\nu^{\star}_{\text{min}} = \nu_{\text{min}}/\ln(1/\delta)$) 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 preexponential lower than ν_{min}.

where, again, $\beta = 1/k_B T$. 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 assumption 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

$$t_{high,stop} \equiv \frac{\ln(1/\delta)}{\nu_{min}} \left(\frac{\nu_{min} t_{low,short}}{\ln(1/\delta)} \right)^{T_{low}/T_{high}}, \tag{14}$$

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 barriers are high and T_{high}/T_{low} is large. However, any anharmonicity error at T_{high} transfers to T_{low}; a rate that is twice the Vineyard harmonic rate due 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 the other methods, the boost is limited by the lowest barrier, although this effect can be mitigated somewhat by treating repeated transitions in a "synthetic" mode [17]. 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.

Temperature accelerated dynamics is particularly useful for simulating vapor-deposited crystal growth, where the typical time scale can exceed minutes. Figure 7 shows an example of TAD applied to such a problem. Vapor deposited growth of a Cu(100) surface was simulated 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 Ref. [18]. Each deposition event was simulated using direct MD 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 propagated the system for the remaining time until the next deposition event was required, on average 0.3 s later. The overall boost factor was $\sim 10^7$. A key feature of this simulation was that, even at this low temperature, many events accepted during the growth process

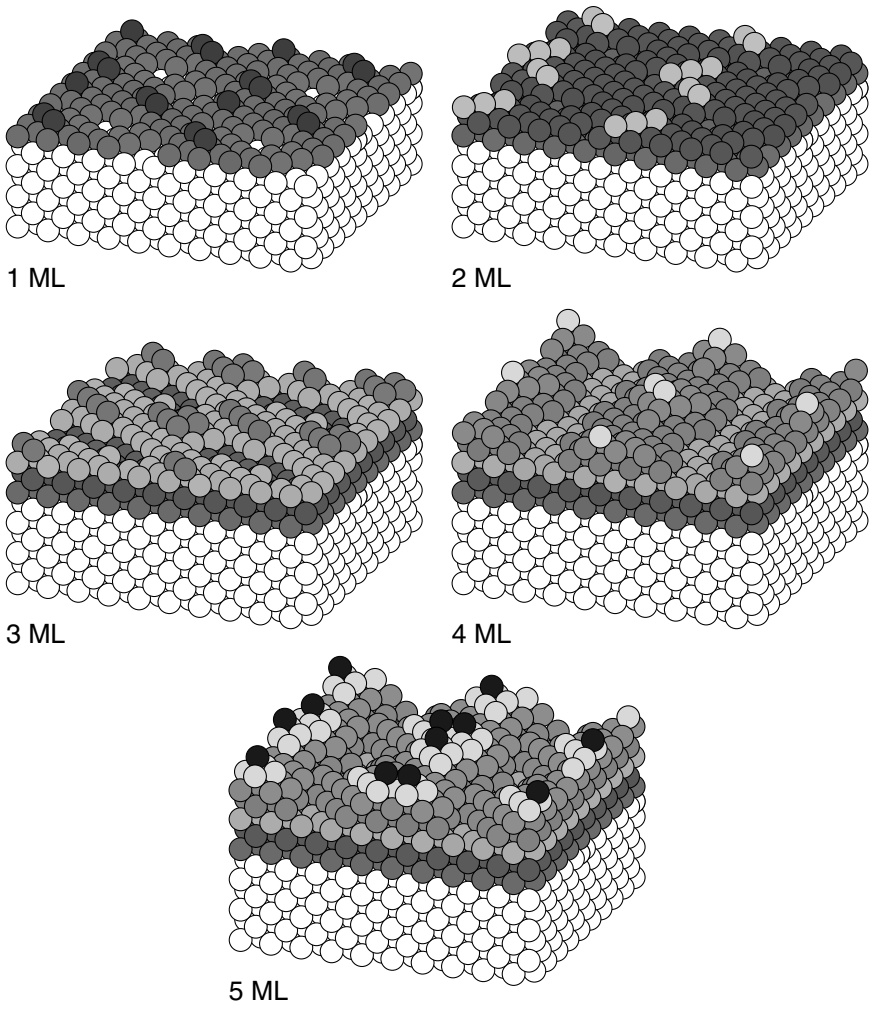

Figure 7. Snapshots from a TAD simulation of the deposition of five monolayers (ML) of Cu onto Cu(100) at 0.067 ML/s and $T = 77$ K, matching the experimental conditions of Egelhoff and Jacob [18]. Deposition of each new atom was performed using direct molecular dynamics for 2 ps, while the intervening time (0.3 s on average for this 50 atom/layer simulation cell) was simulated using the TAD method. The boost factor for this simulation was $\sim 10^7$ over direct MD (after Ref. [1]).

involved concerted mechanisms, such as the concerted sliding of an eight-atom cluster [1].

This MD/TAD procedure for simulating film growth has been applied also to Ag/Ag(100) at low temperatures [19] and Cu/Ag(100) [20]. Heteroepitaxial systems are especially hard to treat with techniques such as kinetic Monte Carlo because of the increased tendency for the system to go off lattice due

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

Recently, enhancements to TAD, beyond the "synthetic mode" mentioned above, have been developed that can increase the efficiency of the simulation. For systems that revisit states, the time required to accept an event can be reduced for each revisit by taking advantage of the time accumulated in previous visits [21]. 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

$$\text{boost(ETAD)} = \frac{\overline{t}_{\text{low,short}}}{\overline{t}_{\text{high,stop}}} = \exp\left[E_{\min}\left(\frac{1}{k_{\text{B}}T_{\text{low}}} - \frac{1}{k_{\text{B}}T_{\text{high}}}\right)\right] \tag{15}$$

for that state. The additional boost (when converged) compared to the original TAD can be an order of magnitude or more.

For systems that seldom (or never) revisit the same state, it is still possible to exploit this extra boost by running in ETAD mode with $E_{\min}$ supplied externally. One way of doing this is to combine TAD with the dimer method [22]. In this combined dimer-TAD approach, first proposed by Montalenti and Voter [21], upon entering a new state, a number of dimer searches are used to find the minimum barrier for escape, after which ETAD is employed to quickly find a dynamically appropriate escape path. This exploits 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 might fail to find the lowest barrier correctly, this is a much weaker demand on the dimer method than trying to find all relevant barriers. In addition, the ETAD phase has some chance of correcting the simulation during the BCMD if the dimer searches did not find $E_{\min}$.

5. Outlook

As these accelerated dynamics methods become more widely used and further developed (including the possible emergence of new methods), their application to important problems in materials science will continue to grow. We conclude this article by comparing and contrasting the three methods presented here, with some guidelines for deciding which method may be most appropriate for a given problem. We point out some important limitations of the methods, areas in which further development may significantly increase their usefulness. Finally, we discuss the prospects for these methods in the immediate future.

The key feature of all of the accelerated dynamics methods is that they collapse the waiting time between successive transitions from its natural time (τ_{rxn}) to (at best) a small number of vibrational periods. Each method accomplishes this in a different way. TAD exploits the enhanced rate at higher temperature, hyperdynamics effectively lowers the barriers to escape by filling in the basin, and parallel-replica dynamics spreads the work across many processors.

The choice of which accelerated dynamics method to apply to a problem will typically depend on three factors. The first is the desired level of accuracy in following the exact dynamics of the system. As described previously, parallel-replica is the most exact of the three methods; the only assumption is that the kinetics are first order. Not even TST is assumed, as correlated dynamical events are treated correctly in the method. This is not true with hyperdynamics, which does rely upon the assumptions of TST, in particular the absence of correlated events. Finally, temperature accelerated dynamics makes the further assumptions inherent in the harmonic approximation to TST, and is thus the most approximate of the three methods. If complete accuracy is the main goal of the simulation, parallel-replica is the superior choice.

The second consideration is the potential gain in accessible time scales that the accelerated dynamics method can achieve for the system. Typically, TAD is the method of choice when considering this factor. While in all three methods the boost for escaping from each state will be limited by the smallest barrier, if the barriers are high relative to the temperature of interest, TAD will typically achieve the largest boost factor. In principle, hyperdynamics can also achieve very significant boosts, but, in practice, existing bias potentials either have a very simple form which generally provide limited boosts for complex many-atom systems, or more sophisticated (e.g., Hessian-based) forms whose overhead reduces the boosts actually attainable. It may be possible, using prior knowledge about particular systems, to construct a computationally inexpensive bias potential which simultaneously offers large boosts, in which case hyperdynamics could be competitive with TAD. Finally, parallel-replica dynamics usually offers the smallest boost given the typical access to parallel computing today (e.g., tens of processors or fewer per user for continuous use), since the maximum possible boost is exactly the number of processors. For some systems, the overhead of, for example, finding saddle points in TAD may be so great that parallel-replica can give more overall boost. However, in general, the price of the increased accuracy of parallel-replica dynamics will be shorter achievable time scales.

It should be emphasized that the limitations of parallel-replica in terms of accessible time scales are not inherent in the method, but rather are a consequence of the currently limited computing power which is available. As massively parallel processing becomes commonplace for individual users, and any number can be used in the study of a given problem, parallel-replica should become just as efficient as the other methods. If enough processors are available

so that the amount of simulation time each processor has to do for each transition is on the order of ps, parallel-replica will be just as efficient as TAD or hyperdynamics. This analysis may be complicated by issues of communication between processors, but the future of parallel-replica is very promising.

The last main factor determining which method is best suited to a problem is the shape of the potential energy surface (PES). Both TAD and hyperdynamics require that the PES be relatively smooth. In the case of TAD, this is because saddle points must be found and standard techniques for finding them often perform poorly for rough landscapes. The same is true for the hyperdynamics bias potentials that require information about the shape of the PES. Parallel-replica, however, only requires a method for detecting transitions. No further analysis of the potential energy surface is needed. Thus, if the PES describing the system of interest is relatively rough, parallel-replica dynamics may be the only method that can be applied effectively.

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 faster computers become available, the same system can be studied at a lower, more desirable, temperature. This in turn increases the boost factor (e.g., see Eqs. (12) and (14)), so that, effectively, there is a superlinear increase in the power of accelerated dynamics with increasing computer speed. Thus, the accelerated dynamics approaches will become increasingly more powerful in future years simply because computers keep getting faster.

A particularly appealing prospect is that of accelerated electronic structure-based molecular dynamics simulations (e.g., by combining density functional theory (DFT) or quantum chemistry with the methods discussed here), since accessible electronic structure time scales are even shorter, currently on the order of ps. However, because of the additional expense involved in these techniques, the converse of the argument given in the previous paragraph indicates that, for example, accelerated DFT dynamics simulations will not give much useful boost on current computers (i.e., using DFT to calculate the forces is like having a very slow computer). DFT hyperdynamics may be a powerful tool in 5–10 years, when breakeven (boost = overhead) is reached, and this could happen sooner with the development of less expensive bias potentials. TAD is probably close to being viable for combination with DFT, while parallel-replica dynamics and dimer-TAD could probably be used on today's computers for electronic structure studies on some systems.

Currently, these methods are very efficient when applied to systems in which the barriers are much higher than the temperature of interest. This is often true

for systems such as ordered solids, but there are many important systems that do not so cleanly fall into this class, a prime example being glasses. Such systems are characterized by either a continuum of barrier heights, or a set of low barriers that describe uninteresting events, like conformational changes in a molecule. Low barriers typically degrade the boost of all of the accelerated dynamics methods, as well as the efficiency of standard kinetic Monte Carlo. However, even these systems will be amenable to study through accelerated dynamics methods as progress is made on this low-barrier problem.

A final note should be made about the computational scaling of these methods with system size. While the exact scaling depends on the type of system and many aspects of the implementation, a few general points can be made. In the case of TAD, if the work of finding saddles and detecting transitions can be localized, it can be shown that the scaling goes as $N^{2-T_{\mathrm{low}}/T_{\mathrm{high}}}$ [21] for the simple case of a system that has been enlarged by replication. This is improved greatly with ETAD, which scales as O(N), the same as regular MD. Real systems are more complicated and, typically, lower barrier processes will arise as the system size is increased. Thus, even hyperdynamics with a bias potential requiring no overhead might scale worse than N.

The accelerated dynamics methods, as a whole, are still in their infancy. Even so, they are currently powerful enough to study a wide range of materials problems that were previously intractable. As these methods continue to mature, their applicability, and the physical insights gained by their use, can be expected to grow.

Acknowledgments

We gratefully acknowledge vital discussions with Graeme Henkelman. 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.

References

[1] A.F. Voter, F. Montalenti, and T.C. Germann, "Extending the time scale in atomistic simulation of materials," *Annu. Rev. Mater. Res.*, 32, 321–346, 2002.

[2] D. Chandler, "Statistical-mechanics of isomerization dynamics in liquids and transition-state approximation," *J. Chem. Phys.*, 68, 2959–2970, 1978.

[3] A.F. Voter and J.D. Doll, "Dynamical corrections to transition state theory for multistate systems: surface self-diffusion in the rare-event regime," *J. Chem. Phys.*, 82, 80–92, 1985.

[4] C.H. Bennett, "Molecular dynamics and transition state theory: simulation of infrequent events," *ACS Symp. Ser.*, 63–97, 1977.

[5] R. Marcelin, "Contribution à l'étude de la cinétique physico-chimique," *Ann. Physique*, 3, 120–231, 1915.

[6] E.P. Wigner, "On the penetration of potential barriers in chemical reactions," *Z. Phys. Chemie B*, 19, 203, 1932.
[7] H. Eyring, "The activated complex in chemical reactions," *J. Chem. Phys.*, 3, 107–115, 1935.
[8] P. Pechukas, "Transition state theory," *Ann. Rev. Phys. Chem.*, 32, 159–177, 1981.
[9] D.G. Truhlar, B.C. Garrett, and S.J. Klippenstein, "Current status of transition state theory," *J. Phys. Chem.*, 100, 12771–12800, 1996.
[10] A.F. Voter and J.D. Doll, "Transition state theory description of surface self-diffusion: comparison with classical trajectory results," *J. Chem. Phys.*, 80, 5832–5838, 1984.
[11] B.J. Berne, M. Borkovec, and J.E. Straub, "Classical and modern methods in reaction-rate theory," *J. Phys. Chem.*, 92, 3711–3725, 1988.
[12] G.H. Vineyard, "Frequency factors and isotope effects in solid state rate processes," *J. Phys. Chem. Solids*, 3, 121–127, 1957.
[13] A.F. Voter, "Parallel-replica method for dynamics of infrequent events," *Phys. Rev. B*, 57, 13985–13988, 1998.
[14] J.P. Valleau and S.G. Whittington, "A guide to Monte Carlo for statistical mechanics: 1. highways," In: B.J. Berne (ed.), *Statistical Mechanics. A. A Modern Theoretical Chemistry*, vol. 5, Plenum, New York, pp. 137–168, 1977.
[15] B.J. Berne, G. Ciccotti, and D.F. Coker (eds.), *Classical and Quantum Dynamics in Condensed Phase Simulations*, World Scientific, Singapore, 1998.
[16] A.F. Voter, "A method for accelerating the molecular dynamics simulation of infrequent events," *J. Chem. Phys.*, 106, 4665–4677, 1997.
[17] M.R. Sørensen and A.F. Voter, "Temperature-accelerated dynamics for simulation of infrequent events," *J. Chem. Phys.*, 112, 9599–9606, 2000.
[18] W.F. Egelhoff, Jr. and I. Jacob, "Reflection high-energy electron-diffraction (RHEED) oscillations at 77K," *Phys. Rev. Lett.*, 62, 921–924, 1989.
[19] F. Montalenti, M.R. Sørensen, and A.F. Voter, "Closing the gap between experiment and theory: crystal growth by temperature accelerated dynamics," *Phys. Rev. Lett.*, 87, 126101, 2001.
[20] J.A. Sprague, F. Montalenti, B.P. Uberuaga, J.D. Kress, and A.F. Voter, "Simulation of growth of Cu on Ag(001) at experimental deposition rates" *Phys. Rev. B*, 66, 205415, 2002.
[21] F. Montalenti and A.F. Voter, "Exploiting past visits or minimum-barrier knowledge to gain further boost in the temperature-accelerated dynamics method," *J. Chem. Phys.*, 116, 4819–4828, 2002.
[22] G. Henkelman and H. Jónsson, "A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives," *J. Chem. Phys.*, 111, 7010–7022, 1999.