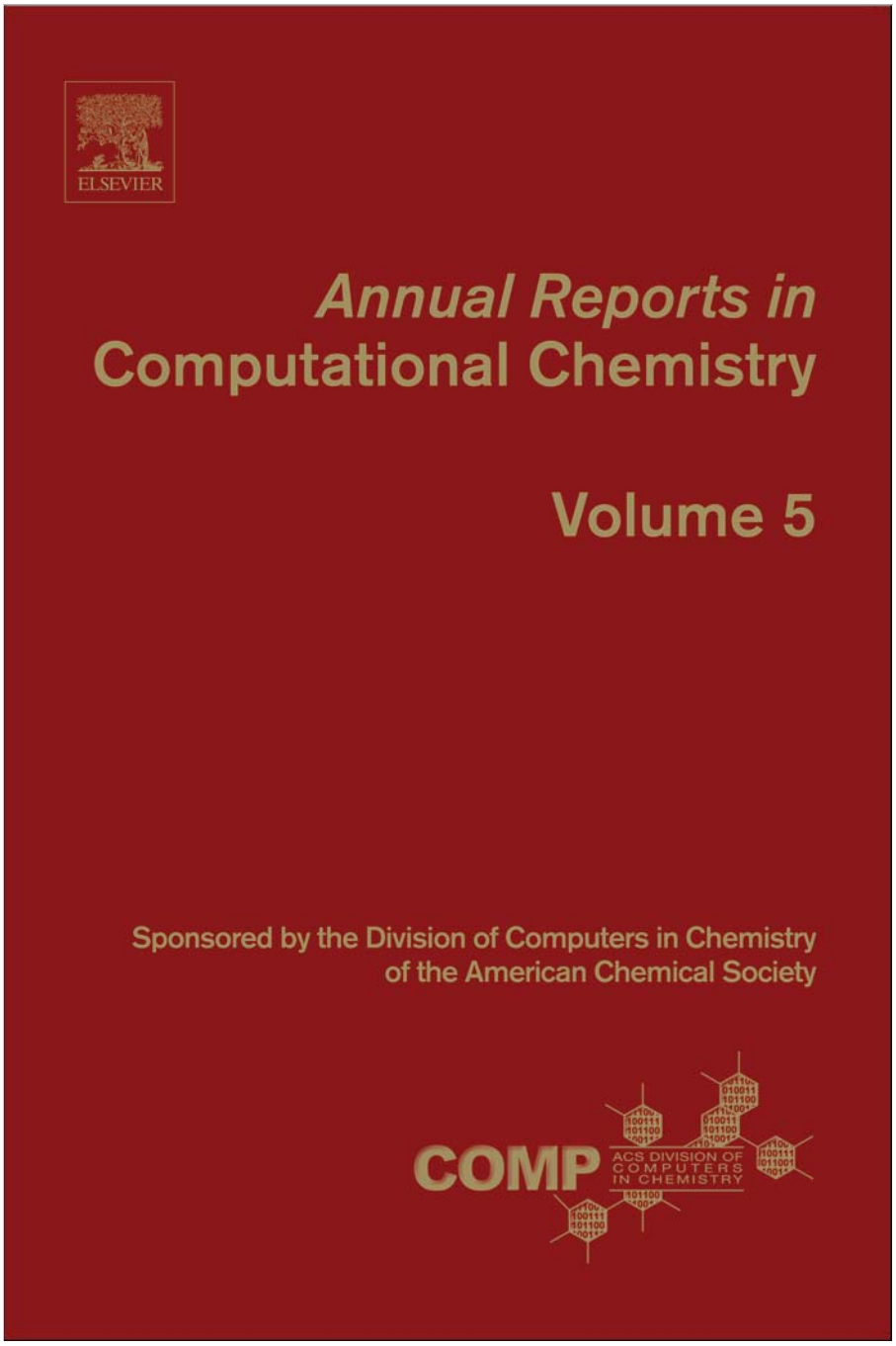
Provided for non-commercial research and educational use only.
Not for reproduction, distribution or commercial use.

This chapter was originally published in the book *Annual Reports in Computational Chemistry*, published by Elsevier, and the attached copy is provided by Elsevier for the author's benefit and for the benefit of the author's institution, for non-commercial research and educational use including without limitation use in instruction at your institution, sending it to specific colleagues who know you, and providing a copy to your institution's administrator.



All other uses, reproduction and distribution, including without limitation commercial reprints, selling or licensing copies or access, or posting on open internet sites, your personal or institution's website or repository, are prohibited. For exceptions, permission may be sought for such use through Elsevier's permissions site at:
http://www.elsevier.com/locate/permissionusematerial

From Danny Perez, Blas P. Uberuaga, Yunsic Shim, Jacques G. Amar and Arthur F. Voter, Accelerated Molecular Dynamics Methods: Introduction and Recent Developments. In: Ralph A. Wheeler, editor: Annual Reports in Computational Chemistry, Vol 5, Annual Reports in Computational Chemistry, Ralph A. Wheeler.
The Netherlands: Elsevier, 2009, pp. 79–98.
ISBN: 978-0-444-53359-3
© Copyright 2009 Elsevier B.V.

CHAPTER 4

Accelerated Molecular Dynamics Methods: Introduction and Recent Developments

Danny Perez[1], **Blas P. Uberuaga**[2], **Yunsic Shim**[3], **Jacques G. Amar**[3] and **Arthur F. Voter**[1]

Contents

1. Parallel-Replica Dynamics 81
2. Hyperdynamics 83
3. Temperature-Accelerated Dynamics 85
4. Recent Advances and Applications 87
   4.1 Superstate parallel-replica dynamics 88
   4.2 Self-learning hyperdynamics 89
   4.3 Spatially parallel TAD 93
5. Conclusion 96
Acknowledgments 97
References 97

Abstract

Because of its unrivaled predictive power, the molecular dynamics (MD) method is widely used in theoretical chemistry, physics, biology, materials science, and engineering. However, due to computational cost, MD simulations can only be used to directly simulate dynamical processes over limited timescales (e.g., nanoseconds or at most a few microseconds), even though the simulation of nonequilibrium processes can often require significantly longer timescales, especially when they involve thermal activation. In this paper, we present an introduction to accelerated molecular dynamics, a class of methods aimed at extending the timescale range of molecular dynamics, sometimes up to seconds or more. The theoretical foundations underpinning the different methods (parallel replica

[1] Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM, USA
[2] Materials Science and Technology Division, Los Alamos National Laboratory, Los Alamos, NM, USA
[3] Department of Physics and Astronomy, University of Toledo, Toledo, OH, USA

Annual Reports in Computational Chemistry, Volume 5
ISSN: 1574-1400, DOI 10.1016/S1574-1400(09)00504-0

© 2009 Elsevier B.V.
All rights reserved.

dynamics, hyperdynamics, and temperature-accelerated dynamics) are first discussed. We then discuss some applications and recent advances, including super-state parallel replica dynamics, self-learning hyperdynamics, and spatially parallel temperature-accelerated dynamics.

Keywords: infrequent events; transition-state theory; accelerated dynamics; hyperdynamics; parallel-replica dynamics; temperature-accelerated dynamics; molecular dynamics; bond-boost hyperdynamics; parallel-accelerated dynamics; Cu(100)

A long-standing limitation in the use of molecular dynamics (MD) simulation is that it can only be applied directly to processes that take place on very short timescales: typically nanoseconds if empirical potentials are employed, or picoseconds if we rely on electronic structure methods. Many processes of interest in chemistry, biochemistry, and materials science require study over microseconds and beyond, due to either the natural timescale for the evolution or the duration of the experiment of interest. The dynamics on these timescales is typically characterized by infrequent-event transitions, from state to state, usually involving an energy barrier. There is a long and venerable tradition in chemistry of using transition-state theory (TST) [1–3] to compute rate constants directly for these kinds of activated processes. If needed, dynamical corrections to the TST rate, and even quantum corrections, can be computed to achieve an accuracy suitable for the problem at hand. These rate constants then allow us to understand the system behavior on longer timescales than we can directly reach with MD. For complex systems with many reaction paths, the TST rates can be fed into a stochastic simulation procedure such as kinetic Monte Carlo (KMC) [4], and a direct simulation of the advance of the system through its possible states can be obtained in a probabilistically exact way.

A problem that has become more evident in recent years, however, is that for many systems of interest there is a complexity that makes it difficult, if not impossible, to determine all the relevant reaction paths to which TST should be applied. This is a serious issue, as omitted transition pathways can have uncontrollable consequences on the simulated long-time kinetics.

Over the past decade or so, we have been developing a new class of methods for treating the long-time dynamics in these complex, infrequent-event systems. Rather than trying to guess in advance what reaction pathways may be important, we return instead to an MD treatment, in which the trajectory itself finds an appropriate way to escape from each state of the system. Since a direct integration of the trajectory would be limited to nanoseconds, while we are seeking to follow the system for much longer times, we modify the dynamics so that the first escape will happen much more quickly, thereby accelerating the dynamics. The key is to design the modified dynamics in a way that does as little damage as possible to the probability for escaping along a given pathway — i.e., we try to preserve the relative rate constants for the different possible escape paths out of the state. We can then use this modified dynamics to follow the system from state to state, reaching much longer times than we could reach with

direct MD. The dynamics within any one state may no longer be meaningful, but the state-to-state dynamics, in the best case (as we discuss below), can be exact. We have developed three methods in this *accelerated molecular dynamics* (AMD) class, in each case appealing to TST, either implicitly or explicitly, to design the modified dynamics. Each of the methods has its own advantages, and we and others have applied these methods to a wide range of problems. The purpose of this article is to give the reader a brief introduction to how these methods work, and discuss some of the recent developments that have been made to improve their power and applicability. Note that this brief review does not claim to be exhaustive: various other methods aiming at similar goals have been proposed in the literature. For the sake of brevity, our focus will exclusively be on the methods developed by our group.

This paper is organized as follows: Sections 1, 2, and 3 discuss the basic theoretical foundations underlying the three main AMD methods, namely parallel-replica dynamics, hyperdynamics, and temperature-accelerated dynamics (TAD), respectively. In Section 4, we present some recent improvements of the original AMD methods, allowing them to tackle more complex, multi-timescale, systems (Section 4.1), to tune themselves automatically to the system at hand (Section 4.2) and to efficiently simulate systems of large sizes (Section 4.3). We then conclude by discussing current and forthcoming challenges that need to be addressed to further extend the applicability and performance of the AMD methods.

1. PARALLEL-REPLICA DYNAMICS

The parallel-replica method [5] is perhaps the least glamorous of the AMD methods, but is, in many cases, the most powerful. It is also the most accurate AMD method, assuming only first-order kinetics (exponential decay); i.e., for any trajectory that has been in a state long enough to have lost its memory of how it entered the state (longer than the correlation time τ_{corr}, the time after which the system is effectively sampling a stationary distribution restricted to the current state), the probability distribution function for the time of the next escape from that state is given by

$$p(t) = ke^{-kt} \tag{1}$$

where k is the rate constant for escape from the state. Parallel-replica allows for the temporal parallelization of the state-to-state dynamics of such a system on M processors. This is to be contrasted with standard parallelizations of MD simulations in which spatial decomposition schemes are used.

We sketch the derivation here. For a state with total escape rate k which is simultaneously explored on M processors, the effective escape rate for the first escape of *any* replica is Mk. If the simulation time accumulated on one processor is t, the total time on the M processors will then be $t_{\text{sum}} = Mt$. Thus, using a

simple change of variable, $p(t)$ can be written as:

$$p(t)dt = Mke^{-Mkt}dt \tag{2}$$

$$= ke^{-kt_{\text{sum}}}dt_{\text{sum}} \tag{3}$$

$$= p(t_{\text{sum}})dt_{\text{sum}} \tag{4}$$

implying that the probability to leave the state per unit MD time is the same whether the simulation is run on one or M processors. While this derivation applies for processors of equal speed, the same conclusion can be shown to be valid if a heterogeneous set of processors is instead used; see ref. 5.

Figure 1 shows a schematic of the algorithm. Starting with a system in a particular state, it is replicated on each of the M processors. Each replica is evolved forward with independent thermostats for a time $\Delta t_{\text{deph}} \geq \tau_{\text{corr}}$ to eliminate correlations between replicas, a stage referred to as dephasing. After dephasing, each processor carries out an independent constant-temperature MD trajectory, together exploring phase space within the particular basin M times faster than a single trajectory would. Once a transition is detected on any processor, all processors are stopped. The simulation clock is then advanced by t_{sum}, the accumulated trajectory time summed over all replicas until the transition occurred.

The parallel-replica method also correctly accounts for correlated dynamical events (there is no requirement that the system obeys TST), unlike the other AMD methods. This is accomplished by allowing the trajectory that made the transition to continue for a further amount of time $\Delta t_{\text{corr}} \geq \tau_{\text{corr}}$, during which recrossings or follow-on events may occur. The simulation clock is then advanced by Δt_{corr}, the new state is replicated on all processors, and the whole process is repeated.

The computational efficiency of the method is limited by both the dephasing stage, which does not advance the system clock, and the correlated-event stage,

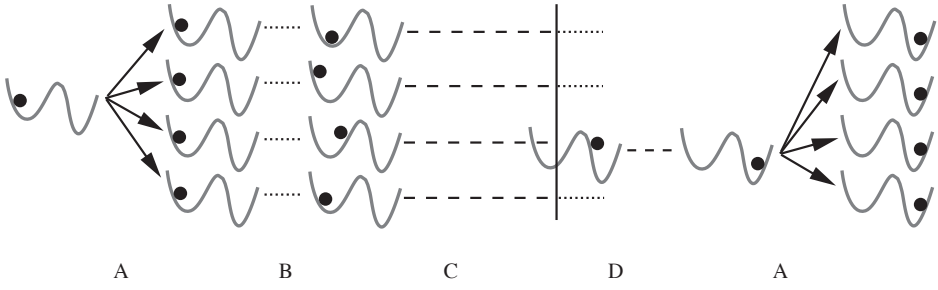


Figure 1 Schematic illustration of the parallel-replica method. The four steps, described in the text, are (A) replication of the system into M copies, (B) dephasing of the replicas, (C) propagation of independent trajectories until a transition is detected in any of the replicas, and (D) brief continuation 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. Reprinted, with permission, from ref. 6. Copyright 2002 by Annual Reviews. www.annualreviews.org

during which only one processor accumulates time. (This is illustrated schematically in Figure 1, where dashed line trajectories advance the simulation clock but dotted line trajectories do not.) Thus, the overall efficiency will be high when

$$\frac{1}{kM} \gg \Delta t_{\text{deph}} + \Delta t_{\text{corr}} \tag{5}$$

An extension to parallel-replica allows the method to be applied to driven systems. To result in valid dynamics, the drive rate must be slow enough that at any given time the rates for the different pathways in the system depend only on the instantaneous configuration of the system [7].

Parallel-replica dynamics has been successfully applied to a number of different problems, including the diffusion of H_2 in crystalline C_{60} [8], the pyrolysis of hexadecane [9], the diffusion of defects in plutonium [10], the transformation of voids into stacking fault tetrahedra in FCC metals [11], the stretching of carbon nanotubes [7], grain boundary sliding in Cu [12], the diffusion of Li through a polymer matrix [13], the fracture process of metals [14], and the folding dynamics of small proteins [15]. As parallel-computing environments become more common, the parallel-replica method will become an increasingly important tool for the exploration of complex systems.

2. HYPERDYNAMICS

Another possible avenue to accelerate the state-to-state evolution of a system of interest is to construct an auxiliary system in such a way that the dynamics of the latter are faster than that of the former while enforcing that one maps onto the other by a suitable renormalization of time. Hyperdynamics [16] realizes this objective by building on the concept of importance sampling [17,18] and extending it into the time domain. In this approach, the auxiliary system is obtained by adding a nonnegative *bias* potential $\Delta V_b(\mathbf{r})$ to the potential of the original system $V(\mathbf{r})$ so that the height of the barriers between different states is reduced, as schematically shown in Figure 2. The relationship between the dynamical evolution of the original and biased systems is recovered using TST. Indeed, according to TST, the rate of escape of the original system out of a given state A is given by

$$k_{A\rightarrow}^{\text{TST}} = \langle |\upsilon_A| \delta_A(\mathbf{r}) \rangle_A \tag{6}$$

where $\delta_A(\mathbf{r})$ is a Dirac delta function centered on the separatrix hypersurface between state A (i.e., the hypersurface is at $\mathbf{r} = 0$) and the neighboring states, υ_A the velocity normal to it, and $\langle P \rangle_A$ the canonical ensemble average of a quantity P for a system confined to state A. By standard importance sampling manipulations, the last equation can be recast in a form where the averages are obtained on the *biased* potential instead. We find:

$$k_{A\rightarrow}^{\text{TST}} = \frac{\langle |\upsilon_A| \delta_A(\mathbf{r}) e^{\beta \Delta V_b(\mathbf{r})} \rangle_{A_b}}{\langle e^{\beta \Delta V_b(\mathbf{r})} \rangle_{A_b}} \tag{7}$$

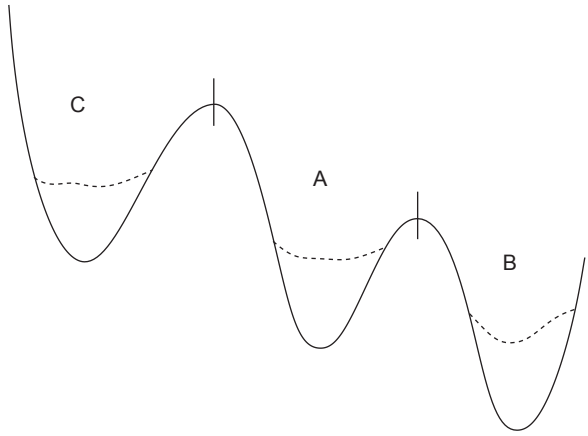


Figure 2 Schematic illustration of the hyperdynamics method. A bias potential ($\Delta V(\mathbf{r})$) is added to the original potential ($V(\mathbf{r})$, solid line). Provided that $\Delta V(\mathbf{r})$ meets certain conditions, primarily that it be zero at the dividing surfaces between states, a 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.

where $\beta = 1/k_B T$ and k_B is the Boltzmann constant. If we impose the condition that the bias potential vanish at the separatrix, the last equation can be rewritten as

$$k_{A\rightarrow}^{\mathrm{TST}} = \frac{\langle |v_A| \delta(\mathbf{r}) \rangle_{A_b}}{\langle e^{\beta \Delta V_b(\mathbf{r})} \rangle_{A_b}} \tag{8}$$

This result is very appealing since the relative rates of escapes from A to other states are invariant under the addition of the bias potential, i.e.,

$$\frac{k_{A_b \rightarrow B}^{\mathrm{TST}}}{k_{A_b \rightarrow C}^{\mathrm{TST}}} = \frac{k_{A \rightarrow B}^{\mathrm{TST}}}{k_{A \rightarrow C}^{\mathrm{TST}}} \tag{9}$$

Thus, the state-to-state dynamics on the biased potential is equivalent to that on the original potential as long as the time is renormalized to account for the uniform relative increase of all the rates introduced by the biased potential. This renormalization is in practice obtained by multiplying the MD timestep Δt_{MD} by the inverse Boltzmann factor for the bias potential, so that n MD timesteps on the biased potential are equivalent to an elapsed time of

$$t_{\mathrm{hyper}} = \sum_{j=1}^{n} \Delta t_{\mathrm{MD}} e^{\Delta V(r(t_j))/k_B T} \tag{10}$$

on the original potential. This renormalization can be shown to be exact in the long-time limit. The overall computational speedup for hyperdynamics is simply

given by the average boost factor

$$\text{boost(hyperdynamics)} = \frac{t_{\text{hyper}}}{t_{\text{MD}}} \tag{11}$$

divided by the relative extra cost of calculating the bias potential and associated forces.

Thus, if both the original and biased systems obey TST so that the above-mentioned derivation holds, hyperdynamics can provide considerable acceleration compared to direct-MD simulations. However, in practice, the applicability of hyperdynamics is limited by the availability of low-overhead bias potentials. Indeed, while some different forms have been proposed in the last few years, often they are computationally expensive, tailored to a limited class of systems or built on sets of restrictive assumptions about the nature of the separatrix. The main challenge, which is the subject of active research in different groups, thus remains the construction of bias potentials that are simple, efficient, generic, and transferable. We present below one recent advance in this area.

Despite the aforementioned difficulties, hyperdynamics has been successfully applied to a variety of systems, including desorption of organic molecules from graphitic substrates [19], surface diffusion of metallic clusters [20], heteroepitaxial growth [21], and the dynamics of biomolecules [22].

3. TEMPERATURE-ACCELERATED DYNAMICS

One natural way of speeding up the dynamics of a system is to simply raise the temperature. However, while the rates of processes will increase with higher temperatures, the relative probabilities of different events occurring will be different than at the original temperature of interest. Correcting for this reordering is the basic idea behind TAD [23]. In TAD, transitions are sped up by increasing the temperature to some T_{high}, but transitions that should not have occurred at the original temperature T_{low} are filtered out. The TAD method assumes that the system obeys harmonic TST and, as a result, is more approximate than the other AMD methods. However, for many applications, especially in solids, this additional approximation is acceptable.

In each basin, the system is evolved at T_{high}. When a transition is detected, the saddle point for that transition is found. The trajectory is then reflected back into the basin and continued. This procedure is referred to as ''basin constrained molecular dynamics'' (BCMD). During the BCMD, a list of escape paths and escape times at T_{high} for each pathway is generated. Assuming that harmonic TST holds, and knowing the saddle point energy for the transition, we can then extrapolate each escape time observed at T_{high} to obtain a corresponding escape time at T_{low}. This extrapolation, which does not require knowing the preexponential factor, can be illustrated graphically in an Arrhenius-style plot ($\ln(1/t)$ vs. $1/T$), as shown in Figure 3. The time for each event seen at T_{high}

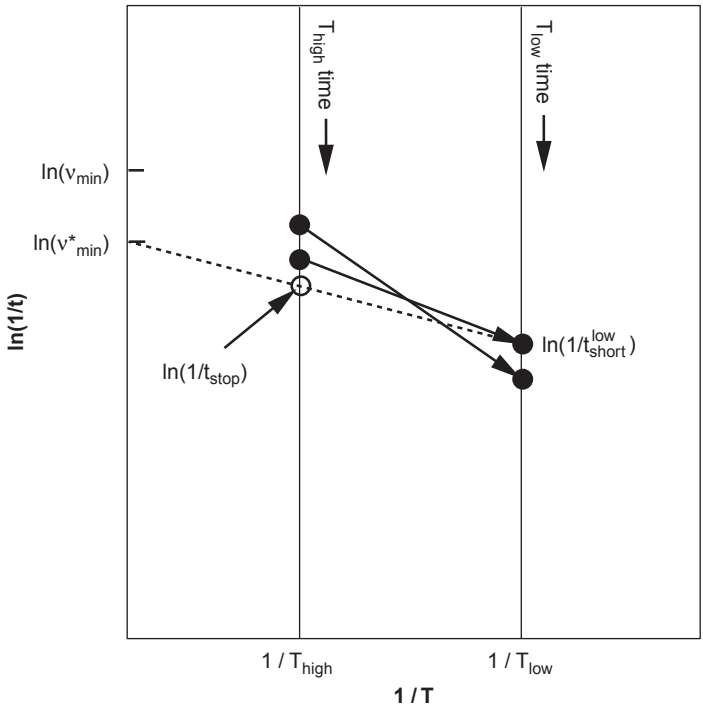


Figure 3 Schematic illustration of the temperature-accelerated dynamics method. Progress of the high-temperature trajectory can be thought of as moving down the vertical timeline at $1/T_{\text{high}}$. For each transition detected during the run, the trajectory is reflected back into the basin, the saddle point is found, and the time of the transition (solid dot on left timeline) is transformed (arrow) into a time on the low-temperature timeline. 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 timeline with the confidence-modified minimum preexponential ($\nu^*_{\text{min}} = \nu_{\text{min}}/\ln(1/\delta)$) on the *y*-axis. The intersection of this line with the high-*T* timeline gives the time (t_{stop}, open circle) at which the trajectory can be terminated. With confidence $1-\delta$, we can say that any transition observed after t_{stop} could only extrapolate to a shorter time on the low-*T* timeline if it had a preexponential factor lower than ν_{min}.

extrapolated to T_{low} is

$$t_{\text{low}} = t_{\text{high}} e^{E_a(\beta_{\text{low}} - \beta_{\text{high}})} \tag{12}$$

where $\beta = 1/k_B T$ and E_a is the activation energy.

As the BCMD is continued, a new shorter-time event may be discovered. With 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. This time has the property that the probability any transition observed later would replace the first

transition at T_{low} is less than δ. This "stop" time is given by

$$t_{\text{high,stop}} \equiv \frac{\ln(1/\delta)}{\nu_{\min}} \left(\frac{\nu_{\min} t_{\text{low,short}}}{\ln(1/\delta)} \right)^{T_{\text{low}}/T_{\text{high}}} \tag{13}$$

where $t_{\text{low,short}}$ is the shortest transition time at T_{low}. When this stop time is reached, the system clock is advanced by $t_{\text{low,short}}$, and the corresponding transition is accepted. The TAD procedure is then started again in the new basin. Thus, in TAD, two parameters govern the accuracy of the simulation: δ and $\nu_{\min}$.

The average boost in TAD can be dramatic when barriers are high and $T_{\text{high}}/T_{\text{low}}$ is large. However, as TAD relies upon harmonic TST for validity, any anharmonicity error at T_{high} will lead to a corruption of the dynamics. This anharmonicity error can be controlled by choosing a T_{high} that is not too high.

A number of advances have led to increased efficiency in particular systems. "Synthetic" mode [23], a KMC treatment of low-barrier transitions, can significantly improve the efficiency in cases where low-barrier events are repeated often. Furthermore, if we know something about the minimum barrier to leave a given state, either because we have visited the state before and have a lower bound on this minimum barrier or because the minimum barrier is supplied *a priori*, we can accept a transition and leave the state earlier than the time given by Equation (13) (see ref. 24 for details).

TAD has been demonstrated to be very effective for studying the long-time behavior of defects produced in collision cascades [25,26]. An MD/TAD procedure has also been applied to the simulation of thin-film growth of Ag [27] and Cu [28] on Ag(100). Heteroepitaxial systems are especially hard to treat with techniques such as KMC due to the increased tendency for the system to go off lattice due to mismatch strain, and because the rate catalog needs to be considerably larger when neighboring atoms can have multiple types. Other applications for which TAD has proven effective include defect diffusion on oxide surfaces [29], the diffusion of interstitial clusters in Si [30] and defect diffusion in plutonium [10].

4. RECENT ADVANCES AND APPLICATIONS

While the basics of the three methods described above were established roughly a decade ago, they provide such a fertile ground for further development that they are still the subject of ongoing research. Currently, this research proceeds along three main axes: (i) generalization of the methods to extend their range of applicability, (ii) algorithmic improvements to make the methods more robust and easier to apply, and (iii) creation of hybrid methods by combining AMD methods together or with other simulation approaches. In the following, we describe some recent advances along these three directions and discuss some successful demonstrations and applications.

4.1 Superstate parallel-replica dynamics

While the derivation of the parallel-replica method in Section 1 does not impose a particular definition of a ''state'' of the system, the operational definition used in practice often corresponds to a single basin of the potential energy surface, i.e., a state is taken to be the ensemble of points of configuration space that converge to the same fixed point under a local minimization of the energy of the system (e.g., using a steepest-descent algorithm). An exponential distribution of escape times is then obtained if the typical timescale for a transition out of the state is long compared to the characteristic vibrational period of the system around that fixed point, i.e., if there is a separation of timescale between vibrations and transitions between basins. While this definition has the virtue of being conceptually and computationally simple, it limits the range of possible applications to systems where the basins are deep enough (relative to k_BT) and well separated from each other and leaves many other, more complex, systems out of reach. There is thus a clear need to develop strategies to capitalize on more general definitions of states and hence higher-level gaps in the characteristic timescales spectrum.

For example, in the case of pyrolysis of hexadecane, it was shown that a state could be defined as the ensemble of all configuration space points that share the same network of covalent bonds [9]. In that case, these ''superstates'' contain a large number of simple energy basins of the potential energy surface, each corresponding to a different conformation of the molecular backbone. There, the method exploited the separation of timescale between the rapid changes of dihedral angles of the backbone (intrasuperstate transitions) and the slow covalent bond breaking process (intersuperstates transitions) rather than between the vibrational timescale and that of sampling of the different dihedral angles. This enables one to ignore the ''irrelevant'' fast transitions that would demand incessant dephasing and decorrelation and concentrate directly on the real kinetic bottlenecks.

Another common situation that arises is one where the infrequent-event system of interest is coupled to a complex and rapidly evolving environment such that, considered as a whole, the combined system makes transitions so rapidly that any computational gain is impossible using a traditional definition of a state. A specific example of this type of system is a solid surface in contact with a liquid. We might be interested in following the evolution of the surface morphology during electrochemical deposition or etching, or discovering the relevant steps of a surface-catalytic reaction or a corrosive process. In this type of system, a single potential energy basin definition of a state would imply that transitions occur every few femtoseconds. However, the vast majority of these would correspond to local changes in the coordination of liquid atoms, and only very rarely would the configuration of the solid atoms, which are really the variables of interest here, change. Similarly to the case of pyrolysis discussed above, a superstate can be here defined as all the points of configuration space that converge to the same configuration of the slow variables of the system — the position of the atoms belonging to the solid phase — under minimization of the energy.

In order to demonstrate that this definition is proper, we carried out superstate parallel-replica simulations of the diffusion of an adatom on a Ag(100) surface (modeled using the embedded atom method) in contact with a film of a prototypical fluid (modeled using a Lennard-Jones potential). Here the interaction strength of liquid atoms with other liquid atoms or with silver atoms was taken to be 10 times weaker than the corresponding silver–silver interaction. The liquid film was left free to expand to its liquid–vapor equilibrium density. The distribution of 500 adatom hopping times at 600 K is shown in the left panel of Figure 4 for a direct simulation using conventional MD and for a superstate parallel-replica simulation. The results clearly show that transition statistics are equivalent for the two approaches, implying that our superstate definition restores the timescale separation essential for the validity of the method. In this case the simulation was quite efficient, with a parallel efficiency of around 0.8 despite the presence of very fast transitions in the liquid.

This good agreement is not due to a negligible effect of the liquid on the adatom dynamics. Indeed, as shown in the right panel of Figure 4, the hopping kinetics are strongly affected by the liquid. Quantitatively, a fit to a standard Arrhenius rate expression $k(T) = \nu_0 \exp(-\Delta E/k_B T)$, where ν_0 is the vibrational prefactor for the transition and ΔE its activation energy, shows that the presence of the liquid significantly increases both ΔE (from 0.53 to 0.70 eV) and ν_0 (from 7.17×10^{13} to $6.28 \times 10^{14}\,\mathrm{s}^{-1}$). Interestingly, while the liquid slows down diffusion in the regime we probed, the results suggest that it could actually assist the adatom's diffusion at temperatures exceeding about 900 K. One must however be careful with such extrapolations because the relevant physics could be significantly modified as the liquid approaches its critical point. Indeed these modifications to the kinetics stem from many, often conflicting, factors, and their effects are extremely difficult to obtain accurately from higher-level models. The development of direct but efficient methods to probe the kinetics of complex systems like this is thus extremely useful.

Note that, in the two cases discussed above, chemical intuition was essential in properly defining superstates such that an appropriate separation of timescales was obtained. For more complex systems, intuition alone will not be sufficient and considerable effort might be required to identify an exploitable gap in the characteristic timescales of the system. There is thus a need to develop on-the-fly methods to appropriately define superstates based on MD data alone. Efforts toward this goal are presently under way [31].

4.2 Self-learning hyperdynamics

As illustrated in the previous section, the AMD algorithms often have to be tuned and sometimes even tailored for particular applications. This might be straightforward in some cases, but might require considerable care in others. It would thus be highly desirable if the methods could be made to automatically adapt themselves to every system, or even to every state of every system, in order to deliver optimal performance while maintaining tight control on accuracy.

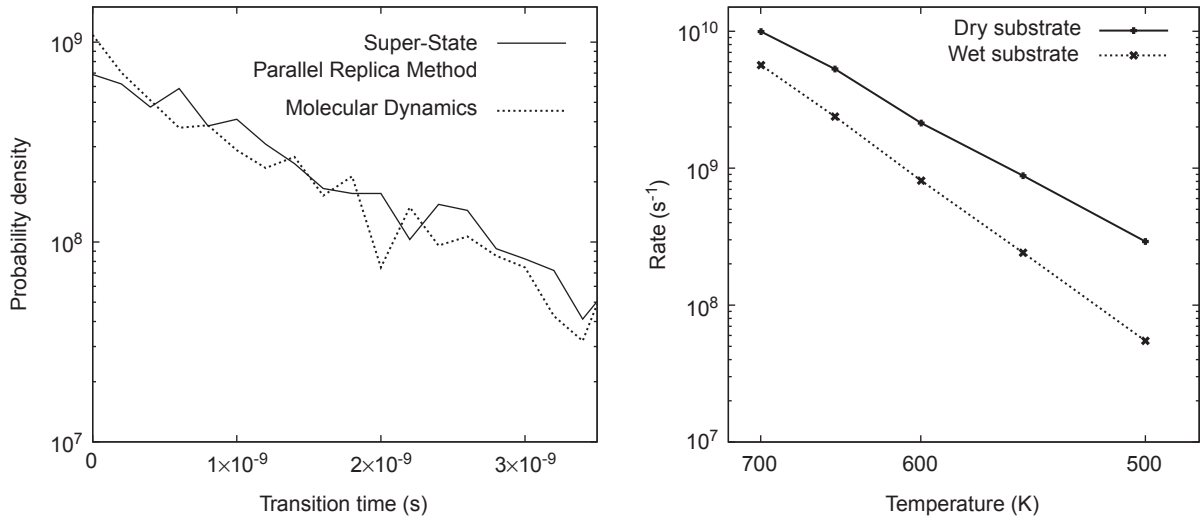


Figure 4 Left: Distribution of hopping times for an adatom at a solid–liquid interface at 600 K for conventional molecular dynamics and for superstate parallel-replica dynamics. Right: Temperature dependence of the hopping rate for an adatom at a dry interface and at a wet interface as obtained using superstate parallel-replica dynamics.

In the current state of affairs, nowhere is this need more pressing than in hyperdynamics.

Indeed, as discussed above, the applicability of hyperdynamics is often hampered by the difficulty in building bias potentials that satisfy all the formal requirements — namely that (i) the bias potential should vanish at any dividing surface between different states and (ii) the kinetics on the biased potential obeys TST — while providing substantial acceleration of the dynamics. Both requirements are very challenging to meet in practice. Indeed, condition (i) must be obeyed for *all* dividing surfaces around *all* states, which, given that the transition pathways or even the possible conformations that the system can adopt are *a priori* unknown, is highly nontrivial to enforce at a reasonable computational cost. An important step in that direction has however been recently taken by Miron and Fichthorn, with the introduction of their ''bond-boost'' bias potential [32]. As the name suggests, the bond-boost potential is composed of pairwise terms that tend to soften the bonds between atoms. The key assumption here is that transitions between states will involve the formation or breaking of some bond so that the proximity to a transition state will be signaled by an unusually large distortion of a bond. If the overall bias potential is then designed to vanish when *any* bond in the system distorts by more than some critical amount (say by more than 20% of its equilibrium length), then it should be possible to safely turn off the bias before a dividing surface is reached. Now, we will make use of the following bond-boost functional form:

$$\Delta V_b(\varepsilon) = \min_{\alpha}[\delta V_\alpha(\varepsilon_\alpha)] \tag{14}$$

with $\varepsilon_\alpha = [r_\alpha - r_\alpha^{\mathrm{eq}}]/r_\alpha^{\mathrm{eq}}$, where r_α and r_α^{eq}, are the current and equilibrium length of bond α, respectively. Following Miron and Fichthorn, we define

$$\delta V_\alpha(\varepsilon_\alpha) = \begin{cases} S_\alpha\left[1 - \left(\dfrac{\varepsilon_\alpha}{q}\right)^2\right], & |\varepsilon_\alpha| \le q \\ 0, & \text{otherwise} \end{cases} \tag{15}$$

where q corresponds to the critical bond deformation and S_α controls the strength of the bias applied along bond α. Note that in this formalism, only the bond that has the minimal value of δV will experience a bias force. Of course, the identity of the biased bond will change frequently in the course of the dynamics. Assuming that a conservative value of q is known (which we acknowledge can be a delicate issue), the bond-boost potential can be used to accelerate the dynamics of systems where transitions involve significant bond distortion.

The remaining question is that of the choice of the local bias strengths S_α. Since a suitable choice of q already guarantees that the bias potential vanishes at dividing surfaces, they should be tuned to yield the largest possible boost while making sure that requirement (ii) is respected, namely that TST is obeyed on the biased potential. Typically, this amounts to requiring that the (biased) potential basin corresponding to each state be sampled on a vibrational timescale, or, in other words, that it is free of local minima where the system could get trapped.

The pairwise nature of the bond-boost makes this task easier since such traps would show up as a non-convexity of some of the biased effective pair potentials, which in the canonical ensemble can be taken to be the pairwise potential of mean force (PMF, denoted as $\mathcal{V}$). Thus, assuming that $\mathcal{V}$ is approximately quadratic for $|\varepsilon| < q$, the safety condition can be enforced by setting $S_\alpha \leq \min[\mathcal{V}_\alpha(-q), \mathcal{V}_\alpha(q)]$, so that $\mathcal{V}_\alpha(\varepsilon_\alpha)+\delta V_\alpha(\varepsilon_\alpha)$ is convex over $[-q, q]$.

The pairwise PMF of a system vibrating in state A can be defined as:

$$\mathcal{V}_\alpha^A(r) = -\frac{1}{\beta}\ln\left[\frac{\langle\rho_\alpha(r)\rangle_A}{\langle\rho_\alpha(r_\alpha^{\mathrm{eq}})\rangle_A}\right] \tag{16}$$

where $\langle\rho_\alpha(r)\rangle_A$ is the canonical distribution function of the length of bond α. The bias potential can thus be parameterized by computing the relative probability density that each bond is at $\varepsilon = \pm q$ over that of being at $\varepsilon = 0$. A direct-MD evaluation of $\mathcal{V}_\alpha(\pm q)$ would however be prohibitively expensive given that the crossing of any $\varepsilon_\alpha = \pm q$ point is by definition a rare event. Since the set of configurations where any of the $\varepsilon_\alpha = \pm q$ forms a hypersurface at which the bias potential vanishes, hyperdynamics can be used to speed up the evaluation of $\mathcal{V}_\alpha$. The key to avoiding the circular problem where $\mathcal{V}_\alpha^A$ is needed to carry out hyperdynamics but hyperdynamics is needed to efficiently compute $\mathcal{V}_\alpha^A$ is to instead aim at estimating a lower bound. This way, a safe but conservative parameterization of the bias can be turned on even before the convergence of $\mathcal{V}_\alpha^A(q)$ is achieved. As the statistics improve, the lower bound on the PMF can be made tighter and thus the bias potential parameterization more aggressive, until convergence is finally achieved. This feedback loop between improvement of the PMF estimate and acceleration of the rate of improvement of this estimate makes the procedure, termed *self-learning hyperdynamics*, extremely efficient. Once a transition to a new state is detected, the S_α is set to zero, and the procedure is repeated. Under self-learning hyperdynamics, the bias potential is thus continually adjusted to provide optimal performances while ensuring accuracy. The complete mathematical and implementation details of the method will be available in an upcoming publication [33].

The efficiency of the method can be appreciated from Figure 5, where the evolution of hyper-time with MD time under self-learning hyperdynamics is shown for two defects — a silver monomer and trimer — on a Ag(100) surface at room temperature. The results show that, after an initial bootstrapping period during which the bias potential is turned off completely, the learning procedure quickly leads to an increase of the bias strengths, thereby increasing the boost factor, until convergence is finally achieved. One can show that in the early part of the learning process, the hyper-time actually increases exponentially with MD time, which makes the procedure very efficient. This example also clearly shows that the bias potential needs to be tuned for each state if optimal performance is to be achieved. Indeed, for a trimer, the maximal safe boost is found to be around 80 while it exceeds 1,000 for the monomer system. A safe nonadaptive approach would thus demand that the minimal safest boost of *any* of the possible states of the system be used, which, as we demonstrated, may be much smaller than the

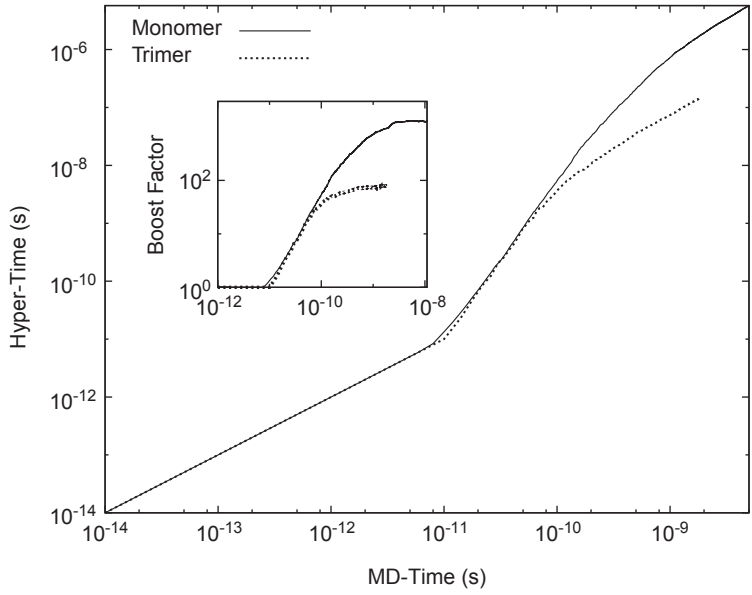


Figure 5 Evolution of the hyper-time as a function of MD time for a monomer and a trimer on a Ag(100) surface at $T = 300$ K under self-learning hyperdynamics. Inset: Evolution of the corresponding boost factors as a function of MD time.

typical safest boost for a given state. Application of this methodology to the study of the mechanical properties of nanowires is currently under way.

4.3 Spatially parallel TAD

While direct-MD simulation with a short-range empirical potential requires O(N) computational work as the system size N is increased, the AMD methods do not exhibit such favorable scaling. This is relatively easy to understand if we consider a system that is made larger in a self-similar way, so that doubling N leads to at least twice as many reactive pathways and twice the total escape rate. A parallel-replica dynamics simulation on the doubled system will reach the first transition in half the wall-clock time, while the dephasing and correlated-event overhead will remain the same. For large enough N, this overhead will dominate the simulation. In hyperdynamics, the bias potential must vanish near every dividing surface, and increasing N increases the fraction of time that the system spends near some dividing surface. Thus, for a valid bias potential, no matter what the form, as N is increased, the boost drops, ultimately approaching unity in the large-N limit.

For TAD, the overall computational work to advance the system by a given time scales at best as $N^{2-\gamma}$, where $\gamma = T_{\text{low}}/T_{\text{high}}$. The power of $-\gamma$ comes from the reduction in $t_{\text{high,stop}}$ from Equation (13) as $t_{\text{low,short}}$ decreases with N, one power of N comes from the cost of each high-T force call, and the other power of N comes from the fact that accepting a transition advances the system by a time inversely

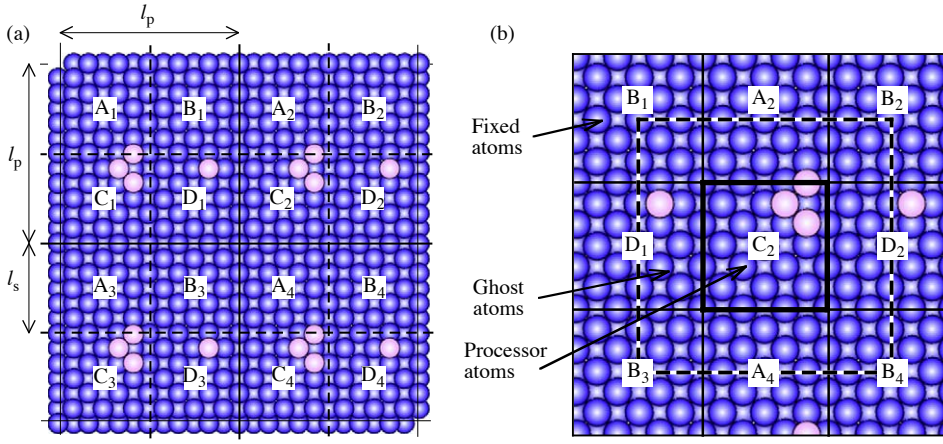


Figure 6 Illustration of the spatial partitioning in the ParTAD method. (a) Lattice and sublattice (SL) partitioning in the SL method on which ParTAD is built. Patches are indicated by numbers, subpatches by letters. (b) Setup for TAD dynamics on one subpatch. Reprinted, with permission, from ref. 34. Copyright 2007 by the American Physical Society. http://link.aps.org/doi/10.1103/PhysRevB.71.125432

proportional to N. If the saddle searches involve all N atoms instead of being localized to a fixed number of atoms, the TAD work at large N is dominated by the cost of these searches, causing an overall scaling of roughly $N^{3+1/3-\gamma}$.

As a consequence of this unfavorable scaling with N, applications of the AMD methods to date have involved at most a few thousand atoms, and have typically been much smaller than that. Since many processes of interest require larger system sizes to capture the essential physics, we are seeking ways to make larger AMD simulations feasible. One important step in this direction is our recent development of a spatially parallel TAD approach [34], which we describe very briefly here.

Spatially parallel TAD, or ParTAD, builds on a parallelization approach originally developed for KMC, the synchronous sublattice (SL) algorithm [35]. In the SL method, the system is spatially divided into a lattice of regions, or patches, each owned by one processor, and each of these regions is further subdivided into SL patches. At any given time, for each of the lattice regions, KMC or TAD dynamics evolve within one, but only one, of the SL patches. After a ''cycle time,'' all processors switch to the next SL patch and again perform dynamics for one cycle time. The simulation then proceeds by repeating this process so that all SL patches cyclically become active.

Figure 6(a) shows an example of a two-dimensional SL definition appropriate for surface diffusion or growth, while Figure 6(b) shows the TAD simulation cell for one processor, consisting of the active SL patch, a ''ghost region'' of additional moving atoms taken from the adjacent SL patches, and a layer of nonmoving atoms further out. During the TAD dynamics, any attempted event determined to be in the ghost region is excluded from consideration for acceptance, since it is not in the dynamically active patch.

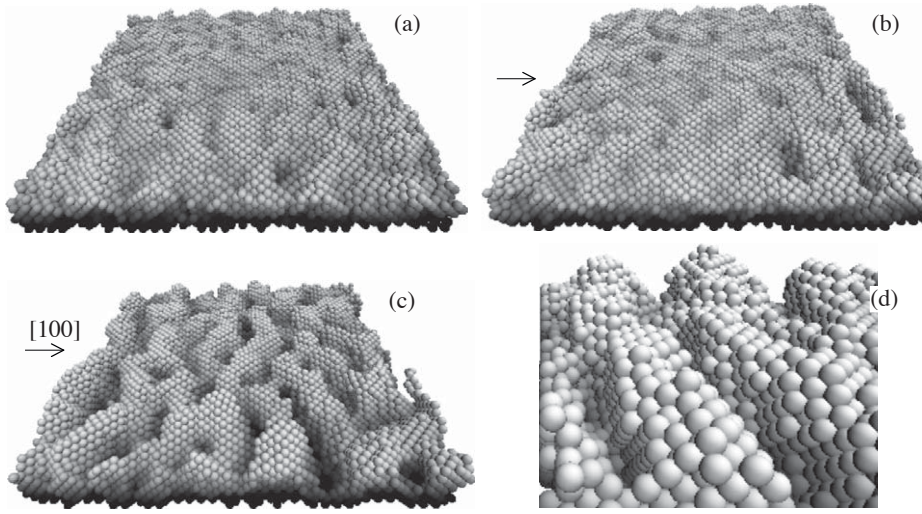


Figure 7 Surface morphology for Cu/Cu(100) films grown to 7-monolayer thickness using ParTAD. (a) Normal-incidence deposition; (b) deposition 30° off normal; (c) deposition 60° off normal; and (d) blow-up of portion of 60° film, showing (100)-oriented cliffs. Arrows indicate the deposition direction. Reprinted, with permission, from ref. 37. Copyright 2008 by the American Physical Society. http://link.aps.org/doi/10.1103/PhysRevLett.101.116101

The important feature in the SL method is that there is always a buffer of dormant SL region separating any two dynamically active regions, which eliminates the difficulty associated with synchronizing events and resolving conflicts at the boundaries between processors. For a finite cycle time, the SL method is not exact, but if the cycle time is comparable to or shorter than a typical reaction time, it is extremely accurate. A detailed discussion and demonstration of the requirements for accuracy in the case of KMC can be found in ref. 35.

Even for vanishing cycle times, the ParTAD method introduces an additional approximation into the dynamics compared to regular TAD; because the subpatch has a fixed size, concerted processes larger than this size, if they exist, will be suppressed. In exchange, the method gives a very favorable computational scaling with system size, which appears to be roughly order $\log(N)$ as processors are added in proportion to N [34].

Metallic film growth is an example where the larger size scale and long-time capability of ParTAD is valuable, since experimental deposition rates are far slower than MD can access, and the morphological features sometimes take on a scale that requires a large simulation cell to prevent artifacts. We have recently studied films of Cu deposited at low temperature onto clean Cu(100) surfaces at varying deposition angles. Our goal was to understand recent X-ray diffraction observations [36] that these films have large surface-normal compressive strain, thought to be due to an enhanced concentration of bulk vacancies. We used MD for the deposition events and ParTAD between deposition events to achieve a growth rate several orders of magnitude slower than we could with MD alone.

Figure 7 shows 7-monolayer (ML) films grown with ParTAD on 36 processors at three different deposition angles at a growth rate of 5,000 ML/s at $T = 40$ K. On this millisecond timescale, activated events do occur, although the experimental growth rate is still much slower (0.02 ML/s).

Our synthesized X-ray diffraction spectrum at a deposition angle of 60° off normal is in very good agreement with experiment, but inspection of the film shows that the large compressive strain arises not from large-scale vacancy incorporation, as had been previously suggested, but rather from nanoscale roughness, as can be seen in Figure 7(c). This roughness, caused by shadowing and the suppression of thermally activated ''downward-funneling'' events at low temperature, also shows interesting vertical (100) ''cliffs'' (Figure 7(d)). An analysis of the relevant activation barriers for the suppressed downward-funneling events also leads to an estimate of the critical temperature for the onset of compressive strain $T_c \sim 120$–150 K, in good agreement with experiment.

5. CONCLUSION

Since their introduction a little more than 10 years ago, the AMD methods have proven useful in a variety of situations where the timescales of interest are out of reach of direct MD and where the kinetics are too rich to be adequately described with a limited list of predetermined pathways. When the activation barriers between the different states are high relative to the thermal energy, any of the AMD methods can yield colossal accelerations, providing a view of atomistic dynamics over unprecedented timescales. Further, by leveraging the particular strength of each of the methods, or, as demonstrated above, by generalizing and combining them with other techniques, a wide variety of situations can be efficiently simulated. More discussion of the specific strengths and weaknesses of the methods can be found in a recent review [38].

If the methods have enjoyed considerable successes, they have also sometimes failed to provide significant acceleration. In most, if not all, of the problematic cases, this failure is related to the presence of large numbers of states connected by very low barriers. In this situation, the low barriers limit the boost, and a huge number of transitions may be required before the timescale of interest can be reached. While some strategies have been put forward to mitigate this issue (e.g., superstate parallel-replica dynamics, synthetic TAD [23], state-bridging hyper-dynamics [39]), more work is required before victory can be claimed. For example, an on-the-fly state definition algorithm that automatically identifies an exploitably large separation of timescales would tremendously extend the reach of parallel-replica dynamics, enabling it to address notoriously difficult problems like protein folding, where the energy landscape is extremely rough. Statistical analysis tools could also be used to identify dynamically ''irrelevant'' states that could be ignored or lumped with others without affecting the long-time dynamics. Many of these ideas are now being explored and will hopefully lead to more general and robust AMD methods in the next few years.

ACKNOWLEDGMENTS

Work at Los Alamos National Laboratory (LANL) was supported by the DOE Office of Basic Energy Sciences and by the LANL Laboratory Directed Research and Development program. LANL is operated by Los Alamos National Security, LLC, for the National Nuclear Security Administration of the US DOE under Contract No. DE-AC52-06NA25396. Work at University of Toledo was supported by NSF grant DMR-0606307.

REFERENCES

1. Eyring, H. The activated complex in chemical reactions. J. Chem. Phys. 1935, 3, 107–15.
2. Pechukas, P. Transition-state theory. Ann. Rev. Phys. Chem. 1981, 32, 159–77.
3. Truhlar, D.G., Garrett, B.C., Klippenstein, S.J. Current status of transition-state theory. J. Phys. Chem. 1996, 100, 12771–800.
4. Voter, A.F. Introduction to the kinetic Monte Carlo method, In Radiation Effects in Solids (eds K.E. Sickafus, E.A. Kotomin and B.P. Uberuaga), Springer, NATO Publishing Unit, Dordrecht, The Netherlands, 2006, pp. 1–24.
5. Voter, A.F. Parallel replica method for dynamics of infrequent events. Phys. Rev. B 1998, 57, 13985–88.
6. Voter, A.F., Montalenti, F., Germann, T.C. Extending the time scale in atomistic simulation of materials. Ann. Rev. Mater. Res. 2002, 32, 321–46.
7. Uberuaga, B.P., Stuart, S.J., Voter, A.F. Parallel replica dynamics for driven systems: derivation and application to strained nanotubes. Phys. Rev. B 2007, 75, 014301-1–9.
8. Uberuaga, B.P., Voter, A.F., Sieber, K.K., Sholl, D.S. Mechanisms and rates of interstitial H_2 diffusion in crystalline C_{60}. Phys. Rev. Lett. 2003, 91(10), 105901–4.
9. Kum, O., Dickson, B.M., Stuart, S.J., Uberuaga, B.P., Voter, A.F. Parallel replica dynamics with a heterogeneous distribution of barriers: application to *n*-hexadecane pyrolysis. J. Chem. Phys. 2004, 121(20), 9808–19.
10. Uberuaga, B.P., Valone, S.M., Baskes, M.I., Voter, A.F. Accelerated molecular dynamics study of vacancies in Pu. AIP Conf. Proc. 2003, (673)213–5.
11. Uberuaga, B.P., Hoagland, R.G., Voter, A.F., Valone, S.M. Direct transformation of vacancy voids to stacking fault tetrahedra. Phys. Rev. Lett. 2007, 99, 135501-1–4.
12. Mishin, Y., Suzuki, A., Uberuaga, B.P., Voter, A.F. Stick-slip behavior of grain boundaries studied by accelerated molecular dynamics. Phys. Rev. B 2007, 75, 224101-1–7.
13. Duan, Y., Halley, J.W., Curtiss, L., Redfern, P. Mechanisms of lithium transport in amorphous polyethylene oxide. J. Chem. Phys. 2005, 122, 054702-1–8.
14. Warner, D.H., Curtin, W.A., Qu, S. Rate dependence of crack-tip processes predicts twinning trends in f.c.c. metals. Nat. Mater. 2007, 6, 876–81.
15. Shirts, M., Pande, V.S. Screen savers of the world unite! Science 2000, 290(5498), 1903–4.
16. Voter, A.F. A method for accelerating the molecular dynamics simulation of infrequent events. J. Chem. Phys. 1997, 106, 4665–77.
17. Berne, B.J., Ciccotti, G., Coker, D.F. (eds), Classical and Quantum Dynamics in Condensed Phase Simulations, World Scientific, River Edge, NJ, 1998.
18. Valleau, J.P., Whittington, S.G. A guide to Monte Carlo for statistical mechanics: 1. Highways. In Statistical Mechanics. A. Modern Theoretical Chemistry (ed. B.J. Berne), Vol. 5, Plenum, New York, 1977, pp. 137–68.
19. Becker, K.E., Mignogna, M.H., Fichthorn, K.A. Accelerated molecular dynamics of temperature-programmed desorption. Phys. Rev. Lett. 2009, 102, 046101-1–4.
20. Voter, A.F. Hyperdynamics: accelerated molecular dynamics of infrequent events. Phys. Rev. Lett. 1997, 78, 3908–11.
21. Miron, R.A., Fichthorn, K.A. Heteroepitaxial growth of Co/Cu(001): an accelerated molecular dynamics simulation study. Phys. Rev. B 2005, 72, 035415-1–7.

22. Hamelberg, D., Mongan, J., McCammon, J.A. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. J. Chem. Phys. 2004, 120, 11919–29.
23. Sørensen, M.R., Voter, A.F. Temperature-accelerated dynamics for simulation of infrequent events. J. Chem. Phys. 2000, 112, 9599–606.
24. Montalenti, F., Voter, A.F. Exploiting past visits or minimum-barrier knowledge to gain further boost in the temperature-accelerated dynamics method. J. Chem. Phys. 2002, 116(12), 4819–28.
25. Uberuaga, B.P., Smith, R., Cleave, A.R., Henkelman, G., Grimes, R.W., Voter, A.F., Sickafus, K.E. Dynamical simulations of radiation damage and defect mobility in MgO. Phys. Rev. B 2005, 71(10), 104102-1.
26. Uberuaga, B.P., Smith, R., Cleave, A.R., Montalenti, F., Henkelman, G., Grimes, R.W., Voter, A.F., Sickafus, K.E. Structure and mobility of defects formed from collision cascades in MgO. Phys. Rev. Lett. 2004, 92(11), 115505-4.
27. Montalenti, F., Sørensen, M.R., Voter, A.F. Closing the gap between experiment and theory: crystal growth by temperature accelerated dynamics. Phys. Rev. Lett. 2001, 87, 126101-1–4.
28. Sprague, J.A., Montalenti, F., Uberuaga, B.P., Kress, J.D., Voter, A.F. Simulation of growth of Cu on Ag(001) at experimental deposition rates. Phys. Rev. B 2002, 66(20), 205415-1–10.
29. Harris, D.J., Lavrentiev, M.Y., Harding, J.H., Allan, N.L., Purton, J.A. Novel exchange mechanisms in the surface diffusion of oxides. J. Phys. Condens. Matter 2004, 16(13), L187–92.
30. Cogoni, M., Uberuaga, B.P., Voter, A.F., Colombo, L. Diffusion of small self-interstitial clusters in silicon: temperature-accelerated tight-binding molecular dynamics simulations. Phys. Rev. B 2005, 71(12), 121203-1–4.
31. Meerbach E., Schütte, C. Sequential change point detection in molecular dynamics trajectories. J. Multivariate Anal. Submitted for publication.
32. Miron, R.A., Fichthorn, K.A. Accelerated molecular dynamics with the bond-boost method. J. Chem. Phys. 2003, 119(12), 6210–6.
33. Perez, D., Voter, A.F. Accelerating atomistic simulations through self-learning bond-boost hyperdynamics. Submitted for publication.
34. Shim, Y., Amar, J.G., Uberuaga, B.P., Voter, A.F. Reaching extended length scales and time scales in atomistic simulations via spatially parallel temperature-accelerated dynamics. Phys. Rev. B 2007, 76, 205439-1–11.
35. Shim, Y., Amar, J.G. Semirigorous synchronous sublattice algorithm for parallel kinetic Monte Carlo simulations of thin film growth. Phys. Rev. B 2005, 71, 1254321-1–14.
36. Botez, C.E., Miceli, P.F., Stephens, P.W. Temperature-dependent vacancy formation during the growth of Cu on Cu(001). Phys. Rev. B 2002, 66, 195413-1–6.
37. Shim, Y., Borovikov, V., Uberuaga, B.P., Voter, A.F., Amar, J.G. Vacancy formation and strain in low-temperature Cu/Cu(100) growth. Phys. Rev. Lett. 2008, 101, 116101-1–4.
38. Uberuaga, B.P., Montalenti, F., Germann, T.C., Voter, A.F. Accelerated molecular dynamics methods. In Handbook of Materials Modeling, Part A: Methods (ed. S. Yip), Springer, Dordrecht, The Netherlands, 2005, p. 629.
39. Miron, R.A., Fichthorn, K.A. Multiple-time scale accelerated molecular dynamics: addressing the small-barrier problem. Phys. Rev. Lett. 2004, 93, 128301-1–4.