LA-UR-15-27574 (Accepted Manuscript)

The Modern Temperature-Accelerated Dynamics Approach

Zamora, Richard James
Uberuaga, Blas P.
Perez, Danny
Voter, Arthur Ford

To be published in: Annual Review of Chemical and Biomolecular Engineering

DOI to publisher's version: 10.1146/annurev-chembioeng-080615-033608

Permalink to record: http://permalink.lanl.gov/object/view?what=info:lanl-repo/lareport/LA-UR-15-27574

The Modern Temperature-Accelerated Dynamics Approach

Richard J. Zamora[1], Blas P. Uberuaga[2], Danny Perez[1], and Arthur F. Voter[1*]

[1]Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM, USA 87545 *email: afv@lanl.gov
[2]Materials Science and Technology, Los Alamos National Laboratory, Los Alamos, NM, USA 87545

Xxxx. Xxx. Xxx. Xxx. YYYY. AA:1–25

This article's doi:
10.1146/((please add article doi))

Keywords

accelerated molecular dynamics, temperature-accelerated dynamics, molecular dynamics, infrequent events, rare events

Abstract

Accelerated molecular dynamics (AMD) is a class of MD-based methods used to simulate atomistic systems in which the metastable state-to-state evolution is slow compared to thermal vibrations. Temperature-accelerated dynamics (TAD) is a particularly efficient AMD procedure in which the predicted evolution is hastened by elevating the temperature of the system and then recovering the correct state-to-state dynamics at the temperature of interest. TAD has been used to study various materials applications, often revealing surprising behavior beyond the reach of direct MD. This success has inspired a number of algorithmic performance enhancements, as well as the analysis of its mathematical framework. Recently, these enhancements have leveraged parallel programing techniques to enhance both the spatial and temporal scaling of the traditional approach. In this paper, we review this ongoing evolution of the modern TAD method and introduce the latest development: Speculatively Parallel TAD (SpecTAD).

Contents

1. INTRODUCTION

Molecular dynamics (MD) is a powerful computational modeling method used within many scientific and engineering disciplines to predict the time evolution of interacting systems of atoms. By integrating the classical equations of motion, MD directly tracks the dynamical evolution of the atomic positions, without making any significant approximations beyond the system-dependent potential energy formulation used to derive the interatomic forces.

Although MD has been used to illuminate many complex phenomena in chemistry, biology, physics, and materials science, its inherently sequential nature often prevents the exploration of time-scales longer than a few microseconds. This is because the discrete time step used to integrate the equations of motion must be small enough to resolve thermal vibrations, i.e., no larger than a few femtoseconds. Billions of MD steps are therefore required to reach microseconds of simulation time. Unfortunately, such simulations are often not long enough to characterize rare-event mechanisms, which are known to govern the macroscopic behavior of many materials over experimentally relevant timescales.

In the case of atomistic-scale material simulation, the rare event designation does not imply that a mechanism occurs infrequently relative to human perception, but infrequently relative to the thermal vibration of atoms. The designation can be more rigorously defined in terms of the (3N)-dimensional potential energy surface (PES) defined by the instantaneous interaction of all N atoms in a canonical system. The PES is typically comprised of local *basins*, each corresponding to a separate metastable state if the barriers leading

to other basins are high enough. Following the assumptions of classical transition state theory, a system starting off in a state A will transition to an adjacent state B when its dynamical trajectory passes through the (3N-1)-dimensional dividing surface located along the potential energy ridgetop between A and B. In a rare-event system, the barrier separating adjacent states is typically $\gg k_B T$, where k_B is the Boltzmann constant and T is the temperature of the system.

Beginning in the late 1990s, with rare-event systems in mind, Voter and coworkers have been developing a family of methods known as accelerated molecular dynamics (AMD) to overcome the timescale limitation of MD. There are currently three AMD methods, each designed to leverage the timescale disparity between thermal vibrations and state-to-state transitions. The first of the methods, hyperdynamics, accelerates the transition rates through modification of the potential surface on which the dynamics evolves (1). The second, parallel replica dynamics (ParRep), accelerates the simulation by parallelizing the accumulation of time (2). The third, temperature-accelerated dynamics, accelerates transitions by elevating the temperature of the system. The common themes among these methods are to (a) reduce the waiting time of a transition from its natural time scale to the order of picoseconds and (b) let MD find appropriate pathways without relying on human intuition or the need to find all relevant pathways. The focus of this paper is the temperature-accelerated dynamics (TAD) method, first presented by Sørensen and Voter (3). For a more comprehensive introduction to AMD methods in general, the reader is referred to recent literature reviews (4, 5, 6, 7).

Since its introduction, TAD has become a popular technique for extending simulation timescales to tackle questions in materials science (8, 9, 10, 11, 12, 13, 14, 15, 16, 17). As a result of this success, the original procedure has been revisited on many occasions in order to further optimize its performance. Many of these variations have aimed at enhancing performance without requiring any parallel computation (18, 19, 20). In addition, three separate algorithmic extensions have been introduced to leverage the growing availability of parallel computing systems. Two of these extensions rely on spatial decomposition to enhance TAD's relatively poor performance when modeling large systems (21, 22). The most recent extension uses a speculative parallelization technique to greatly enhance performance in a way that is even beneficial to small systems.

This paper is intended to review the modern TAD method by introducing the original procedure and useful extensions. In Section 2, we review the basics of TAD and introduce the traditional procedure. This introduction includes a brief mathematical analysis, as well as a description of the kinetic Monte Carlo (kMC) extension to TAD (the so-called synthetic mode). In Section 3, we review selected applications to demonstrate the types of systems in which TAD has proven valuable. In Section 4, we review some modifications that have been made to the TAD formulation in order to improve the serial performance of the method. In Section 5, we review parallel extensions designed to make the TAD procedure better equipped for large systems. In Section 6, we discuss the very recent and ongoing development of Speculatively Parallel TAD (SpecTAD), which uses parallelization to explore multiple concurrent states. Finally, in Section 7 we summarize the current state of TAD and draw conclusions about the future of this valuable AMD approach.

2. TEMPERATURE-ACCELERATED DYNAMICS (TAD) BASICS

The objective of TAD is to hasten the state-to-state evolution of an atomistic system in the canonical ensemble by running molecular dynamics at an elevated temperature and using observed transitions to construct an accurate lower-temperature trajectory (3). TAD assumes that the system of interest obeys the conditions of harmonic Transition State Theory (hTST) (23). In this regard, it is the most approximate of the three AMD methods. The critical hTST assumptions are:

(A.1) There are no correlated barrier re-crossings (i.e., the system immediately reaches equilibrium in a new basin after passing through a saddle plane).

(A.2) The rate law for transitions out of an energy basin along transition path j can be defined by an Arrhenius law,

$$k_j = \nu_{0,j} \exp\left[-E_j/(k_B T)\right], \tag{1}$$

where k_j is the transition rate along path j, E_j is the height of the energy barrier along path j, and $\nu_{0,j}$ is the temperature-independent pre-exponential factor. In typical statistical mechanical systems, $\nu_{0,j}$ is equivalent to the Vineyard prefactor (23), defined as the ratio of relevant vibrational modes at the minimum and saddle states:

$$\nu_{0,j} = \left(\prod_{n=1}^{3N} \nu_n\right) \Big/ \left(\prod_{n=1}^{3N-1} \nu'_n\right). \tag{2}$$

Here, the use of 3N assumes that there are no zero-frequency modes that need to be omitted; i.e. full translational and rotational freedom leads to 3N-6. Note that the original formulation of TAD does not rely on the fact that the Vineyard expression is accurate, but only that the prefactor is temperature independent.

Although the TAD procedure can usually accommodate systems in which (A.1) is not always true, (A.2) is crucial. This is because Eqn. 1 leads to an efficient way to project observed transitions at high temperature (T_{High}) onto a hypothetical low temperature (T_{Low}) time line. Since the pre-exponential factor is assumed to be temperature independent, it cancels out when taking a ratio of rates for the same transition path at different temperatures. This leads to a straightforward definition of the transition time along path j at T_{Low}, $t_{Low,j}$, given an observed transition time at T_{High}, $t_{High,j}$:

$$t_{Low,j} = \frac{k_{High,j}}{k_{Low,j}} = t_{High,j} \exp\left[E_j \left(\frac{1}{k_B T_{Low}} - \frac{1}{k_B T_{High}}\right)\right]. \tag{3}$$

This relationship tells us that we can project a T_{High} transition onto a hypothetical T_{Low} timeline if we can determine E_j for the observed transition. It is important to note, though, that this relationship also tells us that the transition occurring first at high temperature may not be the most probable at low temperature, since, although all transition rates will increase with an increase in temperature, the rate of a high-barrier transition will increase faster than that of a low-barrier transition. Thus, the states visited in a typical high-temperature MD evolution are often unrepresentative of those visited at low-temperature.

2.1. The TAD Procedure

Due to the need to correct for this bias towards high-barrier transitions at high temperature, the basic building block of the TAD procedure is basin-constrained molecular dynamics (BCMD) (24, 3). In BCMD, a system is placed into an initial state, and then MD is performed until the system naturally escapes to a neighboring state. When such a transition occurs, information about the escape path is recorded, and the system is then placed back into the original basin so that the process can be repeated to find alternative escape paths. Before the MD trajectory can be continued to find new events, it must first be thermalized and "blacked-out", such that it has no memory of the previous transitions, a process that is usually performed with another block of MD. TAD efficiently predicts an accurate low-temperature evolution by minimizing the amount of high-temperature BCMD time needed for each state visit before a statistically accurate low-temperature escape path can be chosen.

The length of BCMD time needed in each visited state depends on two critical parameters: The minimum bound on pre-exponential factors expected in the system (ν_{min}), and the level of uncertainty allowed in the selection of the most probable escape path from each visited state (δ). These parameters are used to define ν^*, which can be interpreted as a confidence-corrected lowest possible transition rate at infinite temperature,

$$\nu^* = \frac{\nu_{min}}{\ln(1/\delta)}. \tag{4}$$

Using ν^*, the so-called *stop time* at which the high temperature trajectory can be terminated and the proper escape path (with 1-δ confidence) selected, can be defined as:

$$t_{High,Stop} = \frac{1}{\nu^*}\left(\nu^* t_{Low,Short}\right)^{T_{Low}/T_{High}}, \tag{5}$$

where $t_{Low,Short}$ is the shortest extrapolated low-temperature transition time recorded during the current BCMD run.

The construction of the high-temperature stop time after one observed transition can be understood visually by plotting ln(1/t) vs β, where $\beta = 1/(k_B T)$ such that $\beta_{\mathrm{High}} = 1/(\mathrm{k_B T_{High}})$ and $\beta_{\mathrm{Low}} = 1/(\mathrm{k_B T_{Low}})$. As shown in **Fig. 1a**, the observation of transition 1 at high temperature can be represented as a point at (β_{High}, $ln(1/t_{High,1})$). Following Eqn. 3, the point corresponding to the same transition at low temperature, (β_{Low}, $ln(1/t_{Low,1})$), can be determined by drawing a line with a slope equal to the negative of the transition energy barrier, -$\mathrm{E_1}$. This relationship is shown as a blue line in **Fig. 1a**.

Once (β_{Low}, $ln(1/t_{Low,1})$) is plotted, the ν^* parameter can then be used to define the lower bound on the transition energy barrier for any new escape occurring after time $t_{High,1} + \Delta t$ that can replace $t_{Low,1}$ as the shortest low-temperature transition time. This lower bound, $\mathrm{e}(t_{High,1} + \Delta t)$, is simply the negative slope of the line connecting (0,ln(ν^*)) with (β_{High}, $ln(1/(t_{High,1} + \Delta t))$). Similarly, the upper bound on the energy barrier for any new escape capable of replacing the shortest low-temperature transition time, $\mathrm{E}(t_{High,1} + \Delta t)$, is the negative slope of the line connecting (β_{High}, $ln(1/(t_{High,1} + \Delta t))$) with ($\beta_{\mathrm{Low}}$, $ln(1/t_{Low,1})$). These relationships are both drawn as dashed black lines in **Fig. 1a**, with the hollow point at the intersection representing an arbitrary amount of time, Δt, after the observation of transition 1 at high temperature.

As Δt is increased in the given scenario, there is a shrinking set of energy barriers capable of replacing transition 1 as $t_{Low,Short}$. The point when the upper and lower bounds converge, shown as a red point at (β_{High}, $ln(1/t_{High,Stop})$) in **Fig. 1a**, is the high temperature time

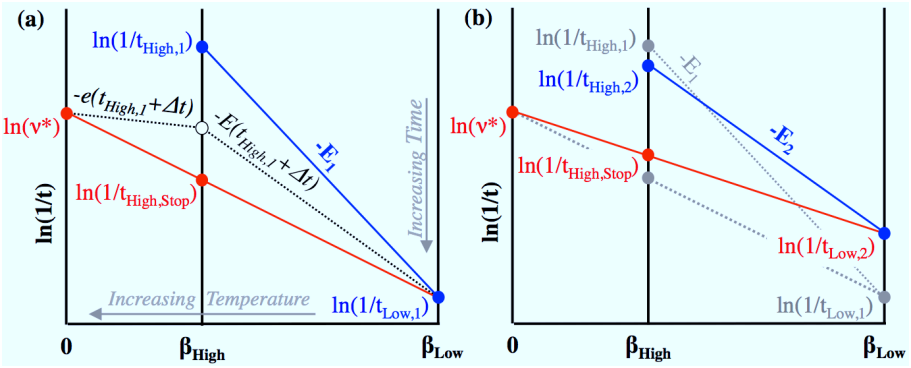

Figure 1

An Arrhenius-like representation of the high-temperature stop time definition in TAD. (a) The stop time construction for the first event at high temperature. (b) The stop time construction for a second high temperature event that would have occurred faster than the first event at low temperature.

at which it is safe to move the system into the next state associated with transition 1 as long as it is still the case that $t_{Low,Short} = t_{Low,1}$. The *stop time* can be updated using Eqn. 5 each time a new high temperature transition leads to a new $t_{Low,Short}$. For example, **Fig. 1b** shows a case in which a second event occurs with $t_{Low,2} < t_{Low,1}$ before $t_{high,Stop}$ is reached. Once the current stop time is finally reached, there is a $< \delta$ probability that an unobserved pathway could still extrapolate to a time shorter than $t_{Low,Short}$.

2.1.1. The Basic TAD Algorithm. We are now ready to introduce the basic TAD algorithm, which predicts the state-to-state evolution of a system at T_{Low} by using thermostated BCMD to explore the PES at T_{High}. Although this introduction is self-contained, additional discussion can be found in references (3, 18, 19, 25, 26).

The TAD procedure starts with the low temperature simulation time for all states (t_{tad}) equal to zero. The following steps are then iterated on for each state visit:

1. Let $t_{High} = 0$, $t_{High,Stop} = \infty$ and $t_{Low,Short} = \infty$. Here t_{High} is the amount of high-temperature MD time that has been accumulated in the current state.
2. Place the system into state i by setting the atom coordinates (X^{atoms}) so that the energy minimizer (X_i^{atoms}) is in the same PES basin.
3. Without advancing t_{High}, perform MD at T_{High} for Δt_{therm}. This step is meant to ensure the system is in proper thermal equilibrium (thermalized) and that the system has no memory of previous state visits or escape attempts.. Go back to **Step 2** if the system escapes state i during Δt_{therm}, otherwise proceed to **Step 4**.
4. Perform MD at T_{High} for Δt_{MD}.
5. Perform a transition check by quenching the system and comparing X^{atoms} to the X_i^{atoms} minimizer. If a transition is detected go to **Step 6**, otherwise, advance t_{High} by Δt_{MD} and then go to step **Step 10**.
6. Compare new escape attempt j to all previous escape attempts made during the current visit to state i. If the path of the new escape attempt was already observed at an earlier time, go to **Step 10**.

7. Set the transition time for the newly observed escape path, $t_{High,j}$, to be a random time between t_{High} and $(t_{High} - \Delta t_{MD})$. Advance t_{High} to be $t_{High,j}$.
8. Perform a two-ended saddle point finding calculation, such as a Nudged Elastic Band (NEB) (27), to locate the appropriate saddle point along path j. use Eqn. 3 to determine $t_{Low,j}$ for the new event.
9. If $t_{Low,j} < t_{Low,Short}$, set $t_{Low,Short} = t_{Low,j}$ and use Eqn. 5 to update the high-temperature stop time.
10. If $t_{High} > t_{High,Stop}$, go to **Step 11**. Otherwise, go to **Step 4**.
11. Increment t_{tad} by $t_{Low,j}$, move the system into the corresponding state and then go to **Step 1**.

After each cycle of these steps, the dynamics of the system are advanced by one state.

2.1.2. The TAD Boost. One of the key purposes of using any AMD method is to simulate a longer period of time than MD in the same amount of wall clock time (WCT). Therefore, the performance of TAD is often discussed in terms of a computational boost factor, B_{TAD}, defined as the increase in the computational speed with respect to the typical performance of MD. For TAD, this boost factor metric is given by

$$B_{TAD} \equiv \frac{WCT_{MD}(t_{TAD})}{WCT_{tad}(t_{tad})} \leq \frac{\bar{t}_{Low}}{\bar{t}_{High,Stop}}. \tag{6}$$

For a generic system, an upper bound on the TAD boost can be written as (18)

$$B_{TAD} \leq \left[\frac{\nu^*}{(\Sigma_i k_{i,Low})}\right]^{1-\beta_{High}/\beta_{Low}} \frac{1}{\Gamma\left(1+\beta_{high}/\beta_{Low}\right)}, \tag{7}$$

where $\Sigma_i k_{i,Low}$ is the total escape rate at low temperature, and Γ is the Euler gamma function. In this definition, the right-hand side corresponds to $\bar{t}_{Low}/\bar{t}_{High,Stop}$, where the overhead caused by detecting transitions and calculating energy barriers is ignored.

2.2. Mathematical Analysis of TAD

Now that we have outlined the basic TAD procedure, we can briefly discuss some modifications recently suggested by Aristoff and Lelievre to generalize the mathematical rigor of the approach (25). Although the reader should refer to the original work for the complete mathematical treatment in terms of quasi-stationary distributions (QSD), the general message is that modifications can be introduced to either partially or completely relax assumptions (A.1) and (A.2).

In an ideal method, rigorous beyond the scope of hTST, additional steps would be needed to relax (A.2). Also, a *de-correlation* phase would be necessary to ensure the system is in thermal equilibrium at T_{Low} before any state is explored at T_{High}, avoiding (A.1). Although Aristoff and Lelievre present a specific *Ideal* TAD procedure that can be used to predict the exact metastable dynamics of a general stochastic system obeying overdamped Langevin dynamics, they caution that this ideal method would require too much computational work to be used in practice. Instead, they show that only three modifications are needed to produce Ideal TAD in the limit that $\beta_{Low}, \beta_{High} \to \infty$ and $\tau_{cor} \to \infty$, where τ_{cor} is the correlation time which will be explained below. The resulting TAD procedure, which achieves a balance of generality and performance, is equivalent to the basic procedure given above, aside from the following modifications:

(M.1) A de-correlation step is accomplished by performing low-temperature MD for τ_{cor} before beginning the standard TAD procedure for a new state visit. This is not a BCMD step, because the system is allowed to move into a new state immediately if a state transition occurs before τ_{cor}.

(M.2) Redefine the high-temperature stop time as $t_{High,Stop} = t_{Low,Short}/C$, where C is a lower bound on $exp[-E_i(\beta_{High} - \beta_{Low})]$ over all paths leaving the current state[1].

Upon closer inspection of these recommended modifications, it is apparent that only (M.1) can be introduced without requiring any previous PES knowledge. As for (M.2), the authors note that the primary motivation for redefining $t_{High,Stop}$ is to avoid the uncertainty, δ, inherent in the traditional ν^* parameter. This modification is already used in the Emin-TAD and Dimer-TAD methods, discussed in Sections 4.2 and 4.3. However, it is not generally the case that C can be known.

It is certainly clear that (M.1) can be used to relax the assumptions associated with (A.1). However, a the proper implementation of (M.1) requires additional low temperature MD time and transition checks. Most importantly, it is a challenge to efficiently prepare an accurate low-temperature trajectory given the corresponding high-temperature trajectory. For these reasons, this modification is most advisable when modeling systems in which assumption (A.1) is expected to introduce significant error. Otherwise, the inevitable reduction in the TAD performance boost may not be worth the limited improvement in the accuracy of the T_{Low} prediction.

2.3. Synthetic Mode Extension

The basic TAD procedure discussed so far assumes that any information collected during a specific state visit (total t_{High} accumulated, E_j for the different observed pathways, etc.) is thrown away once the system moves on to the following state. This practice can be particularly inefficient when exploring a potential energy landscape best characterized as a superbasin of low-barrier states. In such a scenario, the system may revisit the same state many times before moving into a different superbasin. Fortunately, repetitive events can be handled efficiently in TAD by treating them in so-called *synthetic* mode (3, 21). This useful extension to TAD, inspired by kinetic Monte Carlo (kMC), provides a framework for progressively decreasing the amount of BCMD time needed during each revisit to a state.

In contrast to the MD-based approach used in basic TAD, kMC predicts the evolution of an interacting system of atoms by assuming that all transition rate constants, $k_{i,j}$, are known a priori (28, 29). Using such information, the state-to-state evolution can be stochastically generated. The synthetic extension to TAD assumes that we can use the kMC approach for some of the observed pathways. That is, if the rate of the event is known at low temperature, the event can be treated synthetically. This rate can be extracted from a dynamic estimation of the temperature-independent prefactor, using the rate of observed high-temperature events, or a direct Vineyard calculation of Eqn. 2 (23). Although the original TAD paper introduces a synthetic mode extension (3), the specific procedure explained next was first described in Ref. (21).

The general synthetic approach, explained graphically in **Fig. 2**, works as follows: Assume that the system of interest has just entered a state that has been visited before

[1] In the original work (25), this modification is labeled (M.3), while (M.2) corresponds to step 3 in the basic procedure.

(with the previous stop time corresponding to the dashed blue line). During the previous visit, the total amount of accumulated high-temperature MD time had been recorded, as well as the number of attempts and energy barriers for observed escape paths. At this point, it is possible that there is enough information to determine a rate constant for one or more of the previously seen events. If this is so, we can put these events into the so-called synthetic mode by placing stochastic transition times along these pathways on the low-temperature timeline before performing any high-temperature MD. From this point on, we can ignore any transitions observed along the synthetic paths in the T_{High} MD, because we already have low-temperature predictions.

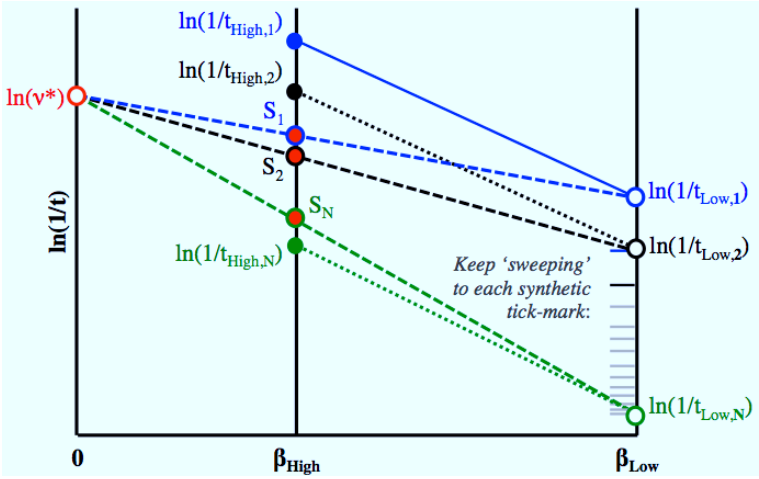

Figure 2

A graphical representation of the kMC-like synthetic mode used in TAD. The solid blue line corresponds to the usual temperature extrapolation of an event observed at high temperature during the first visit to a state (visit 1). The blue dashed line is the usual stop-line corresponding to Eqn. 5, where S_1 corresponds to $\ln(1/t_{\mathrm{HighStop}})$. Here, we assume that this first event is eventually accepted and that its rate-constant can be known before the next visit. Therefore, when the state is revisited, the high and low temperature MD clocks can begin accumulating time at the blue dashed line, rather than at zero. The act of continuing the accumulation of time in this way is sometimes called *sweeping*. This sweeping is allowed because we can use the rate constant from the previously chosen event to predict its next occurrence on the β_{Low} axis (this prediction is shown as a blue tick mark). When the rate constant is used to place a tick mark in this way, the event is said to be in synthetic mode, because MD is no longer needed to predict its occurrence at low temperature. For the second visit to this state, the stop line is shown as a black dashed line. Although this stop line was constructed from a synthetic prediction rather than an MD event, the hypothetical temperature extrapolation is shown as a dotted black line. From here, the synthetic sweeping procedure can be continued indefinitely as long as every accepted event corresponds to a synthetic event (examples shown as grey tick marks on the right-hand axis). At some eventual visit N, the hypothetical high-to-low extrapolation (green dotted line) will predict that a synthetic event should need to run past the sweeping stop line (green dashed line) to actually observe the event using MD. This illustrates why the event time tick mark must be generated synthetically.

Treating repeated events in this way becomes particularly advantageous when a synthetic event is chosen for $t_{Low,Short}$, and then the same state is visited again. In this case we can modify **Step 1** in the basic TAD procedure by setting the initial t_{High} to the high-temperature time reached during previous visit and by offsetting any low-temperature

events by the previously accepted synthetic transition time. Cumulatively tracking the high and low-temperature timelines in this way is often called sweeping. In this sweeping mode, as similar amounts of time are accumulated at the high temperature, increasingly greater amounts of time are "simulated" at the low temperature.

When using synthetic mode in a given state, it is safe to continue sweeping to a longer time upon each revisit, provided that the negative slope of the sweeping line never exceeds any known activation barriers that are not yet in synthetic mode. This requirement is justified in the last two sentences of the **Fig. 2** caption. As the sweeping procedure is allowed to progress, the amount of high-temperature MD time needed to reach $t_{High,Stop}$ becomes shorter and shorter with every revisit to the state. Eventually a single MD step at high temperature may correspond to enough low-temperature time to accept many sequential events into the future, providing TAD with kMC-like efficiency.

3. SELECTED APPLICATIONS

TAD has been extensively applied to a number of problems in materials science and nearly every case has revealed surprising behaviors that significantly altered our understanding of the material. A few examples are highlighted here to demonstrate the types of insights that can be gained from such simulations, and the types of systems that can be studied.

3.1. Annealing of Nanotube Fragments

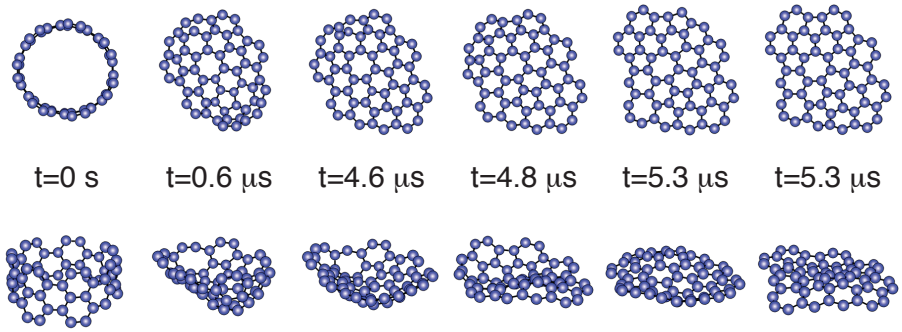

Figure 3

Evolution of a 60-atom fragment of a (7,7) carbon nanotube at T=1500 K as viewed from both an axial and a radial perspective. Relatively quickly in the evolution of the fragment, one of the open ends "zips" closed, which induces strain in the other open end, driving the fragment to flatten further. After 5.3 microseconds, the fragment has formed a graphene-like fragment that is primarily comprised of 6-member rings, but still has a few 5 and 7-member rings within the network.

Inspired by experiments in which buckyballs were found in the soot of laser-ablated carbon nanotubes, TAD was used to examine the annealing of idealized nanotube fragments (17). The goal was to understand how buckyballs might form from small fragments of nanotubes, and what mechanisms at the atomic scale induced this behavior. **Fig. 3** shows an example of one of the simulations (the simulation in the last frame of Fig. 2 in Ref. (17)) in which a 7x7 nanotube fragment, originally cylindrical in shape, flattens over the course of microseconds at T=1500 K into a small graphene fragment. In this particular case, one

end of the original nanotube first closes through a zipper-like mechanism. However, once that happens, the curvature induced by this closure creates tension in the outer edge of the fragment, causing it to further flatten, ultimately leading to the graphene fragment illustrated in the last frame of **Fig. 3**. While these types of graphene fragments are much higher in energy than a closed C_{60} buckyball, the transition from this particular nanotube fragment to a buckyball necessarily passes through these graphene-like states that act as kinetic bottlenecks for the full transformation to a buckyball.

3.2. Cation Dynamics in Complex Oxides

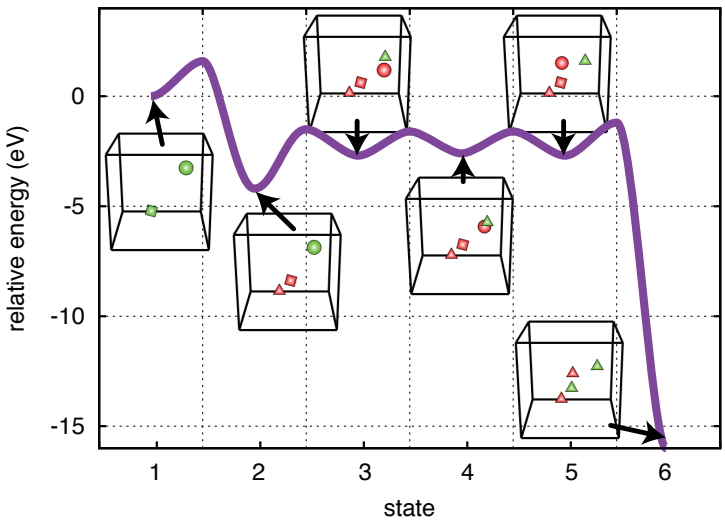

Figure 4

Evolution of a Ti Frenkel pair in $Lu_2Ti_2O_7$ pyrochlore. Immediately, first the Ti vacancy and then the Ti interstitial transform into Lu defects and corresponding antisites. As the Lu interstitial diffuses and then recombines with the Lu vacancy, instead of perfect recombination, two more antisites are formed. Thus, the net result of the recombination of a Ti Frenkel pair is two antisite pairs. The scheme is red for Lu defects and green for Ti defects, and circles for interstitials, squares for vacancies, and triangles for antisites, as determined by comparing the structure to the perfect pyrochlore structure.

Since the first AMD simulations on oxides, involving TAD to examine defect clustering dynamics in MgO (11), TAD has also been used to study defect behavior in complex oxides (30, 31, 32, 33). Understanding cation dynamics in these materials is important for predicting performance in radiation environments, during sintering, and in other applications where degradation is linked to cation evolution. **Fig. 4** shows a typical result of these simulations, this one originally presented in Ref. (34). Under irradiation, defects are inevitably created and it is the evolution of those defects that dictate the ultimate fate of the material. In a simple material, such as Cu or MgO, the most likely result of such a simulation would be that the interstitial diffuses around until it feels an elastic interaction with the vacancy and it would then simply recombine, leading to a defect-free system. In complex oxides such as pyrochlore, TAD simulations reveal more complicated behavior. Two important results

are evident from **Fig. 4**. Result 1: the very first event evolving either the Ti interstitial or the Ti vacancy involves an immediate transformation to an antisite defect and a Lu defect. This implies that B cation defects (B=Ti in this case) are kinetically unstable and transform into other defects. Result 2: as Lu Frenkel pairs recombine, instead of perfect annihilation, even more antisites are created. This implies that perfect recombination is not possible in such materials; antisite disorder will always be generated. This has important ramifications for our understanding of radiation damage evolution in complex oxides.

3.3. Defect Interactions with Grain Boundaries

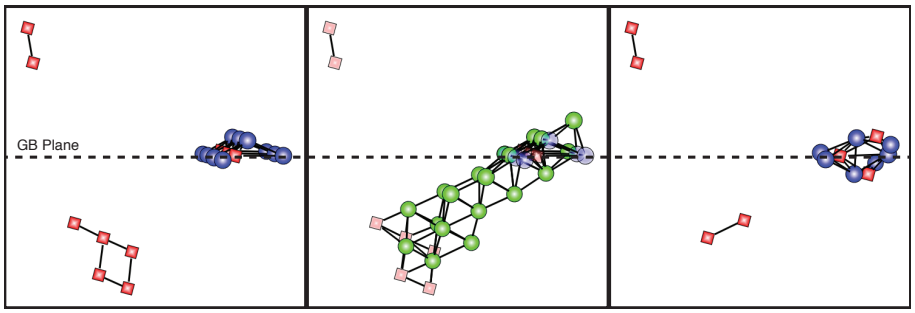

Figure 5

Interaction of collision-cascade defects near a grain boundary in Cu, as simulated by TAD. In the initial state, representative of the defect structure after a collision cascade, interstitials (blue circles) are packed into the grain boundary plane while an excess of vacancies (red squares) remain in the grain interior. Through complex concerted events involving one relatively low barrier (0.17 eV in this case), interstitials can "emit" from the grain boundary and annihilate some of those vacancies (central panel). The green circles highlight atoms that moved more than 1 angstrom and are superimposed on the initial structure. The resulting defect structure, as illustrated in the last frame, contains fewer bulk vacancies and fewer interstitials within the grain boundary. The "interstitial emission" mechanism has led to substantial healing of defects.

As a final example, TAD has been used to understand how irradiation-induced defects interact with grain boundaries, using Cu as a model system (35, 36). These simulations have revealed several surprises, not least of which is the so-called interstitial emission mechanism. When collision cascades occur near grain boundaries, the typical result is that the fast moving interstitials quickly diffuse to the boundary, leading to a damage state in which there is a high concentration of interstitials at the boundary and an excess number of vacancies in the grain interior (more than would form if there were no grain boundary) (37). Starting from such a configuration, as illustrated in **Fig. 5**, TAD simulations revealed that GB and bulk interstitials could still interact. Through highly concerted mechanisms with lower barriers than vacancy migration, the interstitials can emit from the boundary and directly annihilate the vacancies; in the case shown in **Fig. 5**, three vacancies are annihilated by such a mechanism. This highlights the complex defect behavior that can occur near microstructural features. More recent simulations (both TAD and adaptive kMC) have revealed that as interstitials cluster in boundaries, their mobility dramatically decreases (35). Together, these results have significantly altered the theory of defect evolution in nanostructured materials.

4. SERIAL TAD EXTENSIONS

Although the TAD procedure in its basic form still enjoys popularity today, researchers have been developing ways to increase the available boost factor ever since the method was first introduced. In more recent years, these modifications have aimed to parallelize the traditionally serial method. However, significant improvement can be achieved using serial modifications alone. We discuss these first.

4.1. Learning From the Past

The earliest modification to the basic TAD procedure, sometimes referred to as *learning from the past* TAD (LFP-TAD), provides a simple way to reduce the amount of BCMD time needed in a state that has been previously visited (18). LFP-TAD acts as an intermediate step between the basic TAD procedure and the synthetic mode extension, because it does not require the calculation of rate constants. Instead, LFP-TAD provides an alternative way to calculate $t_{High,Stop}$. The key point is that the usual definition of t_{stop} is often too conservative and can be reduced when the minimum energy barrier leading out of the state, or a lower bound on it, is known.

Recall from **Fig. 1** in Section 2.1 that the traditional stop time can be defined for event 1 as the high-temperature time at which $\mathrm{e}(t_{High,1}+\Delta\tau)$ converges with $\mathrm{E}(t_{High,1}+\Delta\tau)$. In this construction, e represents a lower bound on the energy barrier of a path that did not yet occur, while E is the upper bound on the energy barrier of an event that can still redefine $t_{Low,Short}$. Now assume that, when revisiting a given state, τ_{prev} has already been accumulated at high temperature. Therefore, we can reformulate e to be the minimum possible barrier for an event that did not already occur at $t_{High,1}+\tau_{prev}+\Delta\tau$. Since e cannot be allowed to exceed E_{min}, where E_{min} is the minimum energy barrier observed so far in the current state, it becomes

$$\mathrm{e(t)} = \min\left[\mathrm{E_{min}}, \frac{\ln(\mathrm{t}\nu^*)}{\beta_{\mathrm{High}}}\right]. \tag{8}$$

By defining e this way, the stop time can be constructed by finding the the time $\Delta\tau$ when $\mathrm{e}(t_{High,1}+\tau_{prev}+\Delta\tau) = \mathrm{E}(t_{High,1}+\Delta\tau)$. It can be shown that this reformulated equation converges at a stop time given by

$$\mathrm{t_{High,Stop}} = \mathrm{t_{High}} \exp\left[(\mathrm{E_i} - \mathrm{E_{min}})(\beta_{\mathrm{Low}} - \beta_{\mathrm{High}})\right]. \tag{9}$$

For a generic system used to define B_{TAD} in Eqn. 7, an upper bound on the ratio, R_{new}, between the total MD time required by the traditional TAD procedure and that by LFP-TAD can be written as

$$R_{new} = \left[\frac{\nu^*}{(\Sigma_i k_{i,Low})}\right]^{\beta_{High}/\beta_{Low}-1} \Gamma\left(1+\frac{\beta_{high}}{\beta_{Low}}\right) exp\left[E_{min}(\beta_{Low}-\beta_{High})\right]. \tag{10}$$

Although the LFP-TAD procedure cannot achieve the full kMC-like synthetic mode efficiency available for highly repetitive events inside a superbasin, the method does have the potential to provide a near-R_{new} improvement upon the traditional TAD method during the first few revisits to a state (18). For this reason, a modern TAD implementation generally incorporates the LFP-TAD stop-time definition until one or more events have been seen enough times that the synthetic mode performance becomes more advantageous.

It is also interesting to recognize that during the first few revisits, the LFP-TAD method is actually equivalent to the *sweeping* of time in synthetic mode, in the following sense. The LFP-TAD does not require the direct computation of a pre-exponential factor to define the low-temperature rate, relying instead on extrapolated times. However, if the system is perfectly harmonic (as assumed in TAD), so that the temperature-extrapolated times are statistically equivalent to the synthetic-mode kMC times, the amount of high-temperature MD time needed to accept each new event is equivalent in the two methods up until the time at which the slope of the sweeping synthetic stop line saturates at the minimum-barrier value E_{min}. For revisits after that, the boost of the synthetic mode sweeping method continues to increase (see **Fig. 2**), while the LFP-TAD boost becomes constant at $\exp[E_{min}(\beta_{Low} - \beta_{High})]$. While this makes the synthetic mode more powerful for large numbers of revisits, this minimum-barrier limiting boost characteristic of the LFP-TAD method turns out to be a valuable concept that can be exploited in other ways, as we discuss next.

4.2. Minimum-barrier TAD

As just discussed, in the LFP-TAD approach, once the increasingly negative slope of the stop line reaches the value of the minimum barrier for escape from the state, its slope is fixed from then on. Graphically, the stop criterion for all future revisits is found by drawing a line, with slope $-E_{min}$, that connects the low-temperature event time with the high-temperature time line. For example, this stop time would correspond to the hollow point in **Fig. 1a** if it were hypothetically true that the $E(t_{High,1} + \Delta t)$ shown happen to be E_{min}.

As pointed out in (18, 38), that the stop time can be defined in this way has some interesting consequences. While E_{min} is automatically discovered in LFP-TAD after revisiting a state multiple times, the E_{min}-based stop criterion can be employed from the very first visit if the minimum barrier is somehow known in advance. We are thus free to consider ways to determine E_{min} externally and supply that value to the TAD algorithm. We refer to this as minimum-barrier TAD (Emin-TAD) (18, 38).

In Emin-TAD, the boost factor is a simple function of T_{High}, T_{Low}, and E_{min}. The direct dependence on ν_{min} and δ has been eliminated (although an indirect dependence may remain, depending on how E_{min} is determined). We also note that the boost factor is a constant; if the minimum barrier remains unchanged as the system size N is increased, then for a short-ranged interatomic potential, Emin-TAD becomes an N-scaling computational method, the same as direct MD. One can imagine various ways to determine, or approximate, the minimum barrier for escape from a state. For example, one can use physical intuition or prior results in similar materials. A particularly powerful approach is to take the lowest barrier found in a set of saddle searches, as discussed in the next section.

Finally, we mention that one can even employ a "gambling" approach, as was done in a study by Montalenti et al(38) of "surface shuffling" behavior on a perfect Ag(100) surface. A defect-free metal surface, left alone, will occasionally create a Frenkel pair (an adatom and a surface vacancy), usually via a two-atom exchange event. Before the adatom and vacancy annihilate to recover the perfect surface state again, there can be intervening surface diffusion events such as hopping or exchange of the adatom or vacancy. This leads to a shuffling of the atom ordering in the perfect-surface state, something that in principle could be measured in an experiment with isotopically labeled atoms. An obstacle to studying

this phenomenon with TAD is that the rate constant for the initial step, the formation of the Frenkel pair from the perfect-surface state, is extremely low, with a barrier height of 1.3 eV. Using a suitably high temperature, such an event can be observed relatively easily, but the stop time required to accept the event is extremely long.

To circumvent this problem, a gamble was made. By speculating that the lowest-barrier pathway out of the perfect-surface state was indeed this 1.3 eV exchange mechanism (observed in a quick preliminary study), Emin-TAD was used, setting E_{min}=1.3 eV. This gave dramatically increased boost for the perfect-surface state. For all other states, LFP-TAD was employed. Because the study examined the long-time evolution of the Ag(100) surface at various temperatures, by the end of the study, the total amount of high-temperature time spent in the perfect-surface state during all the simulations was great enough to assess whether 1.3 eV was indeed the lowest barrier, using Eq. 9. This assertion was indeed confirmed (no lower-barrier pathway out of the perfect state was observed), thus validating the complete study. If a lower barrier had been discovered, all the TAD simulations up to that point would have had to be redone with the lower value. This novel approach enabled characterization of an interesting and previously unstudied (probably unknown) surface phenomenon on time scales ranging from milliseconds (at T_{Low}=600K) to months (at T_{Low}=300K).

4.3. Dimer TAD

Modern saddle-point search methods offer a good way to find escape pathways out of a state. The *dimer method* search approach developed by Henkelman and Jónsson (39), which requires only first derivatives of the potential, is a good example of such an approach; other methods with good efficiency have been developed as well (40, 41, 42). The dimer method searches uphill along the lowest-eigenvalue mode of the Hessian and downhill along all other directions until a stationary point is found. Henkelman and Jonsson showed that by starting a number of dimer searches at configurations randomly displaced from the basin bottom of the state, they could efficiently generate a list of saddle points that could be used to define the set of escape pathways for an adaptive kMC (AKMC) procedure (43).

In the context of TAD, by preceding the TAD simulation for each new state with a dimer-search stage, the lowest barrier height from the set of found saddles can then be supplied to the Emin-TAD algorithm. In addition to the shorter TAD stop time that results from specifying E_{min}, the saddle points found during the dimer searches can be used to eliminate some of the expensive NEB calculations for escapes encountered during the BCMD at T_{High}. This *dimer-TAD* approach was first applied to study the evolution of small interstitial clusters in MgO (11). This is a good example of a system for which dimer-TAD is powerful; the barriers are particularly high, so the stop time is significantly reduced by knowing E_{min}. This speedup allowed discovery of new mechanisms, relevant to radiation damage annealing, that occur on the time scale of seconds.

As noted above, once E_{min} is specified, the uncertainty arising from δ, and from the required specification of ν_{min}, is eliminated. In dimer-TAD, the new approximation becomes the uncertainty that the set of dimer searches did not find the lowest barrier. While this is an important issue, as it is difficult or impossible to absolutely guarantee that the lowest saddle will be found, dimer-TAD has some interesting characteristics that make it an appealing approximation. For example, if there are multiple pathways with the same lowest barrier (as often happens), it is sufficient to find just one of these, and it may even be good enough

to find only the second- or third-lowest barrier, if that barrier is very similar to the lowest barrier. Moreover, if the dimer search procedure is not reliable, so that lowest barrier is being routinely missed, it will not be long before this becomes evident, as a pathway with a lower barrier than E_{min} will show up during the Emin-TAD high-temperature BCMD. Finally, we note that the dimer searches can be completely parallelized, as they begin from independently randomized starting points, so the additional wall-clock time to implement dimer-TAD in a parallel environment need not be longer than the time to complete one dimer search.

4.4. Adaptive High Temperature

Following ideas suggested in the original TAD paper (3), additional modifications have been recently proposed to improve the serial performance of TAD by adaptively seeking an optimal high-temperature setting (19). Although it may be expected that a general increase in the T_{High}/T_{Low} ratio will produce a general increase in B_{TAD}, the relationship between the two quantities is generally non-monotonic. As the BCMD temperature is increased, an increasing number of high-barrier transitions that have a negligible probability of being accepted at the low temperature will be observed before the stop time is reached. When T_{High}/T_{Low} becomes too large, the time needed to characterize these irrelevant transitions can strongly reduce the overall performance. The adaptive TAD procedures provide a practical means to search for an optimal value of T_{High} at run-time. Algorithms allowing for adaptive T_{High}-learning can be very useful extensions to implement in practice. Such approaches have also proven useful within other simulation methods related to TAD (44); and we note that other parameters, such as the MD block time δt_{MD}, can be optimized in this way. Next, we discuss the use of parallel computing techniques within TAD.

5. PARALLEL TAD EXTENSIONS: EXPLORING LARGE SYSTEMS

Following the analysis in reference (21), the computational cost of a traditional TAD simulation scales approximately as $N^{3+1/3-\gamma}$, where $\gamma = T_{Low}/T_{High}$ and N is the number of moving atoms in the system. This scaling law emerges from multiplying $N^{2-\gamma}$ by $N^{4/3}$. As shown in Ref. (4), $N^{2-\gamma}$ is the approximate scaling of both the high-temperature MD time as well as the number of activated high-temperature events per unit of low temperature simulated time, and $N^{4/3}$ is the approximate lumped scaling of the NEB calculations typically used to find transition saddle points for each of the activated events. For the scaling of saddle point searches, the 4/3 term assumes that the work per convergence iteration scales as N, while the the number of iterations increases due to the presence of long-wavelength modes. Even if transition checks were free and localized saddle point search were implemented (20), the method would still approximately scale as $N^{2-\gamma}$, e.g., leading to $N^{3/2}$ when $T_{High} = 2T_{Low}$.

Although the serial TAD procedure can give extremely large boost factors, e.g., exceeding 10^9 in some cases, its poor scaling performance generally prevents the method from being used to explore systems with more than a few thousand moving atoms. Since many more atoms are often needed to capture long range stress fields and to avoid periodic boundary condition effects, recent efforts have been made to improve the $N^{2-\gamma}$scaling of the serial TAD procedure by exploiting spatial decomposition-based parallelization.

5.1. Spatially Parallel TAD (ParTAD)

The earliest spatial decomposition-based parallelization of TAD, developed by Shim and coworkers, is known as Spatially Parallel TAD (ParTAD) (21). The basic idea of this method is to decompose the system into overlapping spatial domains, and to concurrently run serial TAD within each of these domains on different cores. Although the reader should consult the original reference for a comprehensive explanation of the ParTAD method, here we will briefly review the sublattice algorithm at the heart of the method.

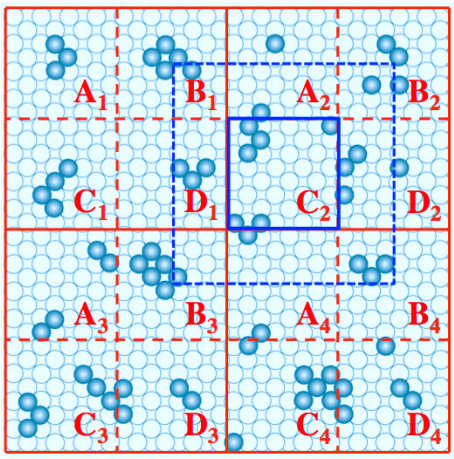

Figure 6

Illustration of the sublattice algorithm used in ParTAD for a surface simulation.

The ParTAD method uses a quasi-2D domain decomposition in combination with the semirigorous synchronous sublattice (SL) algorithm (45). When used to parallelize the kMC simulation of multilayer growth and island coarsening, the SL approach was found to predict accurate results as long as the governing parameters were chosen appropriately. As shown in **Fig. 6**, the approach works by dividing the geometry into n_p domains, where n_p is the number of processors ($n_p = 4$ in the illustration). Within each of the domains, the local geometry is then further divided into subdomains. For a two-dimensional decomposition, there will be $n_s = 4$ subdomains (A,B,C,D), as shown in the figure.

With the SL decomposition defined, the ParTAD algorithm works by allowing each of the n_p processors to perform concurrent TAD procedures in their own distinct domains. In order to ensure that neighboring domains are dynamically synchronized, this procedure is broken into n_s sequential steps. First, all processors will perform TAD in the A subdomain until they have all reached $\Delta t_{tad} \geq \tau$; the same procedure is then repeated in subdomains B through D. At the end of this 4-step cycle, the system has advanced by a time τ. As illustrated with solid and dashed blue lines in **Fig. 6** for subdomain C_2, the simulation cell used to perform TAD contains atoms from all neighboring subdomains. Referring to this illustration, all atoms within the dashed blue box are allowed to move within the TAD simulation, while only events centered within the solid blue box are allowed to be accepted at low temperature. All remaining atoms in the neighboring subdomains are held fixed to avoid unrealistic boundary effects.

Although ParTAD provides a stable and fairly scalable approach to parallelize the tra-

ditional TAD procedure, some limitations still remain. One limitation is that the method is incapable of capturing concerted events spanning multiple sub-lattice domains. For this reason, ParTAD typically would only be applied to the simulation of small-scale diffusion processes. The second limitation is that a growing system size may result in a growing probability that a very fast low barrier event is experienced somewhere in the system during a unit of MD time. Since the total ParTAD boost is highly sensitive to the lowest boost being achieved within all subdomains, the typical presence of low-barrier superbasin mechanisms is likely to dominate the long-range scaling behavior.

5.2. Extended TAD for Large Systems (XTAD)

More recently, the Extended TAD method (XTAD) was introduced with the intention of mitigating the scaling limitations present in ParTAD (22). In XTAD, full-scale parallel MD is used to perform the typical BCMD steps used in TAD. In contrast to the SL procedure discussed above, the use of parallel MD allows for large-scale concerted events to occur across the entire simulated system. When transitions are detected during quenching, XTAD applies an adaptive algorithm to produce localized simulation cells to perform NEB calculations using a reduced number of atoms.

The XTAD approach to partially mitigate the low-barrier scaling limitation is to allow for the low-temperature acceptance of multiple escape events from a given state, as long as the events are spatially independent and there is no direct evidence that a faster event is likely to prevent a subsequent event from occurring. This is enforced in practice by accepting all events with a low-temperature waiting time less than t^*_{cut}. Here, t^*_{cut} is set equal to the shortest low temperature time for that state visit, plus the shortest low-temperature event observed so far. Unfortunately, this choice of t^*_{cut} is somewhat arbitrary, as it is not always true that the rate of future events can be accurately predicted from experience in other states, especially in the early stages of a simulation.

Even with the t^*_{cut} issue in mind, the aggressive approach used by XTAD is a valuable step in the quest to study large material systems with TAD. Unfortunately, it is also important to note that the spatial scaling is still likely to break down for very large systems unless accuracy is further sacrificed. For example, given a fixed transition check frequency, Δt_{MD}, there will be a system size at which it is most likely that multiple events will occur at $t_{High} < \Delta t_{MD}$. Although XTAD has mechanisms to deal with this scenario, the high probability of observing a transition quickly means that a large number of thermalization/blackout steps will be needed to ensure that the system is in proper equilibrium prior to performing BCMD. In such a scenario, the waiting time for fastest event ever seen in the system can quickly approach zero, barring the possibility of accepting multiple events at once.

6. CURRENT STATE OF DEVELOPMENT: SPECULATIVE PARALLELIZATION OF STATES

So far, we have summarized the parallel computing strategies recently used to improve the spatial scaling of TAD, but we have yet to mention any ways that parallelism has been used to improve the TAD algorithm itself. The search for such strategies is on the leading edge of AMD development, so all such implementations are very new. In order to start exploring for the possibility of such a parallel algorithm variation, Mniszewski and coworkers recently

used a discrete event simulation (DES) framework to perform rapid sampling of parallel TAD variations (26). More specifically, they performed a parametric investigation with the most critical algorithm modifications being the spawning (parallel offloading) of transition checks, the spawning of saddle point searches, and the parallelization of the either spawned or in-line NEB saddle searches.

Although the initial DES investigation came up short of finding an extension offering both scalability and dramatically improved performance, it did inspire promising follow-on work. For example, the same approach has since been used to demonstrate that significant performance gains can be achieved by using ParRep to perform accelerated BCMD within TAD and/or speculatively running concurrent TAD processes in distinct states (to be published) (46). The latter of these promising approaches, which is at the leading edge of ongoing development, has been coined Speculatively Parallel TAD (SpecTAD).

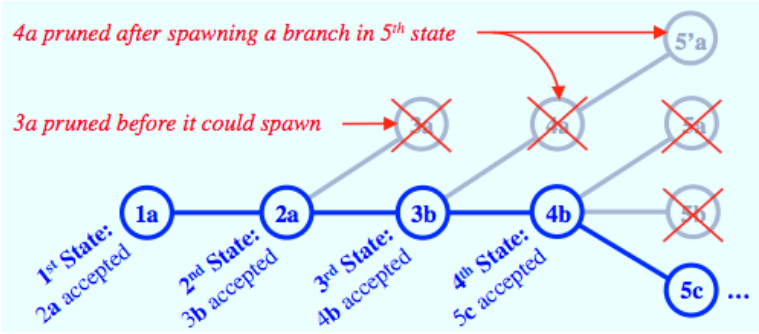

Figure 7

Illustration of SpecTAD tree representation.

The general idea of SpecTAD is to parallelize the BCMD exploration of states by spawning (offloading) a new TAD process to explore the next state as soon as its corresponding escape path is first observed. This approach can be represented through the dynamic construction of a tree, as shown in **Fig. 7**, where each node represents a TAD process exploring a single state and each edge represents a speculation. When a TAD process observes a transition that redefines $t_{Low,Short}$, that TAD process can immediately spawn a new speculative branch and prune any branches it may have spawned earlier. In **Fig. 7**, the numbering (1-5) represents the number of sequential states visited, while the lettering (a, b, c) represents the order of observed events resulting in the redefinition of $t_{Low,Short}$. As illustrated, a TAD process may end up accepting the first event (see state 1a), but it may also need to spawn and then prune multiple branches before spawning a process in the correct next state (see state 4b). Either way, it is likely that many sequential states can be explored concurrently, resulting in a parallel speedup relative to serial TAD.

6.1. The Basic SpecTAD Procedure

Although some aspects of asynchronous parallel programing can be complex in practice, the basic SpecTAD procedure can be formally described as a minor extension to the basic TAD procedure presented in Section 2.1.1. As with TAD, SpecTAD starts with the total low-temperature simulation time (t_{TAD}) equal to zero. However, there is now a separate manager process involved which must activate a free core to begin the execution of a TAD

process in an initial state. This TAD process executes the steps listed in Section 2.1.1, with the following modifications:

(S.1) The usual **Step 12** is not performed. When the TAD process has finished exploring a state, it cleans up its memory and waits for the manager to assign it a new state.

(S.2) Instead of performing **Step 10** as usual, the following is executed: If $t_{Low,j} > t_{Low,Short}$, do nothing and continue to the next step. Otherwise, set $t_{Low,Short} = t_{Low,j}$ and use Eqn. 5 to update $t_{High,Stop}$. If $t_{Low,Short}$ was updated, spawn a new TAD process to explore the state corresponding to path j by sending a spawn request to the manager process. Once the manager process activates the new child process, it will begin at the usual **Step 1**, but with $t_{TAD} = t_{Low,j}^{parent} + t_{TAD}^{parent}$.

(S.3) If the manager process receives a spawn request, it must also kill (prune) any previously activated ancestors of the initiating process.

6.2. Basic SpecTAD Performance

When applied to the basic TAD procedure, the speculative parallelization technique can be expected to use approximately $\bar{t}_{High,Stop}/\bar{t}_{High,Short}$ cores efficiently, where the $\bar{t}$ quantities represent average values. Since this ratio is highly sensitive to the ν^* parameter, it is clear that the parallel scaling is better when conservative values of ν_{min} and δ are chosen. In this sense, SpecTAD offers both a means to accelerate the serial TAD boost, as well as a means to improve the accuracy of a TAD simulation given a fixed simulation time.

It is important to note that the total number of accepted transitions can have a significant effect on parallel scaling. Clearly, the number of cores used at any given time in a SpecTAD simulation cannot exceed the maximum height of the speculative tree generated during the entire simulation. For this reason, the available performance boost ($WCT_{MD}(t_{sim})/WCT_{SpecTAD}(t_{sim})$) must typically increase until the height of the tree has exceeded some number on the order of $\bar{t}_{High,Stop}/\bar{t}_{High,Short}$. At this point, however, the boost will still continue to grow, because the time needed to reach $\bar{t}_{High,Stop}$ ($WCT(\bar{t}_{High,Stop})$) will become a smaller and smaller fraction of the total simulation time ($WCT_{SpecTAD}(t_{sim})$). As $WCT(\bar{t}_{High,Stop})/WCT_{SpecTAD}(t_{sim}) \to 0$, the boost converges to the limiting SpecTAD boost

$$B_{SpecTAD}^{Lim} \leq \frac{\bar{t}_{Low,Short}}{\bar{t}_{High,Short}}. \tag{11}$$

Although it has been shown that speculative parallelization can offer more than an order of magnitude improvement over the traditional TAD performance (46), the SpecTAD procedure just introduced is not optimized for the exploration of superbasin systems in which many repetitive low-barrier events are likely. Next, we discuss how the synthetic SpecTAD extension can be used to treat such systems.

6.3. Synthetic SpecTAD

Since the SpecTAD procedure requires little deviation from the traditional TAD procedure, it is straightforward to perform a synthetic treatment of repetitive events as long as a fairly robust approach is used to share state information across all computational resources. This can be achieved, for example, by requiring all TAD processes to occasionally synchronize their list of official states and transition rate constants with the SpecTAD manager process.

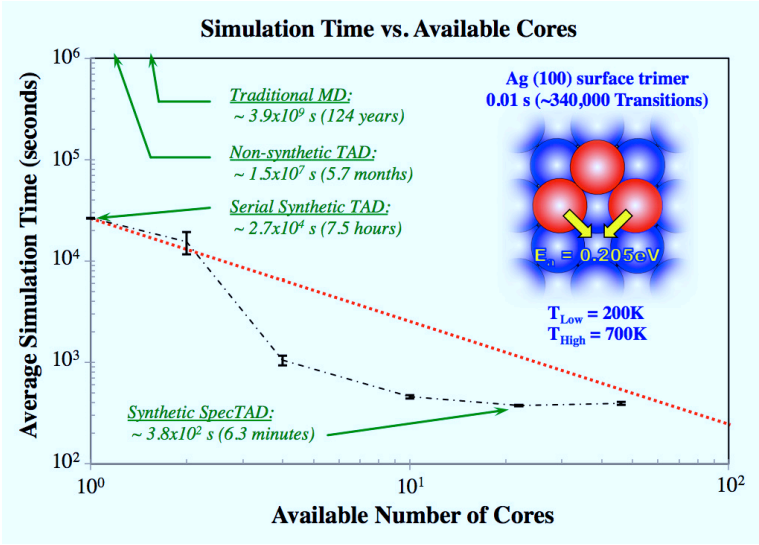

Figure 8

Synthetic SpecTAD results for 0.01s (∼340,000 transitions) of a spinning trimer on a (100) Ag surface at 200K. The dashed red line corresponds to linear scaling, implying super-linear scaling for SpecTAD. As explained in the text, the super-linear scaling is a result of an early over-accumulation of BCMD time relative to serial TAD, leading to an earlier onset of state-to-state kMC skipping. Note that the core count leaves out two manager processes used to minimize communication overhead in the current implementation.

The SpecTAD analogue of synthetic TAD works exactly as explained in Section 2.3, but with the following advantage. Upon entering a previously visited state, a typical synthetic TAD simulation will quickly check if any events are already in synthetic mode, or can be promoted into synthetic mode. In SpecTAD, the neighboring state corresponding to the synthetic event with the shortest low-temperature waiting time can be spawned immediately. This means that a very large core count, potentially much larger than $\bar{t}_{High,Stop}/\bar{t}_{High,Short}$, can be used to escape from the suberbasin, because $\bar{t}_{High,Short}$ becomes effectively negligible. Once a complete set of rate-constants is known for the enclosed set of states, every TAD process will simply draw synthetic waiting times and then spawn a new process to explore the next expected state. For the synthetic SpecTAD implementation discussed below, direct Vineyard calculations are used to determine rate constants following the first observation of each event.

Once a superbasin has been explored for long enough to accumulate significant high-temperature time, unnecessary communication can be avoided by *skipping* through any states in which a synthetic event can be accepted using BCMD time already accumulated by running past the a previous stop time to complete an MD block (Δt_{MD}). That is, a single TAD process can be allowed to move from state-to-state, without spawning a new process, until it lands in a state where additional BCMD time is needed. Although SpecTAD and TAD should both approach kMC performance in the limit that the number of intra-superbasin transitions is infinite, the parallel accumulation of high-temperature time in SpecTAD can result in a kMC-like performance much more quickly.

The advantage of synthetic SpecTAD is illustrated in **Fig. 8**, where various available core counts have been used to simulate the spinning of a Ag trimer on a (100) surface at 200K. At this temperature, the Ag trimer is effectively contained within a superbasin of 12 symmetrically identical states, with two symmetrically identical escape paths leading out of each state. For such a system, a preliminary implementation of synthetic SpecTAD exhibits super-linear parallel scaling (note that the red dashed line in **Fig. 8** denotes linear scaling).

The apparent super-linear scaling of SpecTAD has two main causes: First, the current implementation requires a computational overhead to thermalize a new trajectory before performing BCMD. Second, for simplicity, the current implementation requires that the stop time is reached by a given TAD process before its accumulated BCMD time can be used in later visits to the same state. For this reason, sometimes concurrent TAD processes simultaneously sweep from the same reference point, leading to an over-accumulation of BCMD time compared to the serial case, in which the states are explored sequentially. While this means that SpecTAD will accumulate more BCMD time than necessary during early revisits, it also means that kMC-skipping can begin sooner. When SpecTAD skips through states in this way, it avoids the thermalization overhead that can sometimes dominate the performance. We note that this thermalization overhead can be reduced in both SpecTAD and serial TAD, but such steps have yet to be implemented. By eliminating all significant overhead, SpecTAD performance should improve for low core counts, leading to a more linear scaling.

6.4. Example Application: Surface Deposition at Experimental Rates

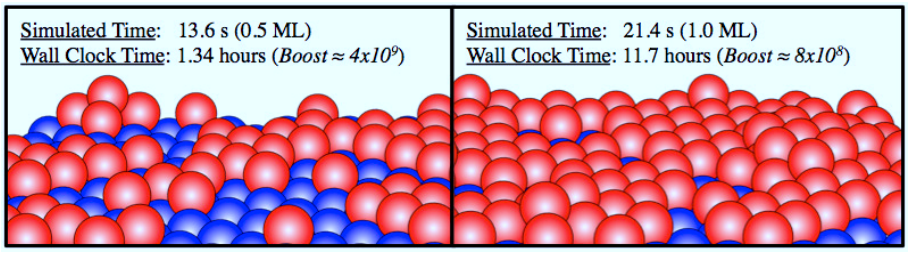

Figure 9

SpecTAD results for the deposition of Cu on a (100) Ag surface at T=77 K and a deposition rate of 0.04 ML/s. Two separate trials were performed to 1.0ML of Cu coverage; the faster of the two is shown. The initial simulation geometry was composed of seven 98-atom (100) layers of Ag, with the top four layers free to move. The simulation details were designed to match Ref. (10).

SpecTAD can offer significant performance enhancements when simulating many types of systems, but the method is particularly well suited to surface deposition simulations. Since a TAD process can speculate from a list of known escape paths, a deposition event can be added to this list by drawing a random waiting time, relative to the last deposition event, from a Poisson distribution. This speculative deposition can then be spawned at any point, and then pruned if or when it is replaced by a faster event.

As an example, we used SpecTAD to simulate the deposition of Cu on a (100) Ag surface at T=77 K, with an experimental deposition rate of 0.04 ML/s. When previously modeling the same system with synthetic TAD, Sprague et. al. needed months of wall clock time

to reach 1.0ML of Cu coverage (10). By providing 144 cores, the same was achieved by SpecTAD in 12 and 25 hours for the two trials considered (see **Fig. 9**). Although part of the speedup relative to reference (10) can be attributed to faster computational resources, much of the improved performance is clearly the result of parallel speedup. In both trials, for example, concurrent TAD processes were already exploring at least fifty to-be-accepted sequential deposition events by the time the initial process had reached its stop time. As low-barrier mechanisms became more dominant at higher ML coverages, the maximum available core count was consistently used to accelerate the synthetic mode treatment of repetitive events.

Although SpecTAD can deliver significant performance boosts over the traditional MD and TAD procedures, there is still room for improvement. For example, we estimate that more than an order of magnitude of further speedup can be obtained by using ParRep to perform BCMD within SpecTAD. This future step of SpecTAD development is still in progress, as an adaptive procedure for optimizing the allocation of computational resources is still under exploration.

7. CONCLUSION

The modern TAD method comprises a broad set of statistical mechanics-based insights that can be used to predict low-temperature evolution using high-temperature MD. Whether the system of interest contains many distinct states separated by high barriers, or it contains a collection of low-barrier superbasins, each extension has its own strengths and weaknesses. In any case, the goal is to push the simulations toward longer and longer timescales, so that new physical behaviors can be identified. Given the growing availability and reliability of parallel computing systems, the leading edge of TAD development is now focusing on the incorporation of advanced parallel computing concepts. Although there is room for further improvement, this trend has already led to significant gains in performance, and will undoubtably lead to new scientific understandings.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The authors are grateful for help from Sriram Swaminarayan and Stephan Eidenbenz, and for helpful discussions with Francesco Montalenti, Jacques Amar and Yunsic Shim. This work was supported by the United States Department of Energy (US DOE) Office of Science, Office of Basic Energy Sciences, Division of Materials Sciences and Engineering. Los Alamos National Laboratory is operated by Los Alamos National Security, LLC, for the National Nuclear Security administration of the US DOE under contract DE-AC52-06NA25396.

LITERATURE CITED

1. Voter AF. 1997. Hyperdynamics: Accelerated molecular dynamics of infrequent events. *Phys. Rev. Lett.* 78:3908–3911

2. Voter AF. 1998. Parallel replica method for dynamics of infrequent events. *Phys. Rev. B* 57:R13985
3. Sørensen MR, Voter AF. 2000. Temperature-accelerated dynamics for simulation of infrequent events. *J. Chem. Phys.* 112:9599–9606
4. Voter AF, Montalenti F, Germann TC. 2002. Extending the time scale in atomistic simulation of materials. *Annu. Rev. Mater. Res.* 32:321–346
5. Uberuaga BP, Montalenti F, Germann TC, Voter AF. 2005. Accelerated molecular dynamics methods. In *Handbook of Materials Modeling*. Springer, 629–648
6. Uberuaga B, Voter A. 2007. Accelerated molecular dynamics methods. In *Radiation Effects in Solids*, eds. KE Sickafus, EA Kotomin, BP Uberuaga, vol. 235 of *NATO Science Series*. Springer Netherlands, 25–43
7. Perez D, Uberuaga BP, Shim Y, Amar JG, Voter AF. 2009. Accelerated molecular dynamics methods: introduction and recent developments. *Annu. Rep. Comp. Chem.* 5:79–98
8. Montalenti F, Voter A. 2001. Applying accelerated molecular dynamics to crystal growth. *Phys. Status Solidi B* 226:21–27
9. Montalenti F, Sørensen MR, Voter AF. 2001. Closing the gap between experiment and theory: Crystal growth by temperature accelerated dynamics. *Phys. Rev. Lett.* 87:126101
10. Sprague JA, Montalenti F, Uberuaga BP, Kress JD, Voter AF. 2002. Simulation of growth of Cu on Ag(001) at experimental deposition rates. *Phys. Rev. B* 66:205415
11. Uberuaga BP, Smith R, Cleave AR, Montalenti F, Henkelman G, et al. 2004. Structure and mobility of defects formed from collision cascades in MgO. *Phys. Rev. Lett.* 92:115505
12. Cogoni M, Uberuaga BP, Voter AF, Colombo L. 2005. Diffusion of small self-interstitial clusters in silicon: Temperature-accelerated tight-binding molecular dynamics simulations. *Phys. Rev. B* 71:121203
13. Cogoni M, Mattoni A, Uberuaga BP, Voter AF, Colombo L. 2005. Atomistic study of the dissolution of small boron interstitial clusters in c-Si. *Appl. Phys. Lett.* 87
14. Uberuaga BP, Valone SM, Baskes M. 2007. Accelerated dynamics study of vacancy mobility in δ-plutonium. *J. Alloy. Compd.* 444:314–319
15. Shim Y, Borovikov V, Uberuaga BP, Voter AF, Amar JG. 2008. Vacancy formation and strain in low-temperature Cu/Cu(100) growth. *Phys. Rev. Lett.* 101:116101
16. Ichinomiya T, Uberuaga BP, Sickafus KE, Nishiura Y, Itakura M, et al. 2009. Temperature accelerated dynamics study of migration process of oxygen defects in UO_2. *J. Nucl. Mater.* 384:315 – 321
17. Uberuaga BP, Stuart SJ, Windl W, Masquelier MP, Voter AF. 2012. Fullerene and graphene formation from carbon nanotube fragments. *Comp. Theor. Chem.* 987:115 – 121
18. Montalenti F, Voter AF. 2002. Exploiting past visits or minimum-barrier knowledge to gain further boost in the temperature-accelerated dynamics method. *J. Chem. Phys.* 116:4819–4828
19. Shim Y, Amar JG. 2011. Adaptive temperature-accelerated dynamics. *J. Chem. Phys.* 134:054127
20. Shim Y, Callahan NB, Amar JG. 2013. Localized saddle-point search and application to temperature-accelerated dynamics. *J. Chem. Phys.* 138
21. Shim Y, Amar JG, Uberuaga BP, Voter AF. 2007. Reaching extended length scales and time scales in atomistic simulations via spatially parallel temperature-accelerated dynamics. *Phys. Rev. B* 76:205439
22. Bochenkov V, Suetin N, Shankar S. 2014. Extended temperature-accelerated dynamics: Enabling long-time full-scale modeling of large rare-event systems. *J. Chem. Phys.* 141
23. Vineyard GH. 1957. Frequency factors and isotope effects in solid state rate processes. *J. Phys. Chem. Solids* 3:121–127
24. Chekmarev S, Krivov S. 1998. Confinement of the molecular dynamics trajectory to a specified catchment area on the potential surface. *Chem. Phys. Lett.* 287:719–724
25. Aristoff D, Lelivre T. 2014. Mathematical analysis of temperature accelerated dynamics. *Mul-*

tiscale Model. Sim. 12:290–317
26. Mniszewski SM, Junghans C, Voter AF, Perez D, Eidenbenz SJ. 2015. TADSim: Discrete event-based performance prediction for temperature-accelerated dynamics. *ACM Trans. Model. Comput. Simul.* 25:15:1–15:26
27. Henkelman G, Uberuaga BP, Jónsson H. 2000. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.* 113:9901–9904
28. Bortz A, Kalos M, Lebowitz J. 1975. A new algorithm for Monte Carlo simulation of ising spin systems. *J. Comp. Phys.* 17:10 – 18
29. Voter AF. 2007. Introduction to the kinetic Monte Carlo method. In *Radiat. Eff. Sol.* Springer, 1–23
30. Uberuaga BP, Pilania G. 2015. Effect of cation ordering on oxygen vacancy diffusion pathways in double perovskites. *Chem. Mater.* 27:5020–5026
31. Uberuaga BP, Perriot R. 2015. Insights into dynamic processes of cations in pyrochlores and other complex oxides. *Phys. Chem. Chem. Phys.*
32. Uberuaga BP, Andersson DA, Stanek CR. 2013. Defect behavior in oxides: Insights from modern atomistic simulation methods. *Curr. Opin. Solid. St. M.* 17:249–256
33. Uberuaga BP, Vernon LJ. 2013. Interstitial and vacancy mediated transport mechanisms in perovskites: A comparison of chemistry and potentials. *Solid State Ion.* 253:18–26
34. Li YH, Uberuaga BP, Jiang C, Choudhury S, Valdez JA, et al. 2012. Role of antisite disorder on preamorphization swelling in titanate pyrochlores. *Phys. Rev. Lett.* 108:195504
35. Uberuaga BP, Vernon LJ, Martinez E, Voter AF. 2015. The relationship between grain boundary structure, defect mobility, and grain boundary sink efficiency. *Sci. Rep.* 5
36. Bai XM, Voter AF, Hoagland RG, Nastasi M, Uberuaga BP. 2010. Efficient annealing of radiation damage near grain boundaries via interstitial emission. *Science* 327:1631–1634
37. Bai XM, Uberuaga B. 2013. The influence of grain boundaries on radiation-induced point defect production in materials: A review of atomistic studies. *JOM-J. Min. Met. Mat. S.* 65:360–373
38. Montalenti F, Voter AF, Ferrando R. 2002. Spontaneous atomic shuffle in flat terraces: Ag(100). *Phys. Rev. B* 66:205404
39. Henkelman G, Jónsson H. 1999. A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. *J. Chem. Phys.* 111:7010–7022
40. Munro LJ, Wales DJ. 1999. Defect migration in crystalline silicon. *Phys. Rev. B* 59:3969–3980
41. Barkema GT, Mousseau N. 1996. Event-based relaxation of continuous disordered systems. *Phys. Rev. Lett.* 77:4358–4361
42. Béland LK, Brommer P, El-Mellouhi F, Joly JF, Mousseau N. 2011. Kinetic activation-relaxation technique. *Phys. Rev. E* 84:046704
43. Xu L, Henkelman G. 2008. Adaptive kinetic monte carlo for first-principles accelerated dynamics. *J. Chem. Phys.* 129:114104
44. Divi S, Chatterjee A. 2014. Accelerating rare events while overcoming the low-barrier problem using a temperature program. *J. Chem. Phys.* 140:184115
45. Shim Y, Amar JG. 2005. Semirigorous synchronous sublattice algorithm for parallel kinetic monte carlo simulations of thin film growth. *Phys. Rev. B* 71:125432
46. Zamora RJ, Voter AF, Perez D, Santhi N, Mniszewski SM, et al. To be published. Discrete event performance prediction of speculatively parallel TAD