Accelerating Molecular Dynamics Simulations

Timothy C. Germann* and Arthur F. Voter**

* Applied Physics and ** Theoretical Divisons,
Los Alamos National Laboratory, Los Alamos, NM 87545
{tcg, afv}@lanl.gov

ABSTRACT

In contrast with the converging length scales of atomistic simulations and experimental nanoscience, large time scale discrepancies still remain, due to the time-scale limitations of molecular dynamics. We briefly review two recently developed methods, derived from transition state theory, for accelerating molecular dynamics simulations of infrequent-event processes. These techniques, parallel replica dynamics and hyperdynamics, can reach simulation times several orders of magnitude longer than direct molecular dynamics while retaining full atomistic detail.

Keywords: Infrequent events, accelerated dynamics, hyperdynamics, parallel replica dynamics

1 INTRODUCTION

Atomistic simulation techniques are playing an increasingly important role as the continually increasing computer power on the one hand, and on the other the nanotechnology drive (exemplified by this conference) which is shrinking device sizes into the nanometer regime, are together eliminating the "length-scale gap" between simulation and experiment. Indeed, micron-sized features can now be directly simulated using classical molecular dynamics (MD) with a billion (10^9) or more atoms. However, the timescale of such simulations is still constrained by the need to resolve individual atomic vibrations, limiting the incremental timestep to a few femtoseconds, at most. Millions of timesteps are thus required to reach even a *nanosecond* timescale, still several orders of magnitude short of the typical millisecond switching time of MEMS devices, not to mention the seconds to hours over which vapor-deposited thin film growth and dopant diffusion take place. The importance of extending the time scale of atomistic simulations to accurately capture long-time dynamical processes is further described elsewhere in these proceedings [1].

Recently, several new approaches have been developed which can extend the accessible time scale by orders of magnitude relative to direct molecular dynamics, while retaining full atomistic detail for systems in which the long-time dynamical evolution is characterized by a sequence of infrequent activated, or diffusive, events [2]. As shown schematically in Fig. 1, a typical trajectory may wander around in a potential basin for many (perhaps a huge number of) vibrational periods. Eventually, when enough energy has been localized into a reactive mode, the trajectory passes through a dividing surface, entering another state. During this brief period of excitation, it may recross the dividing surface, but shortly thereafter it settles into the new state (or the original state if it recrosses), beginning a new session of vibrational wandering, with no memory of how it arrived in that state. Although it may never again visit this state, and "sees" only a single dividing surface as it exits, it nonetheless chooses an escape direction (relative to other possible escape directions) with the correct probability. This property of any infrequent-event system, that a trajectory will automatically choose an appropriate escape path with no prior information, is the basic concept exploited in the accelerated dynamics methods. The key is to coax the trajectory into making this choice more quickly without corrupting the relative escape probabilities (i.e., the rate constants for escape to various states). Two such techniques are briefly reviewed here, parallel replica dynamics [3] and hyperdynamics [4], [5]; a third approach, temperature accelerated dynamics (TAD) [6], is reviewed elsewhere in this volume [7].

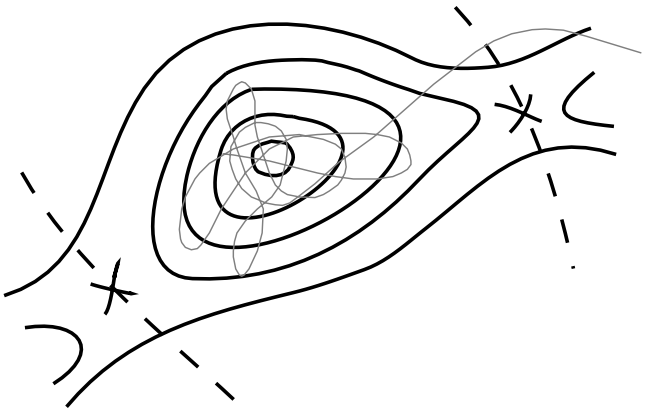

Figure 1: An infrequent-event system.

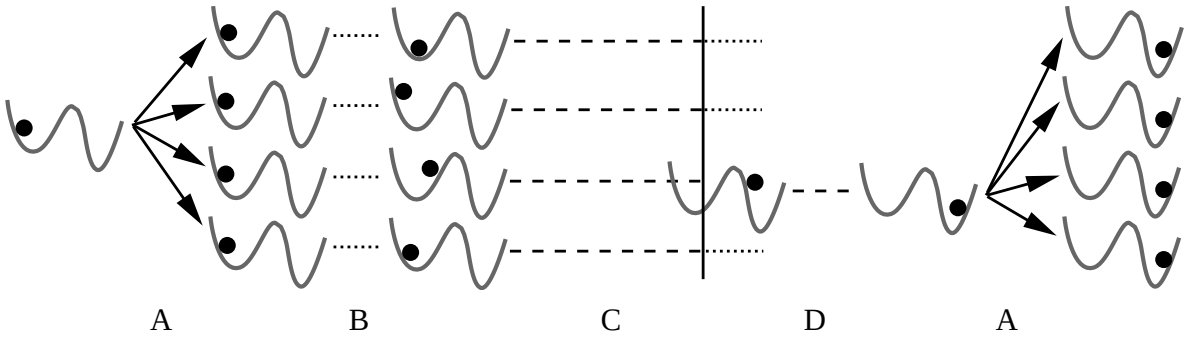

Figure 2: Schematic illustration of the parallel replica method. The four steps, described in the text, are (A) replication of the system into M copies, (B) dephasing of the replicas, (C) independent trajectories until a transition is detected in any of the replicas, and (D) brief continuation of the transitioning trajectory to allow for correlated events such as recrossings or follow-on transitions to other states. The resulting configuration is then replicated, beginning the process again.

2 PARALLEL REPLICA DYNAMICS

The parallel replica method [3] is the simplest and most accurate of the accelerated dynamics techniques, with the only assumption being that of infrequent events obeying first-order kinetics (exponential decay); i.e., for any time greater than τ_{corr} after entering a state, the probability distribution function for the time of the next escape is given by

$$p(t) = k\ exp(-kt) \tag{1}$$

where k is the rate constant for escape. For example, Eq. 1 arises naturally for ergodic, chaotic exploration of an energy basin. The general approach is shown in Fig. 2. Starting with an N-atom system in a particular state (basin), the entire system is replicated on each of M available parallel or distributed processors. After a short dephasing stage (Δt_{deph}) during which momenta are periodically randomized to eliminate correlations between replicas, each processor carries out an independent constant-temperature MD trajectory for the entire N-atom system, thus exploring phase space within the particular basin M times faster than a single trajectory would. Whenever a transition is detected on any processor, all processors are alerted to stop. (One way to recognize transitions is to periodically, say every 10 or 100 timesteps, perform a steepest descent or conjugate gradient quench of the current configuration until it can be ascertained whether this descent is leading towards the original basin or to a different one.) The simulation clock is advanced by the accumulated trajectory time summed over all replicas, i.e., the total time spent exploring phase space within the basin until an escape pathway is found.

The parallel replica method also correctly accounts for correlated dynamical events, in which a trajectory quickly recrosses back to the original state, or skips on to a different state, before equilibrating in the new basin. This is accomplished by allowing the transitioning trajectory to continue on its processor for a further amount of time Δt_{corr}, during which recrossings or follow-on events may occur. The simulation clock is then advanced by Δt_{corr}, the final state is replicated on all processors, and the whole process is restarted. This overall procedure then gives exact state-to-state dynamical evolution, because the escape times obey the correct probability distribution and relative probabilities for different escape paths, and the correlated dynamical events are properly accounted for.

The efficiency of the method is limited by both the dephasing stage, which does not advance the system clock, and the correlated event stage, during which only one processor accumulates time. (This is illustrated schematically in Fig. 2, where dashed line trajectories advance the simulation clock but dotted line trajectories do not.) Thus, the overall efficiency will be high when

$$\tau_{rxn}/M \gg \Delta t_{deph} + \Delta t_{corr}, \tag{2}$$

where τ_{rxn} is the typical event time, although various tricks can lessen this requirement at the expense of somewhat more complicated programming [2], [3].

Parallel replica dynamics has the advantage of being fairly simple to program, with very few "knobs" to adjust – Δt_{deph} and Δt_{corr}, which can be conservatively set at a few ps for most systems. The replica trajectories can themselves be accelerated trajectories, using, for instance, hyperdynamics (discussed next) on each processor, resulting in multiplicative boost factors [8]. As multiprocessing environments become more ubiquitous, with more processors within a node or even on a chip, and loosely linked Beowulf clusters of such nodes, parallel replica dynamics will become an increasingly important simulation tool. Already, Birner et al [10] have used parallel replica dynamics with up to 32 processors to study the growth of silicon interstitial clusters, scanning for trapping and diffusion mechanisms much more quickly than they could have with conventional MD. Parallel replica dynamics is also being used to reduce the accessible strain rate in studies of the tensile loading of carbon nanotubes [11].

3 HYPERDYNAMICS

In hyperdynamics [4], [5], one modifies the potential surface $V(\mathbf{r})$ of the system by adding a nonnegative *bias* potential $\Delta V(\mathbf{r})$, as shown schematically in Fig. 3. A constant-temperature trajectory is then propagated on this modified surface. Assuming that certain conditions hold, namely that the system obey transition state theory (i.e., no correlated events) on both the biased and unbiased potential surfaces and that the bias potential vanish at all of the dividing surfaces, then this trajectory, while relatively meaningless on vibrational time scales, evolves *correctly* from state to state at an accelerated pace. The evolution is correct in the sense that the probability of observing any particular sequence of states is the same for a simulation on the biased potential as for the unbiased potential; it can be readily shown [4] that the bias potential does not change the relative TST rates for different escape paths from a given state.

An important result of hyperdynamics is that the accelerated time is easily estimated as the simulation proceeds. For a regular MD trajectory, the time advances at each integration step by Δt_{MD}, the MD time step (e.g., $\sim$1 fs). In hyperdynamics, the time advance at each step is Δt_{MD} multiplied by an instantaneous boost factor, the inverse Boltzmann factor for the bias potential at that point, so that the total time after n integration steps is

$$t_{hyper} = \sum_{j=1}^{n} \Delta t_{MD}\ exp[\Delta V(\mathbf{r}(t_j))/k_B T], \qquad (3)$$

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

$$t_{hyper}/t_{MD} = \langle exp[\Delta V(\mathbf{r})/k_B T]\rangle_b \qquad (4)$$

(the average is over the trajectory on the biased potential), divided by the extra computational cost of calculating the bias potential and its derivatives. (This computational overhead is typically expressed in the number of MD force evaluations per timestep, which is one for ordinary MD.)

The ideal bias potential should give a large boost factor, should have low computational overhead (although more overhead is acceptable if the boost factor is very high), and should to a good approximation meet the requirements given above. This is very challenging, since we want, as much as possible, to avoid utilizing any prior knowledge of the dividing surfaces or the available escape paths. Two distinct classes of bias potentials have been used to date: "flat" biases and Hessian-based ones.

The simplest, and cheapest, bias potential is the flat one proposed by Steiner et al [13]. In this approach, one specifies in advance a fixed energy V_{flat}, which is higher than the basin minimum but lower than the energy of the lowest saddle point. (In cases where one does not have advance knowledge about the system under investigation, V_{flat} can be slowly ramped up as one becomes more confident about a lower bound on any saddle points exiting the current basin [8].) As the trajectory propagates, if the potential energy $V(\mathbf{r})$ is greater than V_{flat}, the bias potential is zero. However, if $V(\mathbf{r})$ is below V_{flat}, the biased surface becomes flat ($V + \Delta V = V_{flat}$), and the trajectory skates across the ice on the pond. When the bias is turned on, the energy derivatives need not be calculated - the total force is zero, and consequently this approach actually *reduces* the average number of force evaluations per timestep compared with ordinary MD! They demonstrated this approach for a 2-D model system and for a Lennard-Jones realization of adatom diffusion on a small fcc(100) substrate (9 moving atoms).

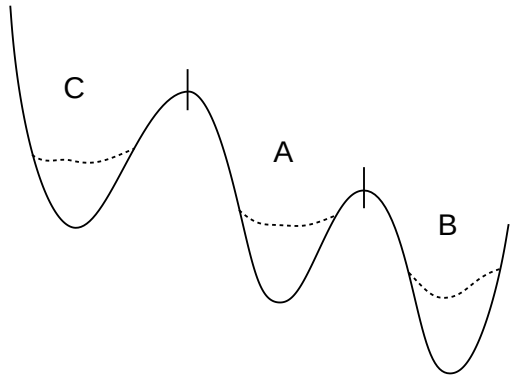

Figure 3: Schematic illustration of the hyperdynamics method. A bias potential ($\Delta V(\mathbf{r})$), is added to the original potential ($V(\mathbf{r})$, solid line). Provided that $\Delta V(\mathbf{r})$ meets certain conditions, primarily that it be zero at the dividing surfaces between states, 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.

Because the average energy (relative to the basin minimum) of a system grows linearly with the number of degrees of freedom, a flat bias potential set below the lowest barrier will give vanishing boost for large systems; the system will only rarely have a potential energy lower than V_{flat}. Indeed, to obtain a substantial boost for their small fcc(100) terrace diffusion system, Steiner et al [13] had to put V_{flat} well above the lowest saddle point, requiring the calculation of a "numerator correction" (involving prior knowledge of the dividing surface) to obtain accurate hopping rates. This is an example of how our intuition can mislead us if we visualize a low-dimensional system. The thermal contribution to the total potential energy in a large system ($\sim(3N-3)k_B T/2$) is almost always greater than the lowest barrier, but the system only goes "over" the barrier when it has enough energy in the particular mode corresponding to the reaction coordinate. For similar reasoning, however, for a system with many dimensions, it becomes safe to set V_{flat} somewhat above the lowest barrier [13], since the transition state subsystem has

an average potential energy of $(3N-4)k_BT/2$. This approach has not been explored in detail, although Choudhary and Clancy [15] have shown in annealing an amorphous silicon system that such a flat bias can be effective for gaining a small but useful boost factor with little programming investment, even for a system with hundreds of atoms.

Greater boosts can be achieved, albeit at the expense of a greater computational overhead, if some local properties of $V(\mathbf{r})$ are taken into account. For example, the bias potentials in the original hyperdynamics paper [4] were based on the lowest eigenvalue (ϵ_1) of the Hessian (the matrix of derivatives $\partial^2 V/\partial x_i \partial x_j$). The bias potential was made positive for regions where $\epsilon_1 > 0$, and zero elsewhere, exploiting the fact that ϵ_1 is positive near the bottom of a basin and negative at saddle points. For a periodic 2-D model system, this gave substantial boosts (in the thousands when k_BT was $\sim$1/20 of the barrier height) and excellent accuracy, even when some recrossings were present. A Ni/Ni(100) adatom diffusion problem at T = 500 K yielded boost factors of 40 to 400 for various bias potentials (only 9 moving atoms were included, to reduce the cost of diagonalizing the Hessian matrix).

However, a bias potential requiring a diagonalization of the full $3N$-dimensional Hessian at every time step becomes prohibitively expensive as N is increased beyond a few tens of atoms. Moreover, except for certain very simple systems, the fraction of configuration space with $\epsilon_1 > 0$ decreases as N increases, so that the boost vanishes for large systems. For this reason, the bias potential and its calculation were modified in a couple of ways to address these issues [5]. First, a more sophisticated form for the bias potential was developed, which goes smoothly to zero in the proximity of any ridgetop. Following the proposal of Sevick et al [12], the system is assumed to be at a ridgetop when $\epsilon_1 < 0$ and $g_{1p} = 0$, where g_{1p} is the projection of the gradient onto the lowest eigenvector. This definition is not guaranteed to coincide with the exact ridgetop between basins, but it is exact at a saddle point and a good approximation nearby. Second, an interative approach was developed to obtain ϵ_1 and g_{1p} and the necessary derivatives by using only first derivatives of the potential, therefore eliminating the need to diagonalize the Hessian, or even to construct it. This bias potential has been successfully applied to systems with hundreds of moving atoms, with reasonable computational overheads of 10 to 100 force evaluations per timestep, although numerical convergence issues (particularly for the derivatives of g_{1p}) remain a subject of ongoing work.

Another promising avenue of research involves the local bias approaches proposed by several groups [9], [13], [14]; in these techniques the bias potential is based on the potential energy or Hessian only in the immediate vicinity of a known defect, such as an adatom or vacancy.

4 CONCLUSIONS

Both of the techniques discussed here, as well as TAD [6], [7], show great promise for reducing and in some cases eliminating the time-scale gap between MD and experiment. The impressive boost factors which have been demonstrated thus far using the various flavors of bias potentials are quite tantalizing, and should motivate future investigations into more powerful bias potentials.

REFERENCES

[1] A. F. Voter, these proceedings.

[2] A. F. Voter, F. Montalenti, and T.C. Germann, "Extending the Time Scale in Atomistic Simulation of Materials," Annu. Rev. Mater. Res., in press (2002).

[3] A. F. Voter, "Parallel replica method for dynamics of infrequent events," Phys. Rev. B **57**, 13985 (1998).

[4] A. F. Voter, "A method for accelerating the molecular dynamics simulation of infrequent events," J. Chem. Phys. **106**, 4665 (1997).

[5] A. F. Voter, "Hyperdynamics: Accelerated molecular dynamics of infrequent events," Phys. Rev. Lett. **78**, 3908 (1997).

[6] M. R. Sørensen and A. F. Voter, "Temperature-accelerated dynamics for simulation of infrequent events," J. Chem. Phys. **112**, 9599 (2000).

[7] F. Montalenti, "Temperature accelerated dynamics: introduction and application to crystal growth," these proceedings.

[8] A. F. Voter and T. C. Germann, "Accelerating the Dynamics of Infrequent Events: Combining Hyperdynamics and Parallel Replica Dynamics to Treat Epitaxial Layer Growth," Mat. Res. Soc. Symp. Proc. **528**, 221 (1998).

[9] W. G. Rudd and A. F. Voter, "Bias Potentials for Hyperdynamics Simulations," Mat. Res. Soc. Symp. Proc. **538**, 485 (1999).

[10] S. Birner, J. Kim, D. A. Richie, J. W. Wilkins, A. F. Voter, and T. Lenosky, "Accelerated dynamics simulations of interstitial-cluster growth," Solid State Comm. **120**, 279 (2001).

[11] B. P. Uberuaga, S. J. Stuart, and A. F. Voter, "Accelerated Dynamics Simulations of Nanotubes," these proceedings.

[12] E. M. Sevick, A. T. Bell, and D. N. Theodorou, "A chain of states method for investigating infrequent event processes occuring in multistate, multidimensional systems," J. Chem. Phys. **98**, 3196 (1993).

[13] M. M. Steiner, P.-A. Genilloud, and J. W. Wilkins, "Simple bias potential for boosting molecular dynamics with the hyperdynamics scheme," Phys. Rev. B **57**, 10236 (1998).

[14] J.-C. Wang, S. Pal and K. A. Fichthorn, "Accelerated molecular dynamics of rare events using the local boost method," Phys. Rev. B **63**, 085403 (2001).

[15] D. Choudhary and P. Clancy, "Novel Methods for Studying the Evolution of Silicon from a Quenched State using Accelerated Molecular Dynamics Schemes," these proceedings.