Introduction Accelerated Molecular Dynamics

Danny Perez

Theoretical Division T-1
Los Alamos National Laboratory
Los Alamos, NM

UBC, July 2017

Plan

Acknowledgements

- This material is based on methods proposed by Arthur Voter at LANL starting about 20 years ago.
- They were further developed, implemented, and used by many people over the years.
- Core LANL AMD team: A.F. Voter, D. Perez, B. Uberuaga, R. Zamora, and a team of collaborator, postdocs and students.

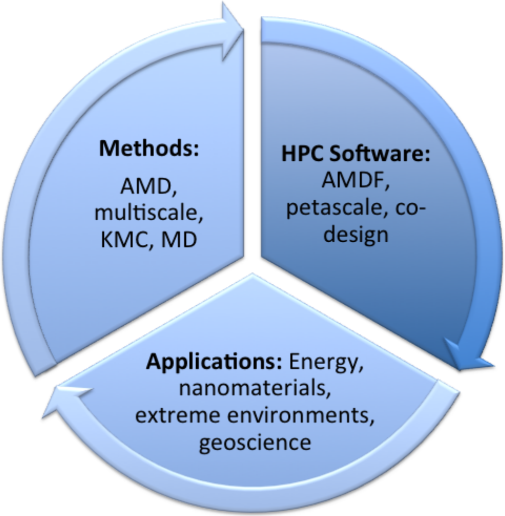
Funding:

- US Department of Energy, Office of Basic Energy Sciences, ASCR, FES, NE
- LANL Laboratory-Directed Research and Development
- US Exascale Computing Project

Los Alamos National Laboratory (LANL)

Los Alamos National Laboratory (LANL)

My Research



Methods:
AMD,
multiscale,
KMC, MD

HPC Software:
AMDF,
petascale, co-design

Applications: Energy,
nanomaterials,
extreme environments,
geoscience

Today

Goal: give an overview of AMD methods and of how they are used in materials science.

I will mostly discuss the "base" methods, not the bleeding edge. I will also concentrate of work carried out at LANL.

Please:

- stop me at any time if I am going too slow, too fast, or if I am not making sense.
- ask questions as we go (don’t wait till the end).
- let me know what interests you.

Plan

1. Molecular Dynamics 101
   - Motivation

Molecular dynamics: What?

Molecular Dynamics is the *numerical solution* of the *equations of motion* of a set of atoms, given an *interatomic potential* V.

E.g., in the microcanonical ensemble (NVE), solve Newton's equations of motion:

$$\begin{aligned} \frac{\partial x_i}{\partial t} &= \frac{p_i}{m} \\ \frac{\partial p_i}{\partial t} &= -\frac{\partial V}{\partial x_i} = f_i \end{aligned}$$

Molecular dynamics: Why?

MD produces $[\mathbf{x}(t), \mathbf{p}(\mathbf{t})]$ trajectories in full atomistic detail.

MD:

- is formally simple
- is classically “exact” (for to a given V)
- naturally handles complexity; the systems does the right thing at the right time *on its own*
- can be used to compute “any” (atomistic) thermodynamical or dynamical property

Molecular dynamics: Why?

With a good potential, we can use MD to interpret experiments, make predictions, and gain physical insight at the atomic level.

MD simulations are experiments in themselves. They can be used to:

- estimate physical quantities
- parameterize higher-level models
- test higher-level models
- simply "see what happens"

Often, MD simulations show unexpected and surprising results and drive the development of new theories.

Molecular Dynamics

This sounds almost too good... So, what’s the catch?

What do we need from MD?

Space:

- Point defects : nm ($\sim 10^2$ atoms)
- Nanostructure: tens of nm ($\sim 10^5$ atoms)
- Microstructure: μm-cm ($> 10^9$ atoms)

Time:

- Vibrations : fs-ps ($\sim 10^2$ timesteps)
- “Unit” transitions: ns-μs ($> 10^6$ timesteps)
- Microstructural evolution : ms-years ($> 10^{12}$ timesteps)

Plan

2. The timescale problem of MD
   - Space and Time Limitations of MD
   - Accelerating the Dynamics

Molecular Dynamics: cost and limits

MD with empirical potentials is relatively cheap:

- LJ potential: $\sim 2\mu$s/atom/timestep
- EAM potential: $\sim 5\mu$s/atom/timestep
- MEAM potential: $\sim 250\mu$s/atom/timestep
- Ab Initio: $\sim$ min/atom/timestep

E.g., for 1000 EAM atoms: $\sim$ 30 ns of MD per day on a single core.

For short-range potentials, MD is linear scaling : the cost of a time-step is proportional to the number of atoms This means that, in principle, you can do 30 ps per day on 10^6 atoms instead.

Molecular Dynamics: cost and limits

What about parallel computers?

MD is inherently timewise serial: you can only distribute work within one timestep.

Parallelization is usually achieved through spatial decomposition: each processor handles atoms in a given region of space.

The current scope of MD simulations

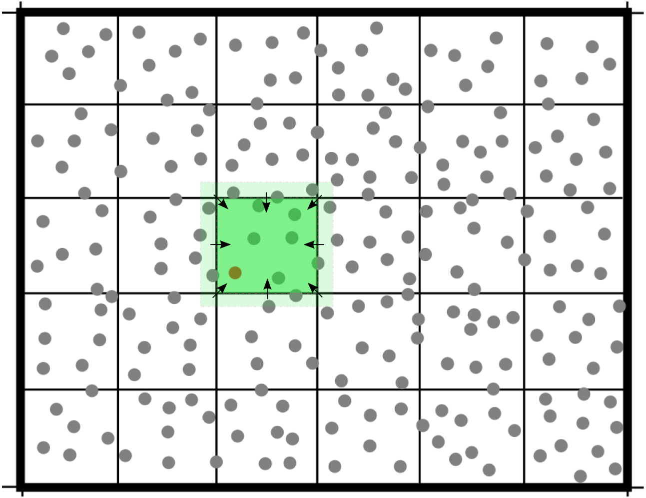


Figure: Spatial decomposition: each cell is handled by a different processor.

Molecular Dynamics: cost and limits

- Weak scaling (problem size scales with number of cores) is easy: can do very large systems (trillions of atoms)
- Strong scaling (scaling at constant problem size) is difficult: **cannot reach long times ($\gg \mu$s)**

Molecular Dynamics: Spatial decomposition

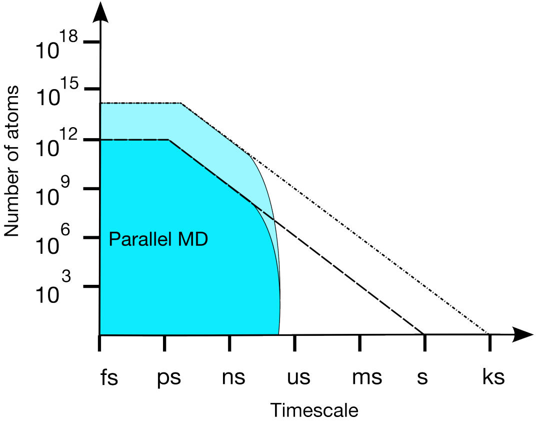
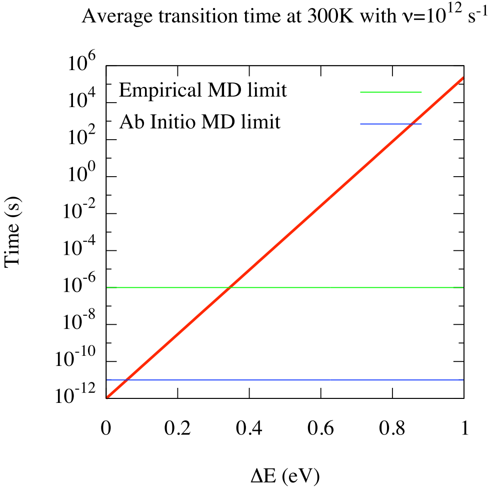


Figure: Scope of empirical MD simulations given a peta- (solid) or exa-flop (shaded) computer for a few days.

The Timescale Problem

When energy barriers $\Delta E \gg k_B T$, MD will **not** provide relevant information about long time, thermally activated, behavior.



MD’s dirty little secret

People rarely admit that their MD runs didn’t do anything interesting. This is because it is relatively easy to force MD to do **something**:

- Increase the temperature
- Increase the stress
- Carefully choose the “right” initial conditions

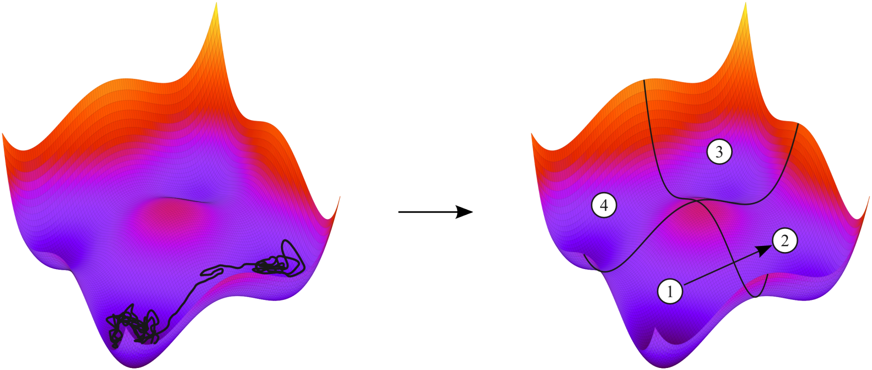
While such runs can be instructive, one should be extremely careful in interpreting them.

Possible solutions to the timescale problem

How can we then generate **long, unbiased, trajectories** with atomistic accuracy?

- Exploit **separation of timescales** between fast and slow components of the dynamics
- Transform **the continuous dynamical problem** of obtaining a long MD trajectory into **a discrete statistical problem** of obtaining correct state-to-state trajectories.

Accelerating the dynamics



Accelerating the dynamics

Traditional path-wise dynamical acceleration methods fall into two broad classes:

1. **The Kinetic Monte Carlo (KMC) way:** Determine relevant states and transitions, compute transition rates, and simulate kinetics.
2. **The Accelerated Molecular Dynamics (AMD) way:** Trick MD into reducing the separation of timescales between fast and slow components of the dynamics and renormalize the MD-time accordingly.

Note that other techniques are emerging: coarse-grained methods, multiscale techniques, etc. More about these later during the workshop.

Plan

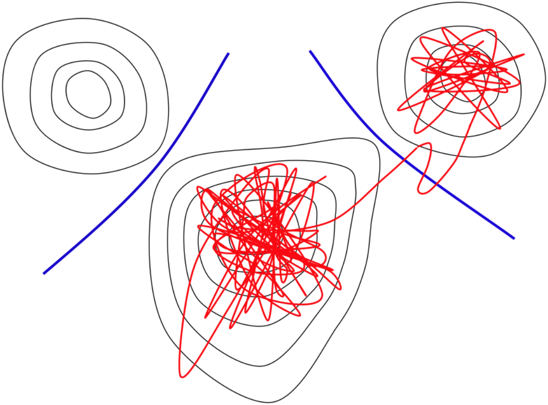
3 Accelerated Molecular Dynamics
- Hyperdynamics
- Temperature Accelerated Dynamics
- Parallel Replica Dynamics

Accelerated Molecular Dynamics

The AMD creed:

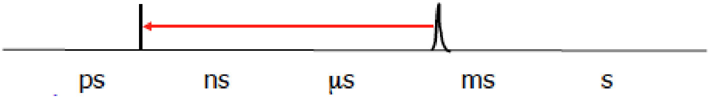
Let MD trajectories find an appropriate way out of each state but coax them into doing so more quickly. Use statistical mechanical concepts (primarily rate theories) to unbias the results.

Accelerated Molecular Dynamics



Accelerated Molecular Dynamics

The goal is to decrease the **separation of timescales** between transitions and relaxation by somehow decreasing the transition time. The amount of compression is the *boost*.



AMD method perform best for strongly metastable systems:

- large separation of timescales = large boost
- small separation of timescales = small boost

Note that some methods, like KMC or phase field crystals, instead move the relaxation times up towards the transition times. In MD, it is not possible (or at least very difficult) to do so.

Accelerated Molecular Dynamics: key points

The goal is the be just like MD, but to reach longer timescales.

- Open-ended: specify an initial state and let the system evolve.
- Stochastic: different runs can give different answers.
- (Ideally) require minimal *a priori* knowledge of the system.
- Typically more accurate than KMC.

But:

- Only useful for system evolving through sequences of rare events.
- Typically not as efficient as KMC.
- Not designed for sampling.
- (Currently) most useful for hard matter.

Practical Considerations

You can/should use AMD if:

- Rare event system (typical transition times $>$ ($\gg$) 100 ps)
- Transitions can be automatically detected and characterized
- You are ready to put in some effort to tailor the methods/code. Not always a black box.
- You are typically interested in paths containing more than one specific transition and/or you don't know the relevant final states

Words of wisdom

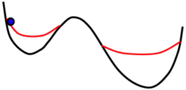
You think you know the relevant final states? You probably don't.

Being able to run long, unbiased, trajectories is invaluable.

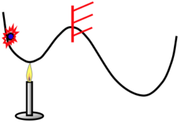
Accelerated Molecular Dynamics

Three AMD methods have been proposed by Arthur F. Voter *et al.* (see Ann. Rep. Comp. Chem. 5, 79, (2009) for a review):

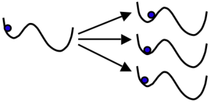
- **Hyperdynamics**



- **Temperature Accelerated Dynamics (TAD)**



- **Parallel Replica Dynamics (ParRep)**



Hyperdynamics

HYPERDYNAMICS

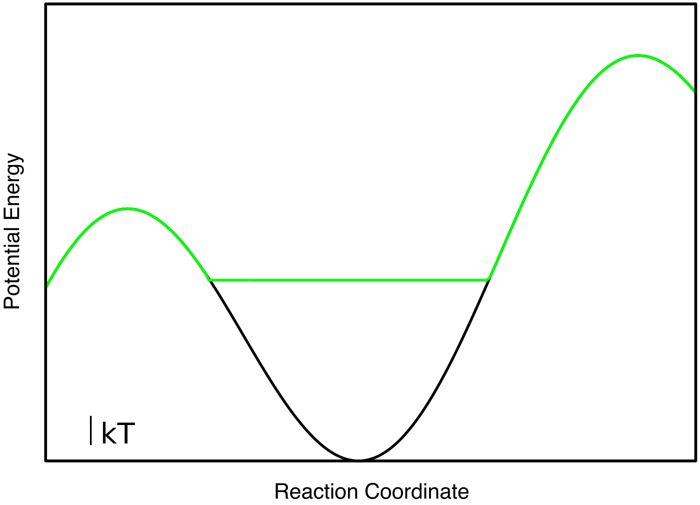
Principle of Hyperdynamics

The Problem: the kinetics are slow because the landscape contains deep potential energy wells
The Solution: modify the potential energy landscape to make the wells shallower

In hyperdynamics [Voter, PRL 78, 3908 (1997)], you:

1. Run MD while adding a non-negative bias potential ΔV_b to the original potential V
2. Map the MD-time t on $V + \Delta V_b$ unto the corresponding hyper-time t^h ($\gg t$) on V

Hyperdynamics: Illustration



Hyperdynamics: Derivation

Assume that the kinetics **obey Transition State Theory (TST)**. The transition rate out of a state *A* is given by:

$$k_{A\to}^{\mathrm{TST}} = \langle |v_A| \delta_A(\mathbf{r}) \rangle_A = \frac{\int d\mathbf{p} \int d\mathbf{r} \Theta_A(\mathbf{r}) |v_A| \delta_A(\mathbf{r}) e^{-\beta H(\mathbf{r},\mathbf{p})}}{\int d\mathbf{p} \int d\mathbf{r} \Theta_A(\mathbf{r}) e^{-\beta H(\mathbf{r},\mathbf{p})}} \tag{1}$$

Introducing a non-negative bias potential ΔV_b, we get:

$$k_{A\to}^{\mathrm{TST}} = \frac{\int d\mathbf{p} \int d\mathbf{r} \Theta_A(\mathbf{r}) |v_A| \delta_A(\mathbf{r}) e^{-\beta[H(\mathbf{r},\mathbf{p})+\Delta V_b(\mathbf{r})]} e^{\beta \Delta V_b(\mathbf{r})}}{\int d\mathbf{p} \int d\mathbf{r} \Theta_A(\mathbf{r}) e^{-\beta[H(\mathbf{r},\mathbf{p})+\Delta V_b(\mathbf{r})]} e^{\beta \Delta V_b(\mathbf{r})}}, \tag{2}$$

or, written in terms of averages on the **biased** landscape:

$$k_{A\to}^{\mathrm{TST}} = \frac{\langle |v_A| \delta_A(\mathbf{r}) e^{\beta \Delta V_b(\mathbf{r})} \rangle_{A_b}}{\langle e^{\beta \Delta V_b(\mathbf{r})} \rangle_{A_b}}. \tag{3}$$

Hyperdynamics: Derivation

Require that the bias vanishes along all dividing surfaces around *A*, i.e.:

$$\Delta V_b(\mathbf{r}) = 0 \text{ when } \delta_A(\mathbf{r}) \neq 0, \tag{4}$$

then:

$$k_{A\rightarrow}^{\mathrm{TST}} = \frac{\langle |v_A| \delta_A(\mathbf{r}) \rangle_{A_b}}{\langle e^{\beta \Delta V_b(\mathbf{r})} \rangle_{A_b}} = \frac{k_{A_b\rightarrow}^{\mathrm{TST}}}{\langle e^{\beta \Delta V_b(\mathbf{r})} \rangle_{A_b}}. \tag{5}$$

Transition rates (and time) on the biased potential are **uniformly** increased by a factor $\langle e^{\beta \Delta V_b(\mathbf{r})} \rangle_{A_b} (\gg 1)$!

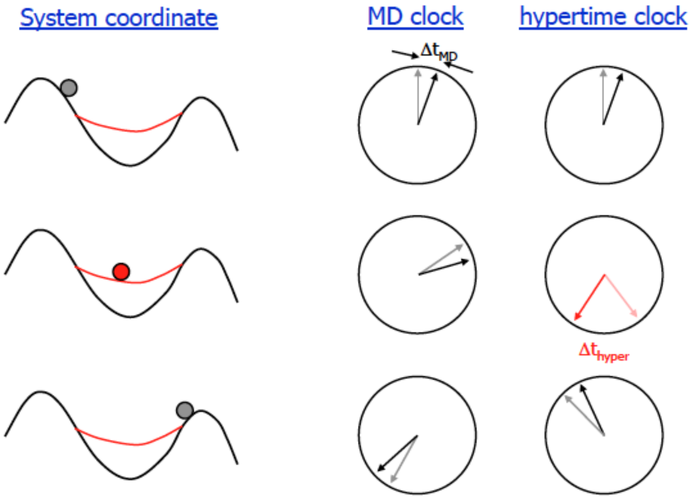
Hyperdynamics: The Algorithm

An hyperdynamics simulation proceeds as follows:

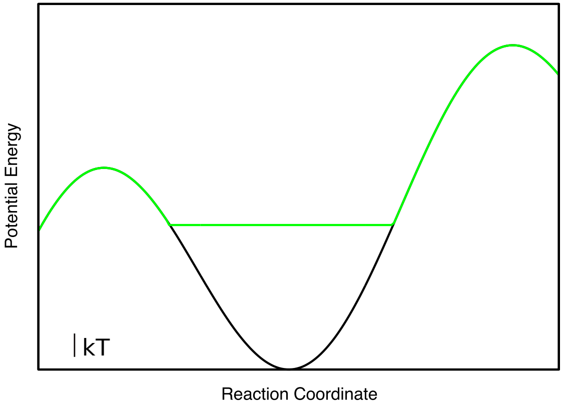
1. $t_{MD} \leftarrow 0$, $t_H \leftarrow 0$
2. Propagate the equations of motion forward by Δt using the biased potential $V + \Delta V_b$
3. $t_{MD} \leftarrow t_{MD} + \Delta t$, $t_H \leftarrow t_H + \Delta t e^{\beta \Delta V_b(\mathbf{r}(\mathbf{t}))}$
4. Go to step 2

When $\Delta V_b \rightarrow 0$, hyperdynamics reduces to molecular dynamics.

Hyperdynamics: The Hyper Clock

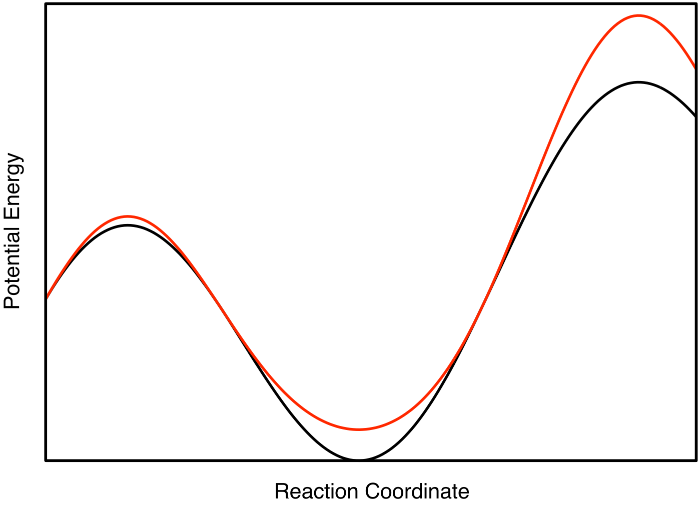


Hyperdynamics: Illustration



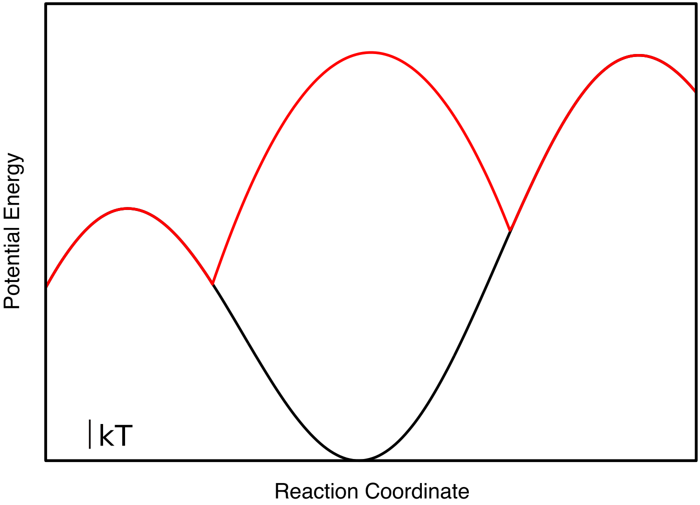
Condition 1: ΔV_b vanishes along all dividing surfaces
Condition 2: The system obeys TST on both V and $V + \Delta V_b$

Hyperdynamics: Illustration



BAD!

Hyperdynamics: Illustration



BAD!

Hyperdynamics: bias potentials

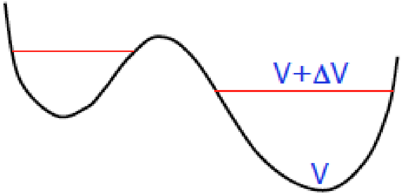
The challenge is to design ΔV_b without having to know the possible transitions pathways and without costing too much.

A few strategies have been proposed over the years:

- Flat bias: $\Delta V_b = V_{th} - V$ if $V < V_{th}$, 0 otherwise. [Steiner, Phys. Rev. B 57, 10236 (1998)]
- Detect ridgetops using local curvature. [Voter, PRL 78, 3908 (1997)]
- Bond-Boost: Assume that transitions signal themselves by a significant change in some bond length. [Miron and Fichthorn, J. Chem. Phys. 119, 6210 (2003)]

Hyperdynamics: bias potentials

The most obvious idea is to simply flatten out the potential below a certain value V_{th}.



This is appealing, as it is safe if V_{th} is below the lowest barrier. However, it does not work in many dimensions.

Hyperdynamics: bias potentials

- To first order, each configurational degree of freedom stores $\frac{1}{2}k_BT \sim 0.025$ eV.
- With 100 degrees of freedom, the potential energy is on average 2.5 eV above the minimum.
- If $V_{th} < 2.5$ eV: no boost
- Else: unsafe if lowest barrier $< V_{th}$ (almost always the case)
- **In many dimensions, a flat bias usually cannot be both safe and provide boost.**

Hyperdynamics: bias potentials

- Be careful with 1D intuitions...
- For most systems, there is always plenty of thermal energy available somewhere in the system. The slow step is for the energy to be channeled into the right degrees of freedom.
- A good bias potential has to be turned on when the system is "far" from the energy minimum.

Hyperdynamics: bias potentials

Hyperdynamics: bias potentials

- But we don’t know where the dividing surface is *a priori*...
- If states are defined using basins of the potential energy surface the following conditions define ridgetops around the state:
    - The lowest eigenvalue of the Hessian ($H_{i,j} = \frac{\partial^2 V}{\partial x_i \partial x_j}$) is negative.
    - The projection of the force on the corresponding eigenvector is 0.
- However, not all basins are fully surrounded by ridgetops...

Hyperdynamics: bias potentials

- One can write a bias potential that becomes 0 when the above conditions are met.
- The bias forces can be obtained through an iterative procedure using ∇V alone [Voter, PRL 78, 3908 (1997)].
- Good convergence can be expensive ($\sim$ 100× more than a standard force calculation).
- That overhead limits the real boost, so this is rarely used in practice.

Hyperdynamics: bias potentials

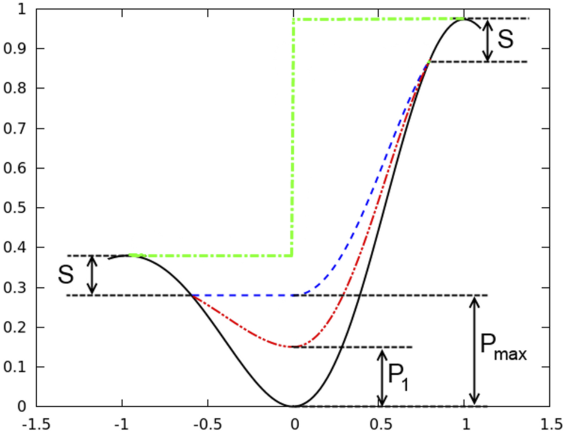
- For materials, transitions usually involve bonds stretching and breaking by some amount $\epsilon_i = \frac{r_i - r_{eq,i}}{r_{eq,i}}$.
- Define q as the minimal amount of stretch required for a transition to occur.
- The simplest bond-boost potential [following Miron and Fichthorn, J. Chem. Phys. 119, 6210 (2003)] is:

$$\begin{aligned}
\Delta V_b &= \min \delta V_{b,i}(\epsilon_i) \\
&\text{with} \\
\delta V_{b,i} &\geq 0 \text{ for } \epsilon_i \leq q \\
&= 0 \text{ otherwise}
\end{aligned}$$

Hyperdynamics: bias potentials

- If **any** bond stretches past its critical value, the bias is zero.
- Turns on even if every bond is stretched a bit.
- If q is chosen properly, it is safe.
- Fast and very easy to implement.
- However, we do not know q a priori. We can often guess pretty well, but we cannot be sure.
- Important: q cannot be validated *a posteriori*.

Hyperdynamics: bias potentials



MEP Hyperdynamics [Huang, Perez, Voter, JCP 143, 074113 (2015)]

Hyperdynamics in practice

Pros:

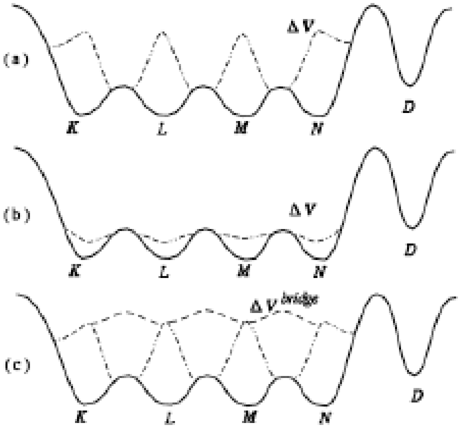
- Formally simple
- Can provide very large boost
- Simple bias functionals are available

Cons:

- Difficult to strickly assess its validity
- Efficiency drops with system size
- Sensitive to low barriers

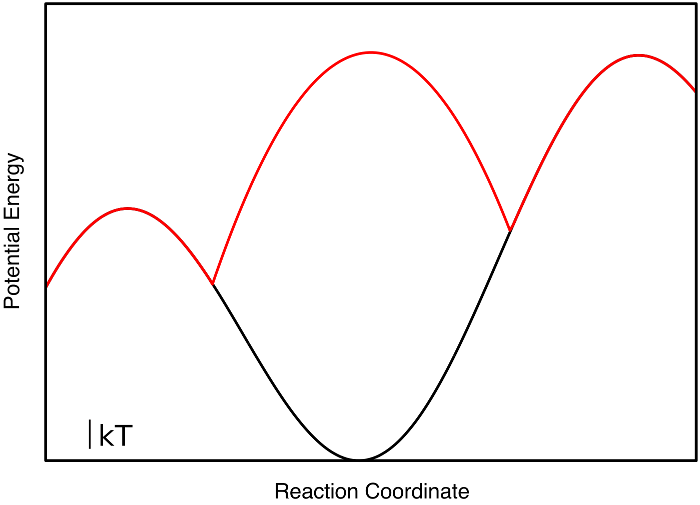
Hyperdynamics has been applied to: point defect diffusion, surface growth, nanowire plasticity, temperature programmed desorption, protein dynamics, etc. See Ann. Rep. Comp. Chem. 5, 79, (2009) for references.

Hyperdynamics: generalizations



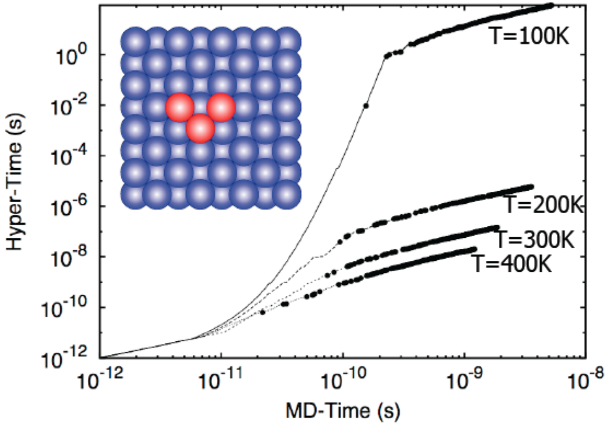
Bridging hyperdynamics [Miron and Fichthorn, Phys. Rev. B 72, 035415 (2005)]

Hyperdynamics



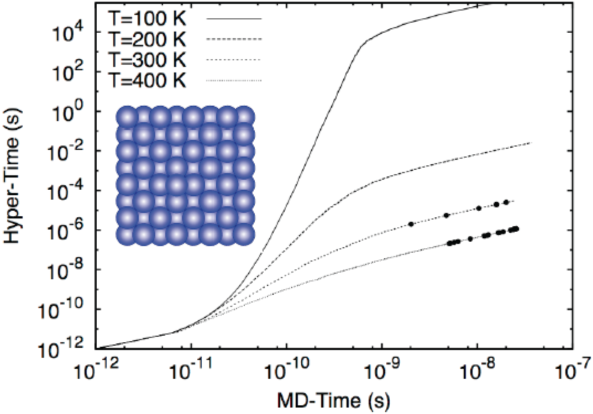
BAD!

Hyperdynamics: generalizations



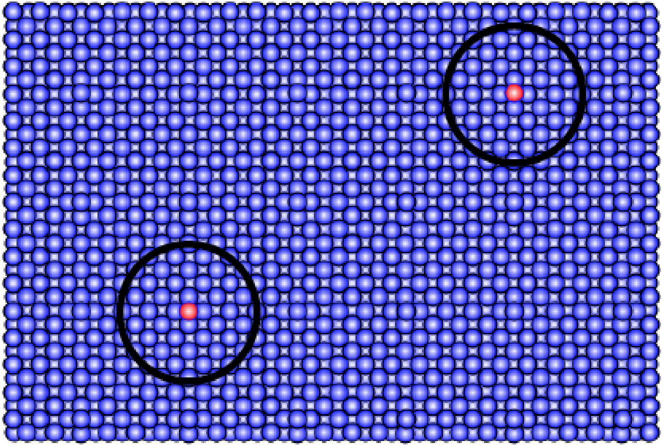
Self-learning hyperdynamics [Perez *et al.*, Ann. Rep. Comp. Chem. 5, 79, (2009)]

Hyperdynamics: generalizations



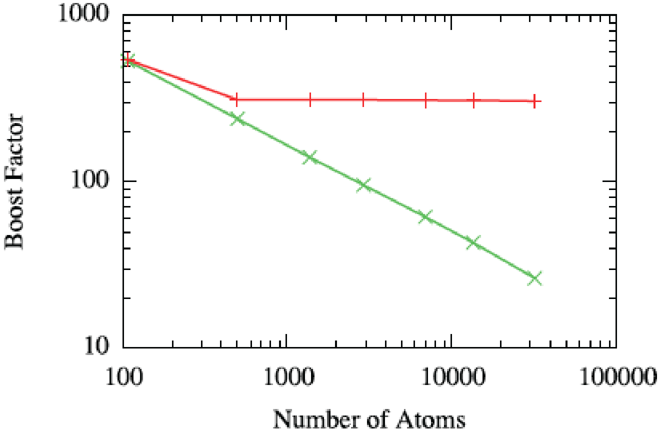
Self-learning hyperdynamics [Perez *et al.*, Ann. Rep. Comp. Chem. 5, 79, (2009)]

Hyperdynamics: generalizations



Local Hyperdynamics [Kim, Perez, Voter, JCP 139, 144110 (2013)]

Hyperdynamics: generalizations

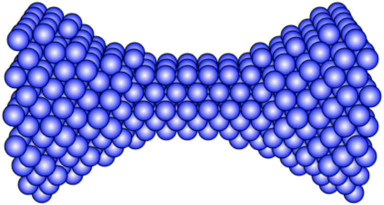


Local Hyperdynamics [Kim, Perez, Voter, JCP 139, 144110 (2013)]

Application to nanowire stretching

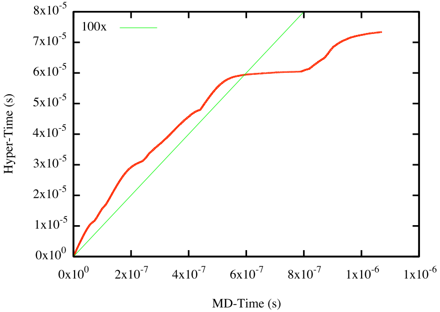
Hyperdynamics: example

We studied the elastic properties of silver nanowires using the Self-Learning Bond-Boost Hyperdynamics (with C.-W. Pao).



- (110) Nanowire
- Voter-EAM Silver
- T=300 K
- Stretching rate: 10^5Å/s (±100x faster than experiment)
- Discrete stretching

Hyperdynamics: example



- Roughly 5000 transitions
- A lot of fast and complex events (edge-running, slip)
- Boost factor varies between 100 and 200
- With more aggressive book-keeping, an additional factor of $3-5$ is possible.

Hyperdynamics: example

Growth of Co/Cu(001) films with bond-boost hyperdynamics [Miron and Fichthorn Phys. Rev. B 72, 035415 (2005)]

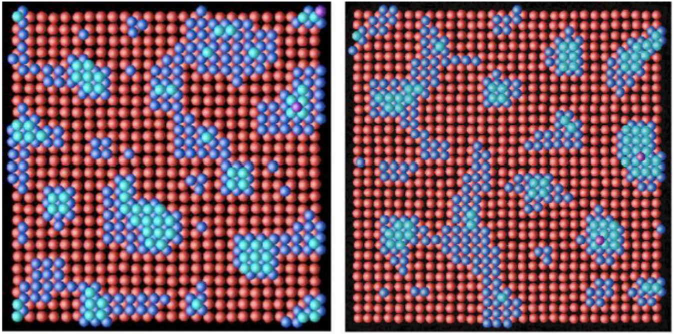


Figure: Left: T=250K; right: T=310K. Simulation (hyper-)time: $\sim$ 0.5s.

Temperature Accelerated Dynamics

TEMPERATURE ACCELERATED DYNAMICS

Principle of Temperature Accelerated Dynamics (TAD)

- One of the most common solution to a sluggish MD simulation is to increase the temperature.
- Makes things happen faster, but also makes the **wrong** things happen.
- This is especially problematic when many processes compete.
- How to unbias the results?

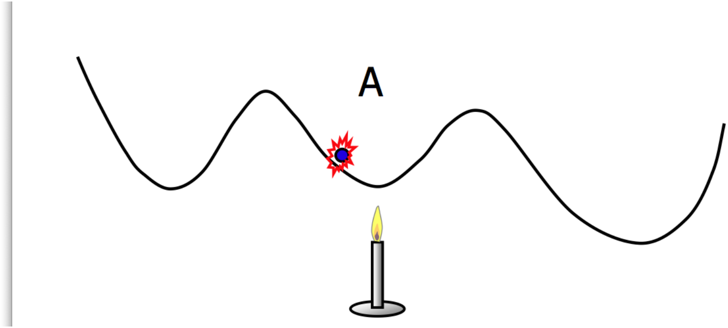
Principle of Temperature Accelerated Dynamics (TAD)

The Problem: the kinetics are slow because the thermal energy is small compared to ΔE.
The Solution: increase the temperature

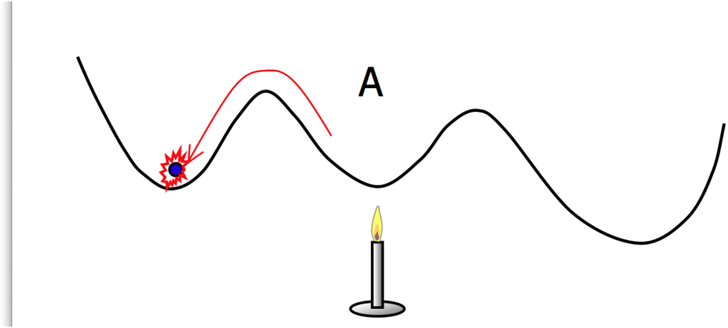
In TAD [Sorensen and Voter, JCP 112, 9599, (2000)], you:

1. Run MD at higher temperature and detect transitions to new states
2. Characterize transitions and place the system back in the initial state
3. Repeat until the proper low-temperature transition is identified
4. Move to the corresponding state

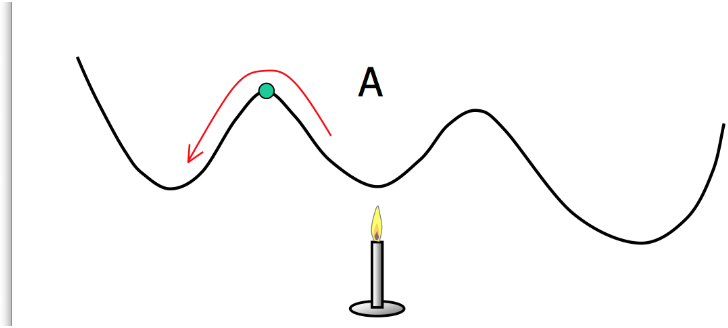
TAD: Illustrations



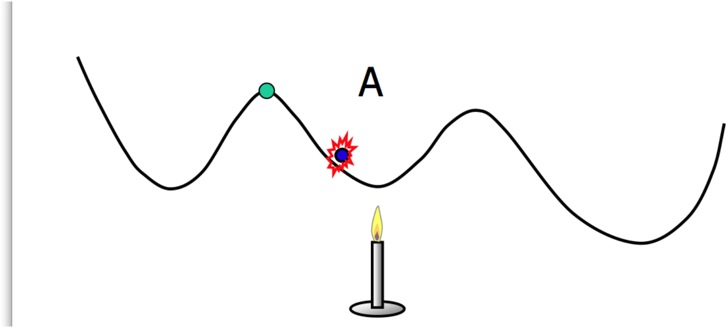
TAD: Illustrations



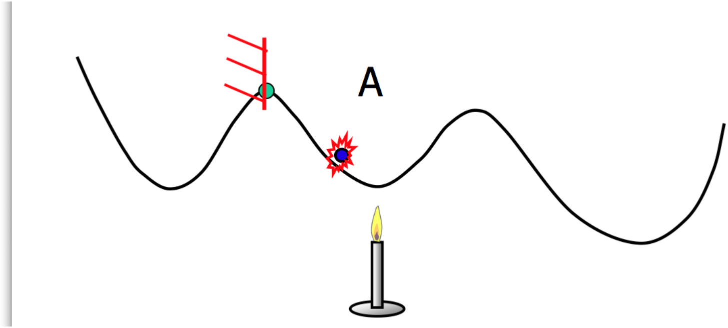
TAD: Illustrations



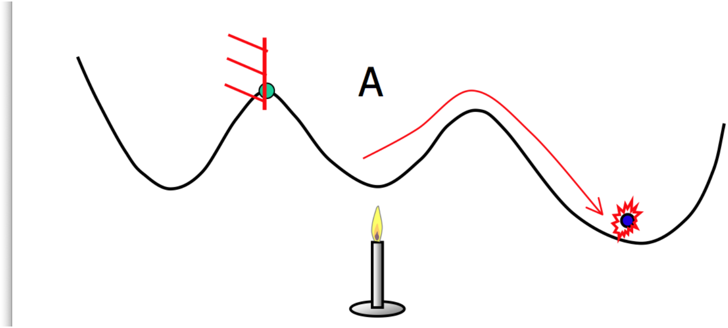
TAD: Illustrations



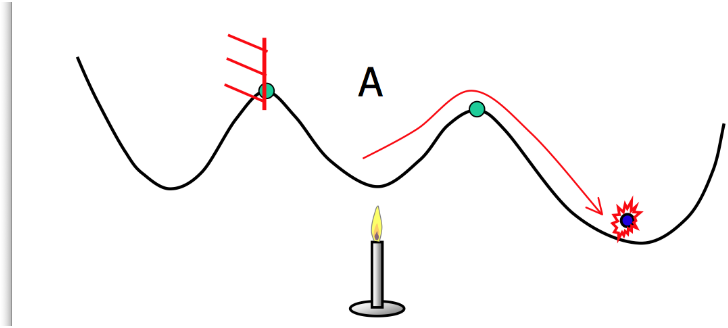
TAD: Illustrations



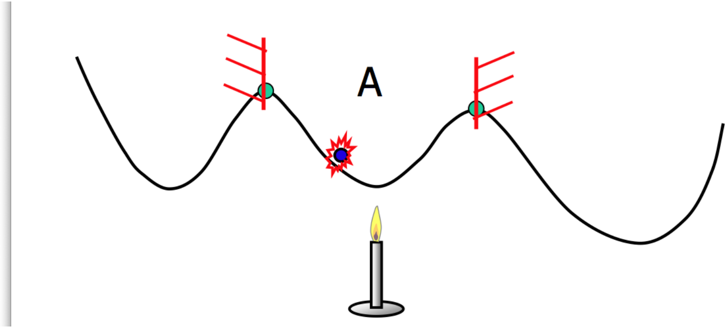
TAD: Illustrations



TAD: Illustrations



TAD: Illustrations



TAD: Derivation

Using transitions observed at high temperature, we need to:

- Choose one of them as a proper representative of *low-temperature* dynamics
- Choose a corresponding *low-temperature* transition time

This is possible under two assumptions:

- Harmonic TST holds
- All prefactors are higher than ν_{min}

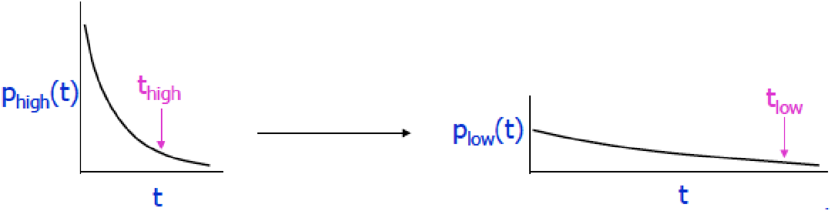
TAD: Sketch

We are really trying to identify a proper sample of the *first* transition out of the current state to occur at $T_{\rm low}$:

- Running at a higher temperature $T_{\rm high}$, observe an event with a barrier ΔE after a time $t_{\rm high}$.
- Generate corresponding escape time $t_{\rm low}$ at $T_{\rm low}$
- Set $t_{\rm low,short} = \min(t_{\rm low,short}, t_{\rm low})$
- If probability of a yet unseen event with $t_{\rm low} < t_{\rm low,short}$ small enough: accept transition corresponding to $t_{\rm low,short}$; else: repeat.

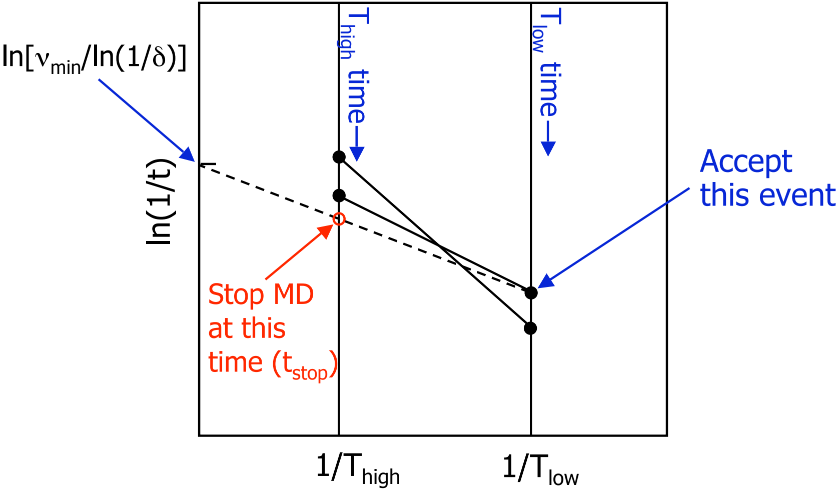
TAD: Illustrations

For rare events, $p(t) = k\exp[-kt]$. A sample from one distribution can be converted into one of another by setting $t_{\text{low}}k_{\text{low}} = t_{\text{high}}k_{\text{high}}$. Using HTST: $t_{\text{low}} = t_{\text{high}}\exp[\Delta E(\beta_{\text{low}} - \beta_{\text{high}})]$.



- Need to find the barrier for each observed process
- High (low) barrier processes extrapolate to long (short) times

TAD: Illustrations



TAD: Derivation

When to stop? Define a **stop time** $t_{\text{high,stop}}$ such that the probability that another event would replace the current $t_{\text{low,short}}$ is less than δ.

Assuming that all transition have a prefactor larger than ν_{min}, one get:

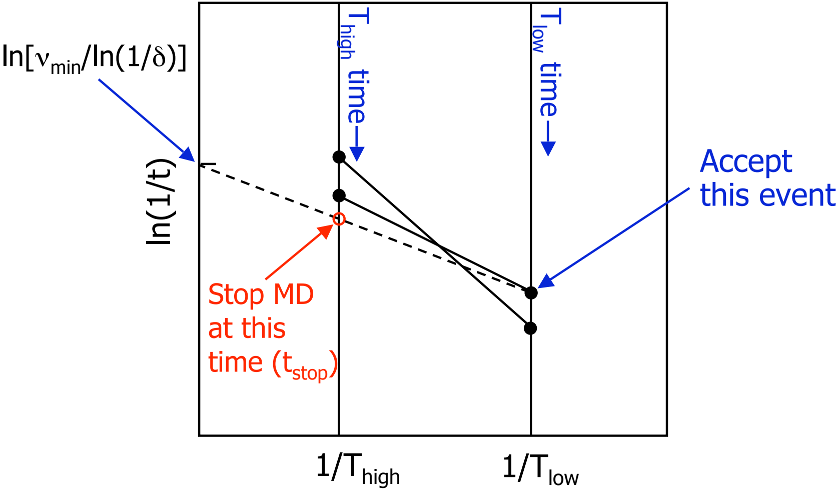
$$t_{\text{high,stop}} = \frac{\ln(1/\delta)}{\nu_{\text{min}}} \left(\frac{\nu_{\text{min}} t_{\text{low,min}}}{\ln(1/\delta)} \right)^{\beta_{\text{high}}/\beta_{\text{low}}} \qquad (6)$$

After having ran MD at T_{high} for a time $t_{\text{high,stop}}$, you can accept the event that occured first at T_{low} and increment your MD clock by $t_{\text{low,short}}$ ($\gg t_{\text{high,stop}}$).

TAD: Derivation

- You are trying to make sure that there is not a low barrier process that could have $t_{\text{low}} < t_{\text{low,short}}$ that is still hiding from you.
- After running MD at T_{high} for a while, you can say with some confidence that if such a process really existed, you should have seen it by now.
- The worst case is a transition with a low barrier, but a very low prefactor. This would extrapolate to a time $t_{\text{low}} \sim t_{\text{high}}$. That is why we have to assume a ν_{min}.

TAD: Illustrations



TAD: Choosing T_{high}

- The performance of the method depends on T_{high}.
- Balance of conflicting requirements:
  - T_{high} too low, you spent too much time doing MD before you see a transition.
  - T_{high} too high, you see transitions that are irrelevant at T_{low}. Each one of these costs you something (e.g., NEB) and you will see many before you accept the right one.
  - T_{high} too high, anharmonicity errors will become significant.
- You might need to guess a bit, but automated adjustment is possible [Shim, Amar, JCP 134, 054127 (2011)].

TAD: Illustrations

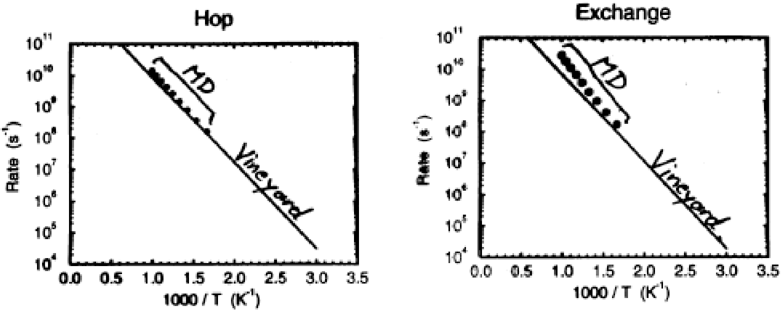


Figure: Diffusion rate of an Ag adatom on Ag(100).

TAD in practice

Pros:

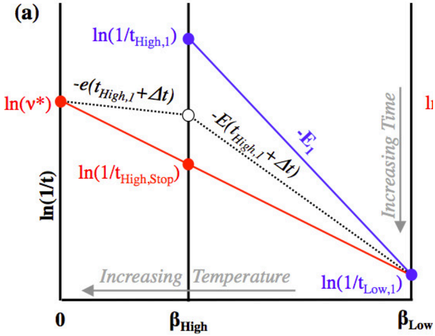
- Can provide very large boost
- Low barriers can be handled to some extent

Cons:

- More approximate than the other methods (correlated events, anharmonicity, minimum prefactor)
- Robust and efficient implementation can be a challenge

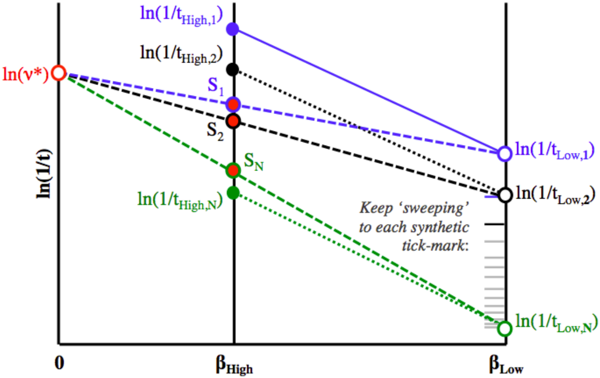
TAD has been applied to: point defect diffusion, surface growth, radiation damage annealing, cluster dynamics, etc. See Annual review of chemical and biomolecular engineering 7, 87 (2016) for examples.

TAD: generalizations



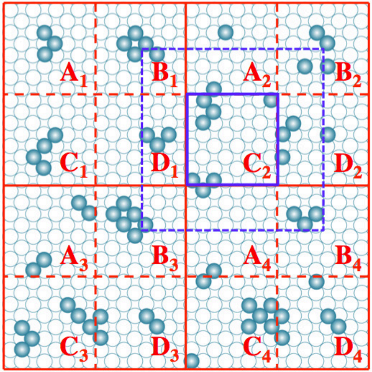
Learning from the past: decrease stop time for revisits to the same state. [Montalenti and Voter, JCP 116, 4819 (2002)]

TAD: generalizations



Synthetic-TAD: treat fast events with KMC but use TAD to keep looking for low-rate events. [Sorensen and Voter, JCP 112, 9599 (2000)]

TAD: generalizations



ParTAD: spatially parallelize TAD using the SSL algorithm. [Shim *et al.*, PRB 76, 205439 (2007)]

TAD: generalizations

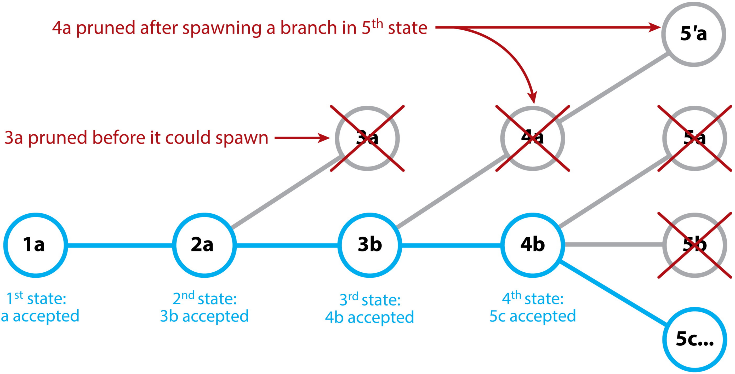


Figure: SpecTAD: speculative parallelization of TAD [Ann. Rev. of Chem. Bio. Eng. 7, 87 (2016)]]

TAD: generalizations

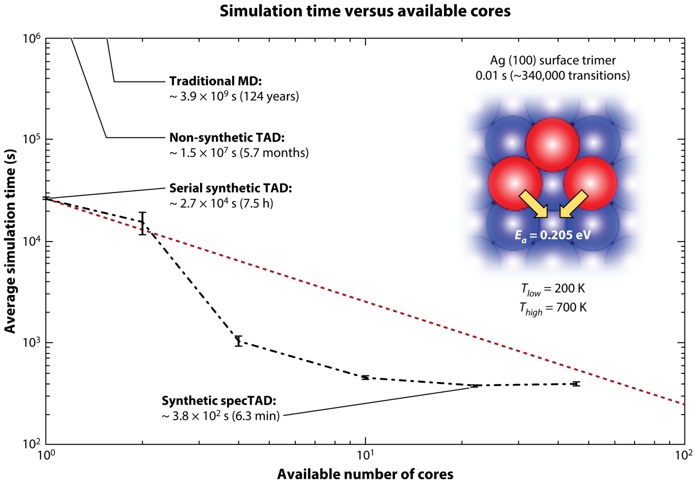


Figure: SpecTAD: speculative parallelization of TAD [Ann. Rev. of Chem. Bio. Eng. 7, 87 (2016)]]

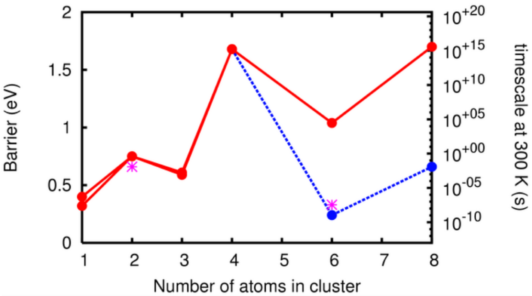
TAD example: Interstitial clusters in MgO

- MgO is a component of nuclear fuel. As such, its tolerance to radiation is of prime interest.
- To first order, radiation causes the formation of Frenkel pairs. Vacancies are pratically immobile, but interstitials diffuse rapidly and coalesce into clusters.
- The behavior of interstitial clusters is very interesting [Uberuaga *et al.*, Phys. Rev. Lett., 92, 115505 (2004)]:
  - Mono-interstitial: diffuse in ns/μs
  - Di-interstitial: diffuse in s
  - Tetra-interstitial: immobile
- In the tetramer a sink for all larger clusters? no!

TAD Simulation: $\mathrm{T}_{\mathrm{high}} = 2000\mathrm{K}$, $\mathrm{T}_{\mathrm{low}} = 300\mathrm{K}$

Figure: Red: O; blue: Mg. Perfect bulk atoms are not shown.

TAD Simulation: $T_{\mathrm{high}} < 2000$K, $T_{\mathrm{low}} = 300$K



- Mobility vs. size pattern is non-trivial
- Metastable clusters can be very mobile
- Metastable clusters can be very long-lived (years)

Parallel Replica Dynamics

PARALLEL REPLICA DYNAMICS

Principle of Parallel-Replica Dynamics (ParRep)

The Problem: the (wall) time between transitions is too long on a single CPU

The Solution: use many CPUs!

Wait....you said this was not possible!

It is possible if you parallelize over time instead of space [Voter, PRB 57, R13985 (1998)].

Principle of Parallel-Replica Dynamics (ParRep)

The Problem: the (wall) time between transitions is too long on a single CPU

The Solution: use many CPUs!

Wait....you said this was not possible!

It is possible if you parallelize over time instead of space [Voter, PRB 57, R13985 (1998)].

Principle of Parallel-Replica Dynamics (ParRep)

The Problem: the (wall) time between transitions is too long on a single CPU

The Solution: use many CPUs!

Wait....you said this was not possible!

It is possible if you parallelize over time instead of space [Voter, PRB 57, R13985 (1998)].

The current scope of MD simulations

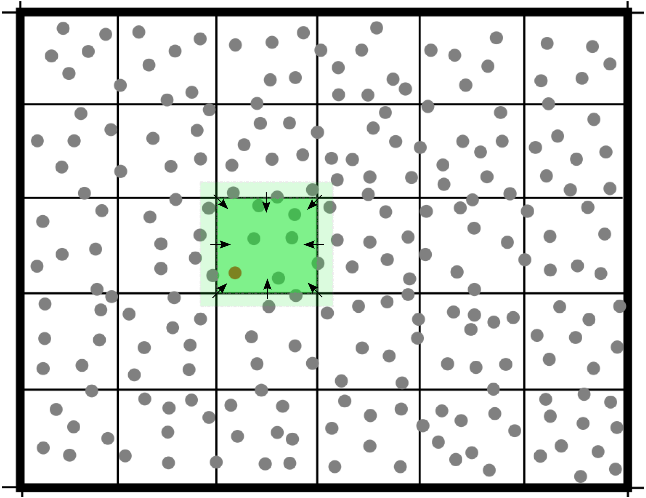


Figure: Spatial decomposition scheme: each cell is handled by a different processor.

Illustration of ParRep

Illustration of ParRep

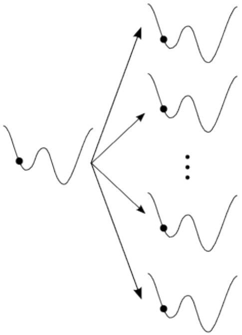


Illustration of ParRep

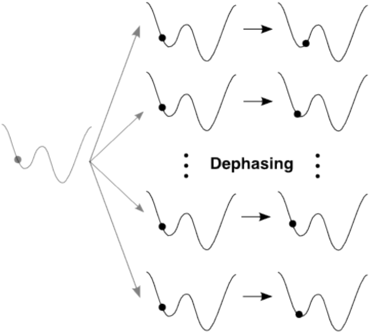


Illustration of ParRep

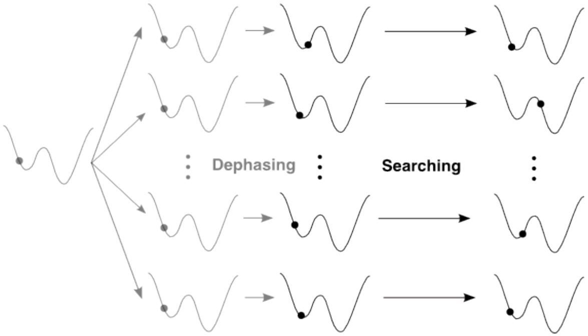
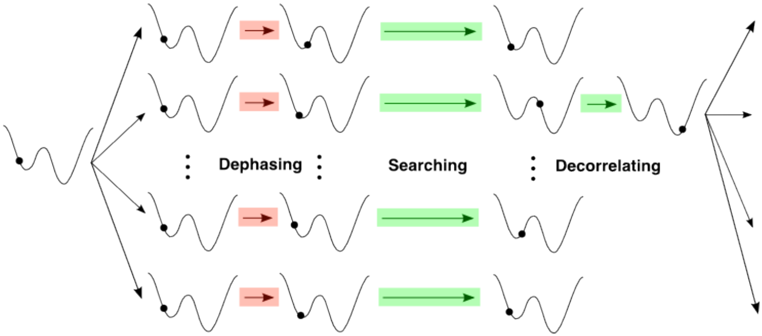


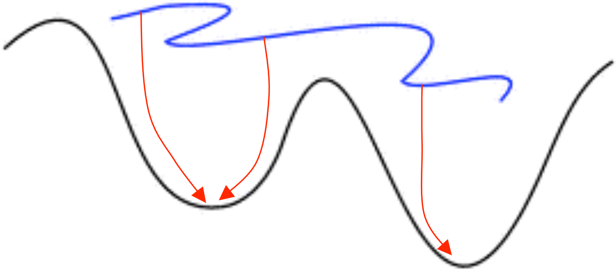
Illustration of ParRep

Illustration of ParRep



Detecting Transistions

Detecting Transistions

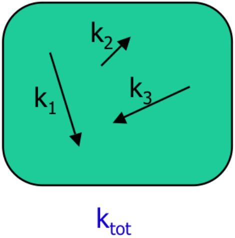


Justification of ParRep

Assume you have a true rare event system (i.e., each process' kinetics is characterized solely by a rate constant). Then, the **(first) escape times are exponentially distributed**, i.e.,

$$p(t) = k_{tot} e^{-k_{tot} t}, \tag{7}$$

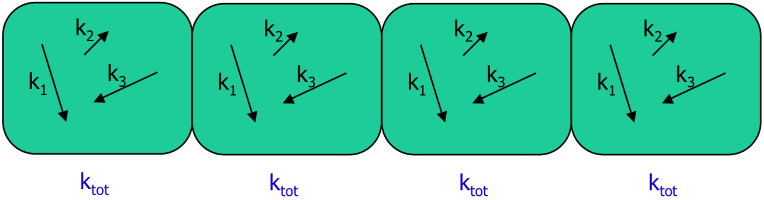
where k_{tot} is the total rate constant for all transitions out of the state ($k_{tot} = \sum k_i$).



Justification of ParRep

Now imagine creating a super-system composed of M **independent** replicas of the system. Then the first escape time distribution of this new system is (since the total rate constant is just scaled by M):

$$p^*(t) = Mk_{tot}e^{-Mk_{tot}t}. \tag{8}$$



Justification of ParRep

Define the summed time as the sum of the times spent in each replica, $t_{sum} = Mt$ before the first escape. Then:

$$\begin{aligned}
p^*(t) &= Mk_{tot}e^{-Mk_{tot}t} \\
p^*(t_{sum}) &= Mk_{tot}e^{-k_{tot}t_{sum}}(1/M) \\
p^*(t_{sum}) &= k_{tot}e^{-k_{tot}t_{sum}} \\
p^*(t_{sum}) &= p(t)
\end{aligned}$$

Justification of ParRep

The distribution of first escape times for the original system and for the composite system (in summed time) are the same!

Since the kinetics on each sub-system are first order, the probability that the system (both the original and composite) first escapes through pathway *i* is simply k_i/k_{tot}.

For first-order kinetics, ParRep is **exact** and accumulates MD time up to *M* times faster than standard MD (in wall-clock time).

Justification of ParRep

In real life, kinetics are **never** truly first-order. Then, what can we say about ParRep?

Traditionally, we thought of this as the main approximation underlying ParRep. We now understand that ParRep does in fact much better.

Even if kinetics are not purely first-order, ParRep can still be made **arbitrarily accurate** [Le Bris, Lelievre, Luskin, and Perez, Monte Carlo Methods and Applications 18, 119 (2012)].

Justification of ParRep

Let us consider overdamped Langevin dynamics:

$$dX_t = -\nabla V(X_t)\,dt + \sqrt{2\beta^{-1}}dW_t$$

The statistical characteristics of the *first escape* are given by the solution of a Fokker-Planck equation with absorbing boundary conditions around the boundary of state.

This equation statistically accounts for the effect of all possible noise sequences on the time evolution of the system.

Justification of ParRep

The probability distribution function of X is given by the solution of:

$$\begin{cases} \partial_t v = Lv \text{ on } A, \\ v = 0 \text{ on } \partial A. \end{cases}$$

with $L = -\nabla V \cdot \nabla + \beta^{-1} \Delta$.

Formally, the solution of this equation can be obtained by spectral decomposition of $-L$:

$$v(X, t) = \sum_k \exp(-\lambda_k t) c_k^0 u_k(X) \tag{9}$$

where the λ_k and u_k are eigenvalues and eigenfunctions of $-L$, respectively and the c_k^0 are set by initial conditions.

Justification of ParRep

At long times ($\gg 1/\lambda_2$), v becomes:

$$v(X,t) \simeq \exp(-\lambda_1 t) c_1^0 u_1(X). \tag{10}$$

with $u_1(x)$ the so-called Quasi-Stationary Distribution (QSD).

If $X_0 \sim u_1$ then, the first exit time T_A from A is exponentially distributed and is a random variable independent of the first hitting point X_{T_A} on ∂A.

If you stay long enough in **any** state, the next escape will be a good first-order process!

Justification of ParRep

We can use these concepts to reinterpret ParRep.

Step 1:

- Take a trajectory that just entered a state.
- Until the initial and final points are in the same state:
    - Run for a decorrelation time τ_{corr}
- Add up the total MD time to the official clock

At the end of that stage, the final point is approximately drawn from u_1 of the last visited state, say A.

This is the decorrelation stage

Justification of ParRep

Step 2:

- Generate $M - 1$ initial points in A
- For each point (on different processors):
  - Run for a decorrelation time τ_{corr}
  - If the final point in is A, break
  - Else, resample a point from A and repeat

At the end of that stage, we have M points approximately drawn from u_1 of A.

This is the dephasing stage

Justification of ParRep

Step 3:

- Until the first escape from A, on each replica:
  - Run MD
- Add up the total MD time until the first escape to the official clock
- Go to step 1 with the first escape trajectory

This is the parallel stage

Since the trajectories were distributed according to u_1, the first escape of a member of that set is (approximately) a first-order process.

Justification of ParRep

The error underlying this procedure is of order $\exp(-(\lambda_2 - \lambda_1)\tau_{\mathrm{corr}})$, i.e., ParRep can be made arbitrarily accurate by adjusting τ_{corr}.

This is true **independently of the definition of states**.

The bigger the spectral gap $\lambda_2 - \lambda_1$, the better the performance.

[Monte Carlo Methods and Applications 18, 119 (2012)]

ParRep in Practice

Pros:

- Very simple
- Arbitrarily accurate
- Processors can have unequal or even variable speeds
- Efficiently exploits parallel computers
- Flexible in terms of the definition of states
- Can handle driven systems

Cons:

- Requires a parallel computer to get some acceleration
- Efficient only if $N\tau_{\text{corr}} \ll \tau_{\text{trans}}$

See Comp. Mat. Sci. 100, 90 (2015) for more examples.

Peta-scale Parallel Replica Dynamics

ParRep provides a unique ability to exploit massively parallel computers.

In order to demonstrate the scalability of ParRep, we implemented it on Roadrunner, the first computer to break the petaflop barrier.

- 1.4 petaflops (1.4×10^{15} operations per second)
- Simulation rates for 1000 EAM atoms: 4 ms/day (10 ns/day for MD on a single core)
- Hybrid architecture
- 12 240 Opteron cores and Cell processors (about 122 400 cores total)

Application to Ag Nanowires

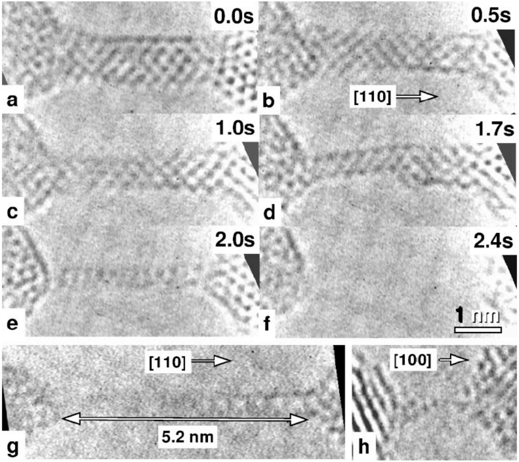


Figure: HRTEM imaging of Ag NWs. (Rodrigues *et al.*, PRB 65, 153402)

Application to Ag Nanowires

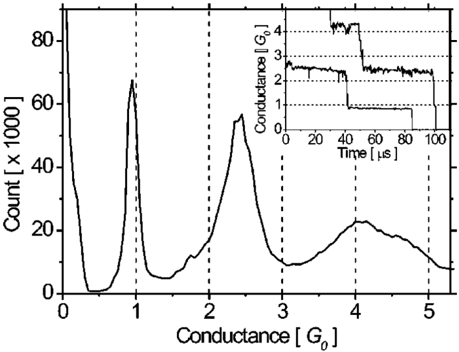
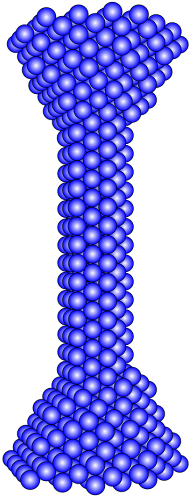


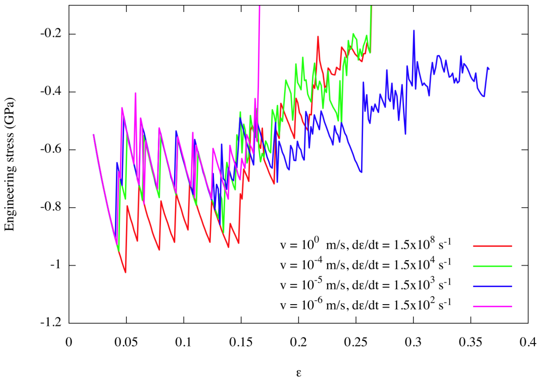
Figure: Break-junction conductance measurement of Ag NWs. (Rodrigues *et al.*, PRB 65, 153402)

Method

- Initial state inferred from HRTEM measurements
- EAM potential (Mishin)
- Canonical Ensemble (Langevin thermostat), T=300K
- $10^{-6} \leq v \leq 1$ m/s, $10^2 \leq \dot{\epsilon} \leq 10^8 \mathrm{s}^{-1}$
- $720 < N_{\mathrm{rep}} < 12000$, 10 cores/replica



Stress-strain Behavior

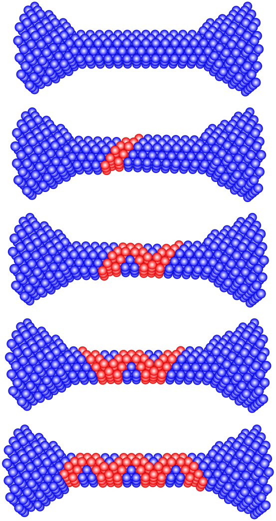


- Engineering yield stress $\simeq$ 1 GPa, Yield stress $\simeq$ 8 GPa
- Early stage behavior is regular and very robust
- Late stage behavior is complex
- Failure occurs for strains between 15% and $>$100%

$v = 10^{-5}\mathrm{m/s}, \dot{\epsilon} = 1.5x10^{3}\mathrm{s}^{-1}$

Early Stage

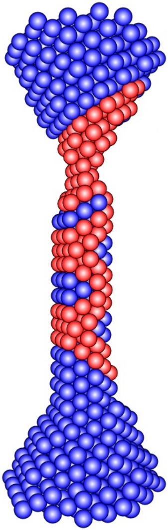
- Plasticity initially mediated by stacking faults along (111) planes
- Transitions more slip-like than dislocation-like
- Formation of a zig-zag pattern
- Uniform thinning of the wire
- Annihilation does **not** directly follow from more strain



Late Stage

Late stage is sensitive to conditions and realization. Possible outcomes are:

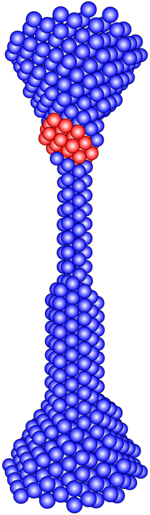
- Necking (all drives, necks earlier for slow drives)



Late Stage

Late stage is sensitive to conditions and realization. Possible outcomes are:

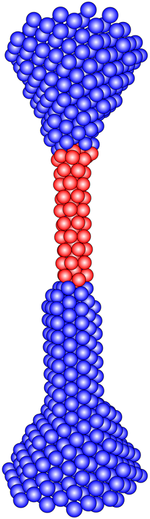
- Necking (all drives, necks earlier for slow drives)
- Non-uniform thinning, transition to an icosahedral structure, gradual conversion from FCC to icosahedral. (intermediate drives)



Late Stage

Late stage is sensitive to conditions and realization. Possible outcomes are:

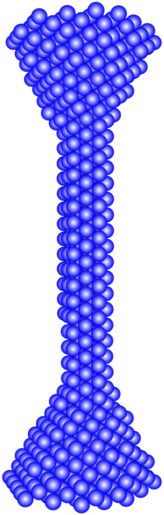
- Necking (all drives, necks earlier for slow drives)
- Non-uniform thinning, transition to an icosahedral structure, gradual conversion from FCC to icosahedral. (intermediate drives)



Late Stage

Late stage is sensitive to conditions and realization. Possible outcomes are:

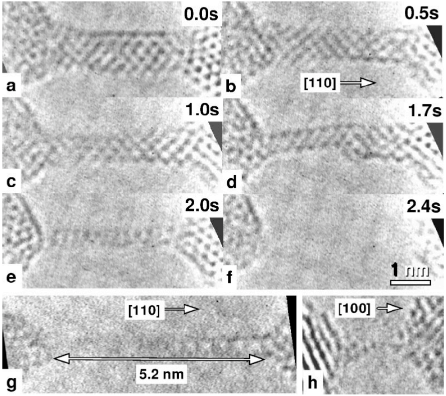
- Necking (all drives, necks earlier for slow drives)
- Non-uniform thinning, transition to an icosahedral structure, gradual conversion from FCC to icosahedral. (intermediate drives)
- Annihilation of the stacking fault network (slow drives)



Comparison with Experiments

Qualitative comparison with HRTEM is excellent.

- Uniform thinning
- Non-uniform thinning (kinks)
- Postulated “super-elastic” state consistent with our observation of an icosahedral phase that can “unwind” an FCC wire
- Atomic chain not very stable for that orientation



ParRep: generalizations

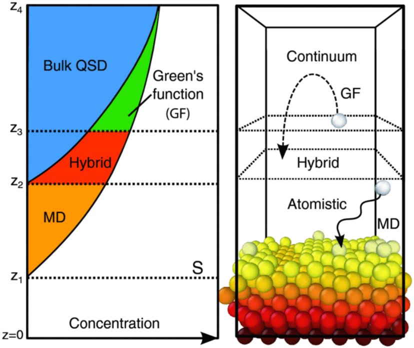


Figure: ParRep for solid-liquid systems: correctly allows diffusion at wet surfaces and even dissolution and precipitation [JCP 140 044116 (2014)]

ParRep: generalizations

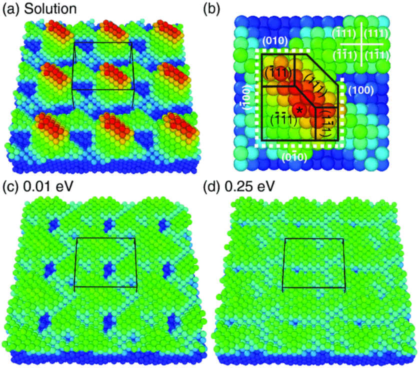


Figure: ParRep for solid-liquid systems: correctly allows diffusion at wet surfaces and even dissolution and precipitation [JCP 140 044116 (2014)]

ParRep: generalizations

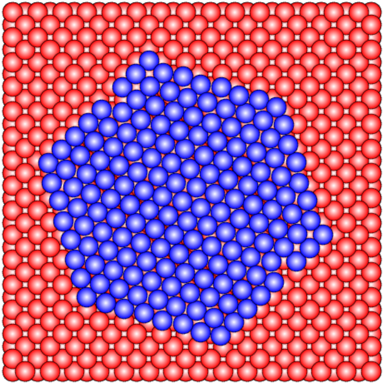


Figure: Super-state ParRep: Change the definition of the states to allow efficient simulations of systems with low barriers [PRL 103, 046101 (2009)]

ParRep: generalizations

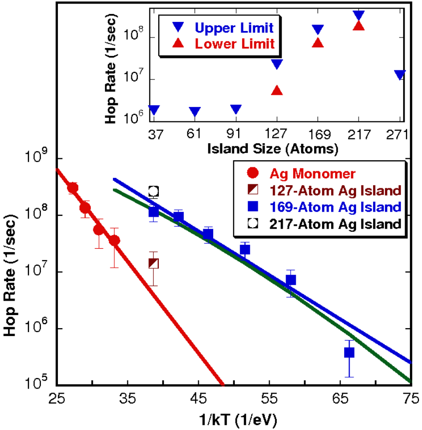


Figure: Super-state ParRep: Change the definition of the states to allow efficient simulations of systems with low barriers [PRL 103, 046101 (2009)]

Definition of States

In standard ParRep, we exploit the separation of timescale between τ_{rel} and τ_{trans} at the level of single basins.

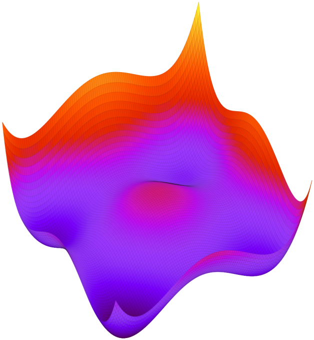


Figure: This is a good strategy when the landscape looks like this

Definition of States

In standard ParRep, we exploit the separation of timescale between τ_{rel} and τ_{trans} at the level of single basins.

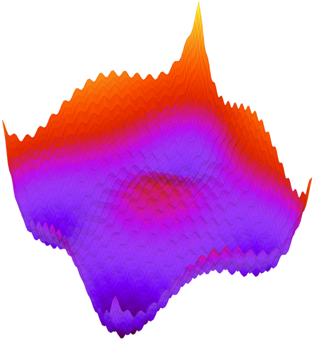


Figure: This is a bad strategy when the landscape looks like this

Definition of States

Any partition of configuration space can be used. The greater the separation of timescales between $\tau_{rel} \sim \lambda_2^{-1}$ and $\tau_{trans} \sim \lambda_1^{-1}$, the better the performance.

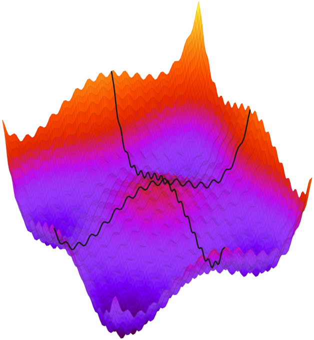
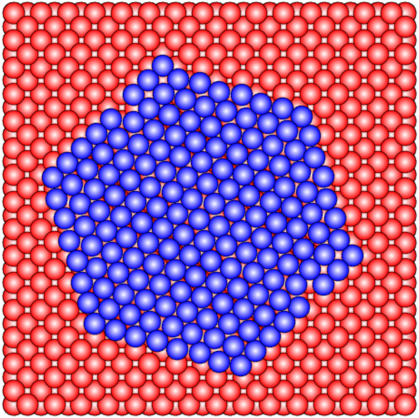


Figure: This is a good definition of states

Application to Ag on Cu Island Dynamics

- Cu (001) substrate
- *c*(10*x*2) hexagonal Ag island (between 37 and 271 atoms)
- 11% lattice mismatch between Ag and Cu



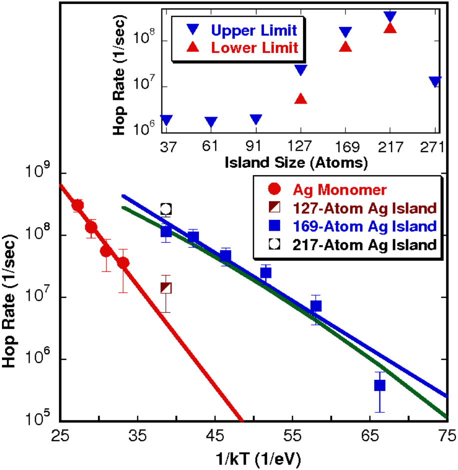
[Uche, Perez, Voter, Hamilton, PRL **103**, 046101 (2009)]

MD Simulation at 300 K

Application to Ag on Cu Island Dynamics

Observations:

- Large islands are mobile on MD timescales!
- The hopping rate is extremely sensitive to the island size
- There is a “magic size” effect whereby only islands with between 127 and 217 atoms diffuse rapidly
- Transitions are extremely frequent



Super-States Parallel Replica Dynamics

Slow Coordinates?

- The system jumps from basin to basin in <1 ps
- Most transitions only affect the core of the island
- At 175K, edge vacancies diffuse only every $\sim$1 ns
- Super-state = all basins with the same edge configuration

Using this definition, we simulated the 169-atoms island at 175K and 200K

- Total MD-time: 25μs (3 years of cpu time)
- Used up to 256 cpus
- Observed around 30 000 transitions and 10 hops

Super-State Network

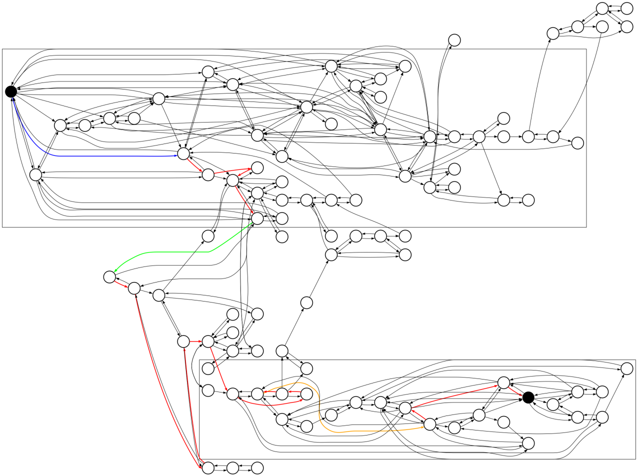


Figure: Observed transitions over $6\mu s$ at $T = 175K$.

Favorable hopping pathway: vacancy and glide

Figure: Low energy transition pathway inferred from T=175K super-state Par-Rep simulations.

Nudged Elastic Band Analysis

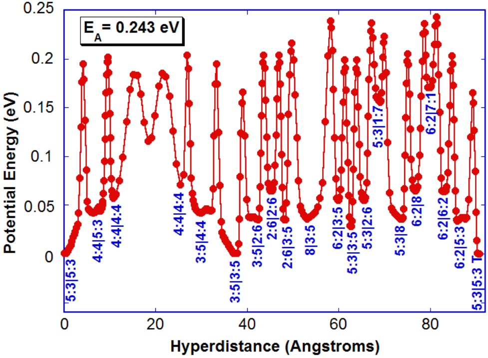


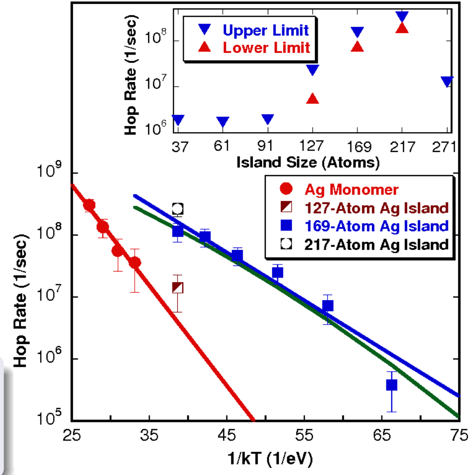
Figure: Minimum energy pathway along the previous path.

- The barrier to glide is small compared to that of vacancy diffusion
- The rate limiting step is the annihilation and creation of vacancies at the edges

Size Dependence of the Hopping Rate

- Core glide is active at all sizes
- The formation energy of vacancies is large and positive (negative) at small (large) sizes
- Only in the intermediate regime is the creation and annihilation of vacancies thermally accessible

The size dependence stems from the sensitivity of the energetics of edge vacancies.



ParRep for driven systems

ParRep is not restricted to “static” system, it can also be applied to driven systems (systems with time-dependent boundary conditions).

In this case one must:

- During the parallel stage: increase the driving rate by a factor of M; enforce synchronization between the replicas.
- During the dephasing stage: set the driving rate to 0.
- During the decorrelation stage: use the normal driving rate.

ParRep for driven systems

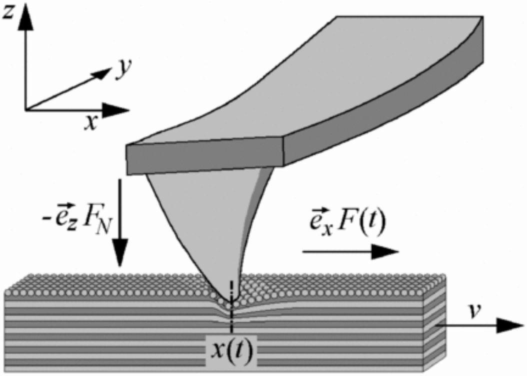
This recipe works as long as the drive is quasi-static, i.e., $k_i(x,t) = k_i(x(t))$, i.e., the rates do not depend on the driving rate, just on the current state of the drive.

In this case, the combined system (with drive $\times M$) spends the same amount of time in a given drive interval than the original system.

The probabilities of transitions per unit time are thus preserved and the dynamics are correct.

ParRep for Driven Systems

This technique can by applied to study nanoscale friction, as measured by a Friction Force Microscope (FFM).



ParRep for Driven Systems

ParRep for Driven Systems

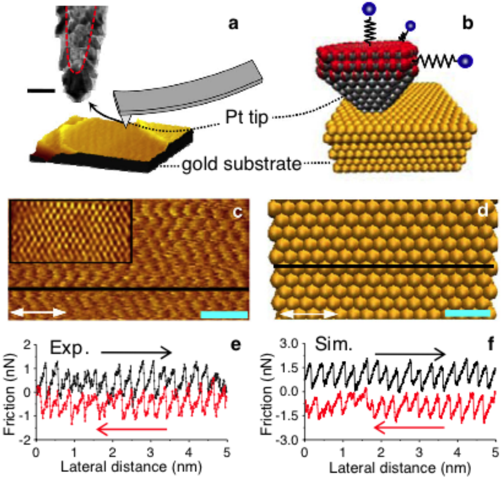
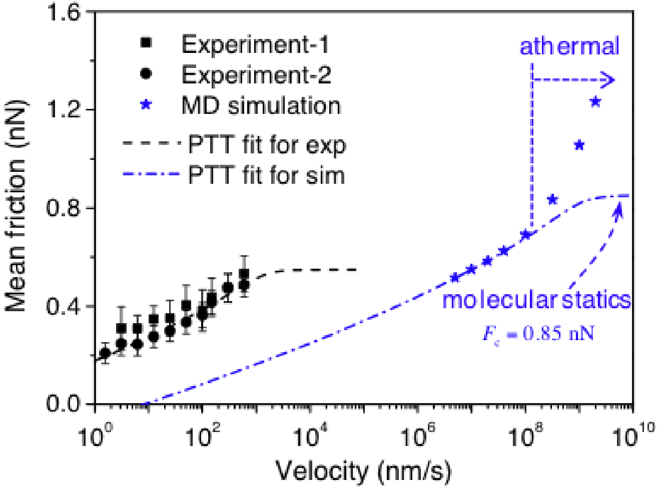


Figure: Li, Dong, Perez, Martini, Carpick, PRL 106, 12 (2011)

ParRep for Driven Systems



Plan

4 Conclusion

Practical Considerations

How do the different methods compare?

- Accuracy: ParRep $>$ Hyper $>$ TAD
- Simplicity: ParRep $>$ TAD $>$ Hyper
- Flexibility: ParRep $>$ TAD $>$ Hyper
- Acceleration: TAD $>$ Hyper $>$ ParRep (might not be true for long)

ParRep is often the best starting point when approaching AMD method for the first time.

Practical Considerations

Can (should) I apply AMD method to my system? Yes if:

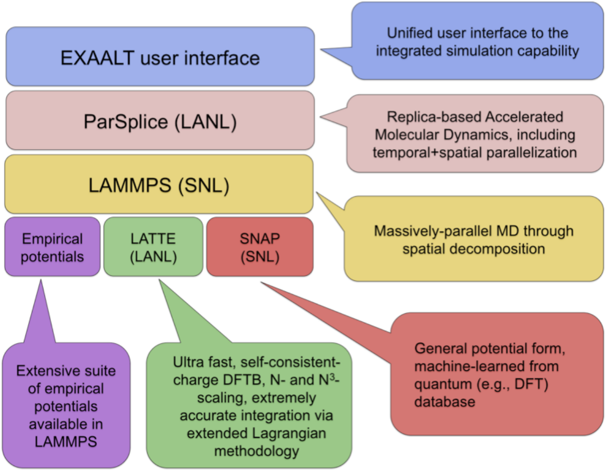
- It is a rare event system (typical transition times $>$ ($\gg$) 100 ps)
- Transitions can be automatically detected and characterized. (often means your energy landscape needs to be sufficiently smooth)
- You are typically interested in paths containing more than one transition.
- You are interested in dynamics, not just in sampling.
- You can afford to run MD for long enough to see events, detect transitions, do NEBs, etc.

Available tools

- LAMMPS (http://lammps.sandia.gov/): ParRep, TAD
- DL POLY (www.ccp5.ac.uk/): TAD, Hyperdynamics

Coming soon

EXAALT: a massively parallel AMD/MD/QMD simulation tool.



EXAALT user interface

Unified user interface to the integrated simulation capability

ParSplice (LANL)

Replica-based Accelerated Molecular Dynamics, including temporal+spatial parallelization

LAMMPS (SNL)

Massively-parallel MD through spatial decomposition

Empirical potentials

LATTE (LANL)

SNAP (SNL)

Extensive suite of empirical potentials available in LAMMPS

Ultra fast, self-consistent-charge DFTB, N- and N^3-scaling, extremely accurate integration via extended Lagrangian methodology

General potential form, machine-learned from quantum (e.g., DFT) database

Open-source: available soon on gitlab.com.

References

Overall review of AMD: Annual Reports in Computational Chemistry 5, 79-98 (2009)

ParRep:

- Original paper: Physical Review B 57, R13985 (1998)
- Review: Computational Materials Science 100, 90-103 (2015)

TAD:

- Original paper: The Journal of Chemical Physics 112, 9599-9606 (2000)
- Review: Annual review of chemical and biomolecular engineering 7, 87-110 (2016)

Hyper:

- Original paper: Physical Review Letters 78 (20), 3908 (1997)
- Pedagogical intro and review by K. Fichthorn: Computational Materials Science 100, 104-110 (2015)

Current Challenges

The main challenge facing AMD methods is the low barrier problem, i.e., the acceleration is limited by the fastest processes. This limit our ability to escape from meta-basins.

Partial solutions:

- State-bridging hyperdynamics: bias over low barriers (Miron and Fichthorn)
- Super-state ParRep: lump shallow states together
- Synthetic-TAD: use KMC to treat events we know about, use TAD to discover new events

Recent approaches based on speculation (SpecTAD, ParSplice) are vey promising ways to address this challenge (see my talk tomorrow). However, truly general solutions has yet to be developed.

Conclusion

- AMD methods can provide considerable acceleration of systems where the dynamics is activated, providing insight on the long-time behavior of materials.
- AMD methods (ideally) do not require *a priori* knowledge about the important processes.

Conclusion

QUESTIONS?