Introduction to Molecular Dynamics and Accelerated Molecular Dynamics

Danny Perez

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

Plan

Plan

1. Molecular Dynamics 101
   - Motivation
   - Interatomic Potentials
   - Boundary Conditions
   - Integrators
   - Ensembles
   - Space and Time Limitations of MD

Molecular dynamics : What ?

Say you want to compute a certain property A, function of the positions and momenta of an atomistic system, e.g., a given canonical average.

One possibility is to approximate :

$$\bar{A} = \frac{\int d\mathbf{x} d\mathbf{p} \exp[-\beta E(\mathbf{x},\mathbf{p})] A(\mathbf{x},\mathbf{p})}{\int d\mathbf{x} d\mathbf{p} \exp[-\beta E(\mathbf{x},\mathbf{p})]} \tag{1}$$

E.g., given a black-box that samples according to the canonical measure :

$$\bar{A} \simeq \frac{1}{N} \sum_{i=1}^{N} A(\mathbf{x}_i, \mathbf{p}_i) \tag{2}$$

This is the **Monte Carlo** way.

Molecular dynamics : What ?

The other possibility is to use time averages of a dynamics that samples the same measure (e.g. Langevin dynamics), i.e. :

$$\bar{A} = \frac{1}{T} \int_0^T dt\ A(\mathbf{x}(t), \mathbf{p}(t)) \tag{3}$$

E.g., given a black-box that generates trajectories from that dynamics :

$$\bar{A} \simeq \frac{1}{N} \sum_{i=1}^{N} A(\mathbf{x}(t_i), \mathbf{p}(t_i)) \tag{4}$$

This is the **Molecular Dynamics** way.

Molecular dynamics : What ?

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

E.g., solve Newton's equations of motion :

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

for 1000 atoms interacting through a Lennard-Jones potential, initially arranged as an fcc crystal in a cubic periodic cell.

Molecular dynamics : Why ?

MD is the largest scale model that gives unbiased dynamics $[\mathbf{x}(t), \mathbf{p}(\mathbf{t})]$ in full atomistic detail.

MD :

- is simple
- is "exact" for classical dynamics (with respect 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 ?

- Structure factor in liquids
- Thermal expansion coefficient
- Free energies
- Defect diffusion : processes and kinetics
- Melting point
- Thermal conductivity
- Yield/instability point
- Chemical reaction rates
- Shock wave propagation
- Molecular conformations
- Phonon dispersion
- Elastic properties
- Plastic properties
- Structure of solvation layers
- Evolution of collision cascades
- Things you know you don't know
- Things you don't know you don't know

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 :

- test higher-level models
- parameterize higher-level models
- simply see what happens !

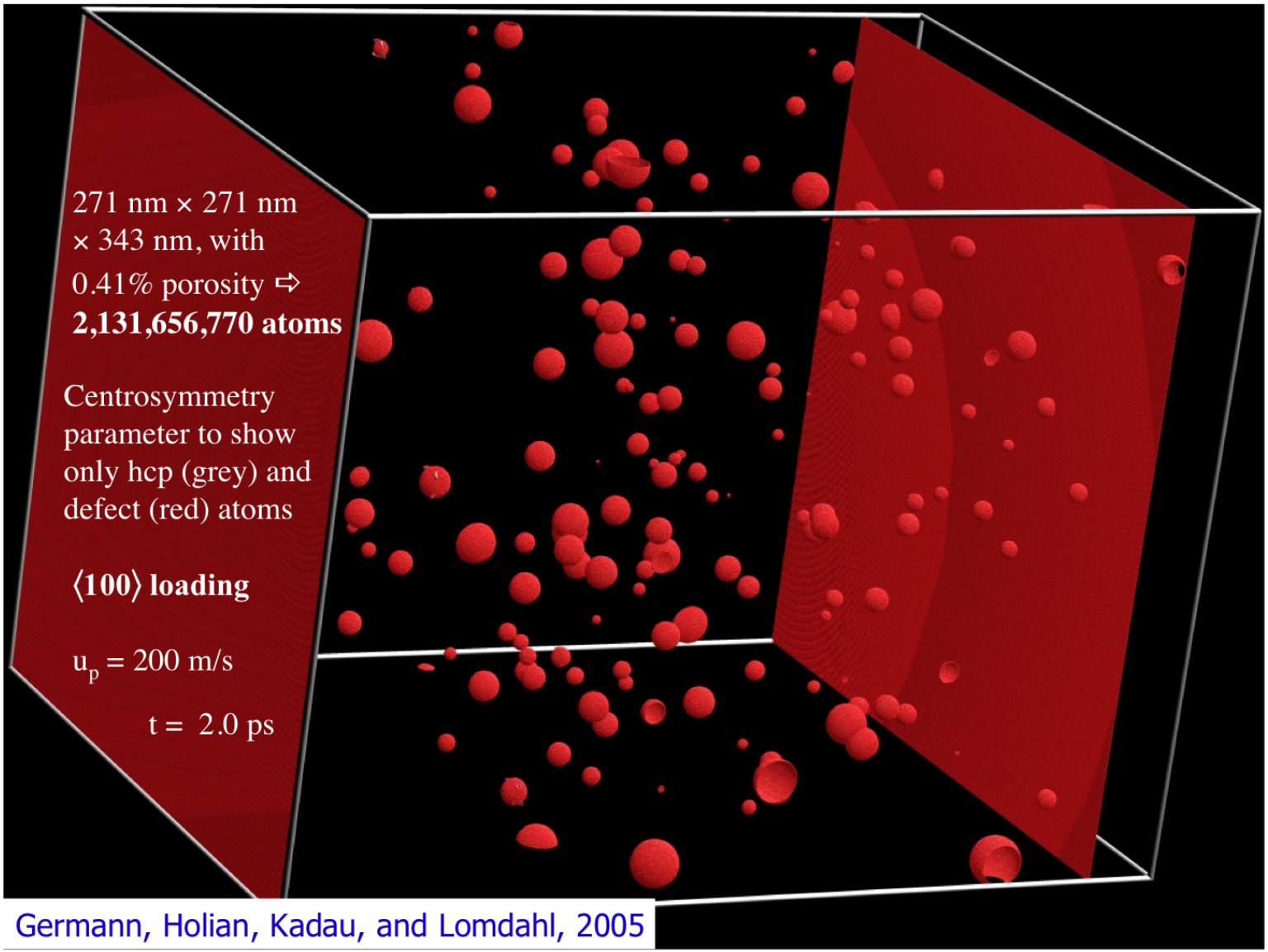
Often, MD simulations show unexpected and surprising results and drive the development of new theories.

MD in Action

MD in Action

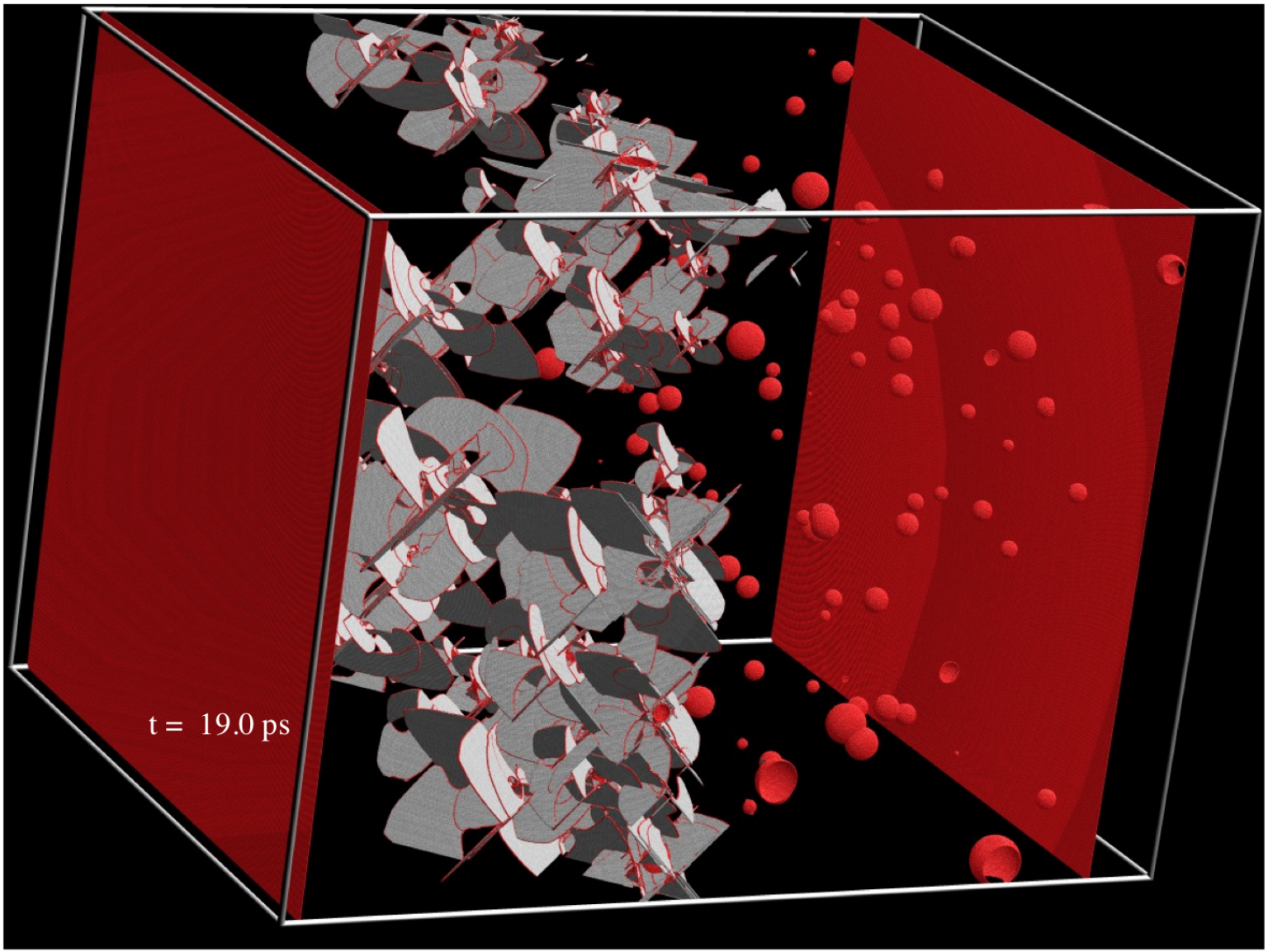
MD in Action

Molecular dynamics : Why ?



Germann, Holian, Kadau, and Lomdahl, 2005

Molecular dynamics : Why ?



MD in Action

Interatomic Potentials

The practical usefulness of MD relies on good (and cheap) interatomic potentials.

In nature, V comes from electronic interactions. In MD, these details are typically swept under the rug.

While *ab initio MD* is possible, its scope is extremely limited. One instead mostly relies on *empirical* approximation to the true quantum-mechanical potential, i.e., on explicit functions of the positions of all atoms $V(\mathbf{r})$.

Interatomic Potentials

Generally speaking, we would like potentials to be :

- accurate : important properties (elastic constants, phase stability, defect formation energies, etc.) are well described
- transferable : accurate for different phases, surfaces, crystal structures, etc.
- short-ranged : good efficiency
- functionally simple : smooth and avoids overfitting

To fit high-quality potentials, one has to compromise between these conflicting requirements.

Interatomic Potentials

The most simple interatomic potentials are *pair potentials* :

$$V = \frac{1}{2}\sum_{i \neq j}^{N} \phi(r_{ij}). \tag{5}$$

Probably the most well-known in this category is the Lennard-Jones (LJ) 6-12 :

$$\phi(r_{ij}) = 4\epsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right] \tag{6}$$

where σ sets the length scale, and ϵ the energy scale. The LJ is appropriate for rare gases, but it is also frequently used as a generic potential.

Interatomic Potentials

Another very popular form is the Embedded Atom Method (EAM).

$$V = \sum_i^N V_i$$

$$V_i = \frac{1}{2} \sum_j \phi(r_{ij}) + F(\bar{\rho}_i)$$

$$\bar{\rho}_i = \sum_j \rho(r_{ij})$$

Physically, you compute the energy required to embed an atom in a background distribution of electrons and nuclei.

When F is non-linear, EAM is a *many-body* potential. EAM works especially well for FCC metals.

Interatomic Potentials

Some materials have strong angular components to their interactions (e.g., Si). These can be accounted for through angular terms :

$$V_i = \sum_j \phi(r_{ij}) + \sum_{j,k} f(\theta_{ijk})$$

Higher order terms or collective many-body terms used (such as bond order terms, or embedding of complex densities) can be added as necessary.

Interatomic Potentials

Potentials come in many shapes and forms and are often specialized for specific classes of materials.

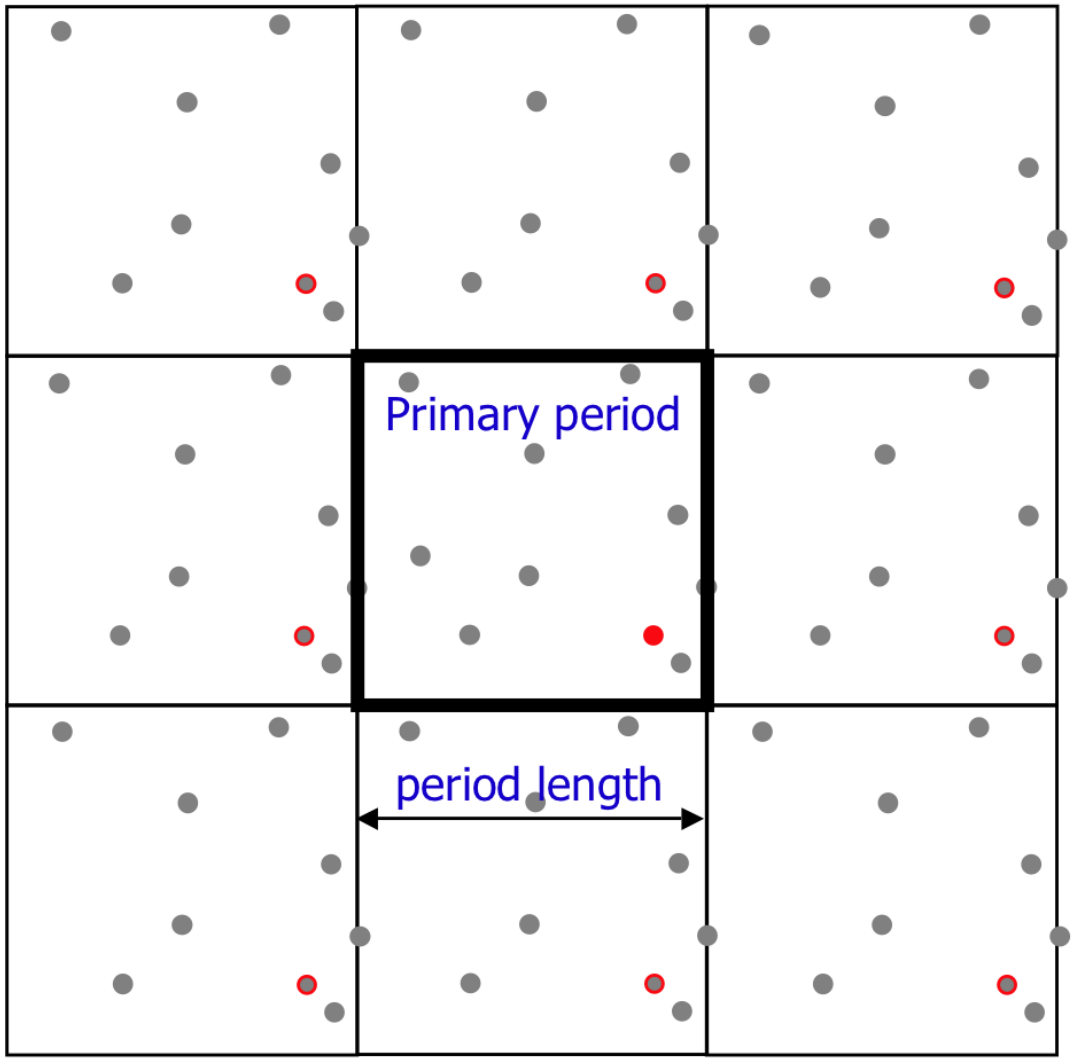
- Bond covalency, angular terms
  - Molecular mechanics (proteins, polymers)
  - Stillinger-Weber (C, Si)
  - Tersoff (C, Si)
  - modified EAM
  - Bond-order
- Ionic systems
  - Buckingham (oxides)
- Charge Transfer
  - ReaxFF

Boundary conditions

Boundary condition are used to specify the environment in which the atoms evolve. These include :

- Open
- Periodic
- Fixed atoms
- Directional constraints (along a line, in a plane, etc.)
- Soft/Hard walls
- Driven (time-dependent)
- Isobaric
- Thermal (including sources and sinks)
- Multiscale (coupling to larger length-scale models)
- Combinations of these

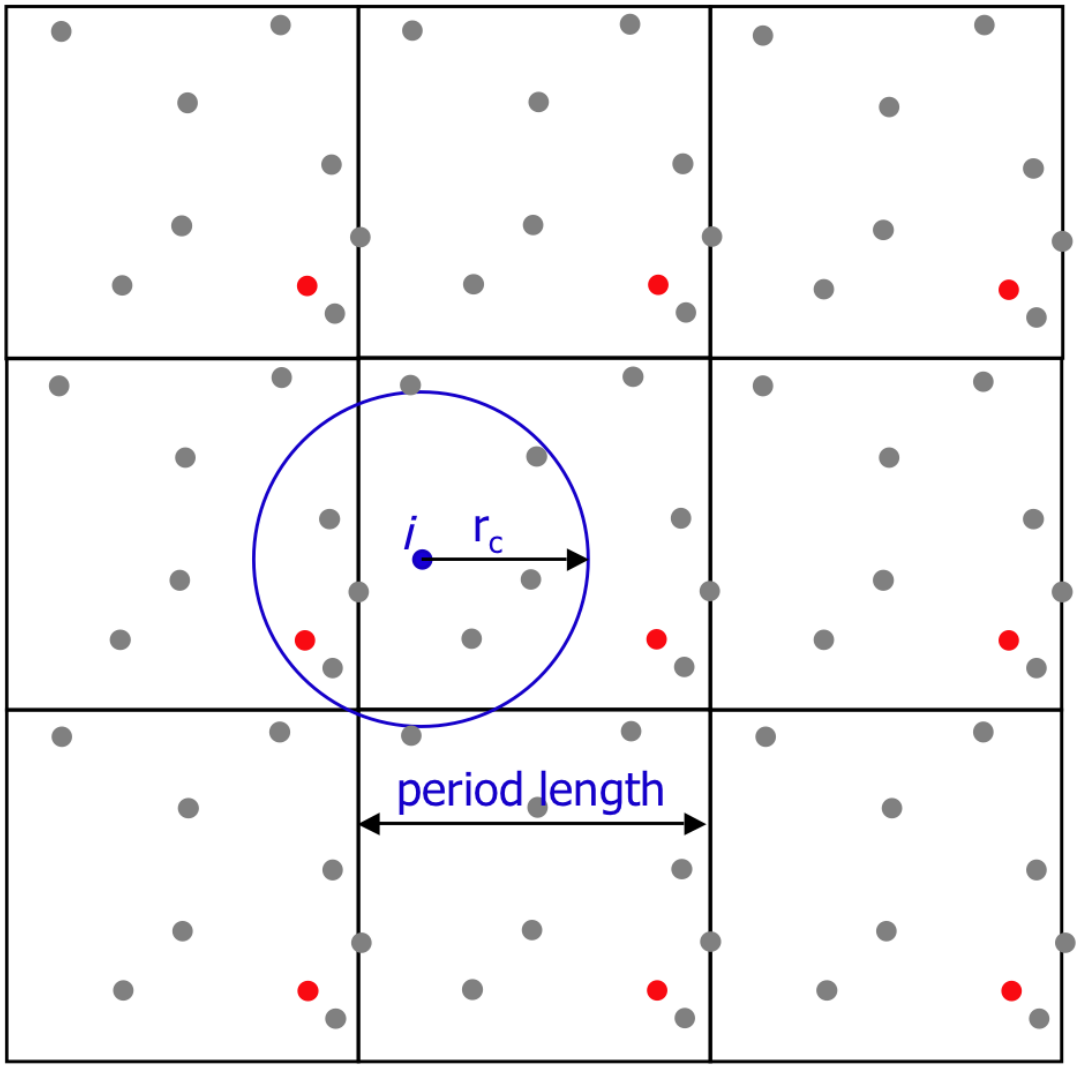
Boundary conditions : Periodic



Periodic boundary conditions are extremely common when simulating bulk systems.

They can be thought of as involving tiling of a primary cell.

Boundary conditions : Periodic



One then adopts the minimum-image convention : atoms interact only with the closest image of other atoms.

Boundary conditions : Time-dependent

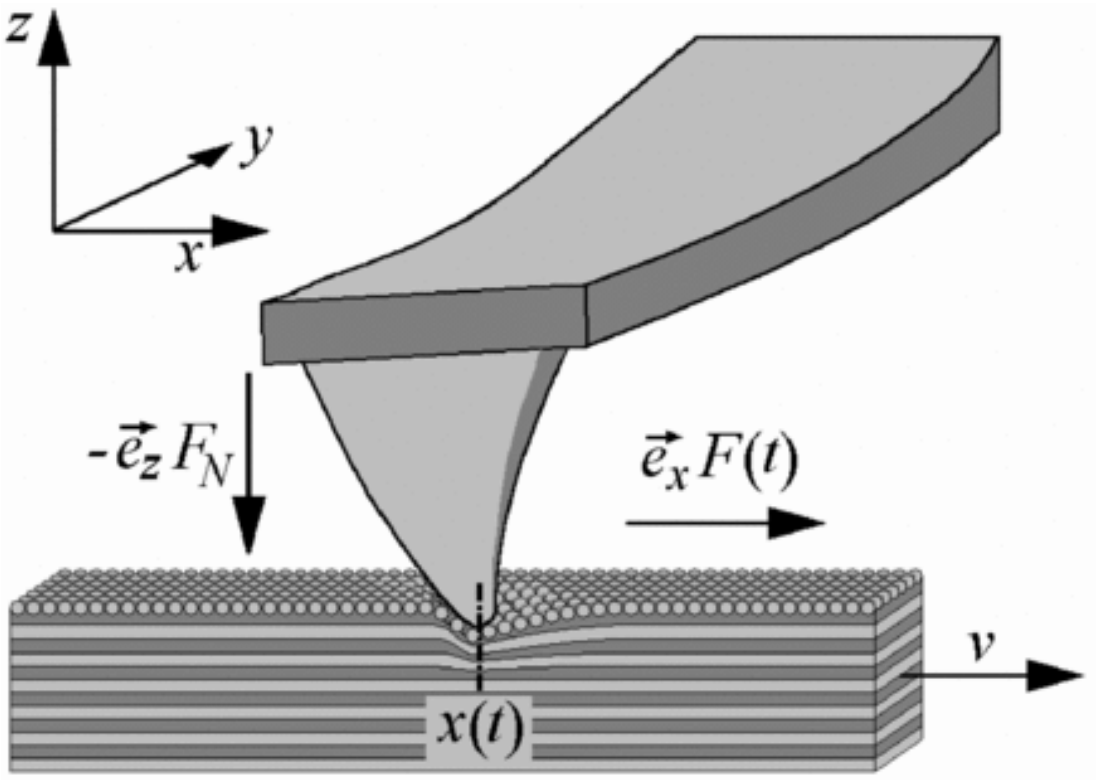
Boundary conditions are not always fixed for the duration of the simulation.

They might evolve according to their own equations of motion :

- Parrinello-Rahman method : basis vectors of the cell evolve to achieve a given target pressure.
- Coupling to finite-element models (e.g., to account for far-field elasticity during crack propagation).
- Imposition of mechanical loading (e.g., displacement- or stress-driven actuators).

Friction Force Microscopy

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



Friction Force Microscopy

Friction Force Microscopy

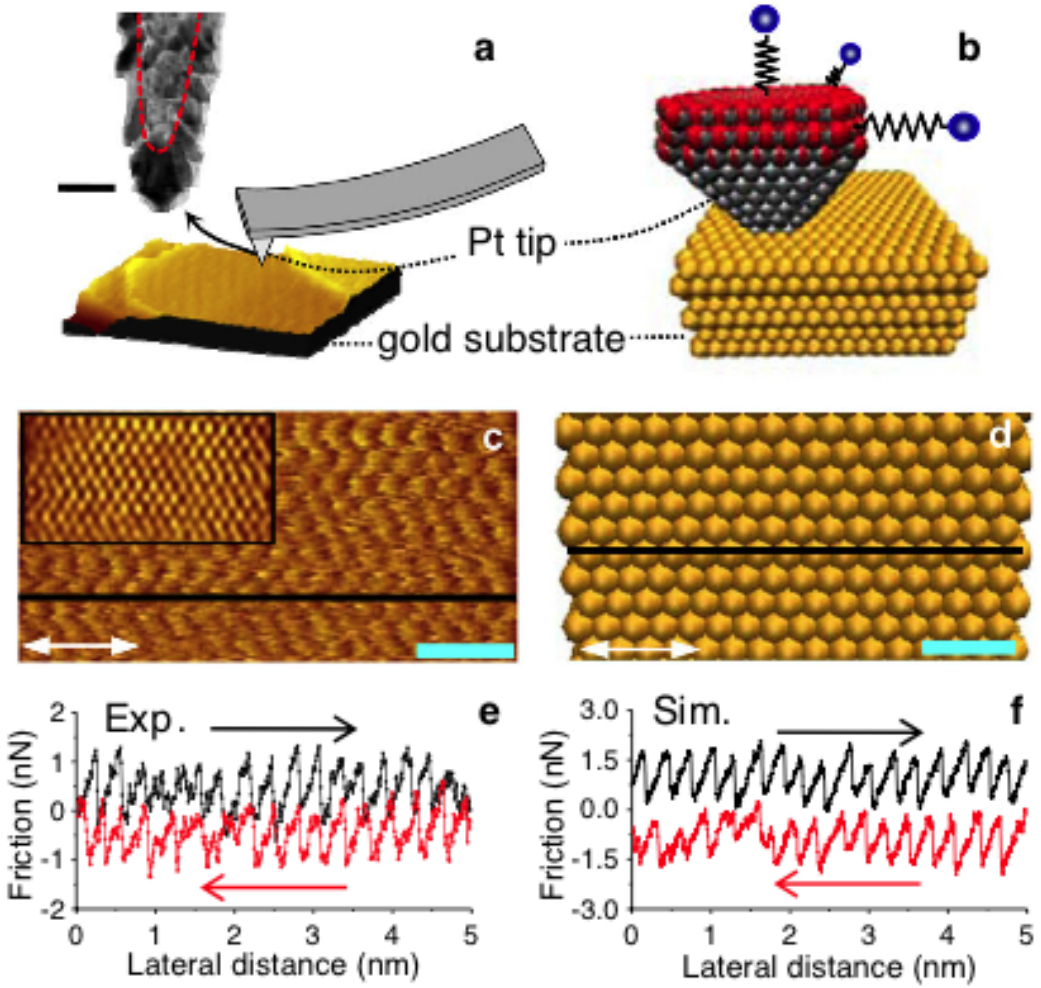


FIGURE: Li, Dong, DP, Martini, Carpick, PRL 106, 12 (2011)

Numerical integration : The Verlet algorithm

MD requires the numerical solution of the equations of motion.

The most common integration algorithm in the micro-canonical ensemble is the velocity-Verlet algorithm.

Given $x(t), p(t)$, do :

$$
\begin{aligned}
p(t + \Delta t/2) &= p(t) - \nabla V(x(t))\Delta t/2 \\
x(t + \Delta t) &= x(t) + p(t + \Delta t/2)\Delta t/m \\
p(t + \Delta t) &= p(t + \Delta t/2) - \nabla V(x(t + \Delta t))\Delta t/2
\end{aligned}
$$

to obtain $x(t + \Delta t), p(t + \Delta t)$. Repeat as needed (i.e., often !).

Numerical integration : The Verlet algorithm

The Verlet algorithm is :

- Simple
- Cheap : one force per timestep
- Globally second order accurate
- Symplectic : Time reversible, approximatively conserves energy for extremely long times
- Typical timestep : 2-6 fs for metals, 0.5-1 fs for systems containing light atoms. Stability set by the fastest vibrational motion.
- Formally interesting (shadow Hamiltonians)

While more sophisticated algorithms (e.g., higher order) do exist, the Verlet algorithm is usually the integrator of choice.

Numerical integration : Ensembles

Dynamics generated by the Verlet algorithm sample an NVE (micro-canonical) measure.

Other integrators allow for sampling of different ensembles, e.g. :

- Canonical (NVT) : Langevin, Andersen, Nose-Hoover
- Isothermal-Isobaric (NPT) : Parrinello-Rahman
- Grand-canonical (μVT) : Pettitt

It is important to remember that the dynamics in these other ensembles is not Newtonian. One must then be careful in how dynamical details are interpreted.

Numerical integration : Canonical Ensemble

One of the most common ensemble in practice is the canonical (NVT) ensemble.

A robust way to sample this ensemble is the Langevin thermostat, where the equations of motion become :

$$\ddot{x} = -\frac{1}{m}\nabla V - \alpha\dot{x} + \frac{A(t)}{m}$$

The second and third terms on the RHS correspond to a viscous drag and to a random thermal force, respectively.

This physically corresponds to having the system immersed in a bath of smaller particles evolving at temperature T.

Numerical integration : Canonical Ensemble

The friction α and the temperature T are related through the fluctuation-dissipation theorem :

$$\langle A_i(t)A_j(t+\Delta t)\rangle = 2\alpha m k_B T \delta(\Delta t)\delta_{ij}.$$

The Langevin method is very appealing because :

- It is simple
- It is provably canonical and ergodic
- The strength of the coupling with the thermostat is adjustable
- The thermalization dynamics quantifiable (at least in the harmonic approximation)
- It is formally interesting (synchronization)

Molecular Dynamics : cost and limits

This all sounds very nice, but how much does it really cost ?

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

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

For short-range potentials, MD is nominally linear scaling : the cost of a time-step is proportional to the number of atoms ($O(N \log N)$ for long-range). This means that, in principle, you can freely trade atoms for time, i.e., 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, i.e., you (usually) cannot start to work on the next timestep before you completed the previous one ! This means you can only distribute work within a timestep, not across.

This is usually achieved with spatial decomposition schemes : each processor handle the atoms in a given region of space. This is efficient only if each processor has enough work to do before communication is required.

The current scope of MD simulations

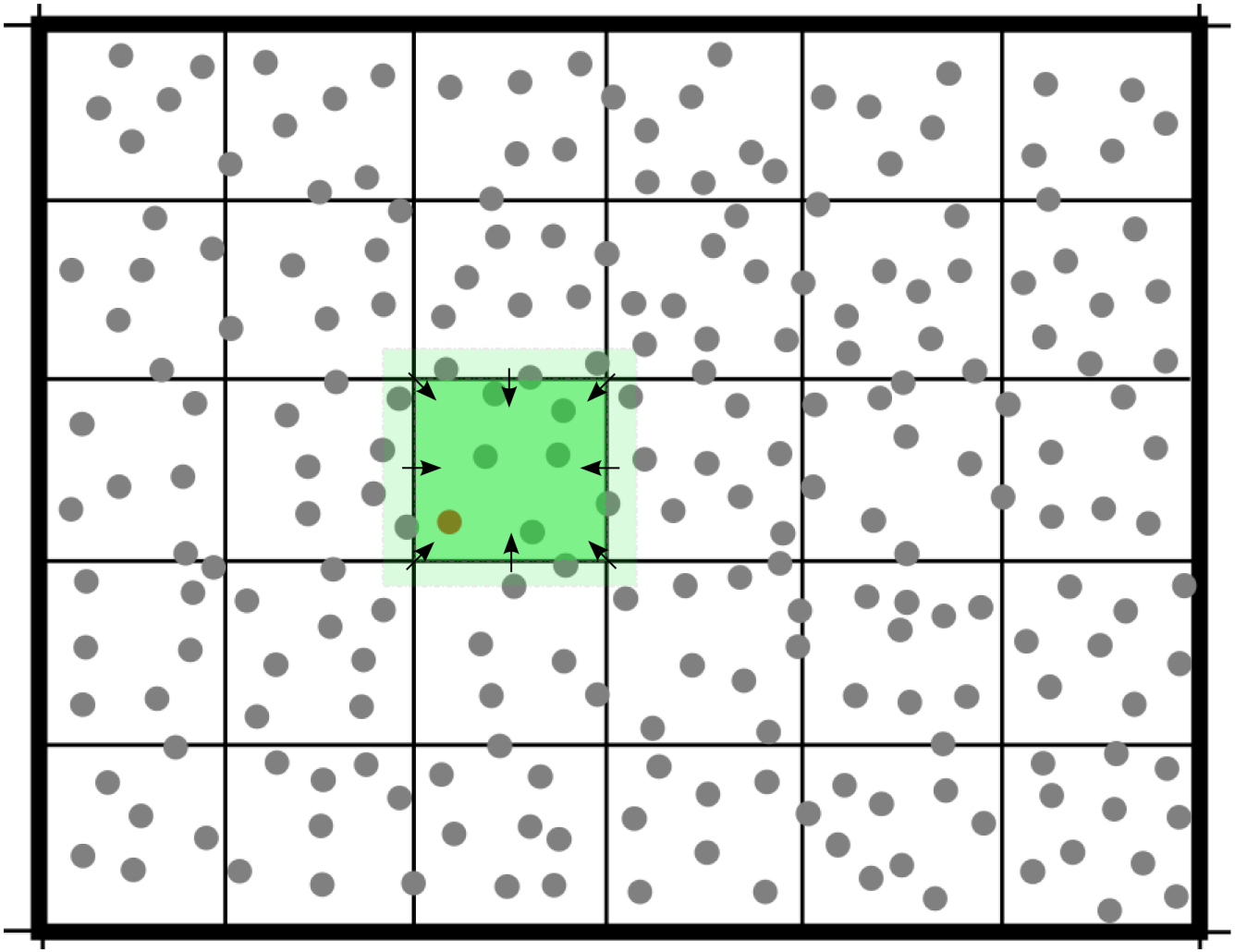


FIGURE: Spatial decomposition scheme : each cell is handled by a different processor.

Molecular Dynamics : cost and limits

This implies :

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

Molecular Dynamics : Spatial decomposition

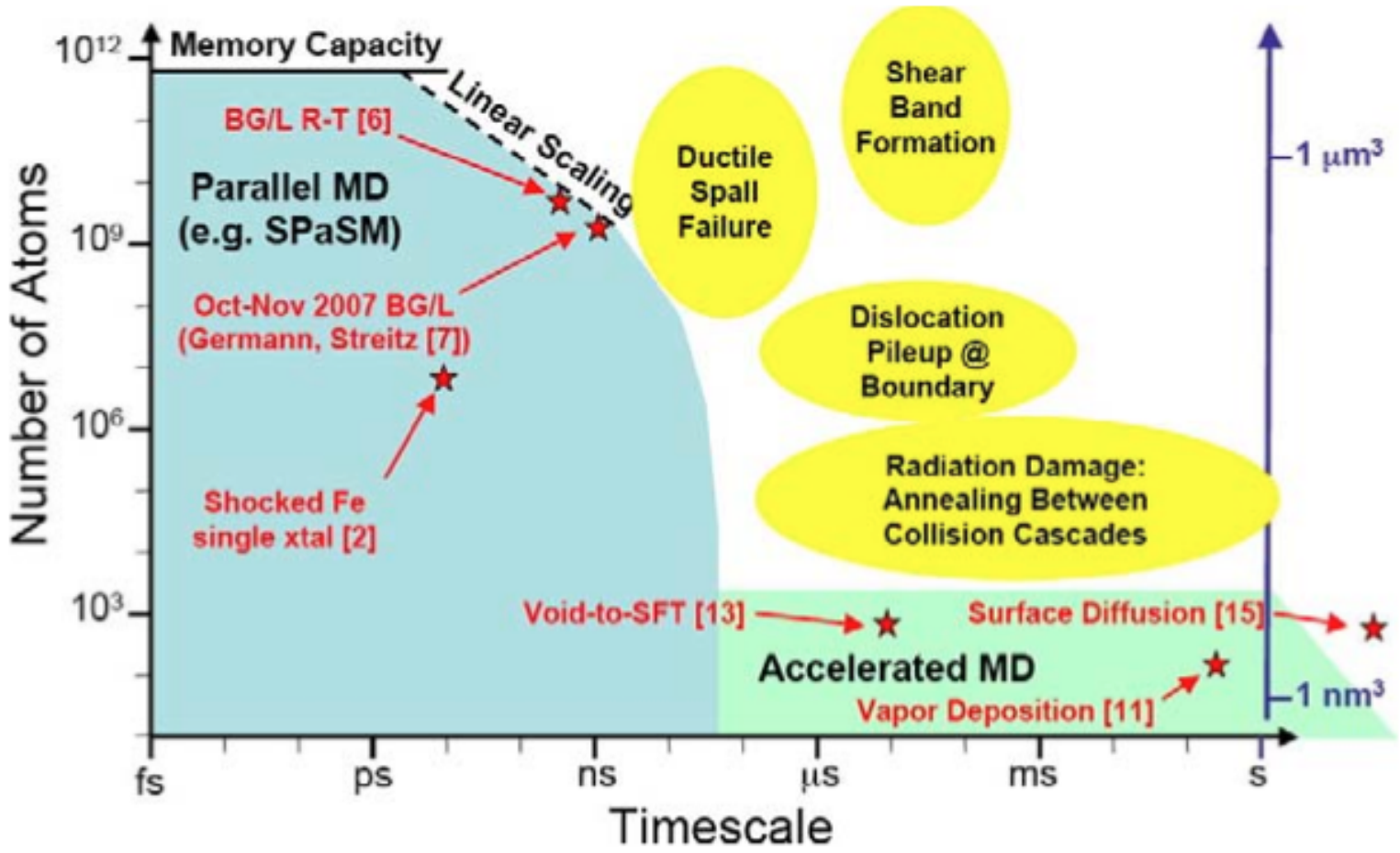
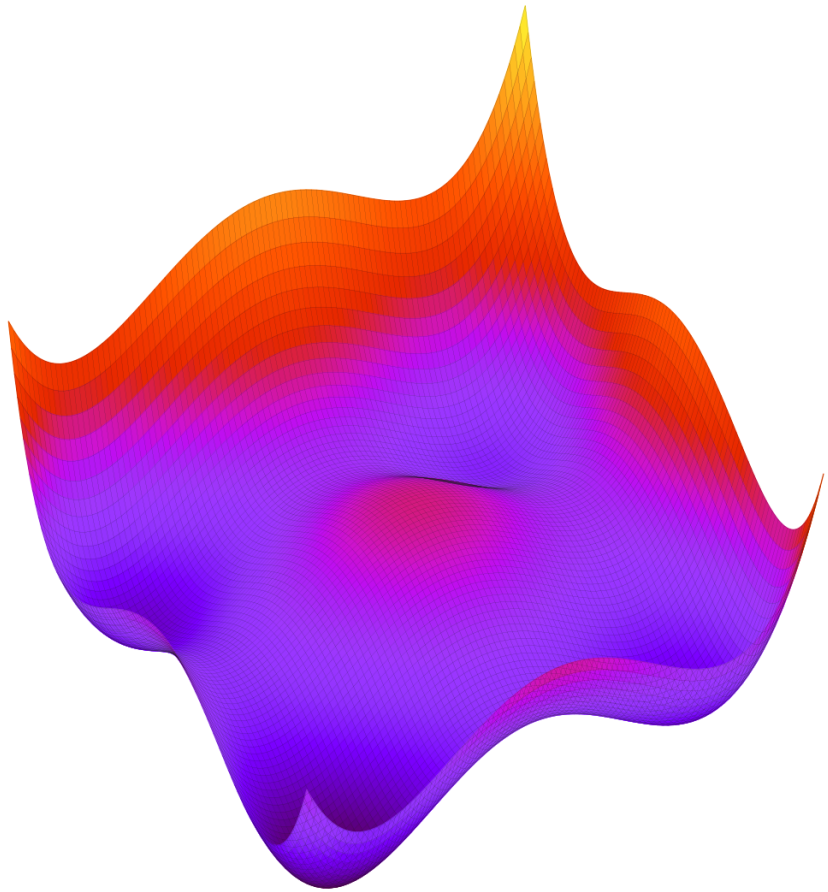


FIGURE: Scope of short-range, empirical potential MD simulations (blue region) given a PetaFlop super-computer.

Time and energy

Due to the rugged structure of their *potential energy landscape*, materials often possess a wide range of characteristic timescales.

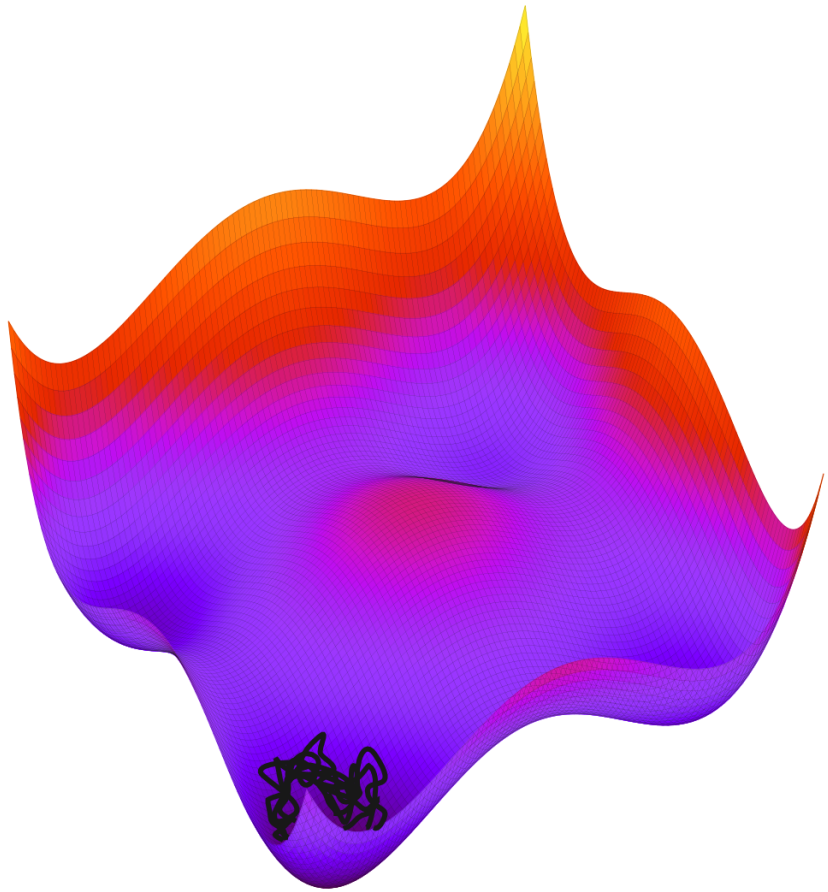
- Intra-basin Vibration/Relaxation timescale ($\tau_{\text{rel}} \sim \omega^{-1} \simeq 1ps$)
- Inter-basin Transition timescale ($\tau_{\text{trans}} \sim e^{\Delta E/k_B T} \tau_{\text{vib}} \simeq [1ps, \infty]$)
- Intra/Inter Meta-basin Relaxation/Transition timescales
- ...



Time and energy

Due to the rugged structure of their *potential energy landscape*, materials often possess a wide range of characteristic timescales.

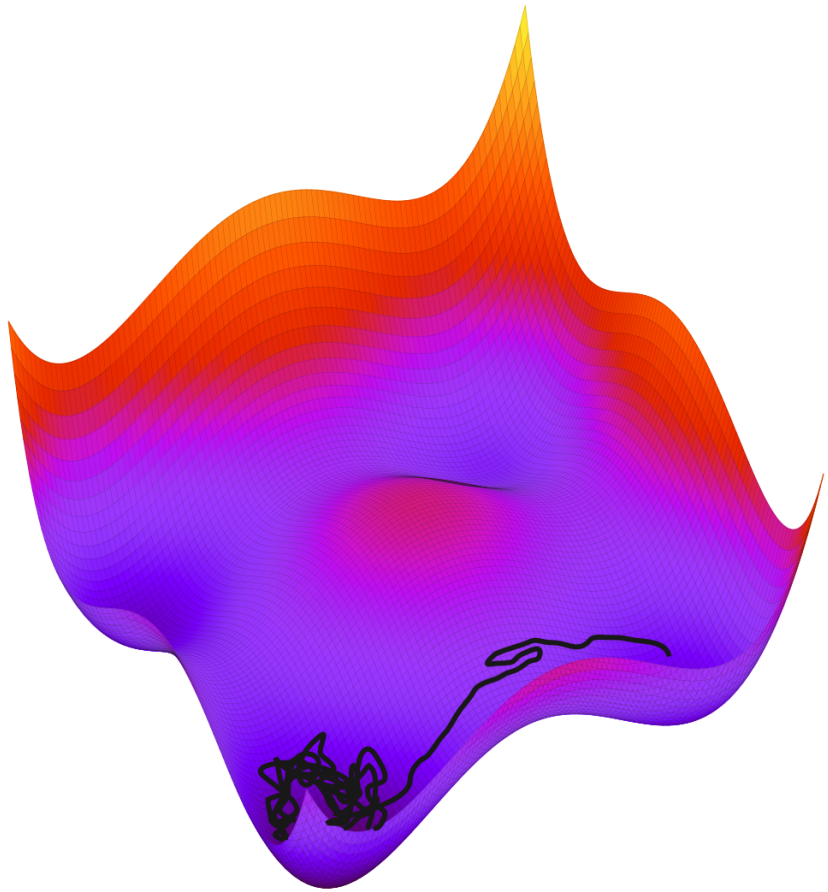
- Intra-basin Vibration/Relaxation timescale ($\tau_{\text{rel}} \sim \omega^{-1} \simeq 1ps$)
- Inter-basin Transition timescale ($\tau_{\text{trans}} \sim e^{\Delta E/k_B T} \tau_{\text{vib}} \simeq [1ps, \infty]$)
- Intra/Inter Meta-basin Relaxation/Transition timescales
- ...



Time and energy

Due to the rugged structure of their *potential energy landscape*, materials often possess a wide range of characteristic timescales.

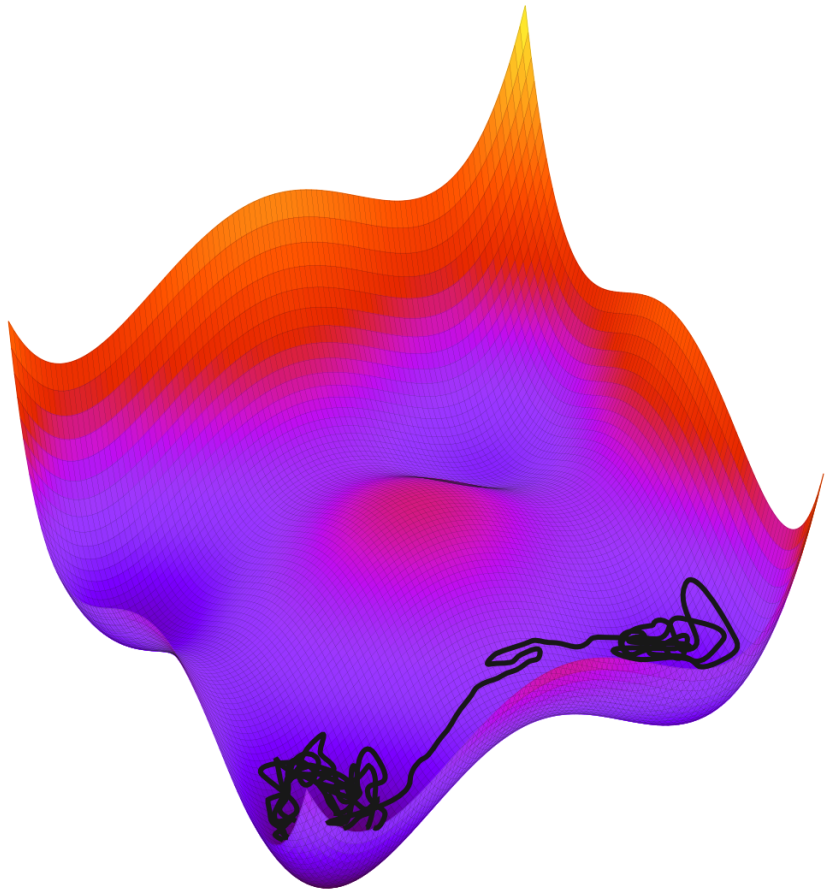
- Intra-basin Vibration/Relaxation timescale ($\tau_{\text{rel}} \sim \omega^{-1} \simeq 1ps$)
- Inter-basin Transition timescale ($\tau_{\text{trans}} \sim e^{\Delta E/k_B T} \tau_{\text{vib}} \simeq [1ps, \infty]$)
- Intra/Inter Meta-basin Relaxation/Transition timescales
- ...



Time and energy

Due to the rugged structure of their *potential energy landscape*, materials often possess a wide range of characteristic timescales.

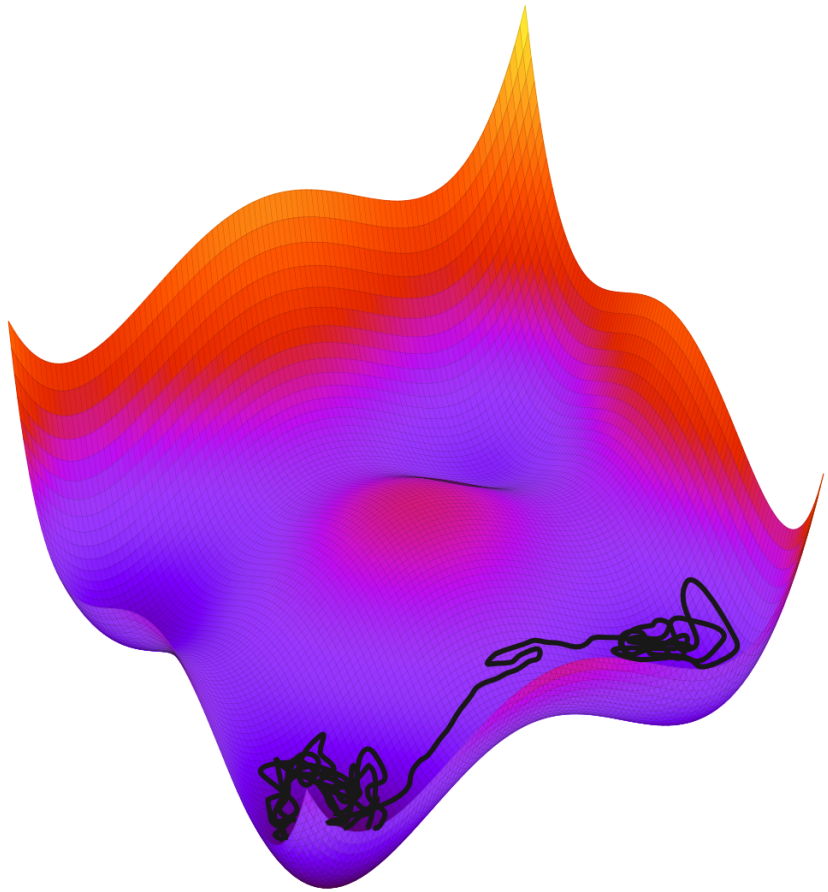
- Intra-basin Vibration/Relaxation timescale ($\tau_{\text{rel}} \sim \omega^{-1} \simeq 1ps$)
- Inter-basin Transition timescale ($\tau_{\text{trans}} \sim e^{\Delta E/k_B T} \tau_{\text{vib}} \simeq [1ps, \infty]$)
- Intra/Inter Meta-basin Relaxation/Transition timescales
- ...



Time and energy

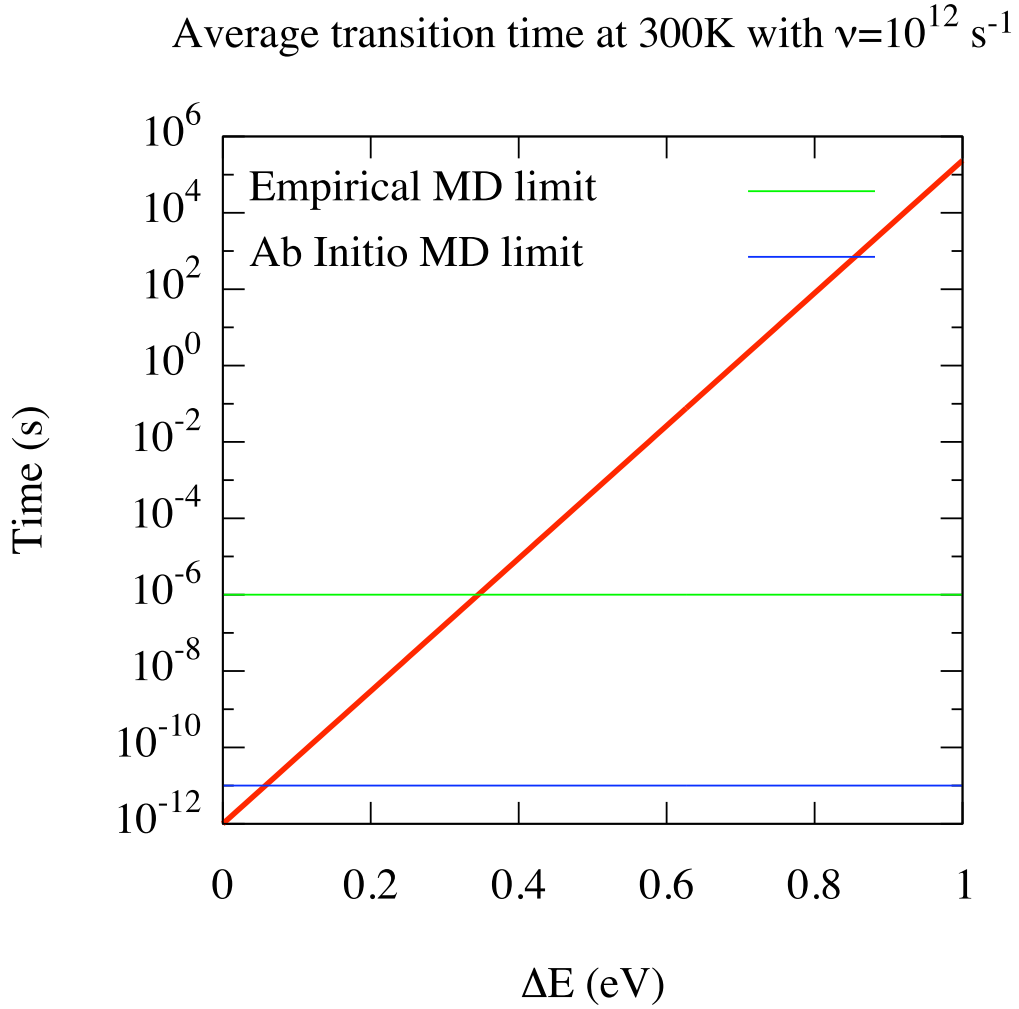
Due to the rugged structure of their *potential energy landscape*, materials often possess a wide range of characteristic timescales.

- Intra-basin Vibration/Relaxation timescale ($\tau_{\text{rel}} \sim \omega^{-1} \simeq 1ps$)
- Inter-basin Transition timescale ($\tau_{\text{trans}} \sim e^{\Delta E/k_B T} \tau_{\text{vib}} \simeq [1ps, \infty]$)
- Intra/Inter Meta-basin Relaxation/Transition timescales
- ...



The Timescale Problem

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



MD's dirty little secret

People rarely (publicly) 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
- Carefully choose the "right" potential

Unfortunately, it is generally not easy to unbias the results of such overdriving.

MD as part of other methods

However, MD can often efficiently be used as the computational kernel of more sophisticated methods, such as :

- Umbrella sampling
- Parallel tempering
- Metadynamics
- Transition path sampling
- Accelerated Molecular Dynamics
- ...

Molecular Dynamics

- MD is an extremely powerful techniques that can generate *unbiased* atomistic trajectories, in a wide variety of conditions.
- These trajectories can be useful in themselves, or they can be used as part of more complex methods.
- MD's reach is limited in terms of **time** and length scales.

Plan

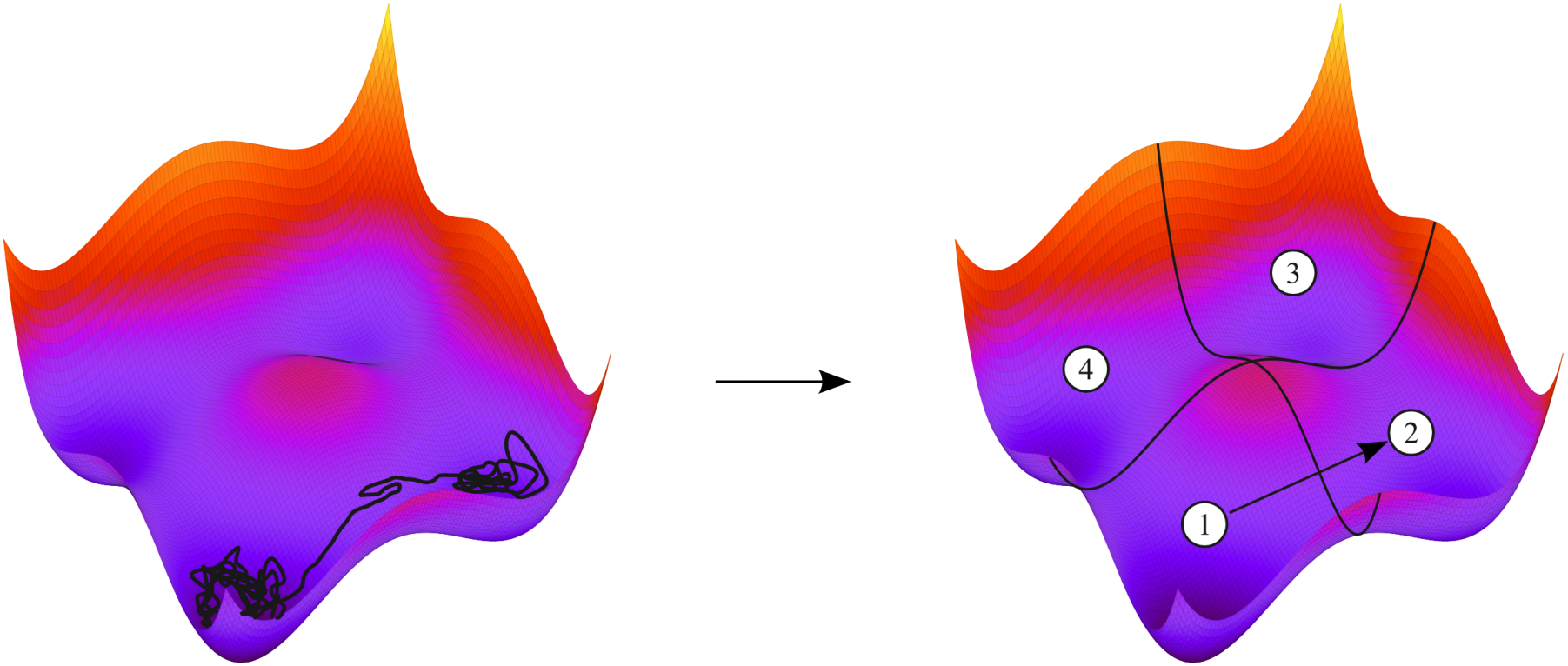
2. Accelerating the Dynamics
   - KMC
   - Temperature Accelerated Dynamics
   - Parallel Replica Dynamics

Possible solutions to the timescale problem

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

- The basic idea of acceleration methods is to exploit **separation of timescales** between fast and slow components of the dynamics (typically between τ_{rel} and τ_{trans}).
- If such a separation exists, it is sometimes possible to 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

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.

Kinetic Monte Carlo

KMC is a stochastic computational procedure that generates statistically correct state-to-state trajectories.

The main steps to setup a KMC simulation are to :

- Identify possible states and transitions pathways
- Compute rates for these transitions
- Randomly select a pathway with a probability proportional to its rate
- Sample an escape time from $p(t) = k_{\text{tot}} \exp(-k_{\text{tot}} t)$
- Repeat

KMC is very efficient, but it needs to know about **all** pathways.

Kinetic Monte Carlo : Finding transitions

To find pathways, one might :

- Guess the process and characterize the corresponding pathway with, e.g., a nudged elastic band. **If you think you can guess all important pathways, you are (probably) wrong.**
- Generate pathways with MD (typically at high-T)
- Use direct saddle point searching algorithms (e.g., the dimer method)

Generating complete (enough) catalogs is still a challenge, but adaptive KMC methods made considerable progress in the last few years.

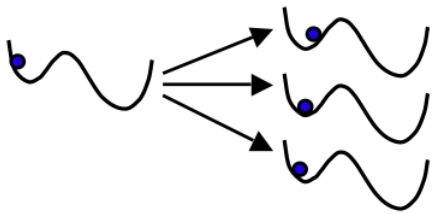
Accelerated Molecular Dynamics

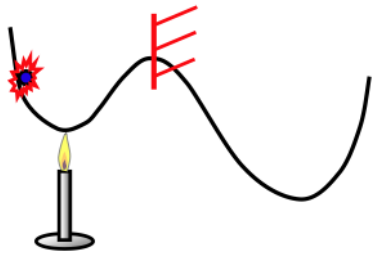
- Another possibility is to let MD trajectories do the hard work.
- The basic idea is that trajectories find appropriate way out (i.e., at rates proportional to the rate constant) without knowing about any of the escape paths except the first one it sees.
- In accelerated molecular dynamics (**AMD**), we try to maintain this key characteristic, while using statistical mechanics to trick MD into making escapes happen sooner.

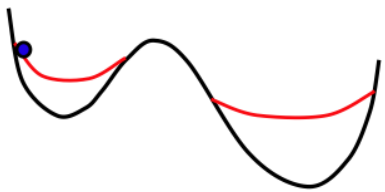
Accelerated Molecular Dynamics

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

- **Parallel Replica Dynamics (ParRep) :** use a parallel computer to parallelize the dynamics in the time domain.
- **Temperature Accelerated Dynamics (TAD)** : increase the temperature to explore more rapidly but select events that should have occurred at the original temperature.
- **Hyperdynamics :** modify the potential energy landscape to decrease the effective depth of potential wells while leaving the relative probability of each event unchanged.







Accelerated Molecular Dynamics : key points

Just like MD, but (often) better.

- Open-ended : specify an initial state and let the system evolve.
- Stochastic : different runs can give different answers.
- Useful for system evolving through sequences of rare events.
- (Ideally) require minimal *a priori* knowledge of the system.
- Typically more accurate than KMC.
- Typically not as efficient as KMC.
- Not optimal for sampling.
- (Currently) most useful for hard matter.

Principle of Temperature Accelerated Dynamics (TAD)

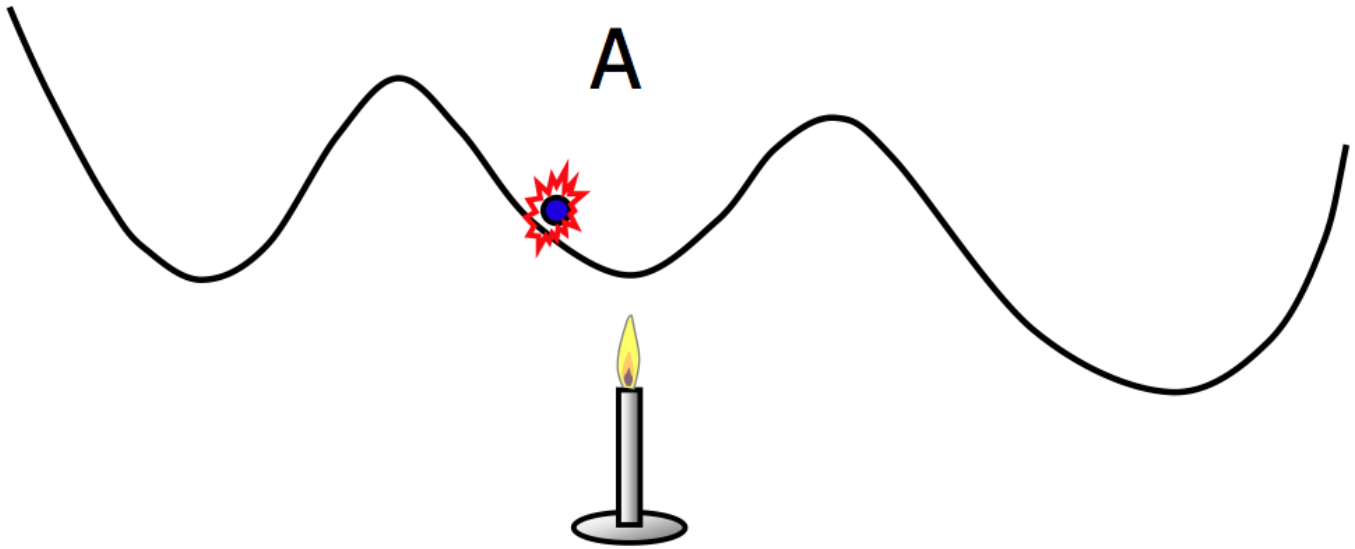
- One of the most common solution to a sluggish MD simulation is simply to increase the temperature.
- This indeed make things happen faster, but it also makes the **wrong** things happen.
- This is especially problematic when many processes compete.
- How to unbias the results ?

Principle of Temperature Accelerated Dynamics (TAD)

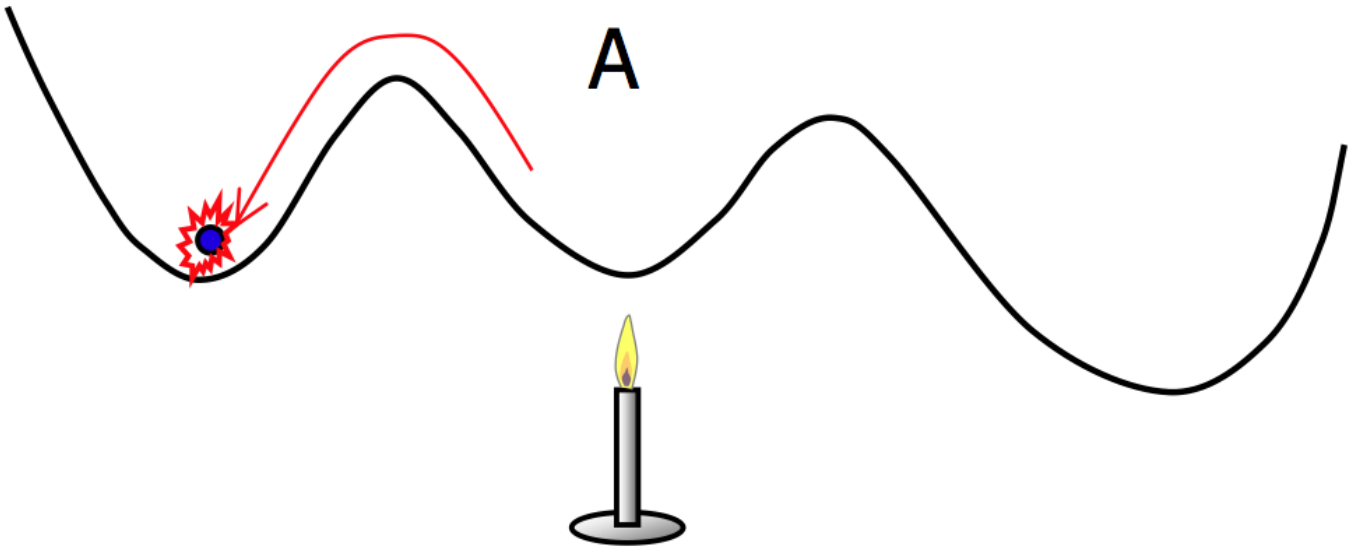
In TAD [Sorensen and Voter, JCP 112, 9599, (2000)], you :

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

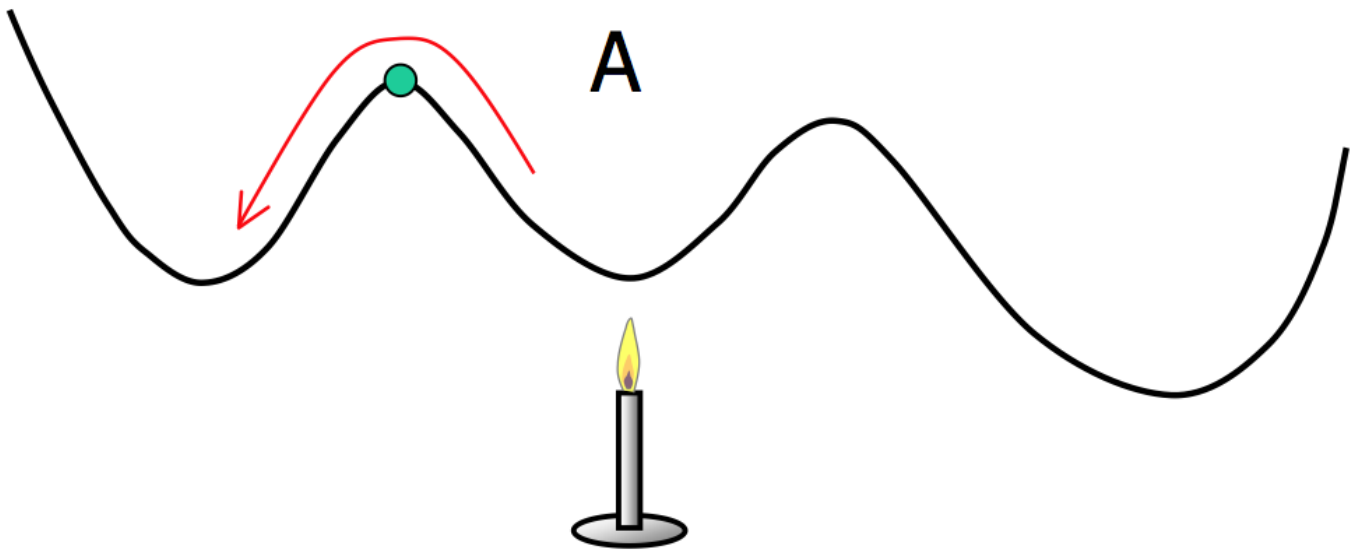
TAD : Illustrations



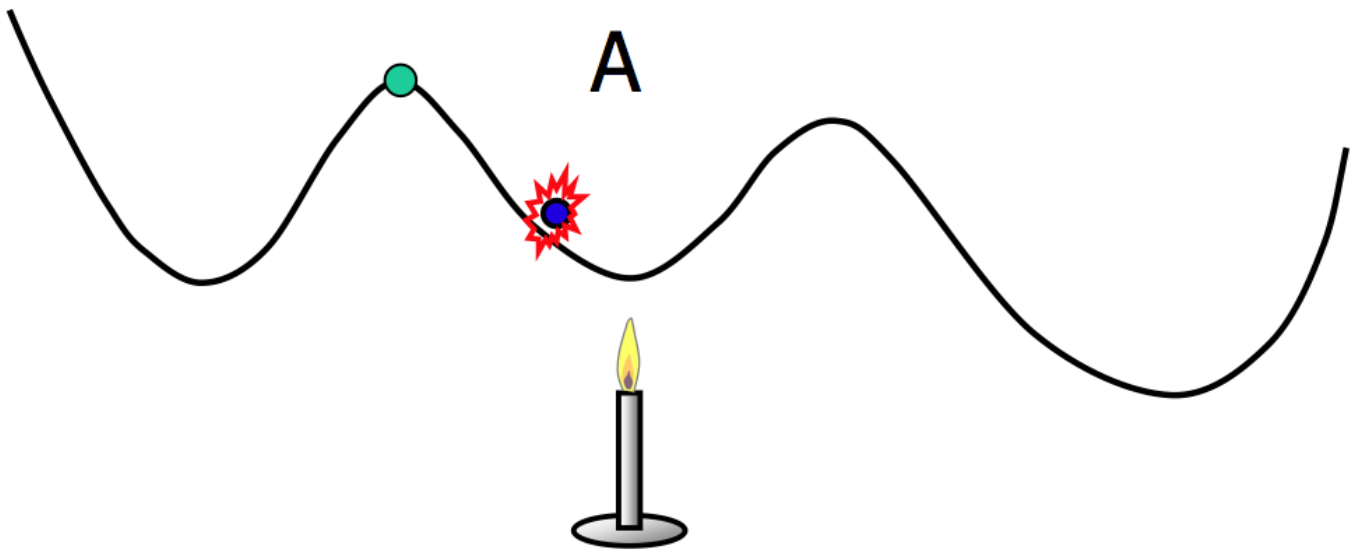
TAD : Illustrations



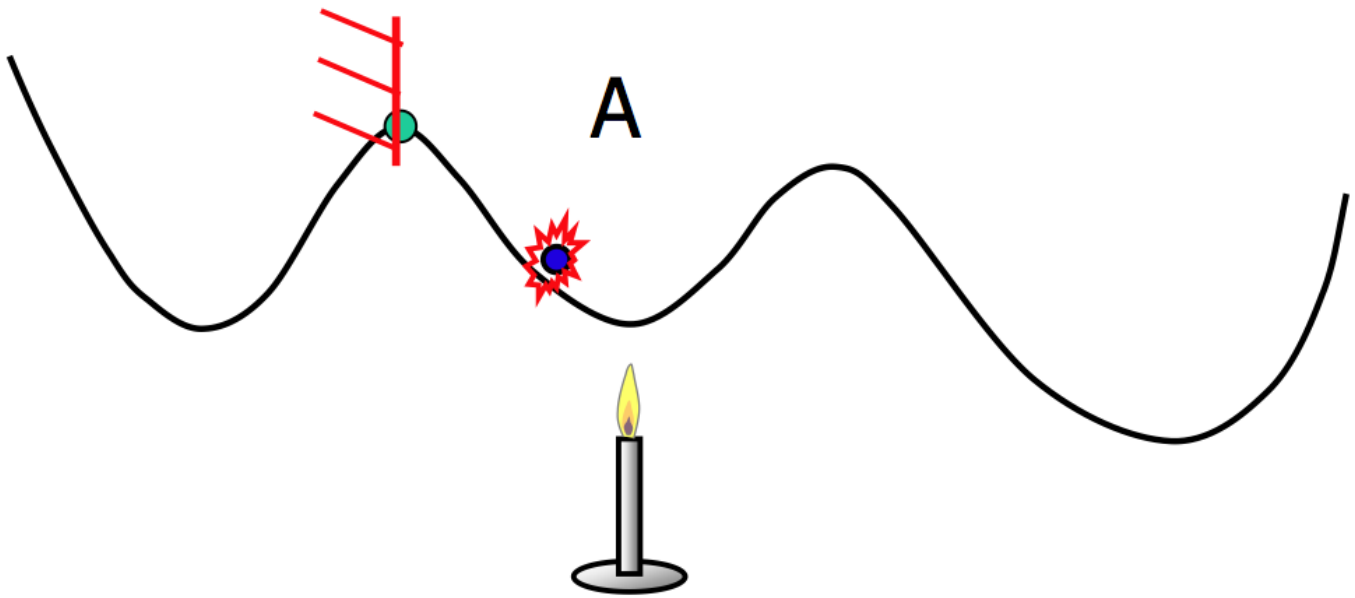
TAD : Illustrations



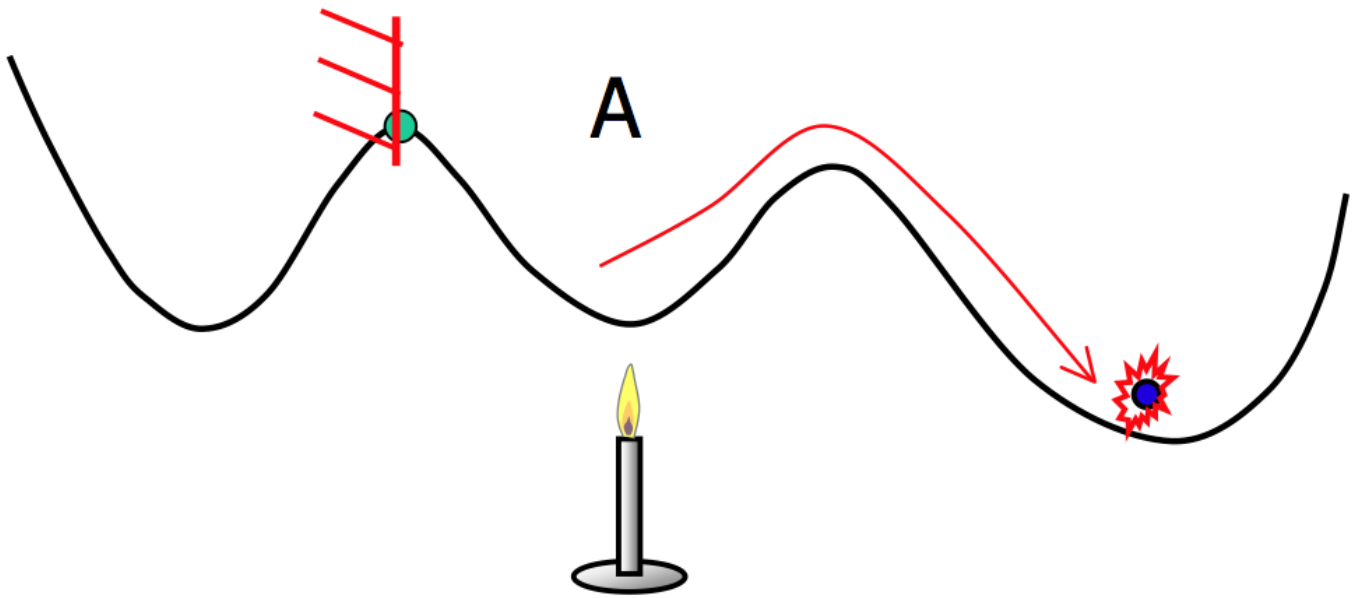
TAD : Illustrations



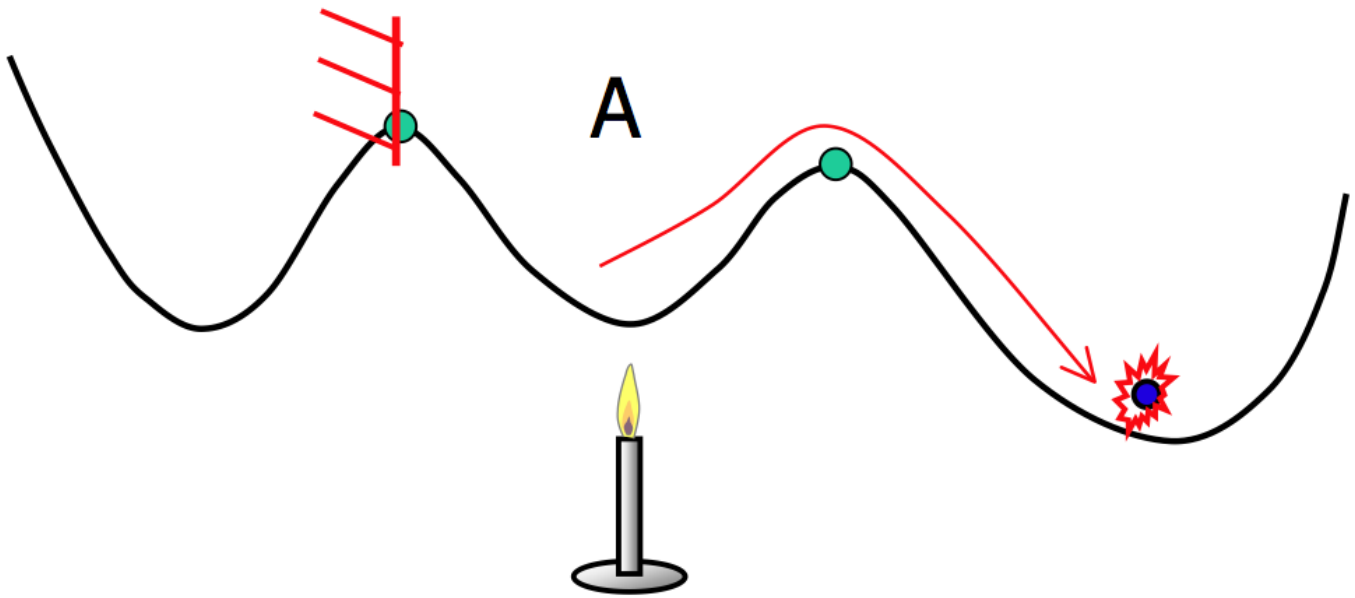
TAD : Illustrations



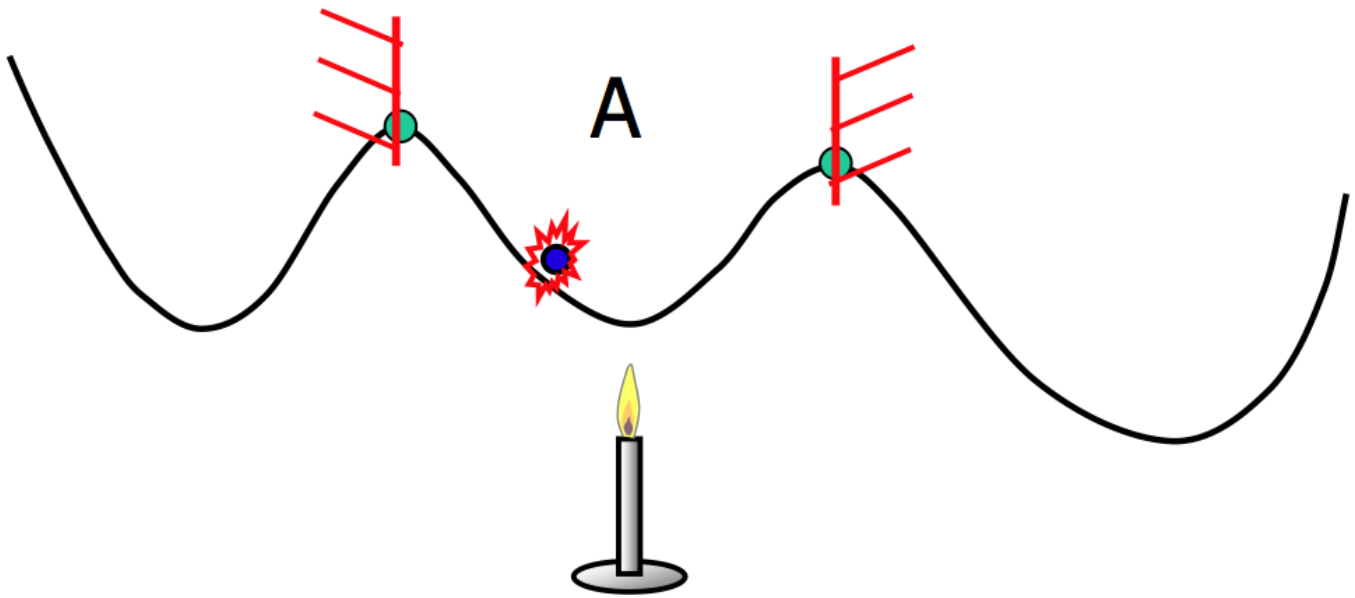
TAD : Illustrations



TAD : Illustrations



TAD : Illustrations



TAD : Derivation

We should accept the transition that would (in a given random realization) have occurred first at T_{low}.

In a TAD simulation, you :

- Assume that the kinetics obey **Harmonic TST** : $k = \nu \exp(-\beta \Delta E)$.
- Running at a higher temperature T_{high}, observe an event with a barrier ΔE after a time t_{high}.
- Generate a statistically correct escape time at T_{low} :

$$t_{\text{low}} = t_{\text{high}} \exp[\Delta E(\beta_{\text{low}} - \beta_{\text{high}})]$$

- Put the system back in the initial state
- Repeat until...

TAD : Derivation

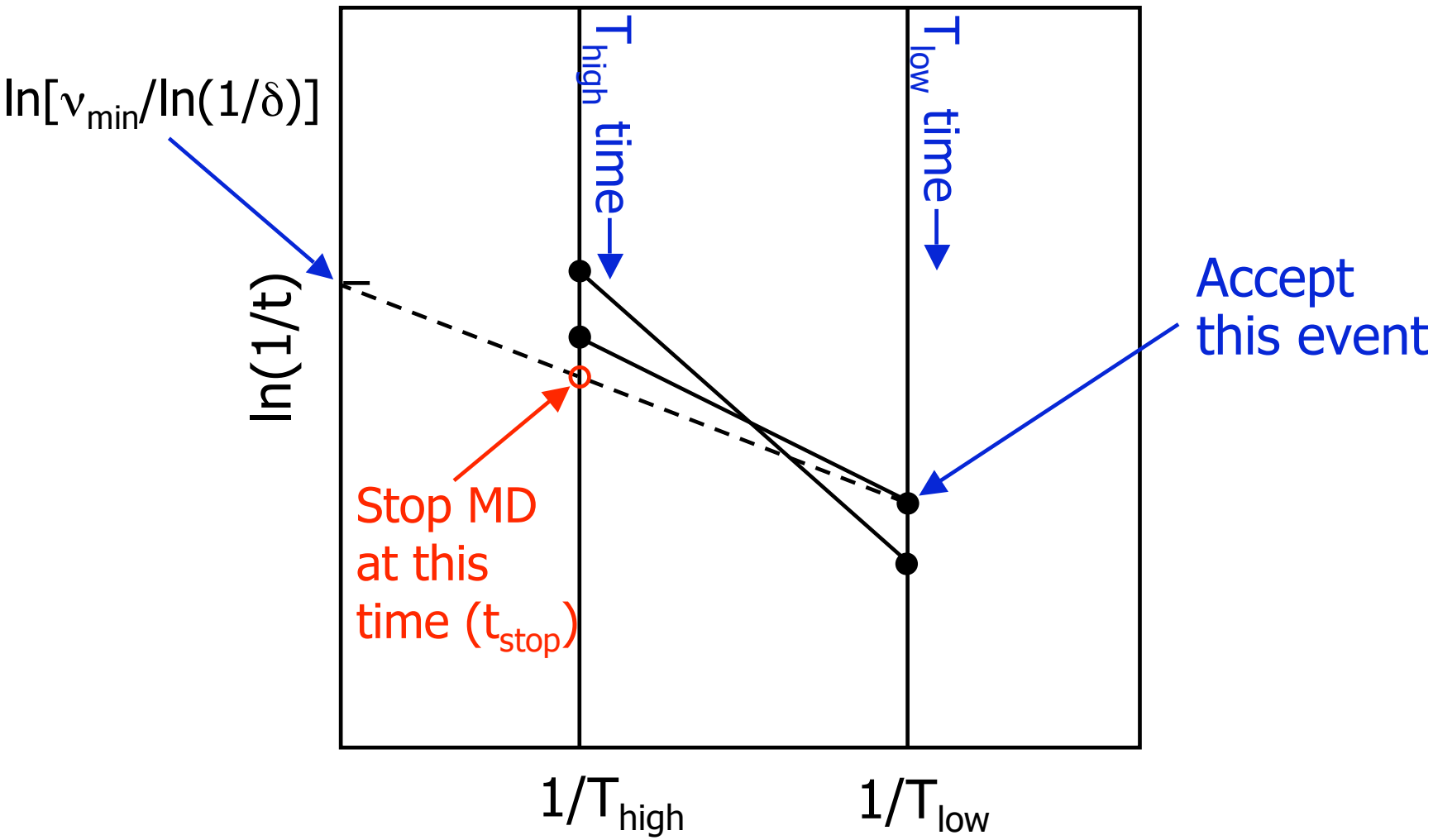
When to stop ? Define a stop time $t_{\text{high,stop}}$ such that the probability that another event replaces the current $t_{\text{low,short}}$ is less than δ.

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

$$t_{\text{high,stop}} = \frac{\ln(1/\delta)}{\nu_{\min}} \left(\frac{\nu_{\min} t_{\text{low,min}}}{\ln(1/\delta)} \right)^{\beta_{\text{high}}/\beta_{\text{low}}} \tag{7}$$

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 : Illustrations



TAD in practice

Pros :

- Can provide very large boost on a single CPU (10x-10^9x)
- 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.

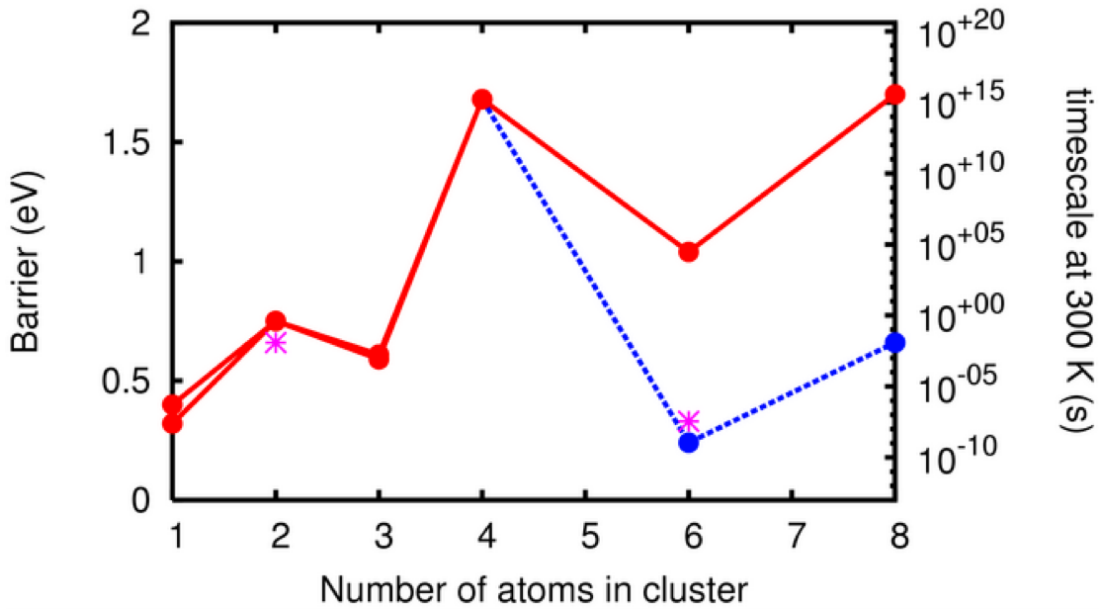
Interstitial clusters in MgO

- MgO is a component of nuclear fuel for fission plants. 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 : $T_{high} < 2000K$, $T_{low} = 300K$

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

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



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

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)].

Illustration of ParRep

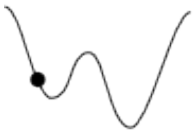


Illustration of ParRep

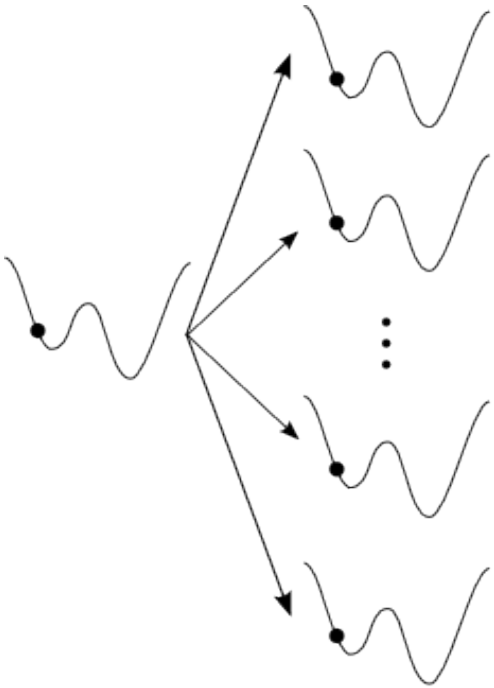


Illustration of ParRep

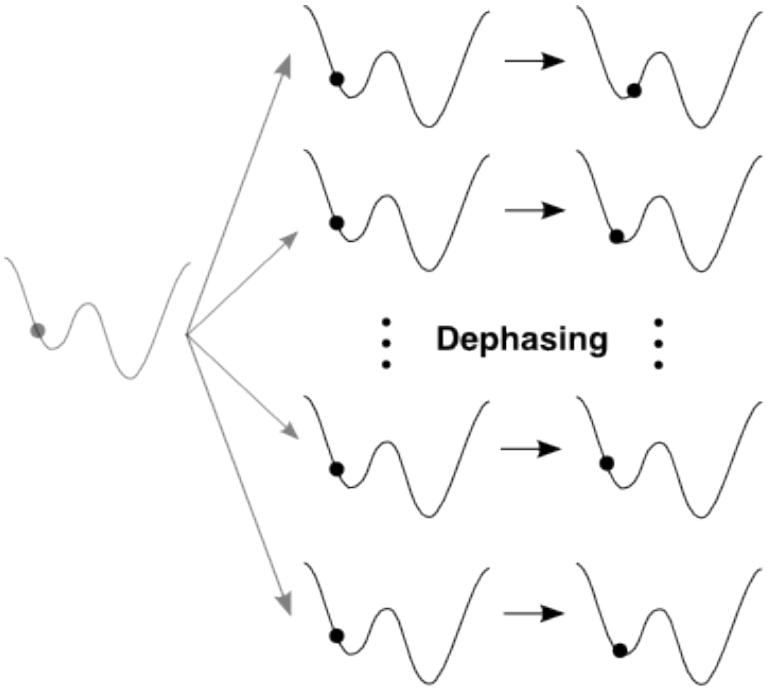


Illustration of ParRep

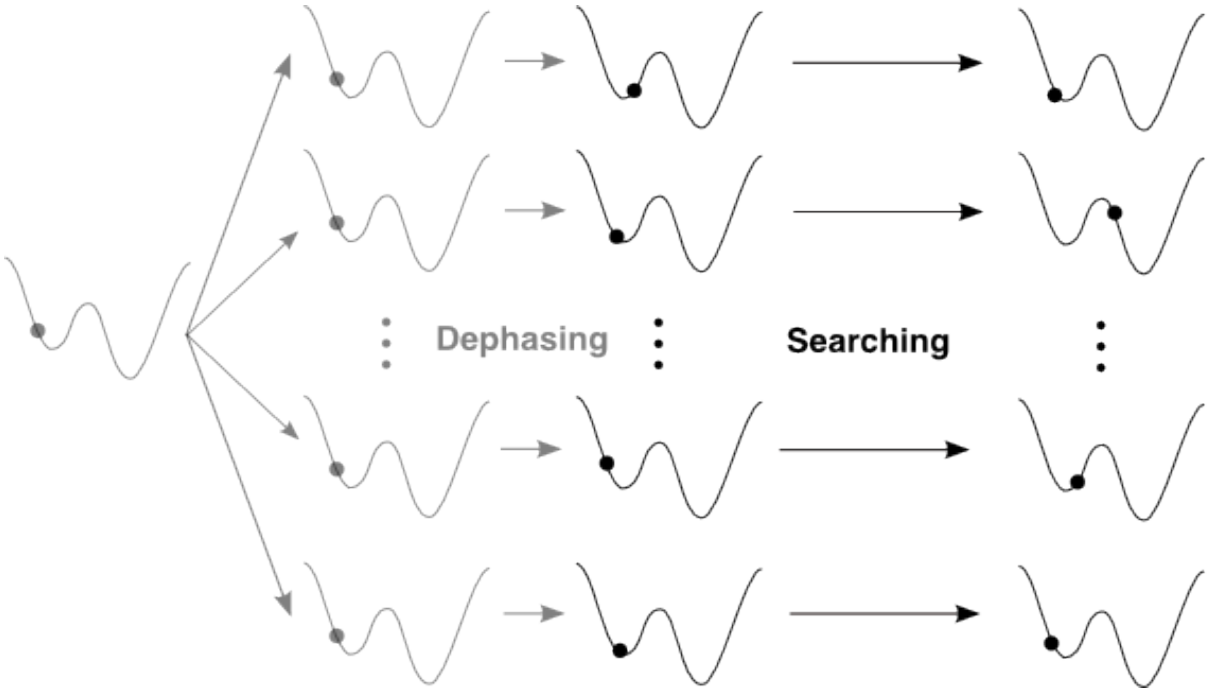


Illustration of ParRep

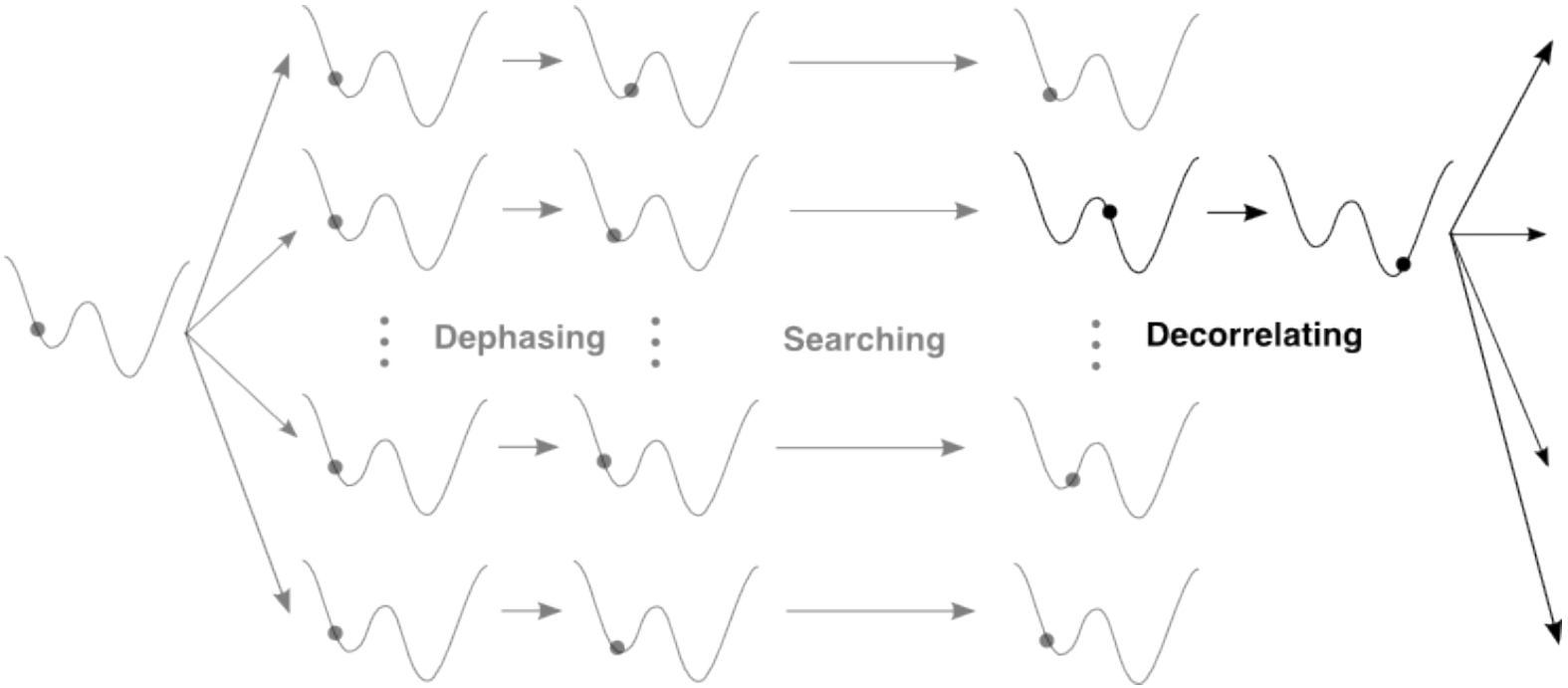
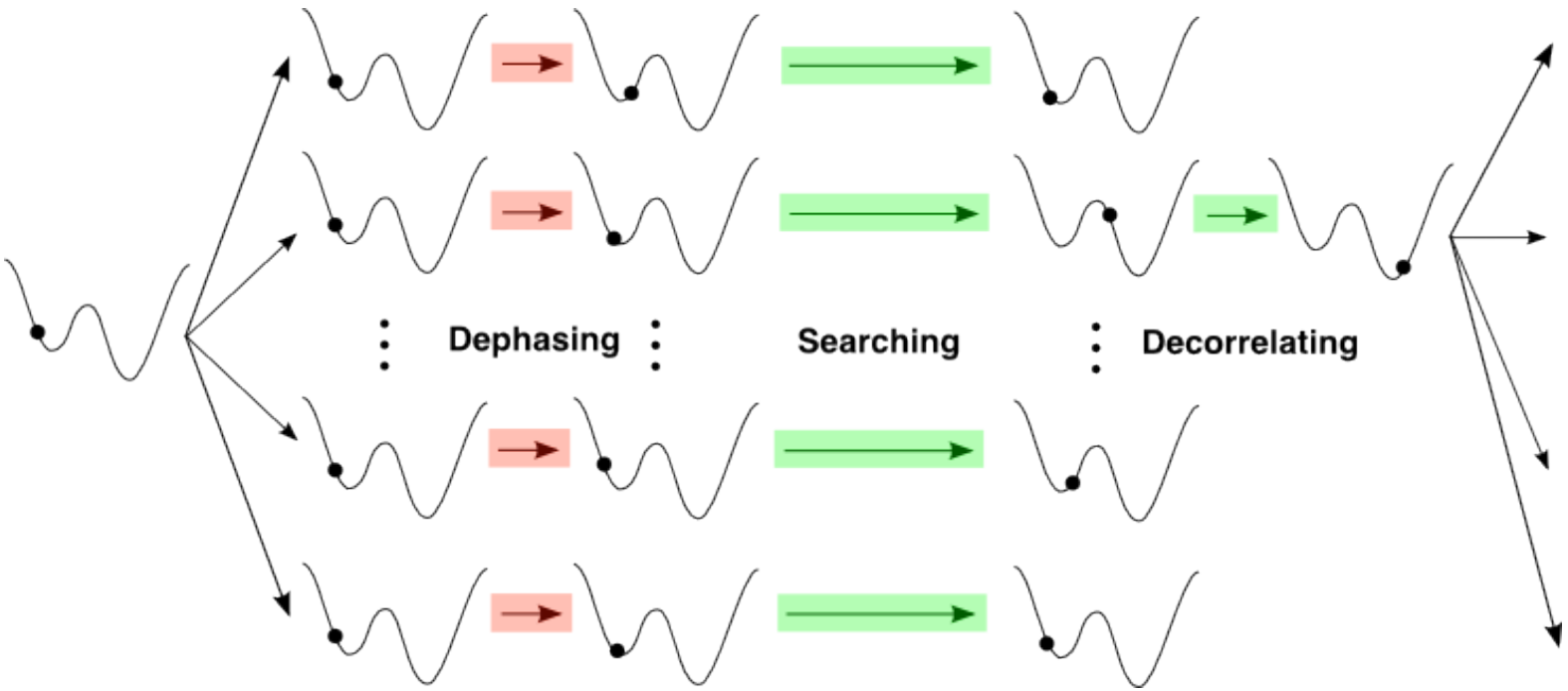
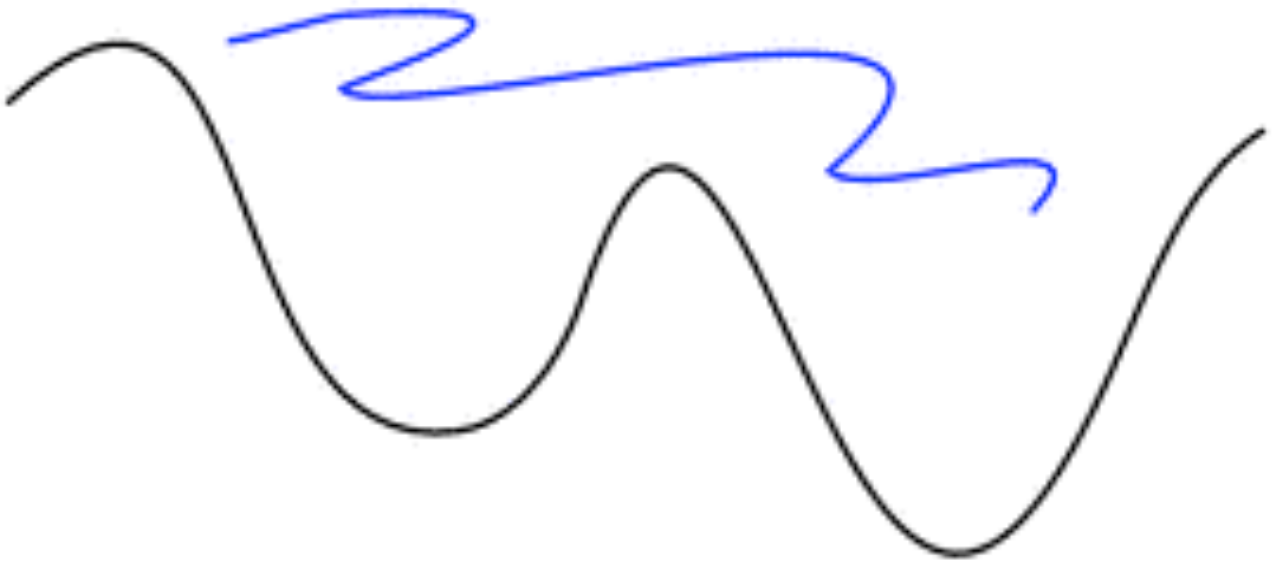


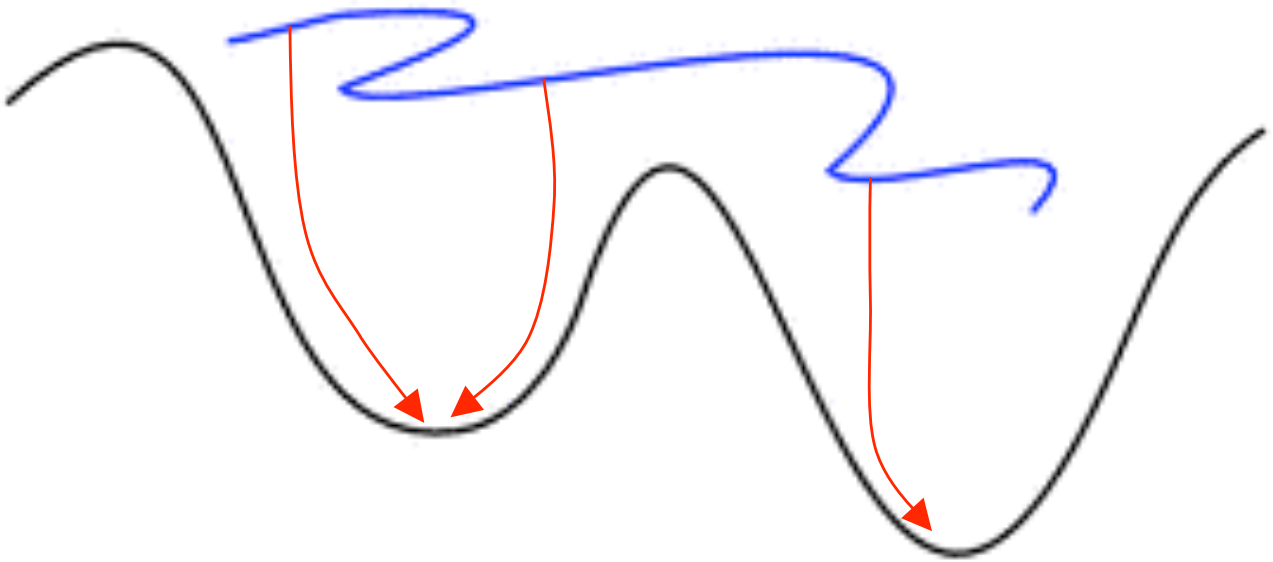
Illustration of ParRep



Detecting Transistions



Detecting Transistions

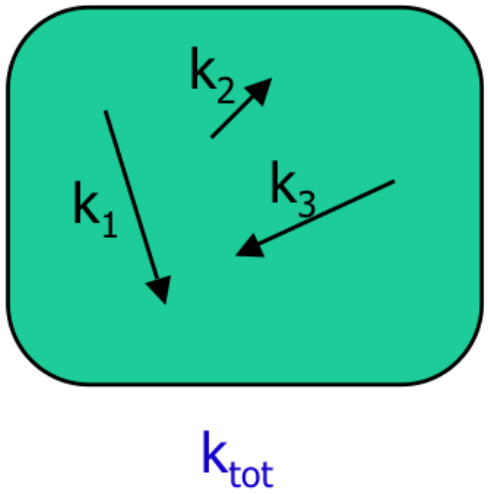


Justification of ParRep

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.,

$$p(t) = k_{tot} e^{-k_{tot} t}, \tag{8}$$

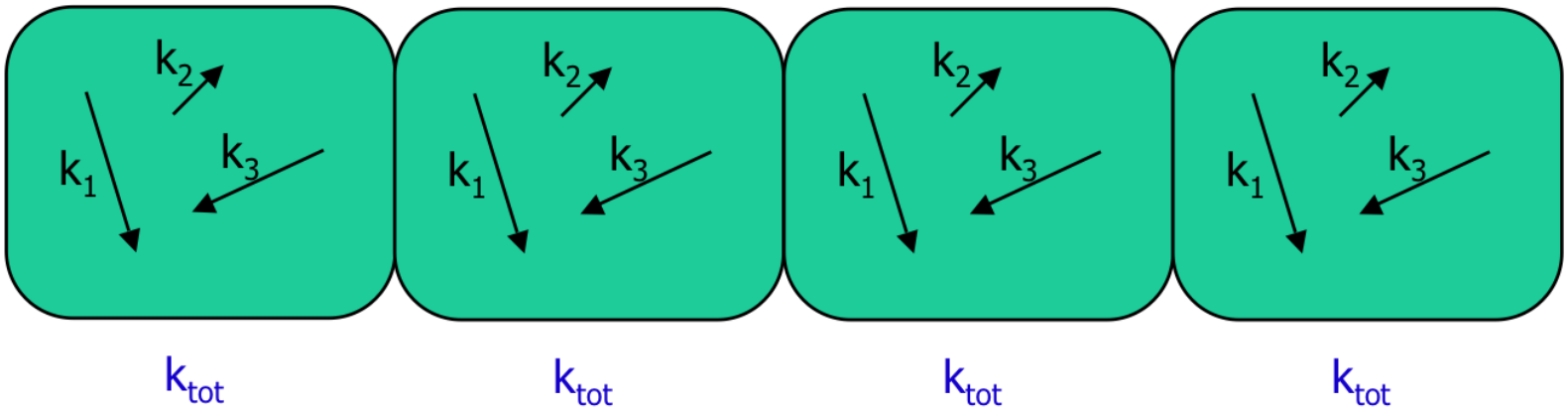
where k_{tot} is the total rate constant for all transitions out of the state ($k_{tot} = \sum k_i$).



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) :

$$p^*(t) = Mk_{tot}e^{-Mk_{tot}t}. \tag{9}$$



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 (this would imply an instantaneous loss of memory upon entering a state). Then, what can we say about ParRep ?

Traditionally, we have been thinking 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 DP, MCMA (2012)].

Justification of ParRep

For simplicity, let us consider overdamped Langevin dynamics :

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

For a given initial position inside of a state, 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 :

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

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

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

$$v(X,t) = \sum_k \exp(-\lambda_k t) c_k^0 u_k(X) \tag{10}$$

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 :

$$v(X,t) \simeq \exp(-\lambda_1 t) c_1^0 u_1(X). \tag{11}$$

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

u_1 is said to be quasi-stationary because
$v(X, t+\Delta t)/v(X,t) = \exp(-\lambda_1 \Delta t)$, i.e., time only renormalizes the distribution. Note that λ_1 is the escape rate from u_1.

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.

Justification of ParRep

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

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

This is true independently of the definition of states.

The parallel efficiency is however dependent of that definition : the bigger the spectral gap $\lambda_2 - \lambda_1$, the better the performance.

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 $N\tau_{\text{corr}} \ll \tau_{\text{trans}}$

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)

Roadrunner

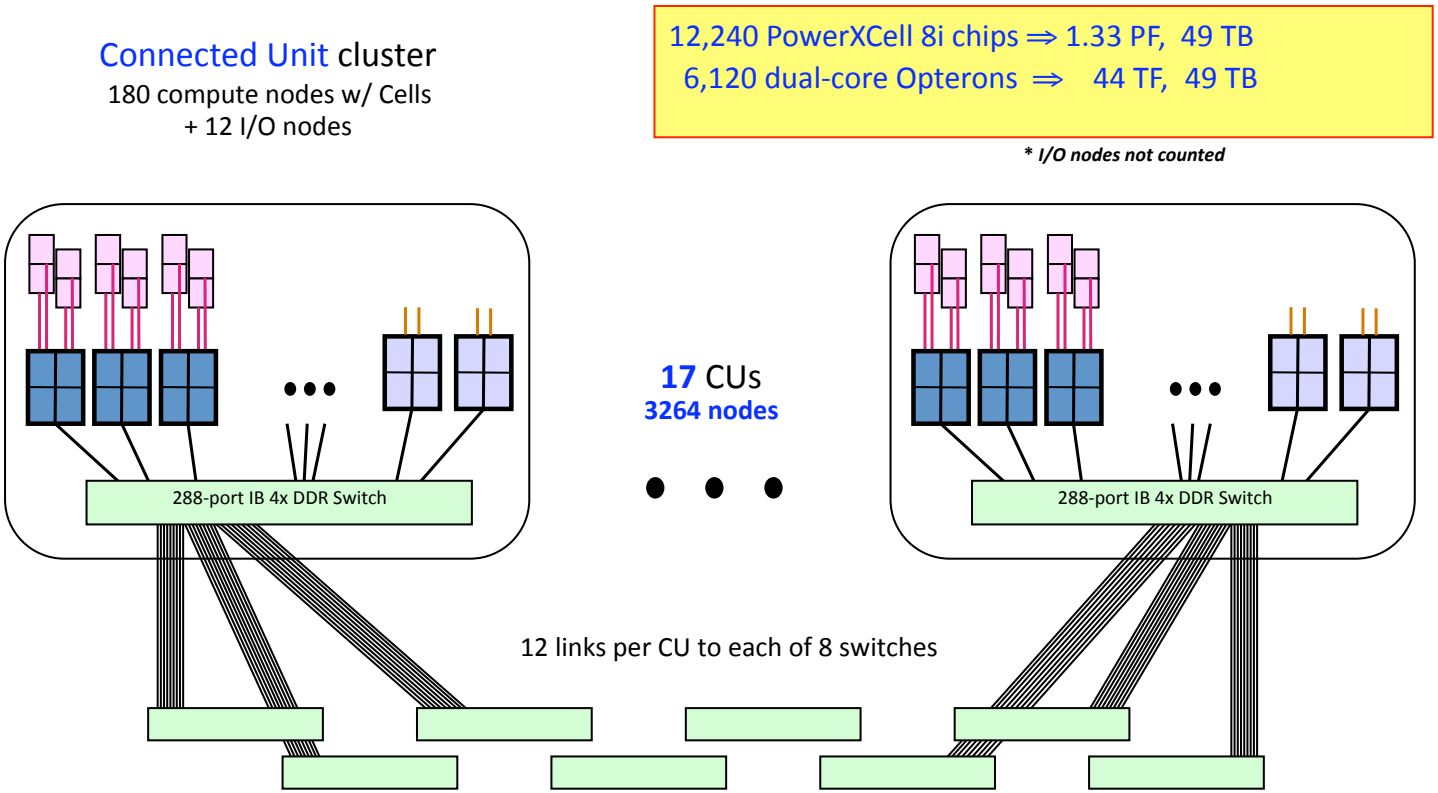


FIGURE: The Roadrunner architecture

Application to Ag Nanowires

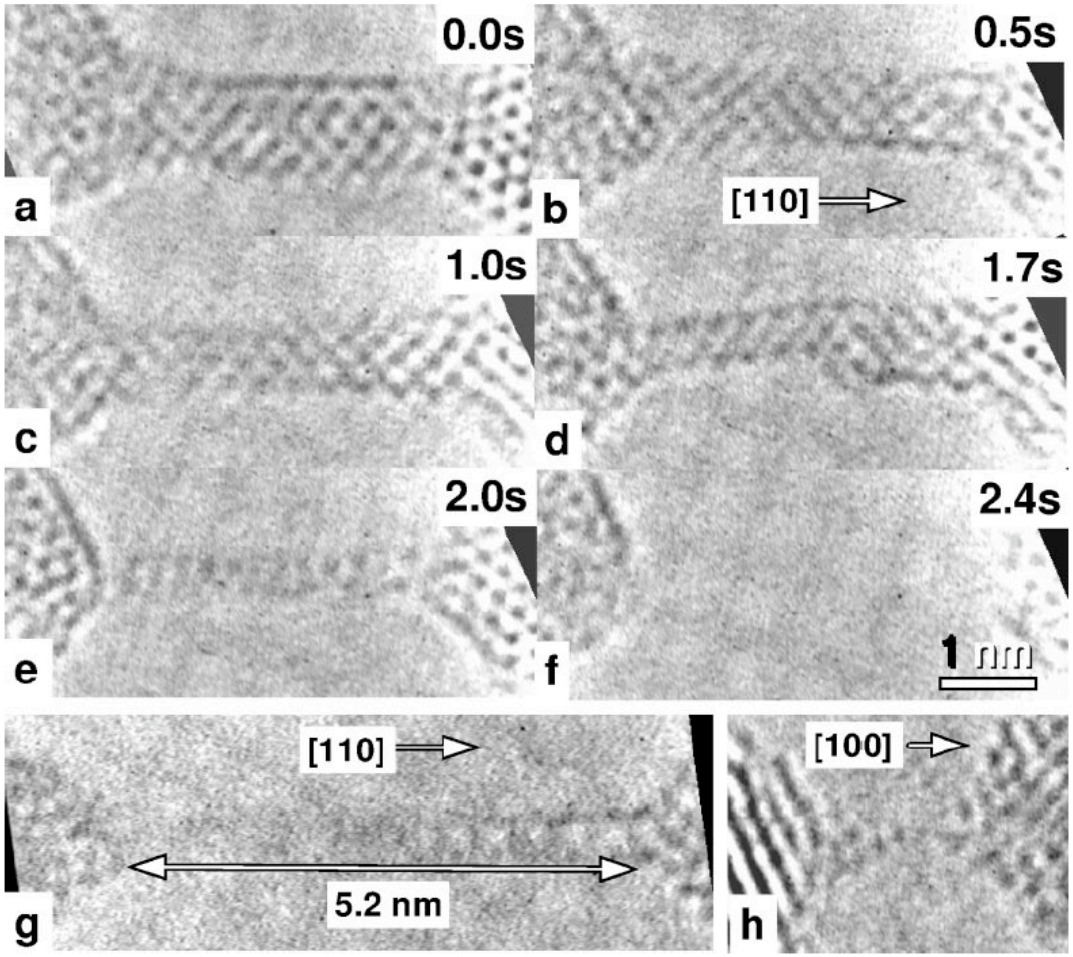


FIGURE: HRTEM imaging of Ag NWs. (Rodrigues *et al.*, PRB 65, 153402)

Application to Ag Nanowires

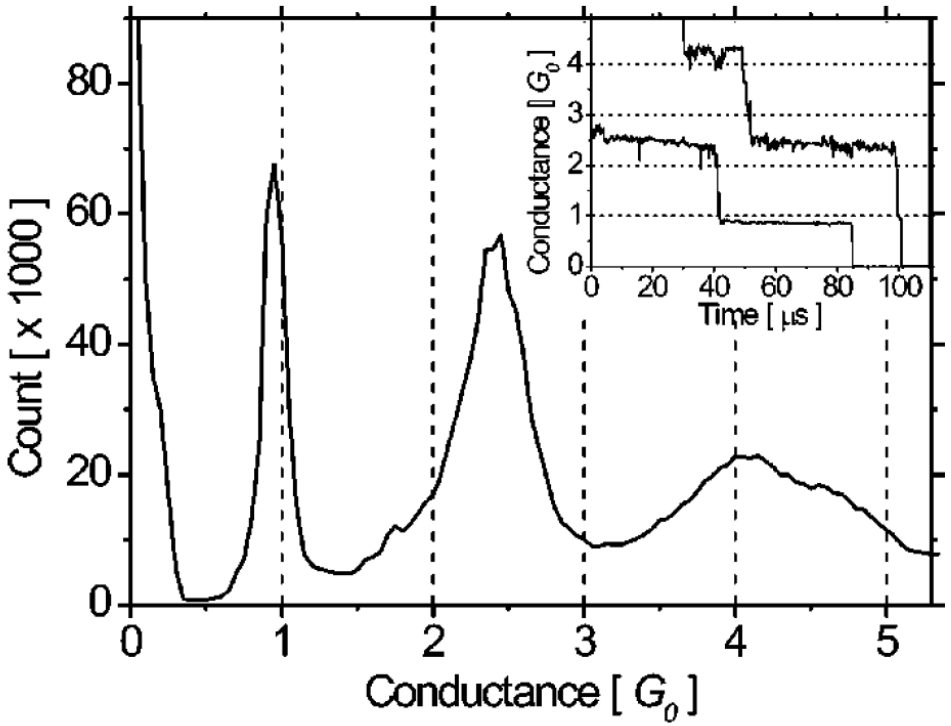
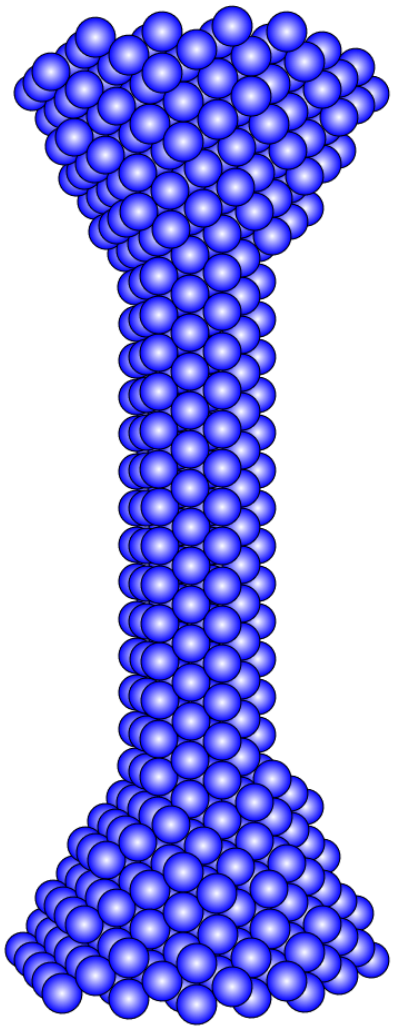


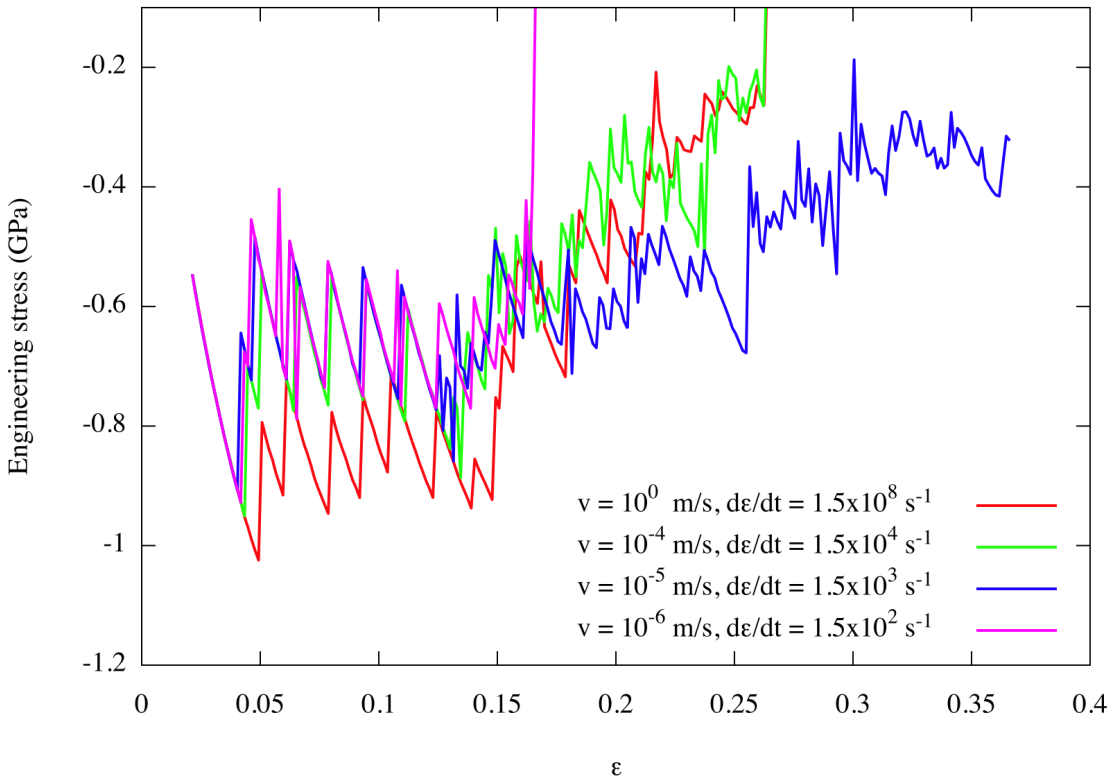
FIGURE: Break-junction conductance measurement of Ag NWs. (Rodrigues *et al.*, PRB 65, 153402)

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

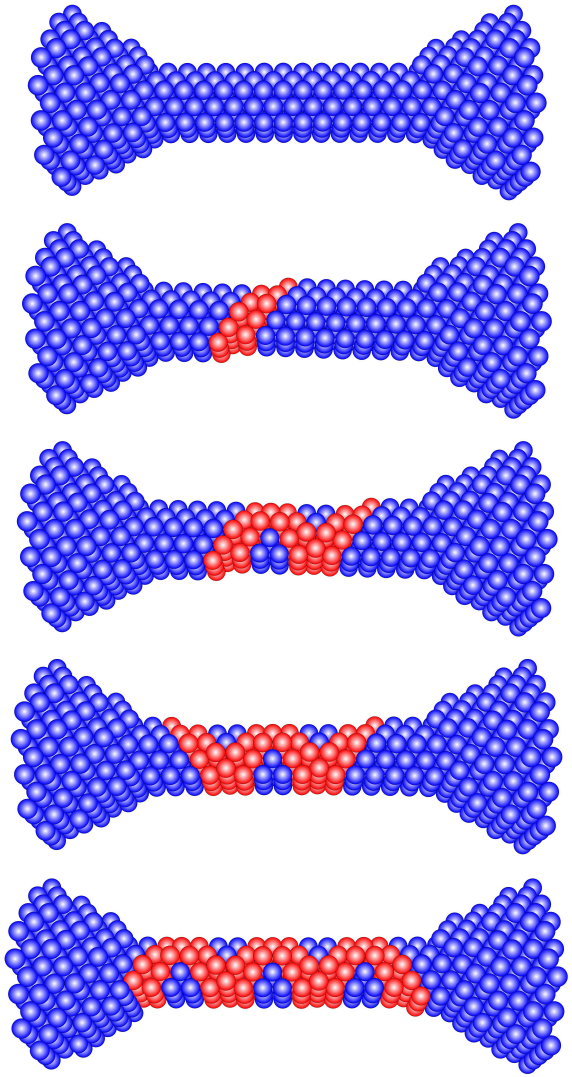


- Engineering yield stress $\simeq$ 1 GPa, Yield stress $\simeq$ 8 GPa
- Early stage behavior is regular and very robust
- Late stage behavior is complex
- Failure occurs for strains between 15% and >100%

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

Early Stage

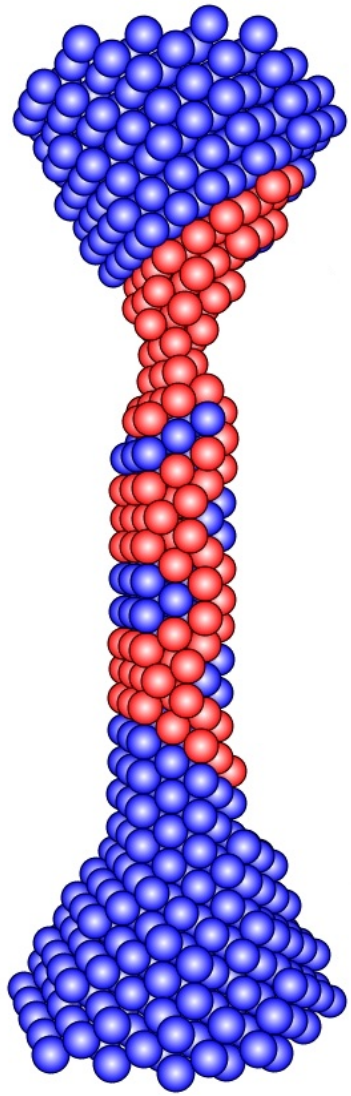
- Plasticity initially mediated by stacking faults along (111) planes
- Transitions more slip-like than dislocation-like
- Formation of a zig-zag pattern
- Uniform thinning of the wire
- Annihilation does **not** directly follow from more strain



Late Stage

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

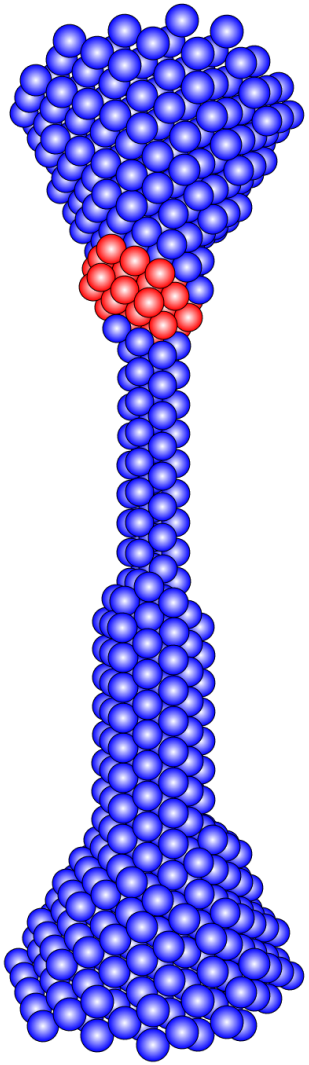
- Necking (all drives, necks earlier for slow drives)



Late Stage

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

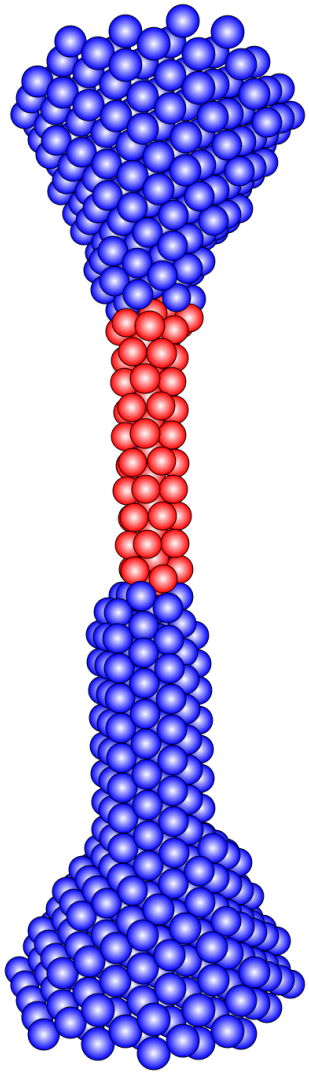
- Necking (all drives, necks earlier for slow drives)
- Non-uniform thinning, transition to an icosahedral structure, gradual conversion from FCC to icosahedral. (intermediate drives)



Late Stage

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

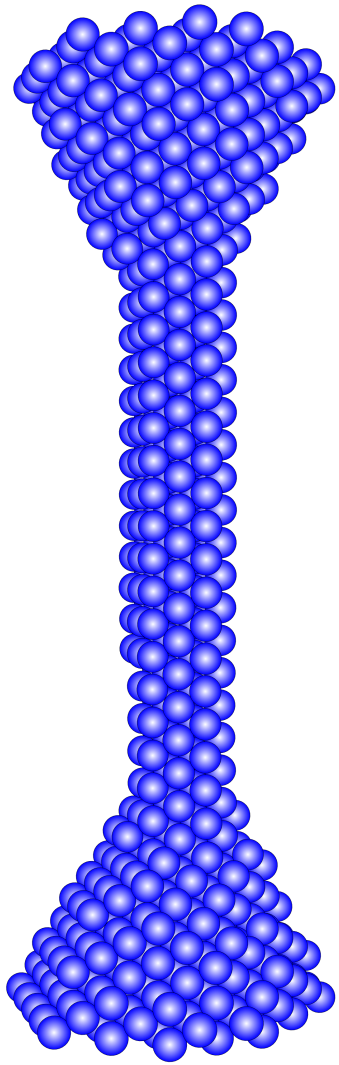
- Necking (all drives, necks earlier for slow drives)
- Non-uniform thinning, transition to an icosahedral structure, gradual conversion from FCC to icosahedral. (intermediate drives)



Late Stage

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

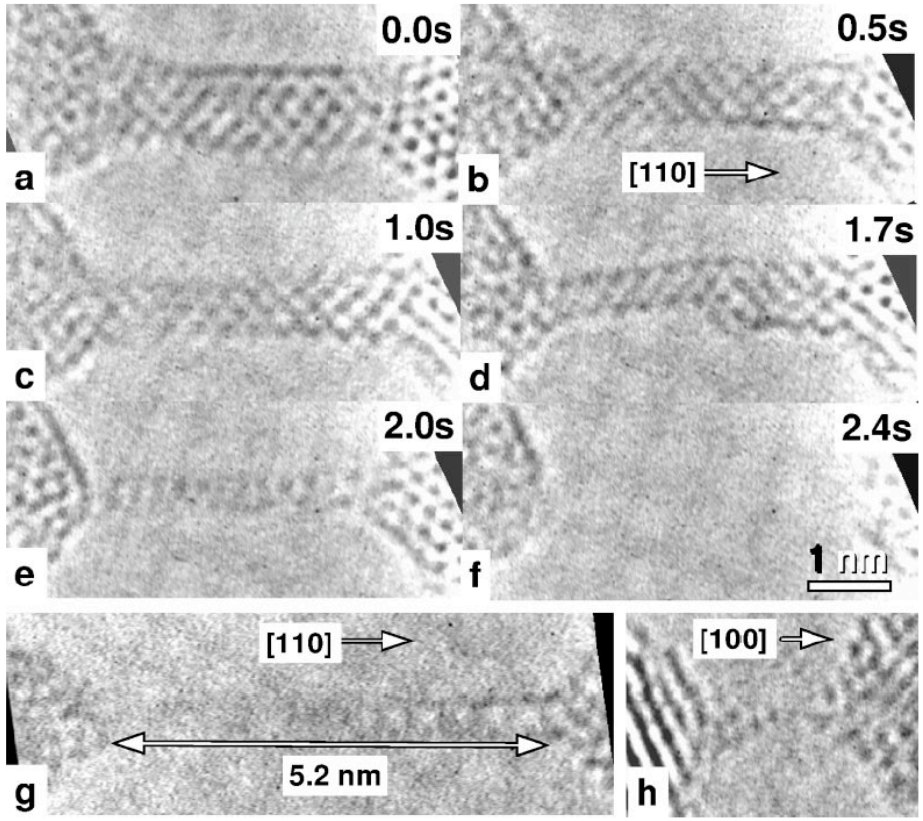
- Necking (all drives, necks earlier for slow drives)
- Non-uniform thinning, transition to an icosahedral structure, gradual conversion from FCC to icosahedral. (intermediate drives)
- Annihilation of the stacking fault network (slow drives)



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



Definition of States

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

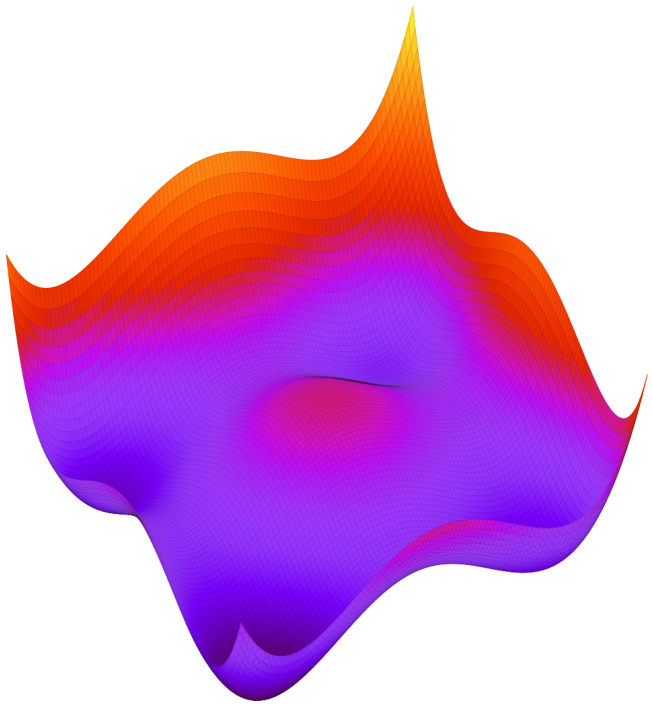


FIGURE: This is a good strategy when the landscape looks like this

Definition of States

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

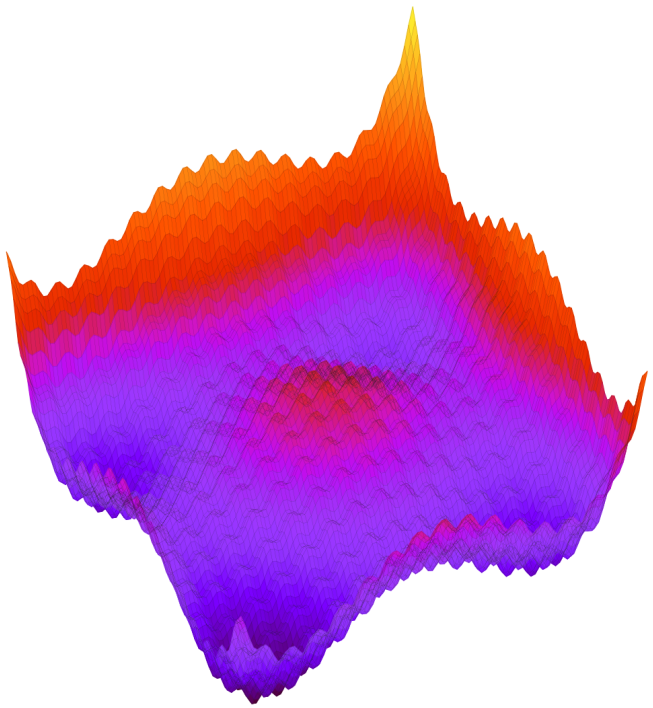


FIGURE: This is a bad strategy when the landscape looks like this

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.

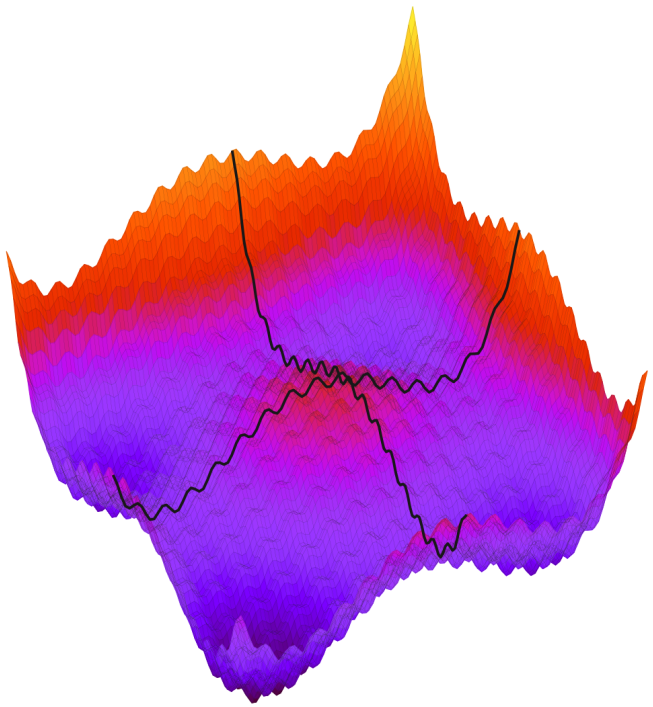
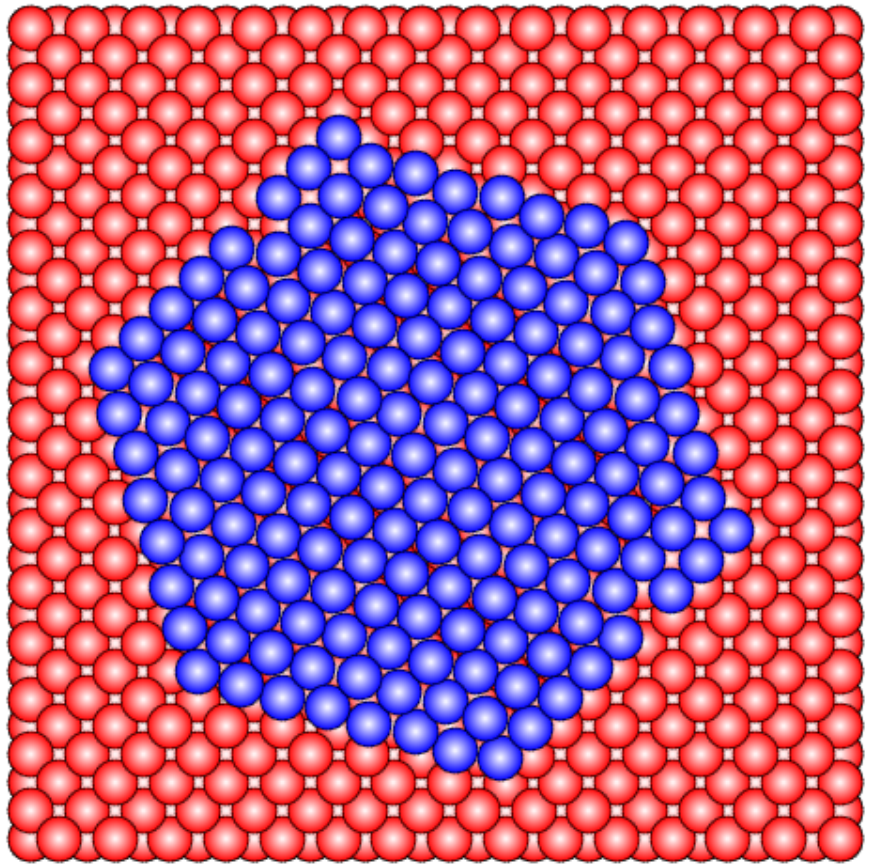


FIGURE: This is a good definition of states

Application to Ag on Cu Island Dynamics

- Cu (001) substrate
- $c(10x2)$ hexagonal Ag island (between 37 and 271 atoms)
- 11% lattice mismatch between Ag and Cu



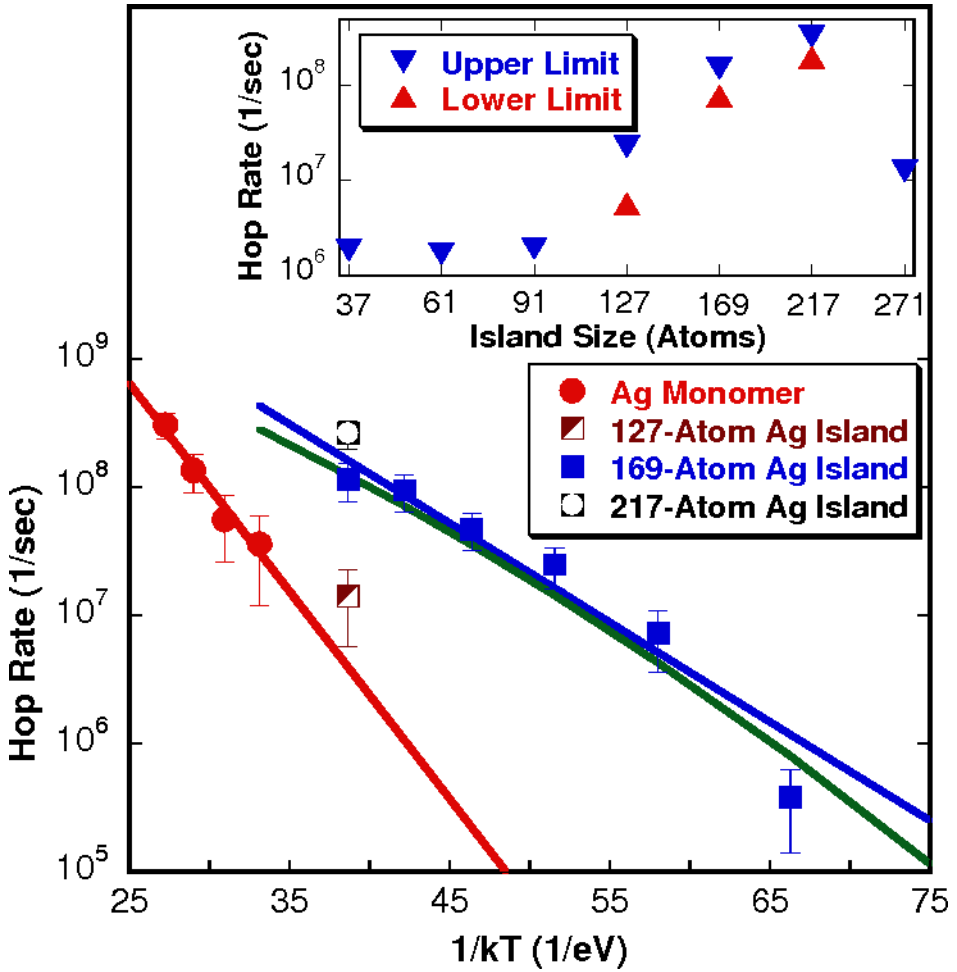
[Uche, DP, Voter, Hamilton, PRL **103**, 046101 (2009)]

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

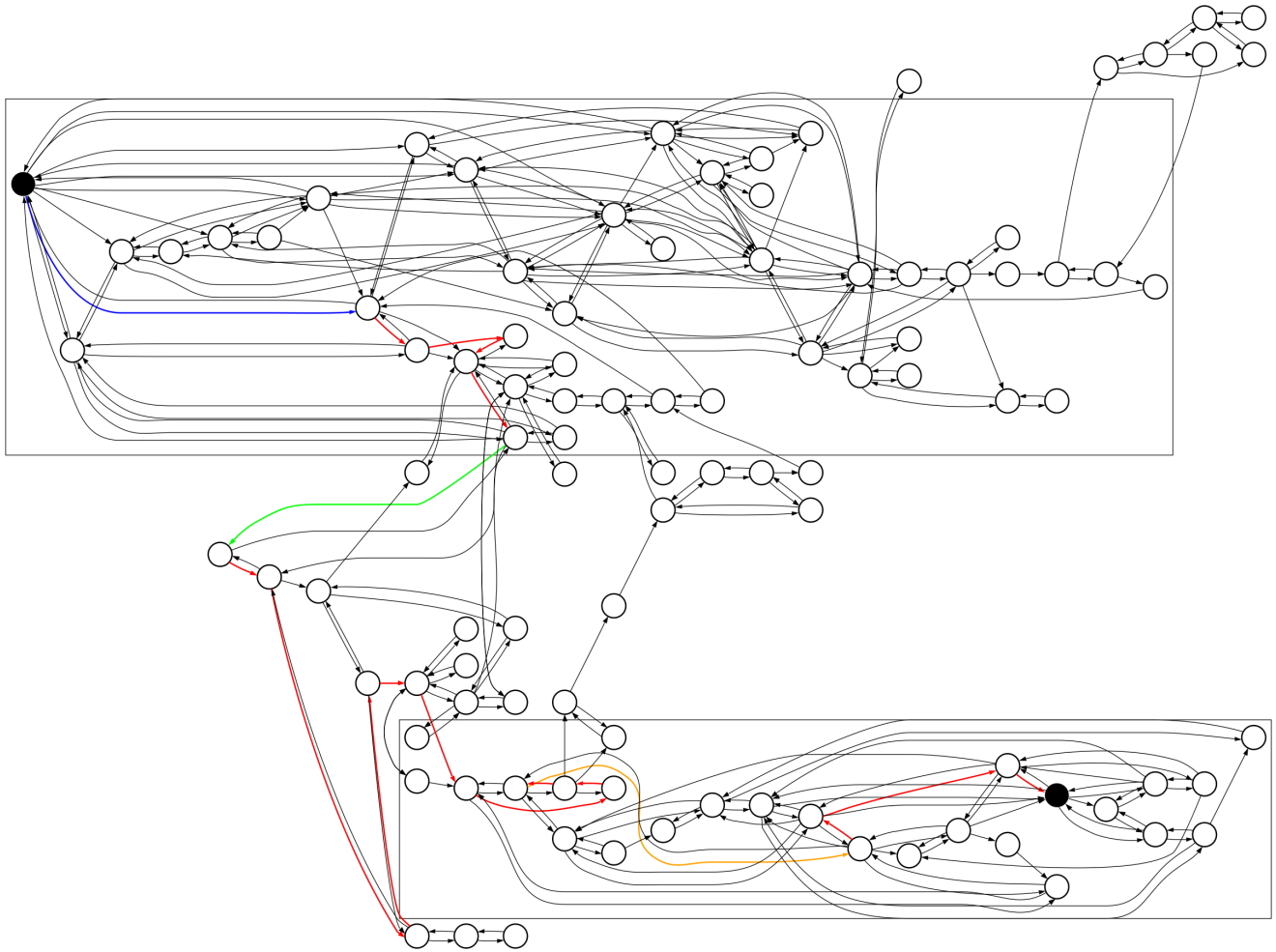


FIGURE: Observed transitions over $6\mu s$ at $T = 175K$.

Favorable hopping pathway : vacancy and glide

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

Nudged Elastic Band Analysis

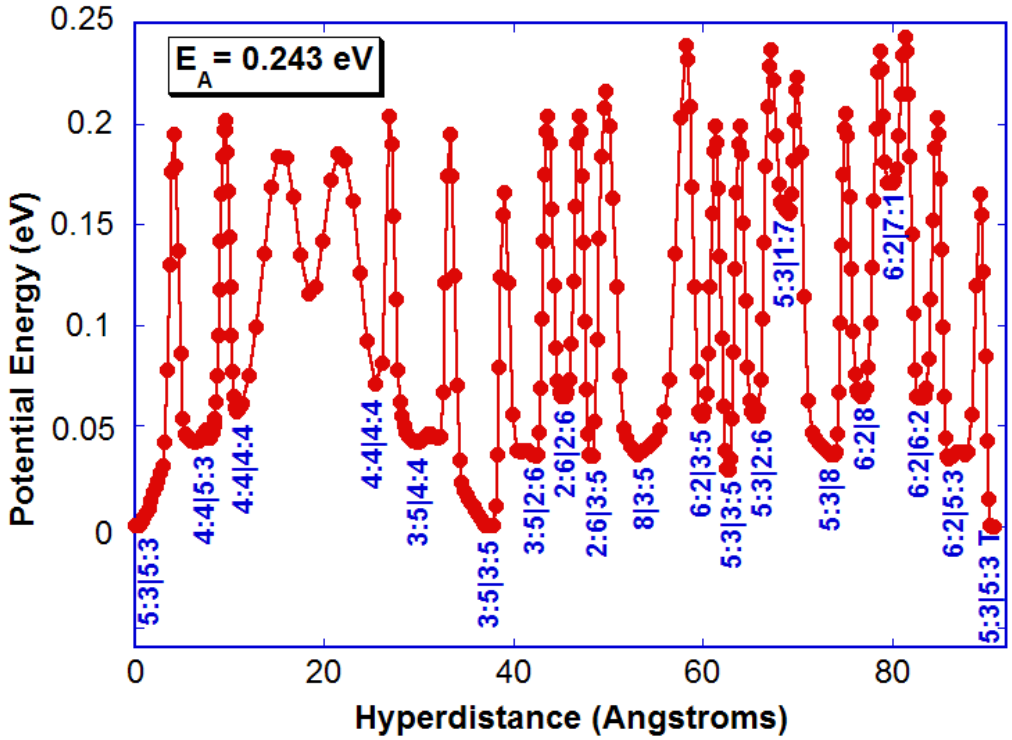


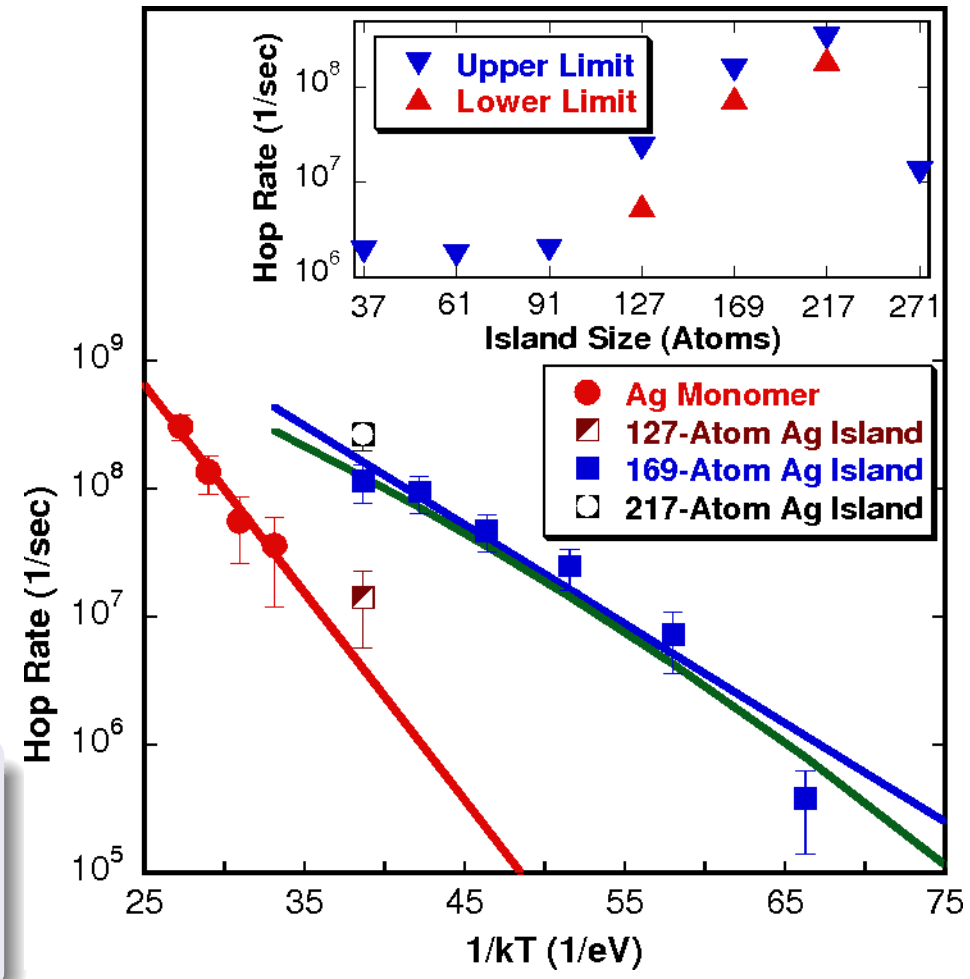
FIGURE: Minimum energy pathway along the previous path.

- 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

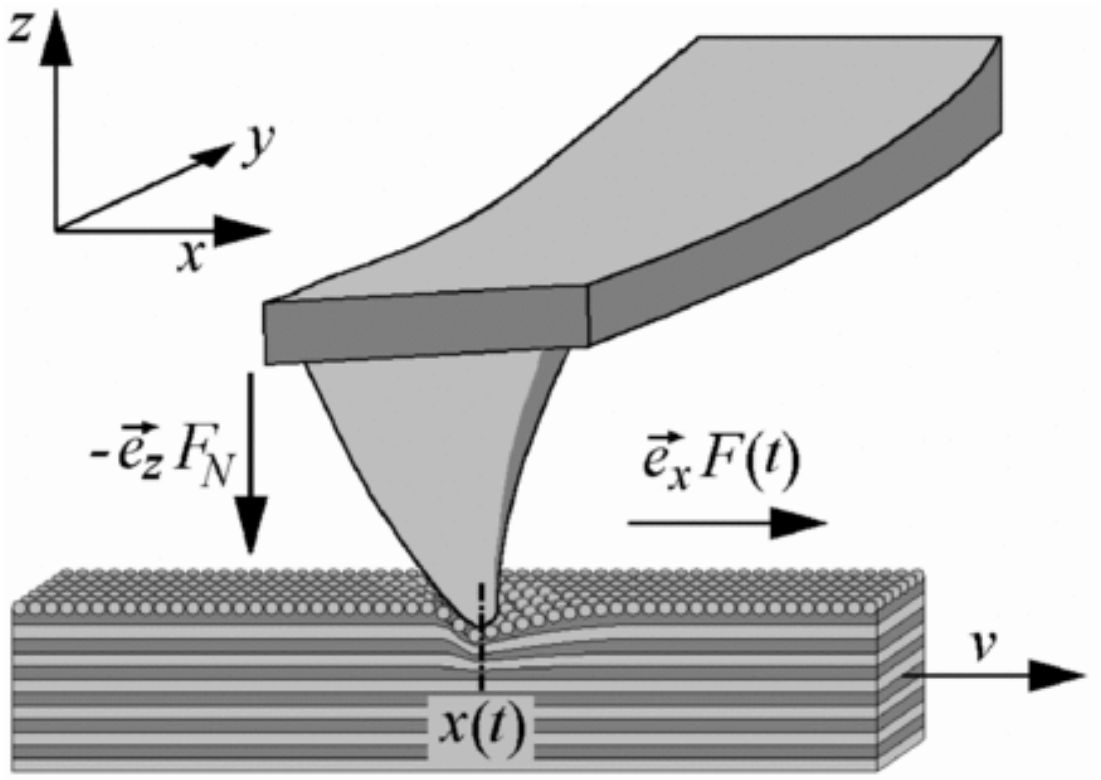
This recipe works as long as the drive is quasi-static, i.e., $k_i(x,t) = k_i(x(t))$, i.e., the rates do not depend on the driving rate, just on the current state of the drive.

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.

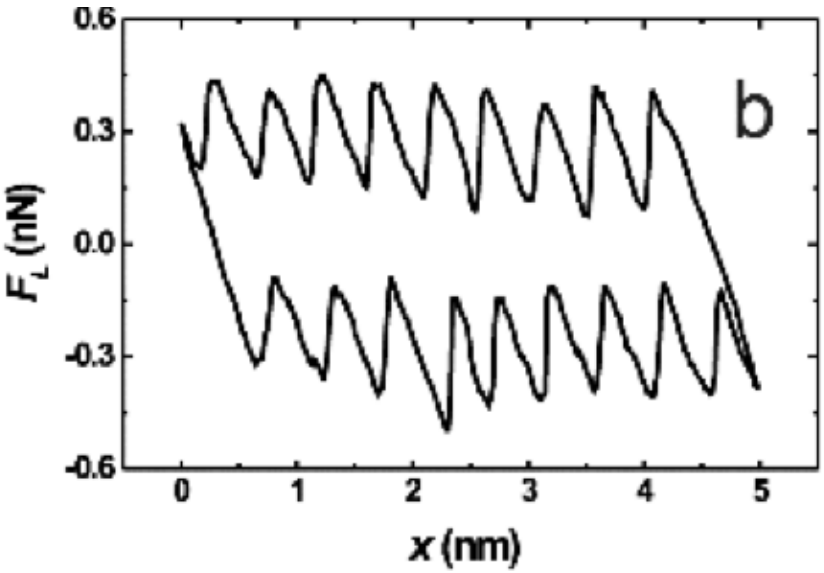
ParRep for Driven Systems

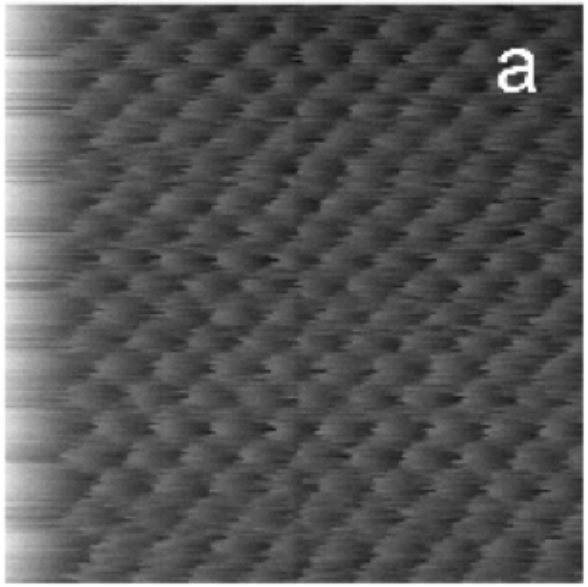
This technique can by applied to study nanoscale friction, as measured by a Friction Force Microscope (FFM).



ParRep for Driven Systems

From the bending of the cantilever, one can mechanically map the surface at the atomic scale and infer the friction (the amount of dissipated energy).





Force trace (left) and friction map (right) of a NaCl (100) surface. Scanning speed : 25 nm/s. Taken from Phys. Rev. Lett., 84 :1172 (2000).

ParRep for Driven Systems

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

ParRep for Driven Systems

Of course, one cannot simulate the whole apparatus atomistically (the cantilever is 100s of μm long).

- Only the contact region is simulated
- The elasticity of the cantilever is introduced by springs tied to a rigid support
- The support is driven at constant velocity (down to 10^6 nm/s)
- Pt tip on Au substrate (EAM)
- We attempted to match the contact size, the relative orientation of the interfaces, and the elasticity of the cantilever, to an FFM experiment.
- Used up to 512 replicas at the lowest speed (for a few μs of total MD time).

ParRep for Driven Systems

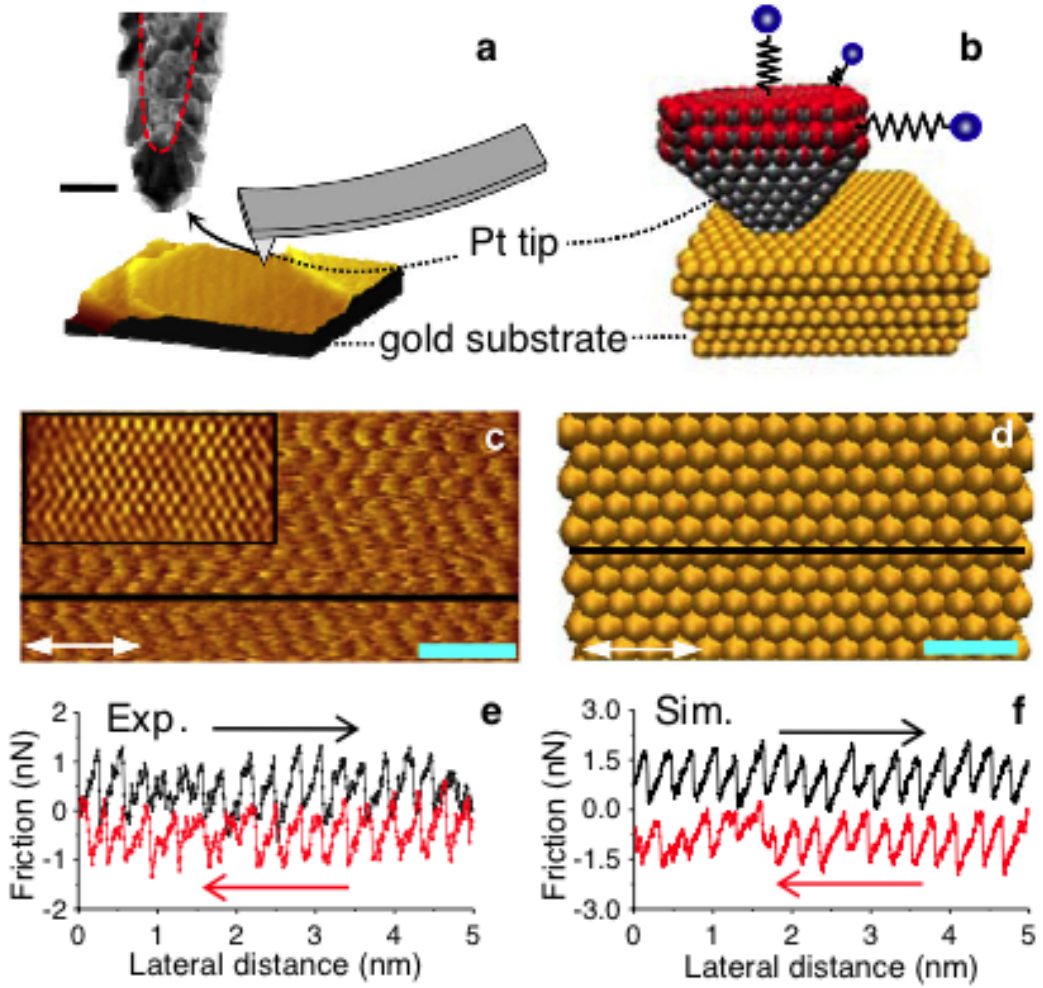
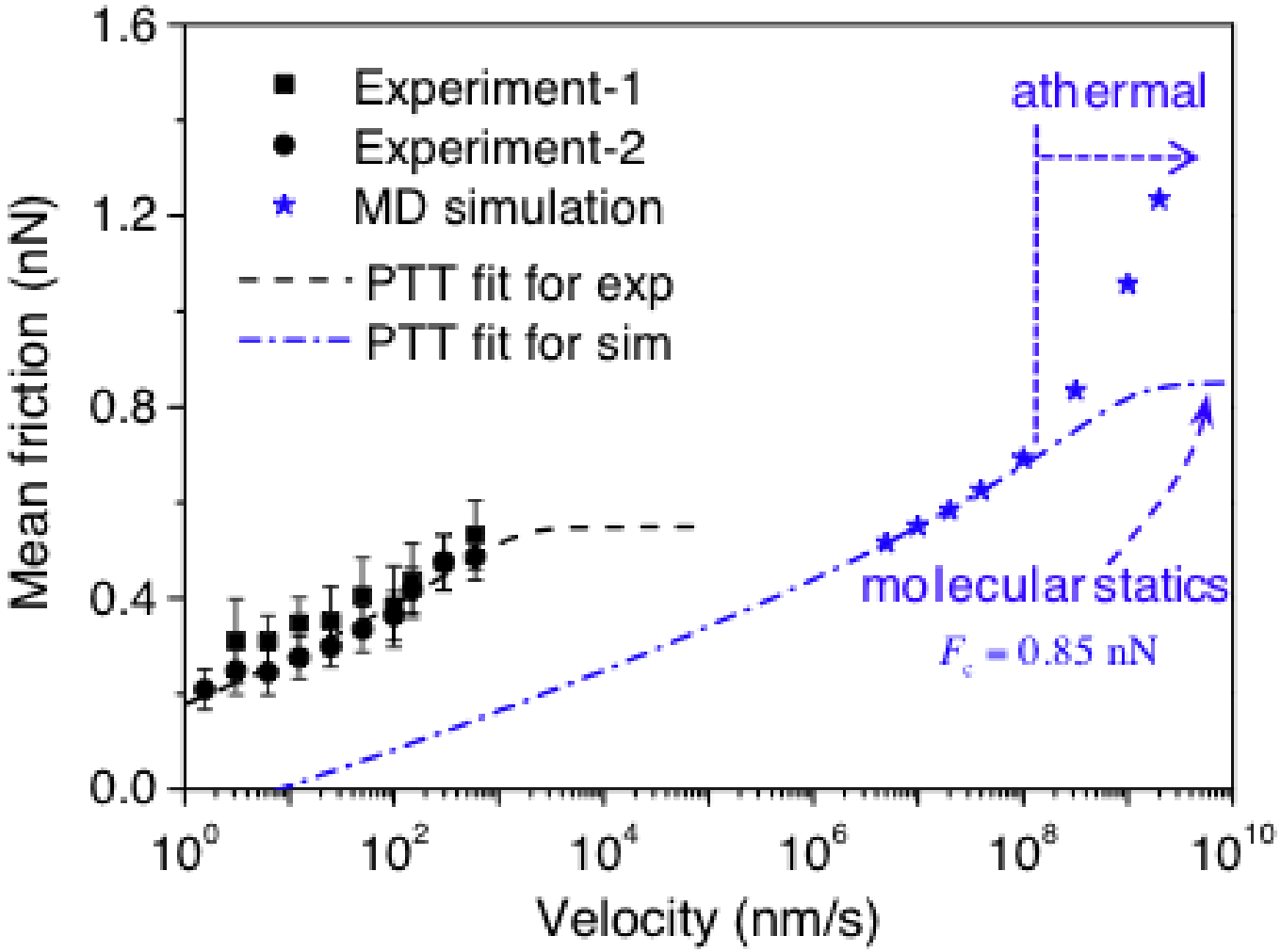


FIGURE: Li, Dong, DP, Martini, Carpick, PRL 106, 12 (2011)

ParRep for Driven Systems



ParRep for Driven Systems

- Qualitative agreement with experiments, despite the remaining gap in velocities.
- Simulations (at low velocity) and experiments follow the trend predicted by the Tomlinson model (ball + spring + corrugation).
- At high velocity (in the regime accessible with MD), simulations deviate considerably from the expected behavior. This marks the onset of non-thermal effects. **AMD is required even for qualitative comparisons with experiment.**

ParRep for Driven Systems

- Analysis of the effective model parameters show good agreement in terms of energetics (peak force and shape of the potential), but poor in terms of the rate theory prefactors for slip (by 6 orders of magnitude !).
- While the elasticity of the cantilever is accounted for, its mass is not. Our tip is too light, and hence vibrating too fast, which affects the prefactors.
- We need a true multiscale treatment, in both time and space, to account for this effect properly.

Plan

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.

At present, your chances are best for “simple” systems with a separation of timescales between vibrations and transitions between individual energy basins.

If this is not the case, ParRep could still work for you.

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.

Current (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

However, truly efficient solutions are often system specific.

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 do not require *a priori* knowledge about the important processes.

Acknowledgements

- Arthur F. Voter, Blas P. Uberuaga, Sriram Swaminarayan, Chun-Wei Pao
- Funding : US DOE/BES,ASCR ; LANL LDRD

Hyperdynamics : generalizations

- State-Bridging Hyperdynamics (Miron, Fichthorn) : bias over "irrelevant" saddles
- Self-learning Bond-Boost Hyperdynamics : automatically adjust the strength of the bias potential for maximal performance and accuracy
- Local Bond-Boost Hyperdynamics : localize the definition of the bias to maintain acceleration for very large systems
- Multiscale hyperdynamics : accelerate the dynamics of concurrent multiscale models (e.g., quasicontinuum method) to reach very long timescales on very large systems

TAD : generalization

- Synthetic-TAD : treat fast events with KMC but use TAD to keep looking for low-rate events.
- Spatially-parallel TAD (Amar, Voter) : Exploit the locality of typical events to parallelize TAD simulation.
- Blocking-TAD : Explicitly block already observed transitions to allow for higher high-T.

ParRep : generalizations

- ParRep for solid-liquid systems : correctly allows diffusion at wet surfaces and even dissolution and precipitation
- Mathematical formalization of ParRep in terms of eigen-properties of generators of the dynamics
- Super-state ParRep : Change the definition of the states to allow efficient simulations of systems with low barriers.