Introduction to Accelerated Molecular Dynamics

Danny Perez
danny_perez@lanl.gov

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

Hands-on DFT And Beyond

Molecular dynamics

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

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)

Scope of Molecular Dynamics

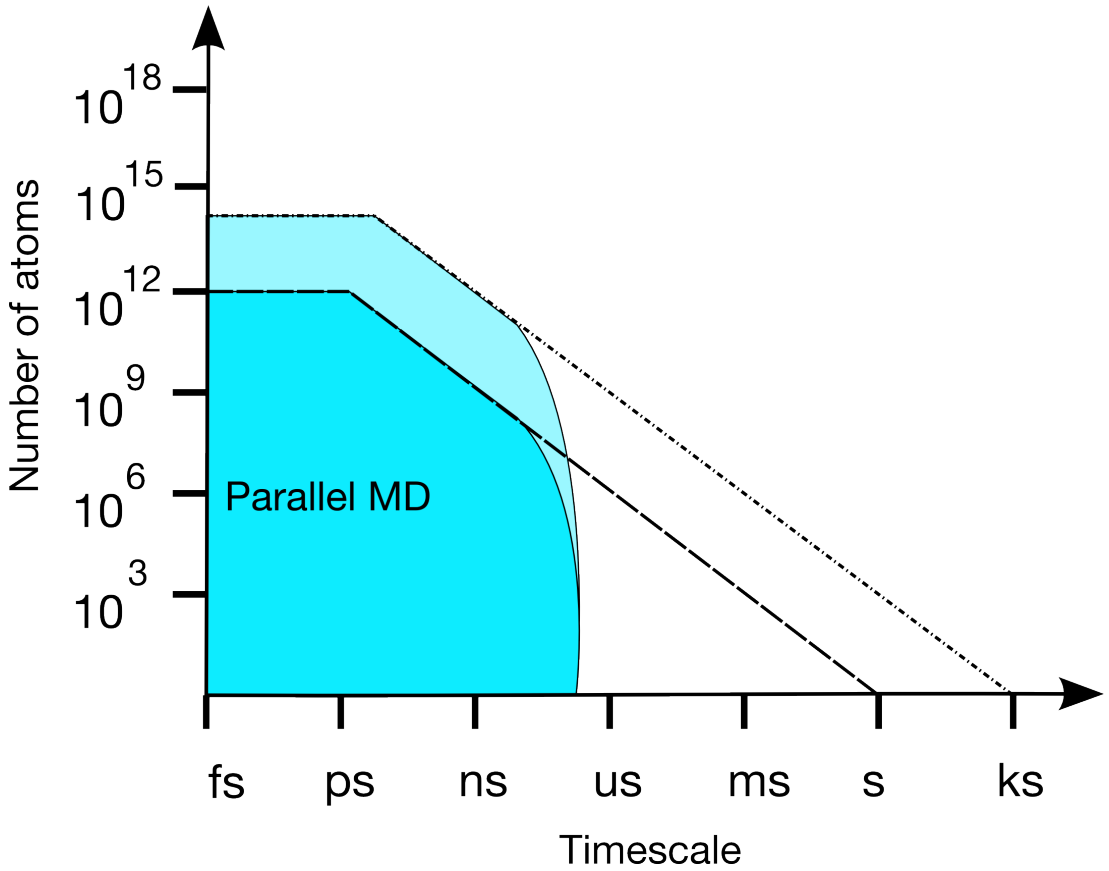
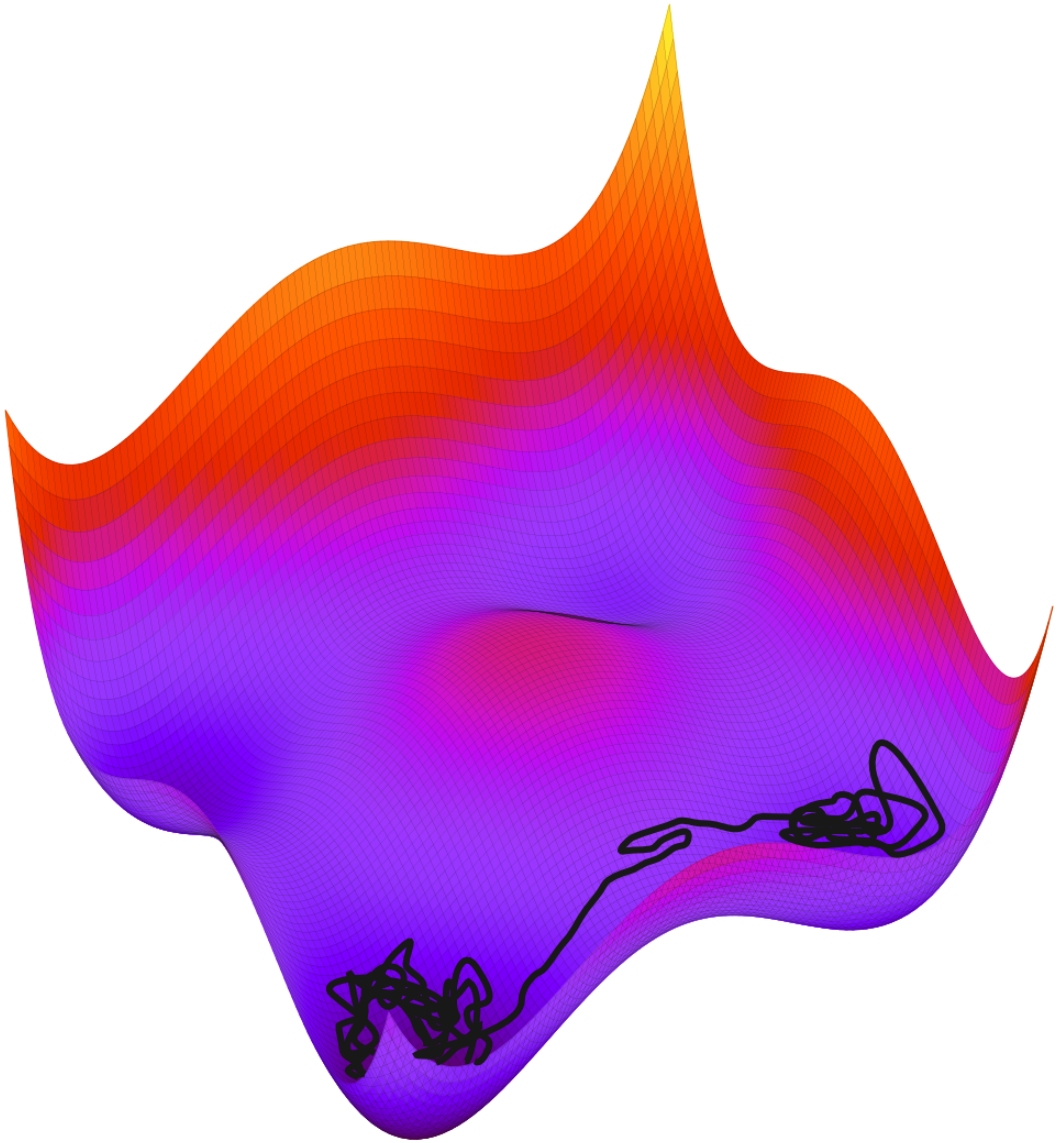


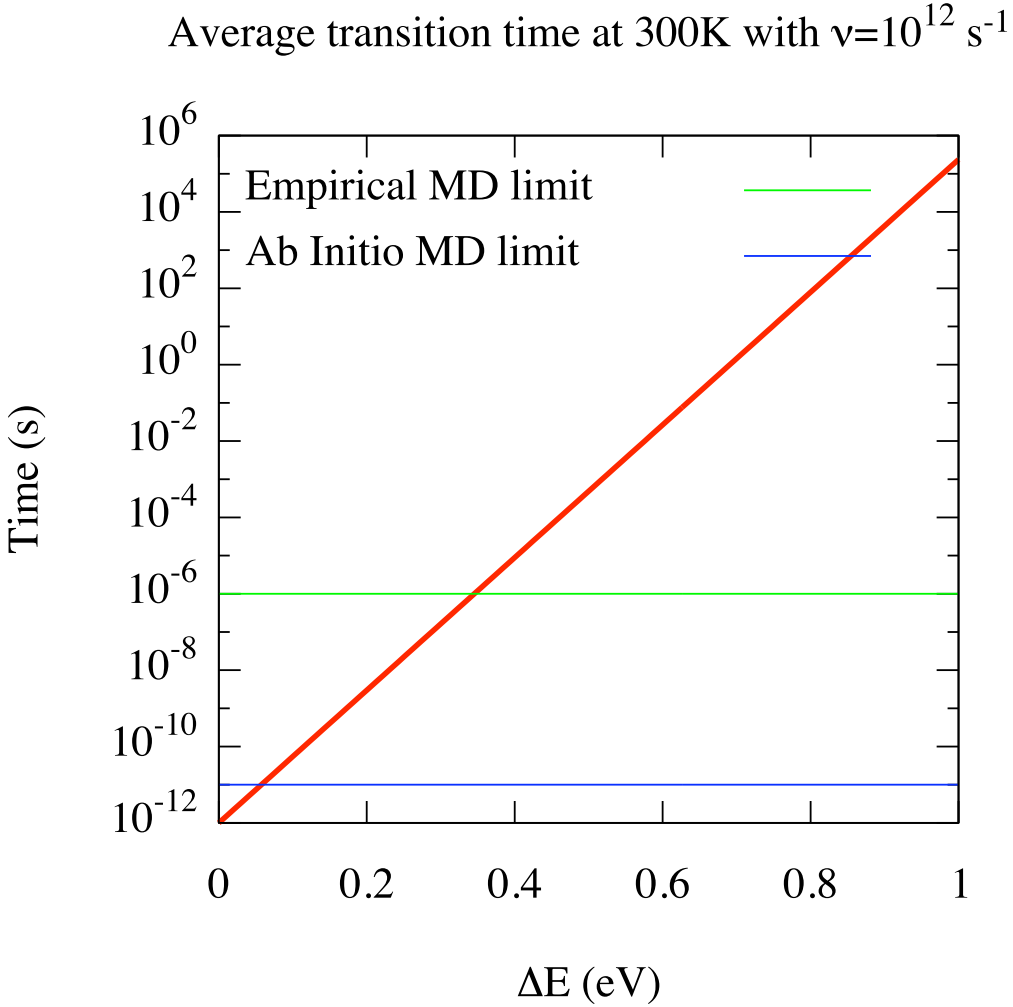
FIGURE – Scope of empirical MD simulations given a peta- (solid) or exa-flop (shaded) computer for a few days.

MD Trajectory

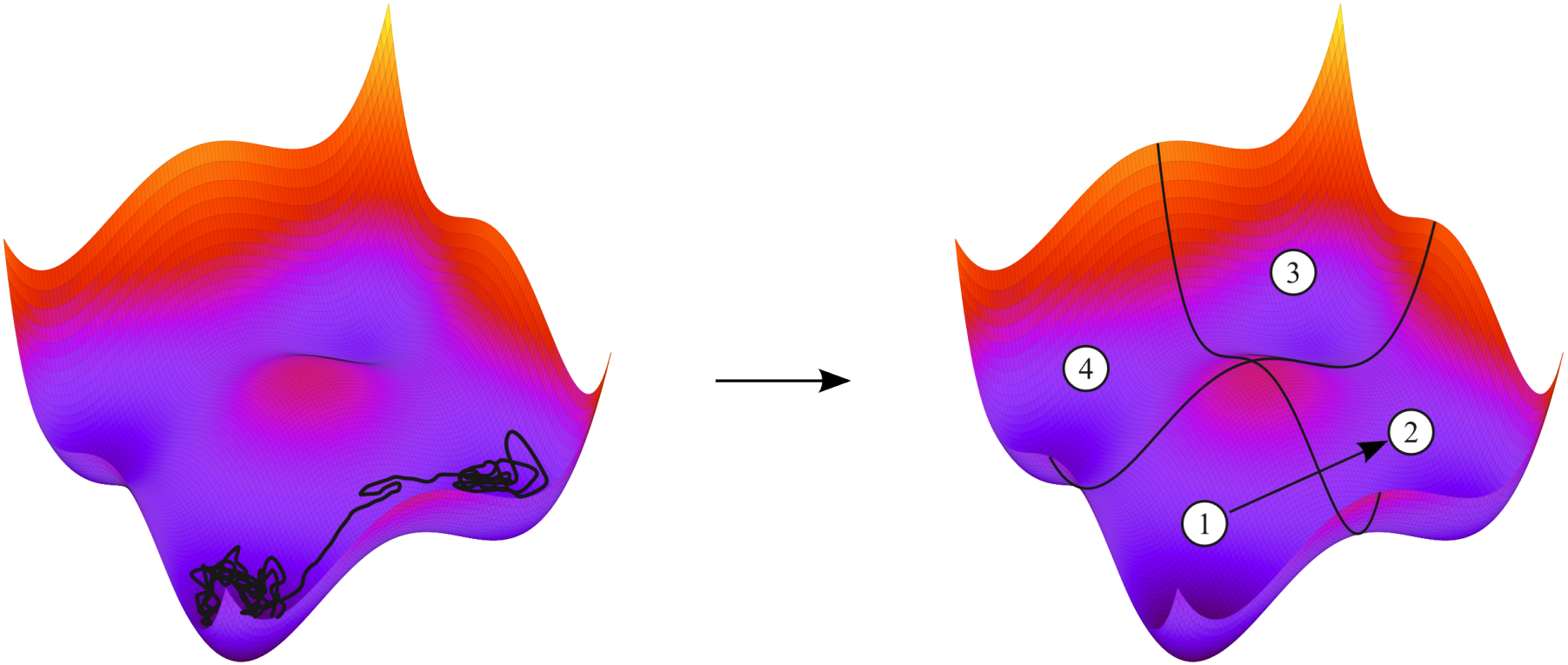


The Timescale Problem

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



Accelerating the dynamics



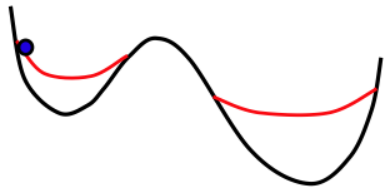
Possible solutions to the timescale problem

- In the following, I **assume** that a separation of timescales exists, i.e., I only consider **rare event dynamics**
- Extremely common situation, but not universal
- Very difficult to accelerate dynamics without such a separation. **You generally cannot speedup what is already fast**.
- Separation is typically (but not exclusively) between vibrations and transitions

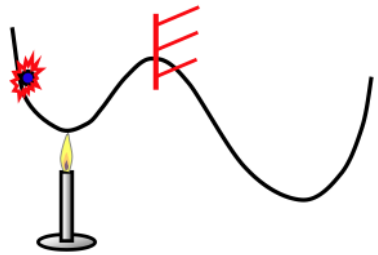
Accelerated Molecular Dynamics

Three AMD methods have been proposed by Arthur F. Voter *et al.* :

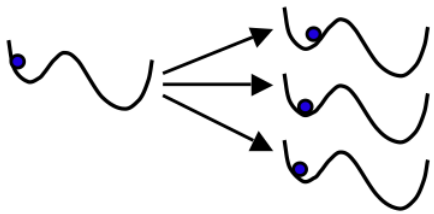
- **Hyperdynamics**



- **Temperature Accelerated Dynamics (TAD)**



- **Parallel Replica Dynamics (ParRep)**



Accelerated Molecular Dynamics : key points

Just like MD, but (often) better.

- Goal is to generate proper dynamics
- Open-ended : specify an initial state and let the system go
- Stochastic : different runs can give different answers
- (Ideally) require minimal *a priori* knowledge. No need to choose a reaction coordinate.
- Typically more accurate than KMC
- Typically not as fast as KMC

HYPERDYNAMICS

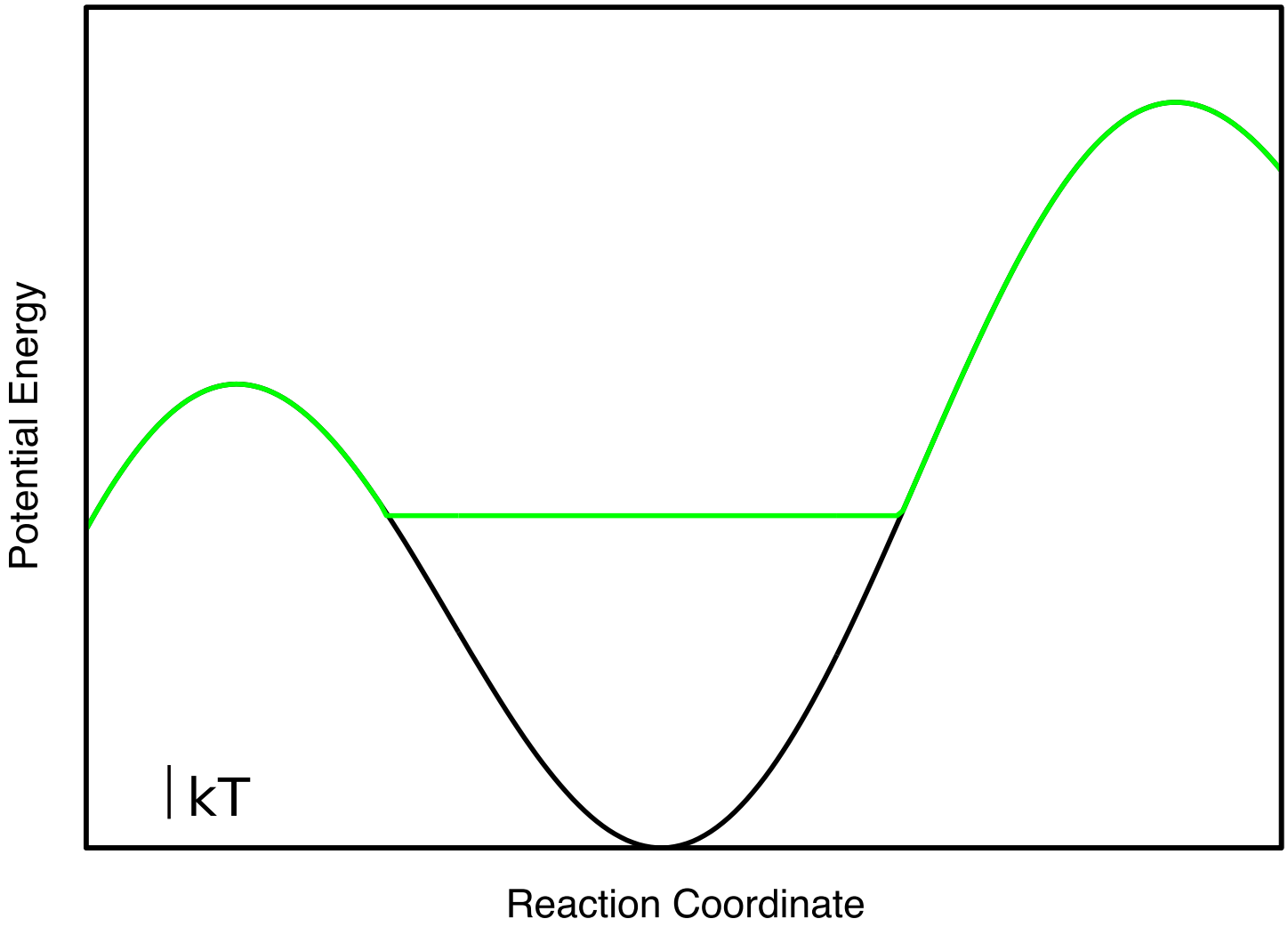
Principle of Hyperdynamics

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 :

$$k_{A\rightarrow}^{\mathrm{TST}} = \langle |v_A| \delta_A(\mathbf{r}) \rangle_A \tag{1}$$

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

$$k_{A\rightarrow}^{\mathrm{TST}} = \frac{\langle |v_A| \delta_A(\mathbf{r}) e^{\beta \Delta V_b(\mathbf{r})} \rangle_{A_b}}{\langle e^{\beta \Delta V_b(\mathbf{r})} \rangle_{A_b}}. \tag{2}$$

Require that the bias vanishes along all dividing surfaces :

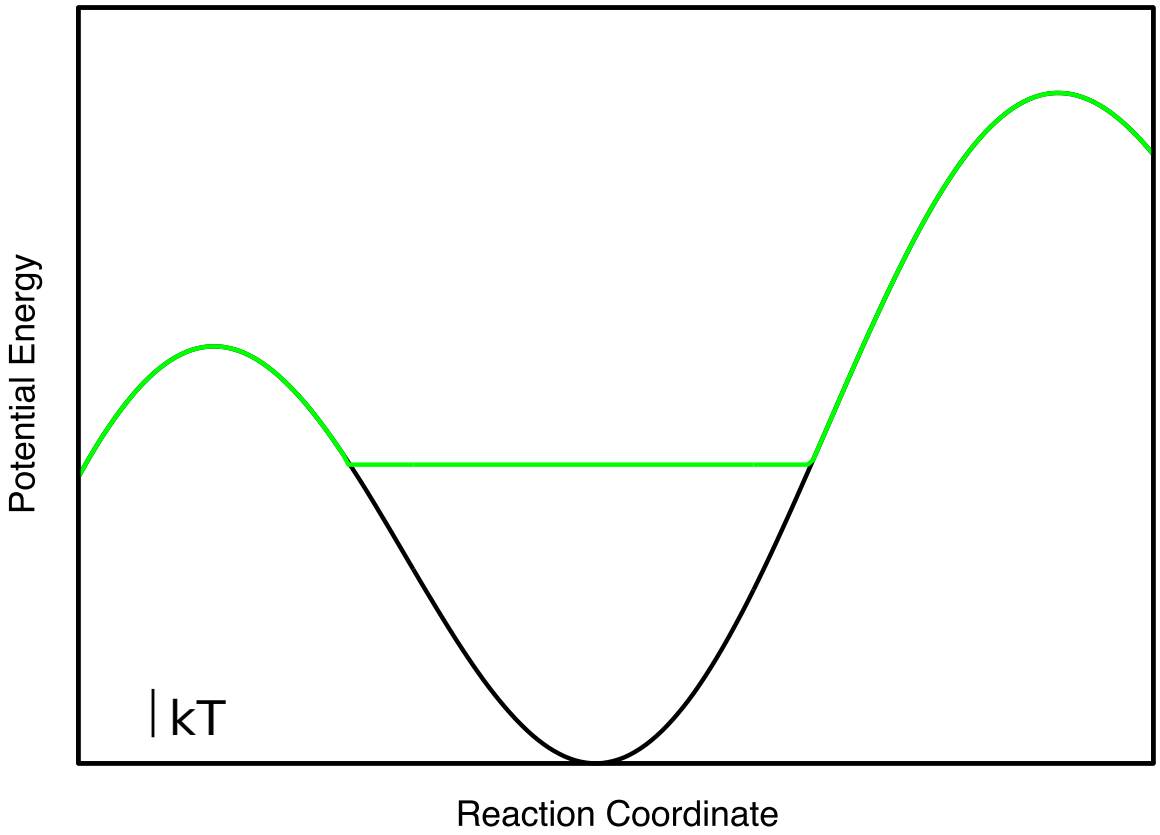
$$\Delta V_b(\mathbf{r}) = 0 \text{ when } \delta_A(\mathbf{r}) \neq 0, \tag{3}$$

then :

$$k_{A\rightarrow}^{\mathrm{TST}} = \frac{\langle |v_A| \delta_A(\mathbf{r}) \rangle_{A_b}}{\langle e^{\beta \Delta V_b(\mathbf{r})} \rangle_{A_b}} = \frac{k_{A_b\rightarrow}^{\mathrm{TST}}}{\langle e^{\beta \Delta V_b(\mathbf{r})} \rangle_{A_b}}. \tag{4}$$

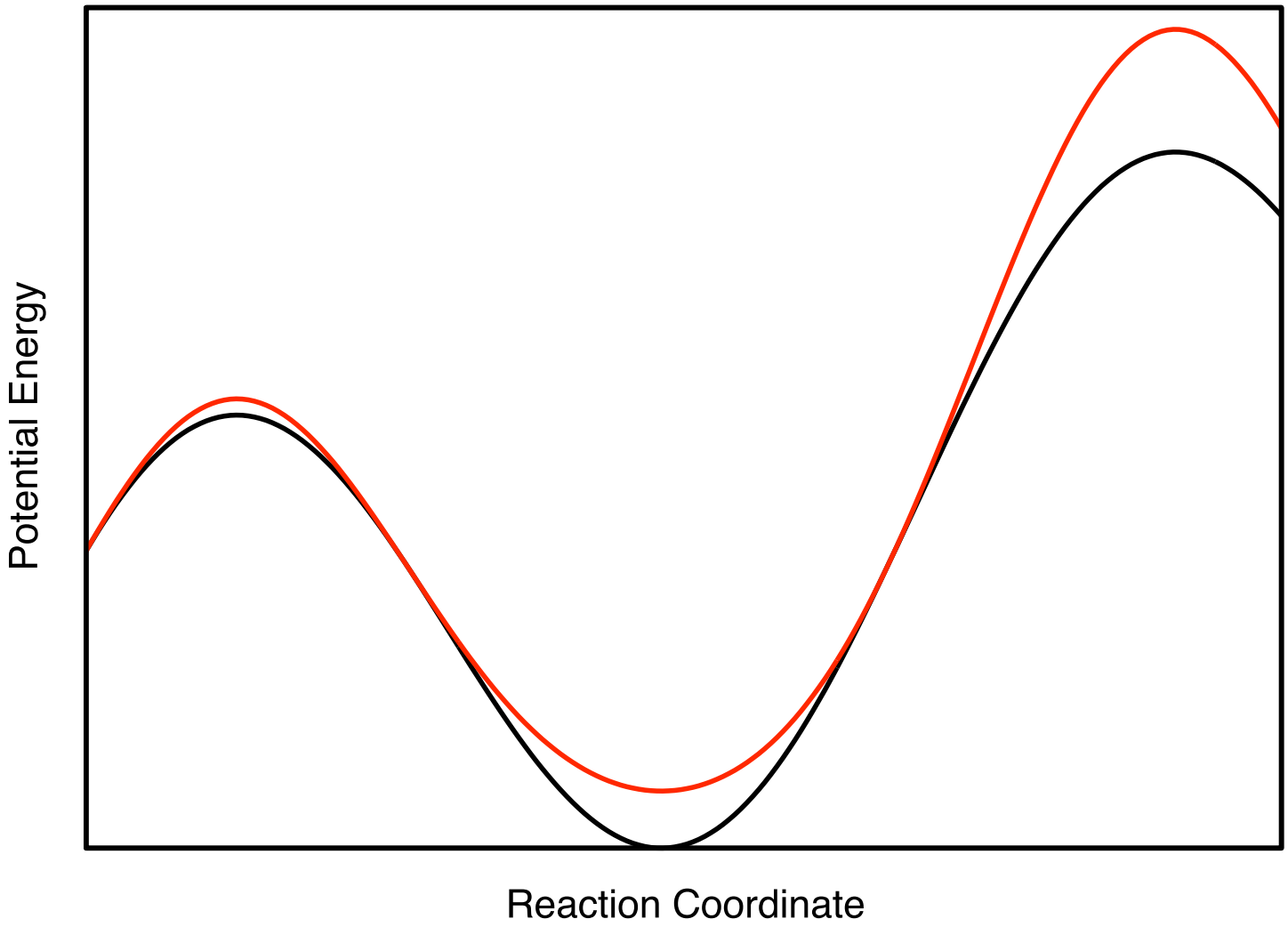
Time flows $\langle e^{\beta \Delta V_b(\mathbf{r})}$ times faster on the biased surface than on the original one !

Hyperdynamics : Illustration



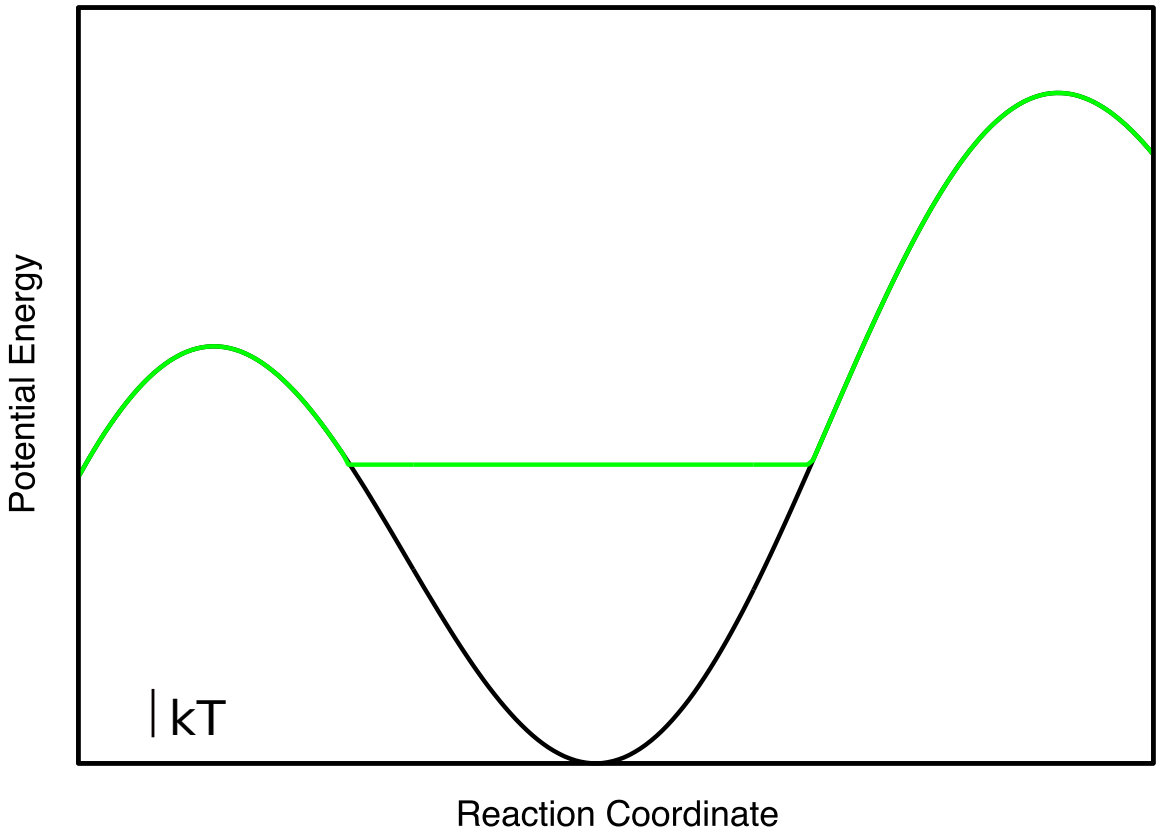
Condition 1 : ΔV_b vanishes along all dividing surfaces
Condition 2 : The system obeys TST on both V and $V + \Delta V_b$

Hyperdynamics : Illustration



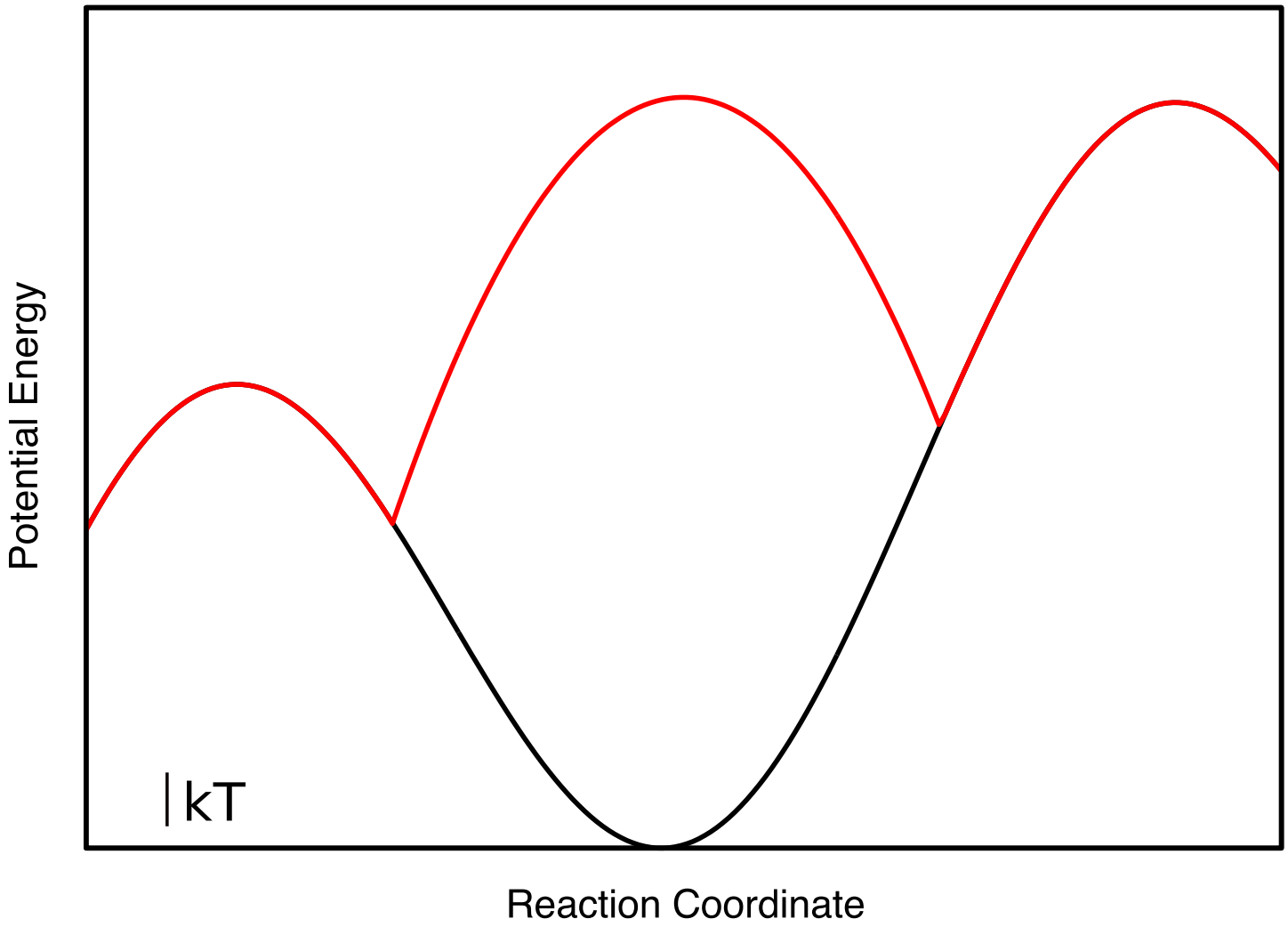
BAD !

Hyperdynamics : Illustration



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 : 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 in practice

Pros :

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

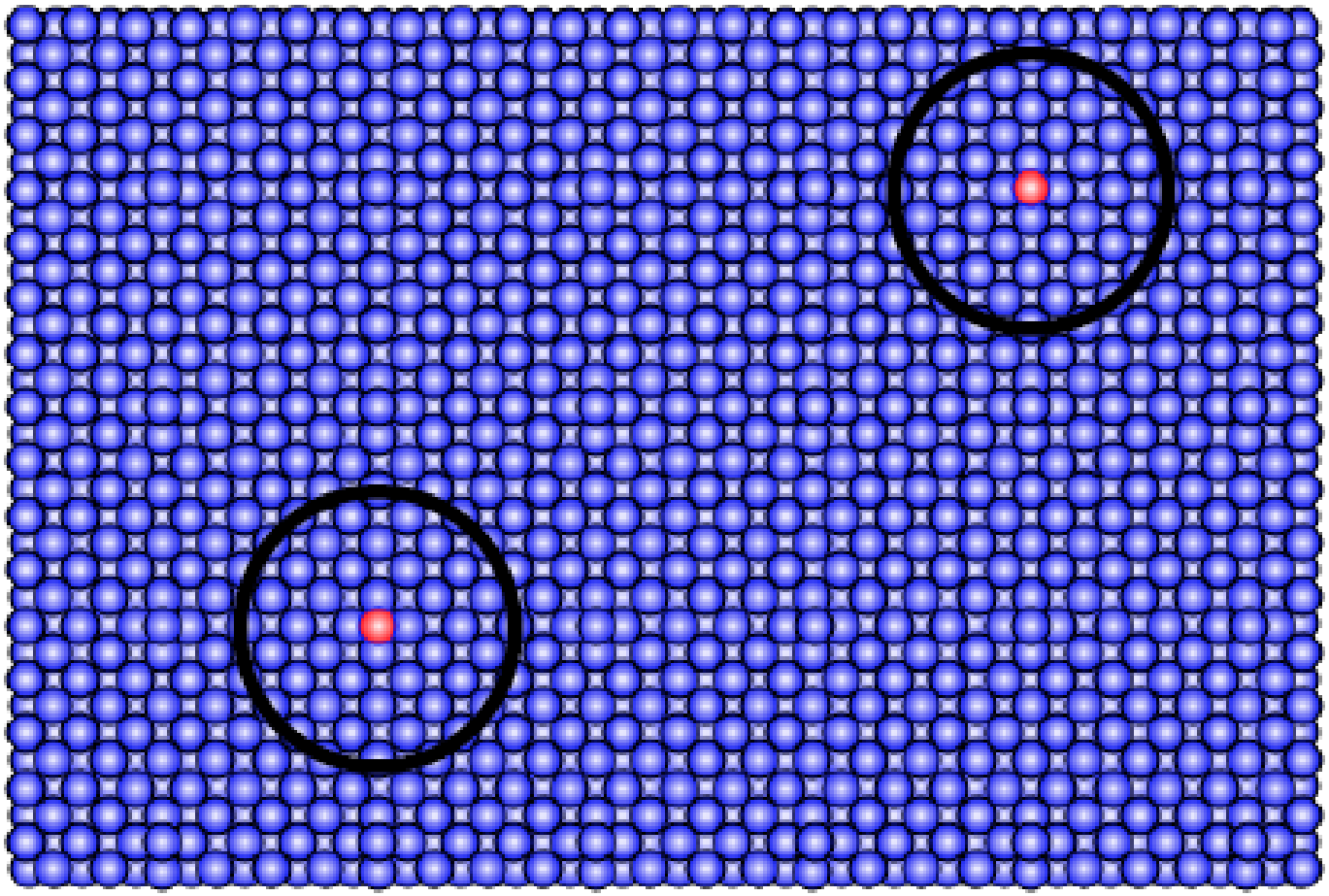
Cons :

- 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.

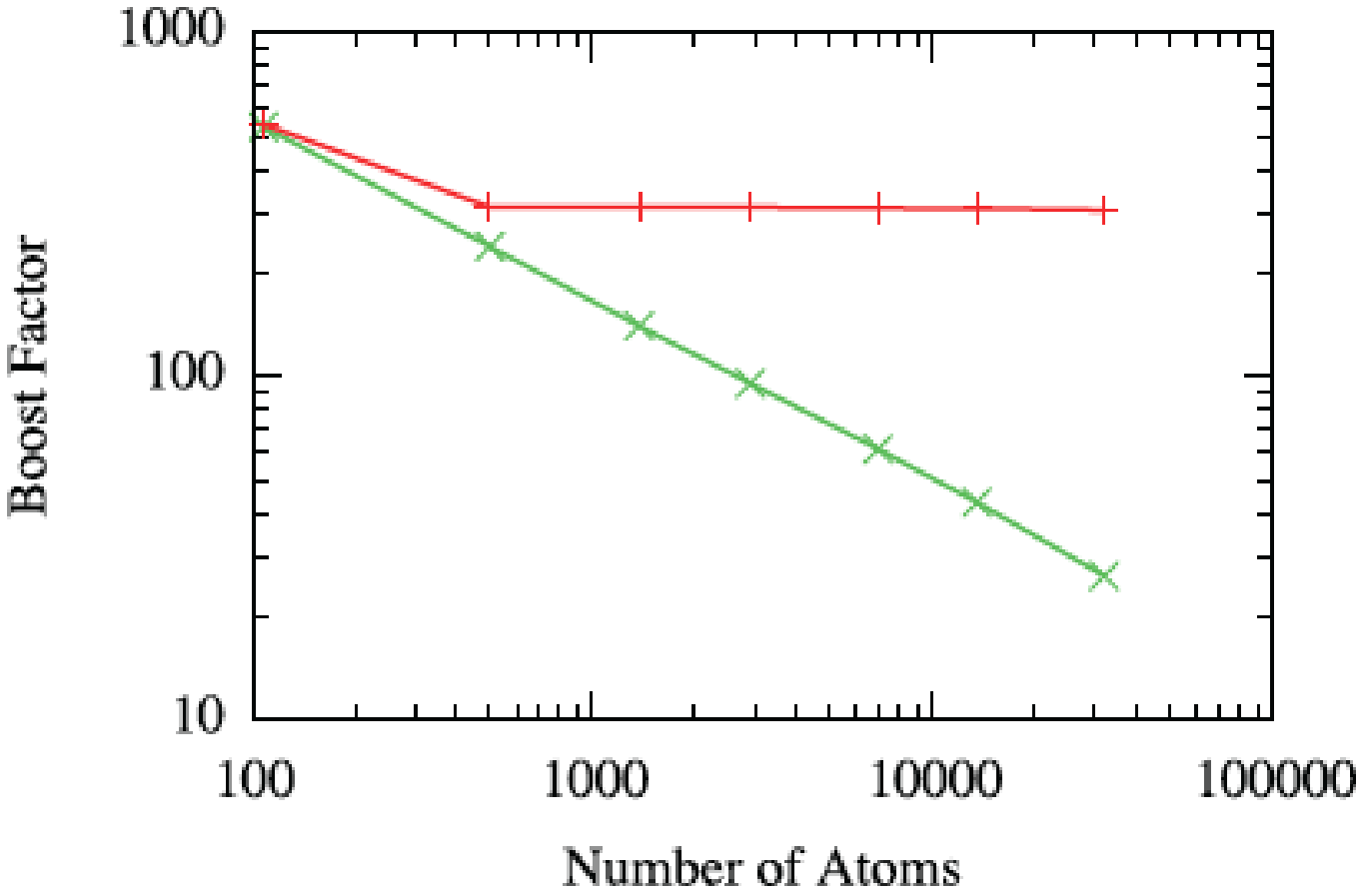
See Ann. Rep. Comp. Chem. 5, 79, (2009) for a review.

Hyperdynamics : generalizations



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

Hyperdynamics : generalizations



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

Massively parallel local hyperdynamics

- Local hyperdynamics is ideally suited to massively-parallel architectures.
- It allows for simultaneously reaching long times and large scales.

Demonstration using LAMMPS (S. Plimpton, SNL) :

- Pt/Pt(100), 4% adatom coverage
- 10^6 atoms
- 24h on 4096 cores
- Simulation time : 1 ms (4000x boost)

Hyperdynamics : generalizations

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 ?

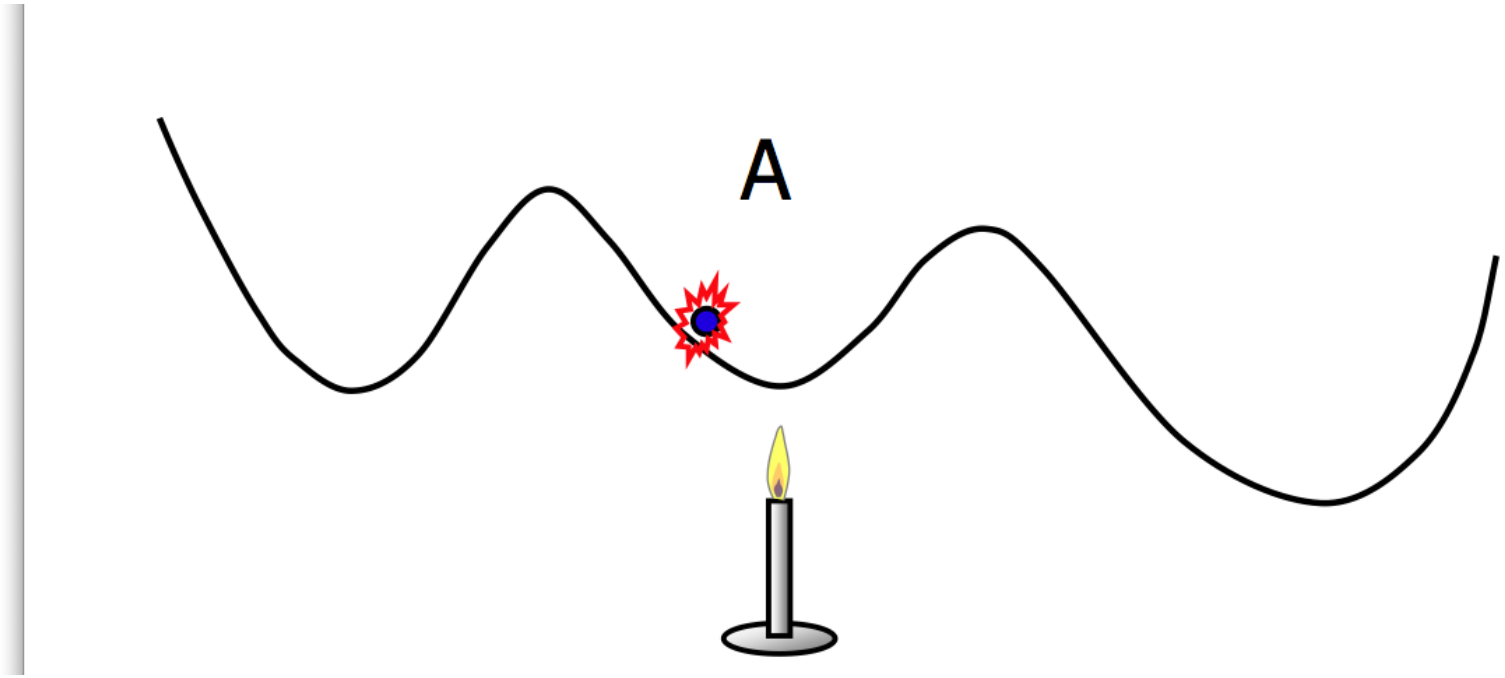
Principle of Temperature Accelerated Dynamics (TAD)

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

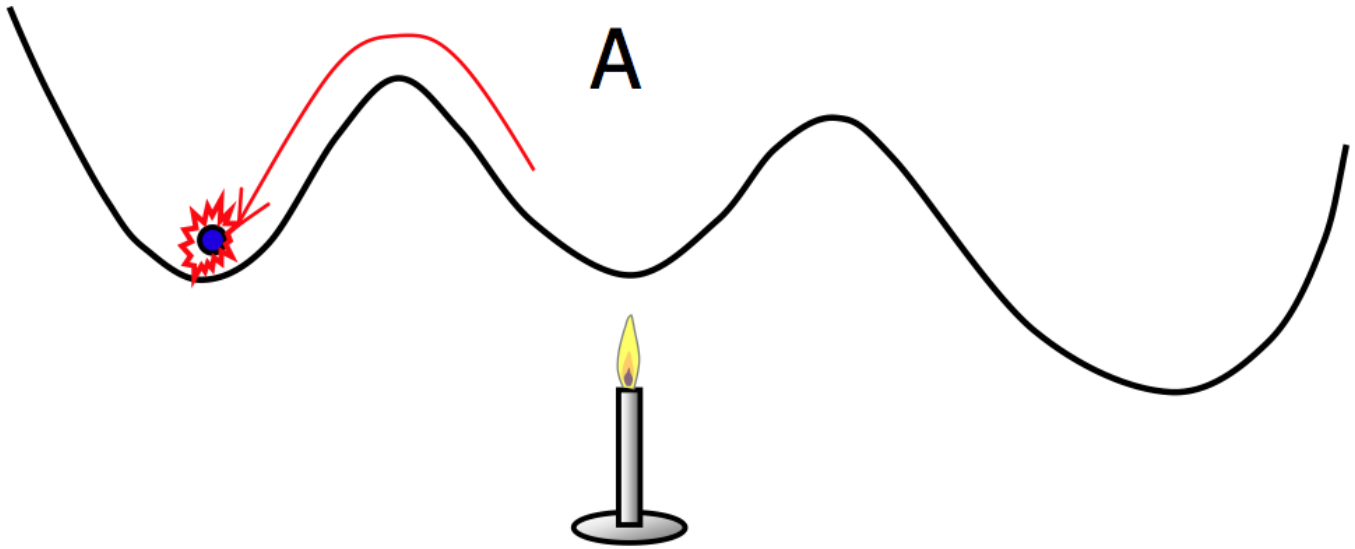
In TAD [Sorensen and Voter, JCP 112, 9599, (2000)], you :

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

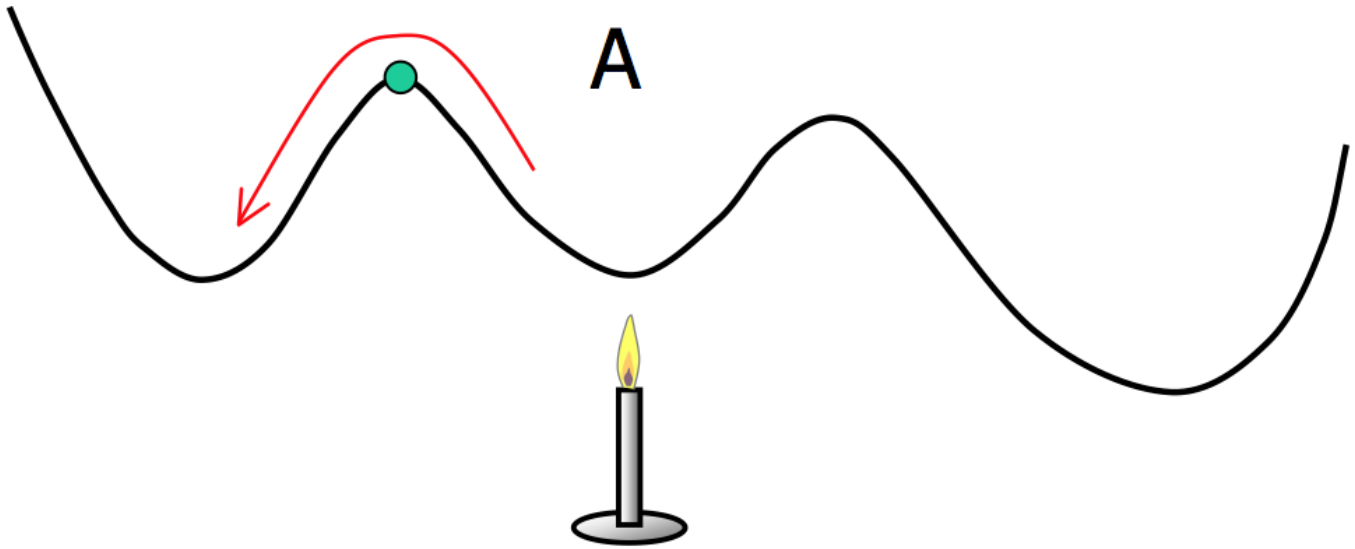
TAD : Illustration



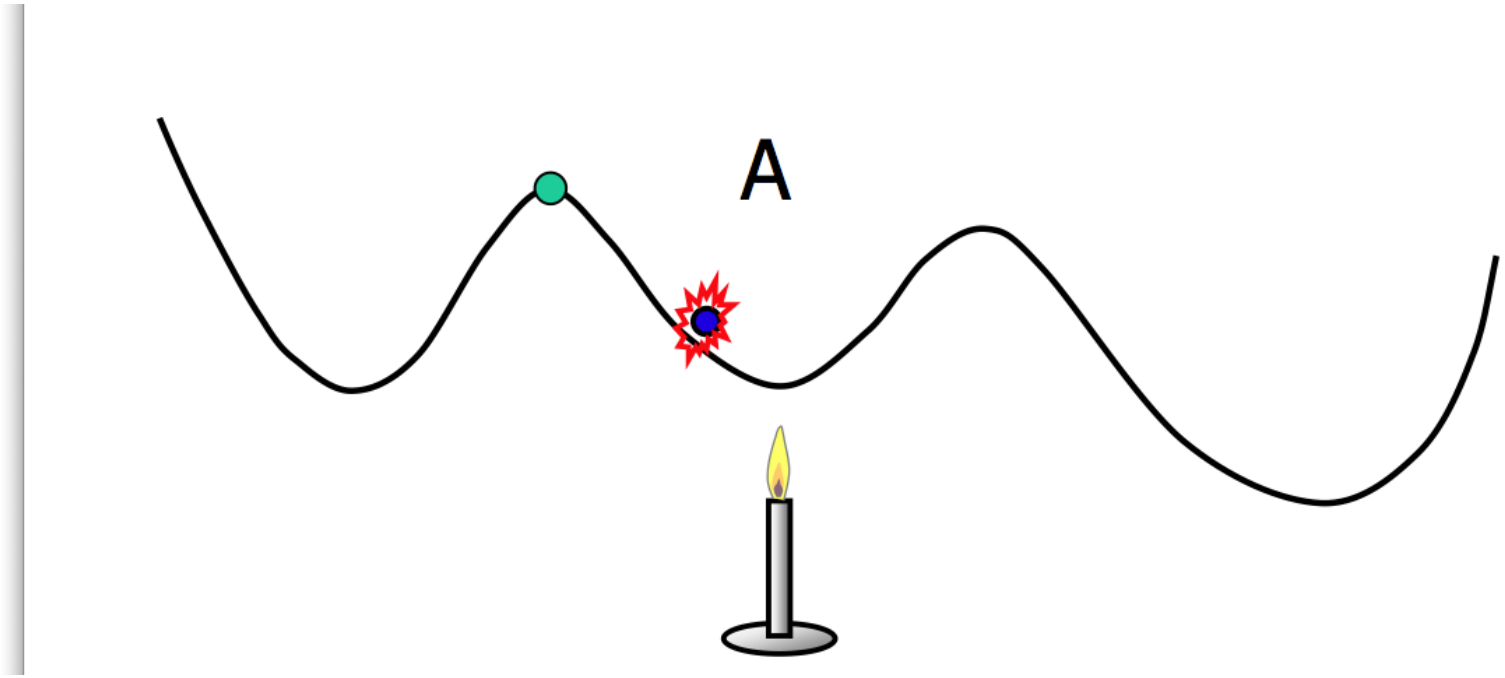
TAD : Illustration



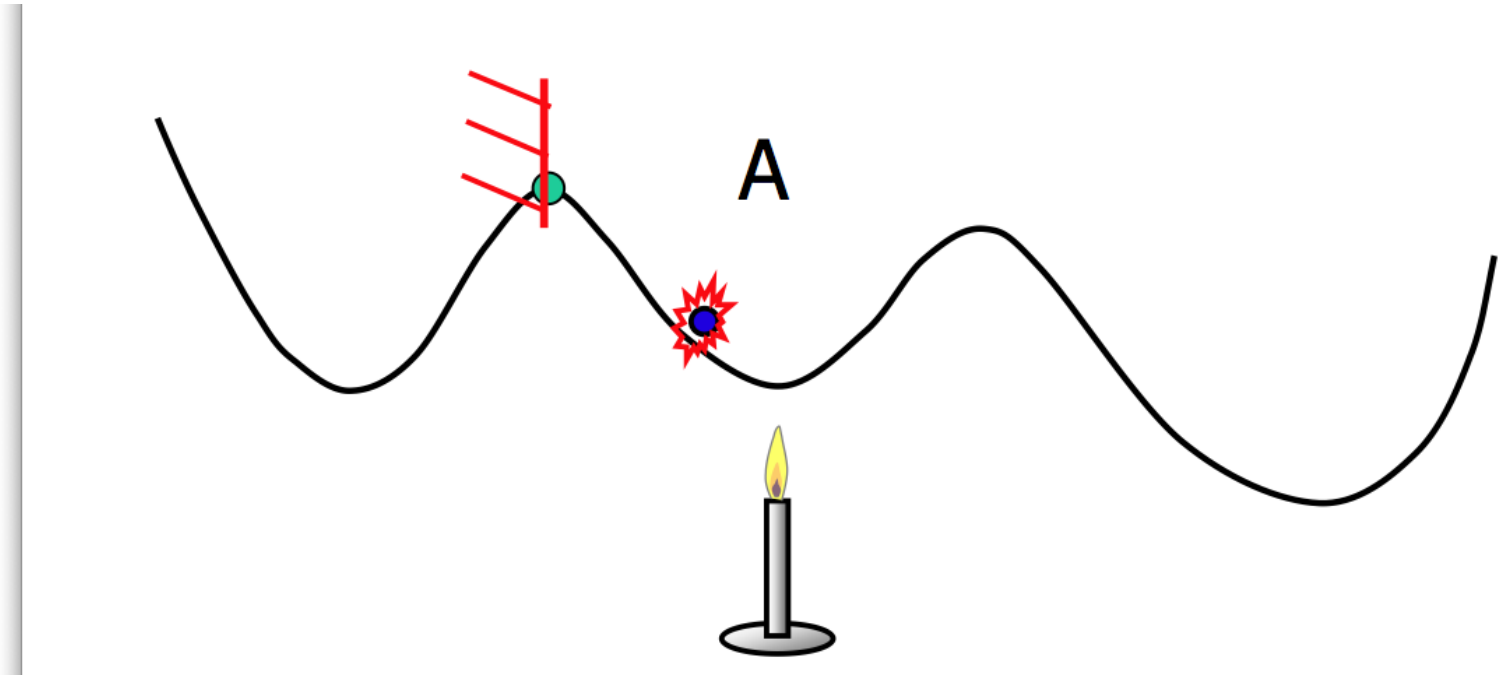
TAD : Illustration



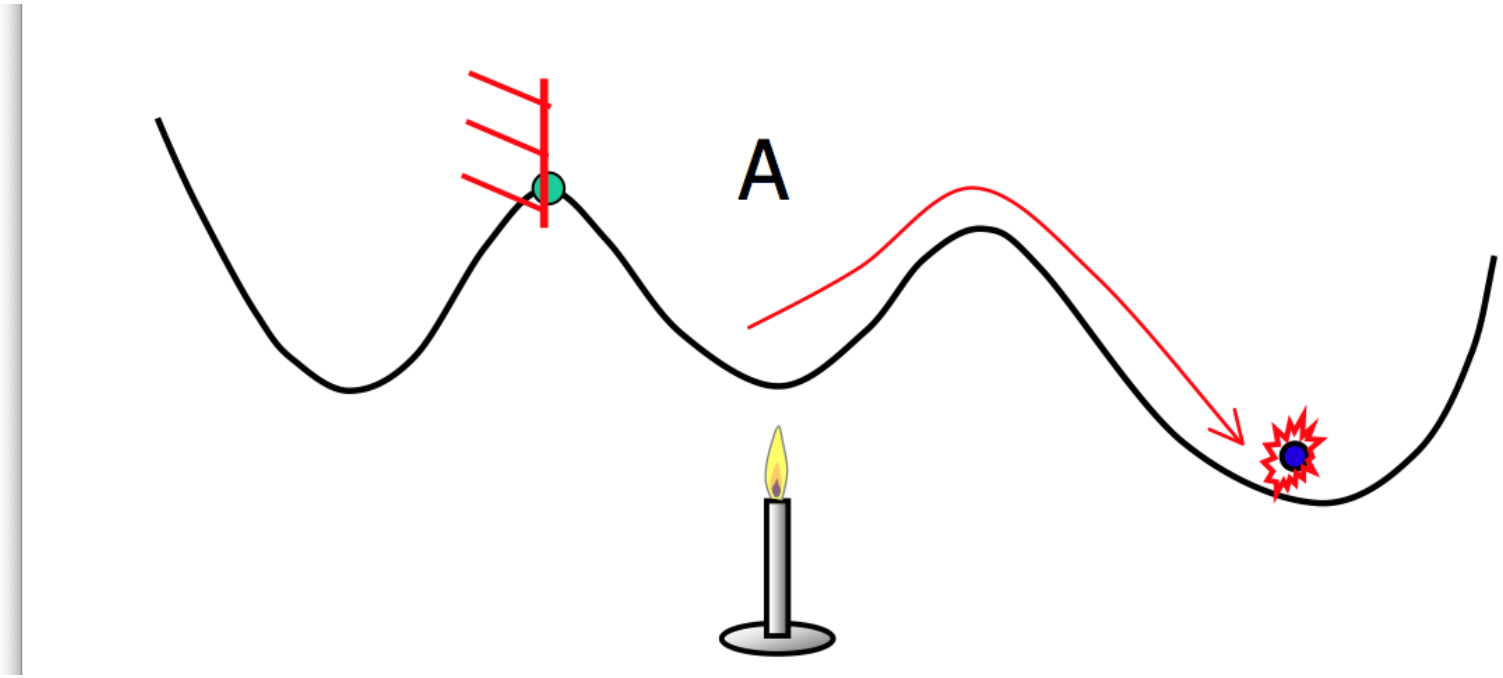
TAD : Illustration



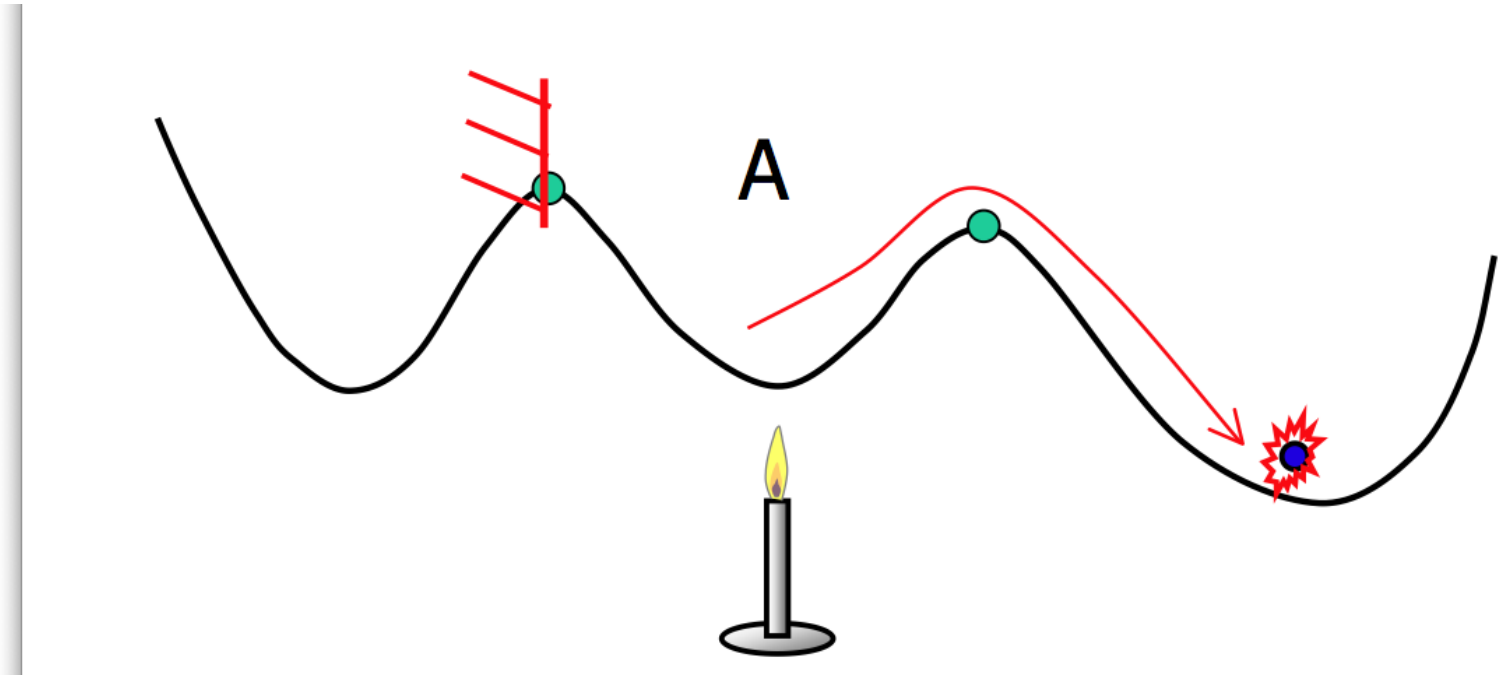
TAD : Illustration



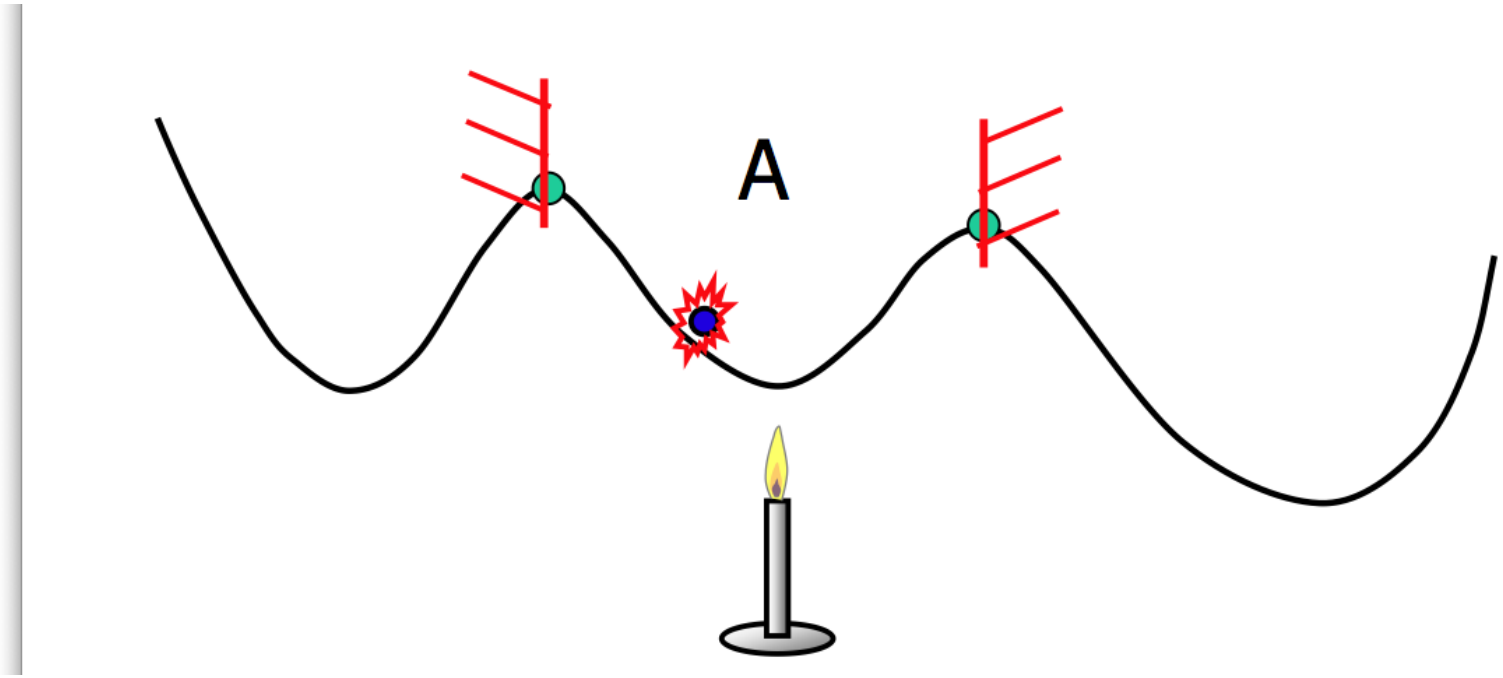
TAD : Illustration



TAD : Illustration



TAD : Illustration



TAD : Derivation

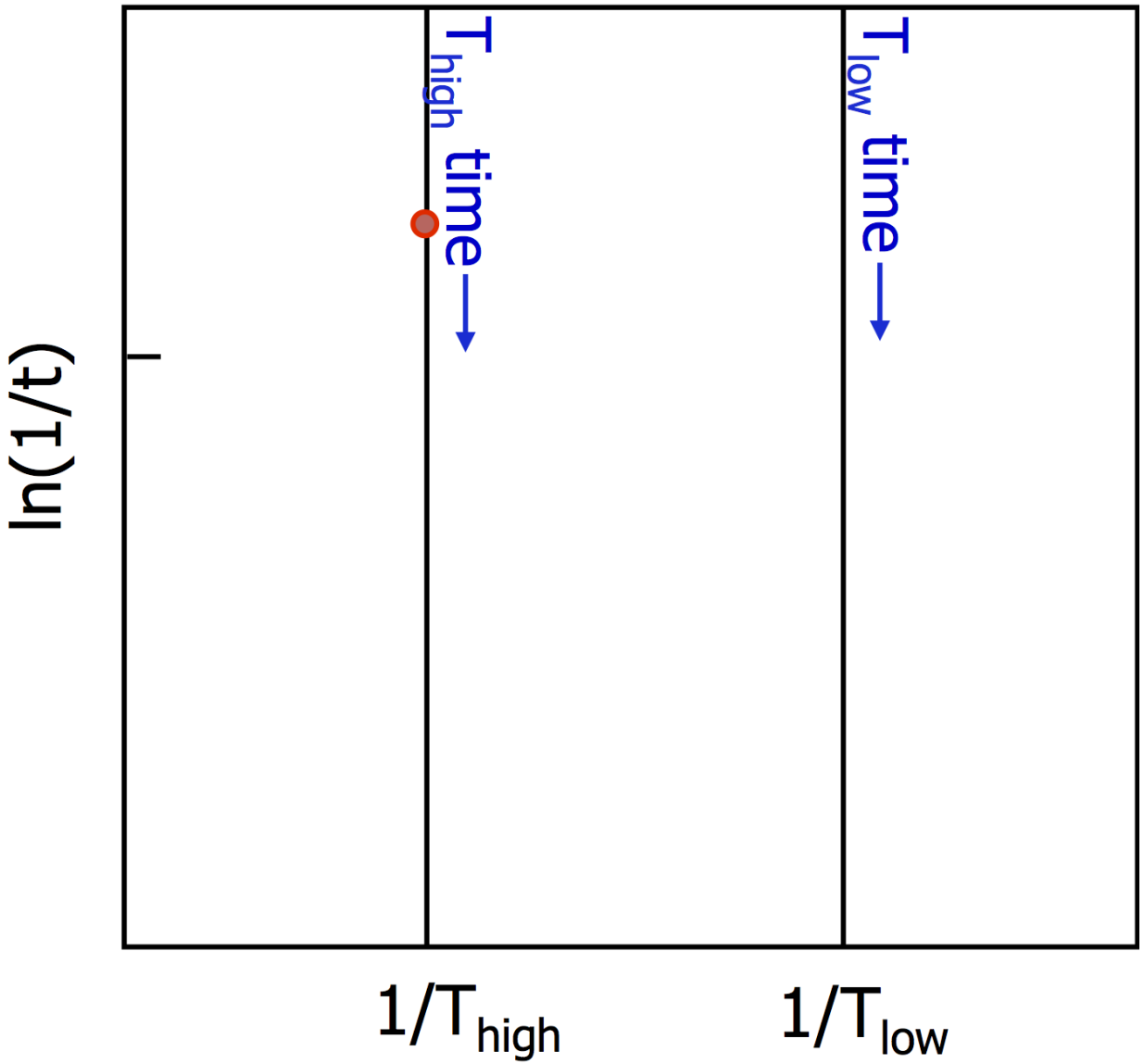
When should we stop and accept a transition ?

When we can statistically establish that we have observed the "right" first event that should have occurred first at low temperature, i.e., that running longer is unlikely to change our conclusion.

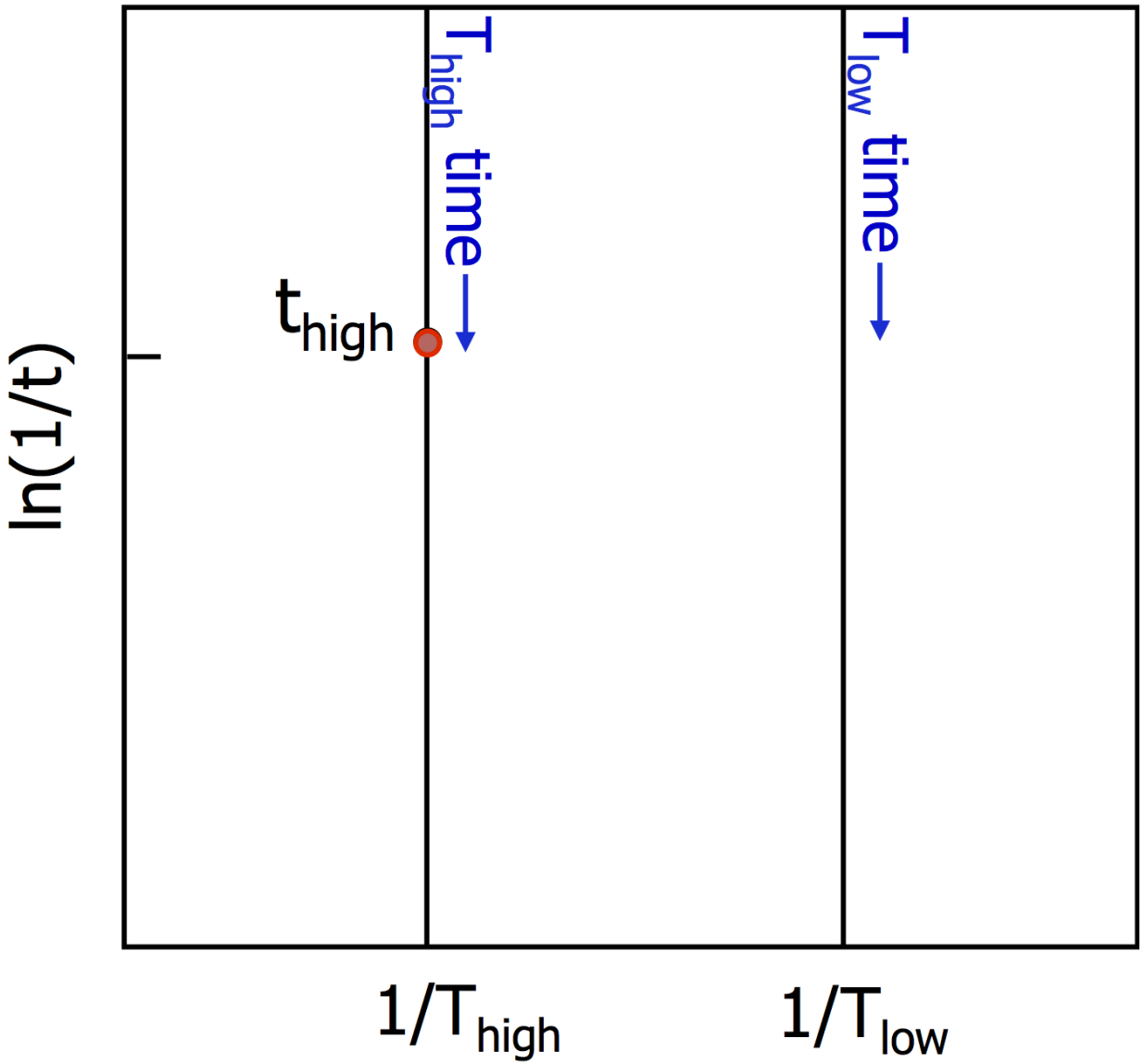
This is possible under two assumptions :

- Harmonic TST holds : $k = \nu \exp(-\beta \Delta E)$
- All prefactors are higher than ν_{min}

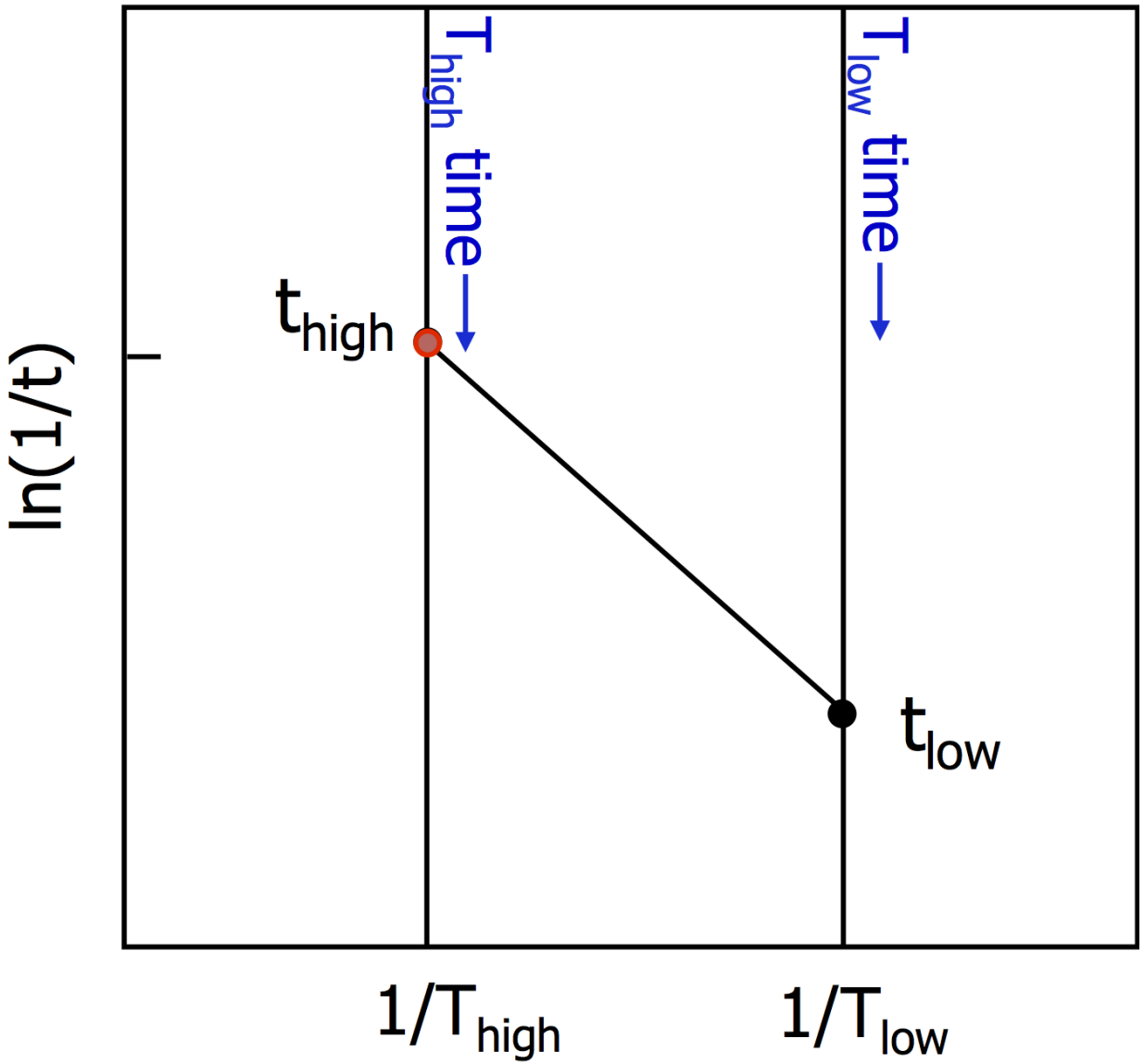
TAD : Illustrations



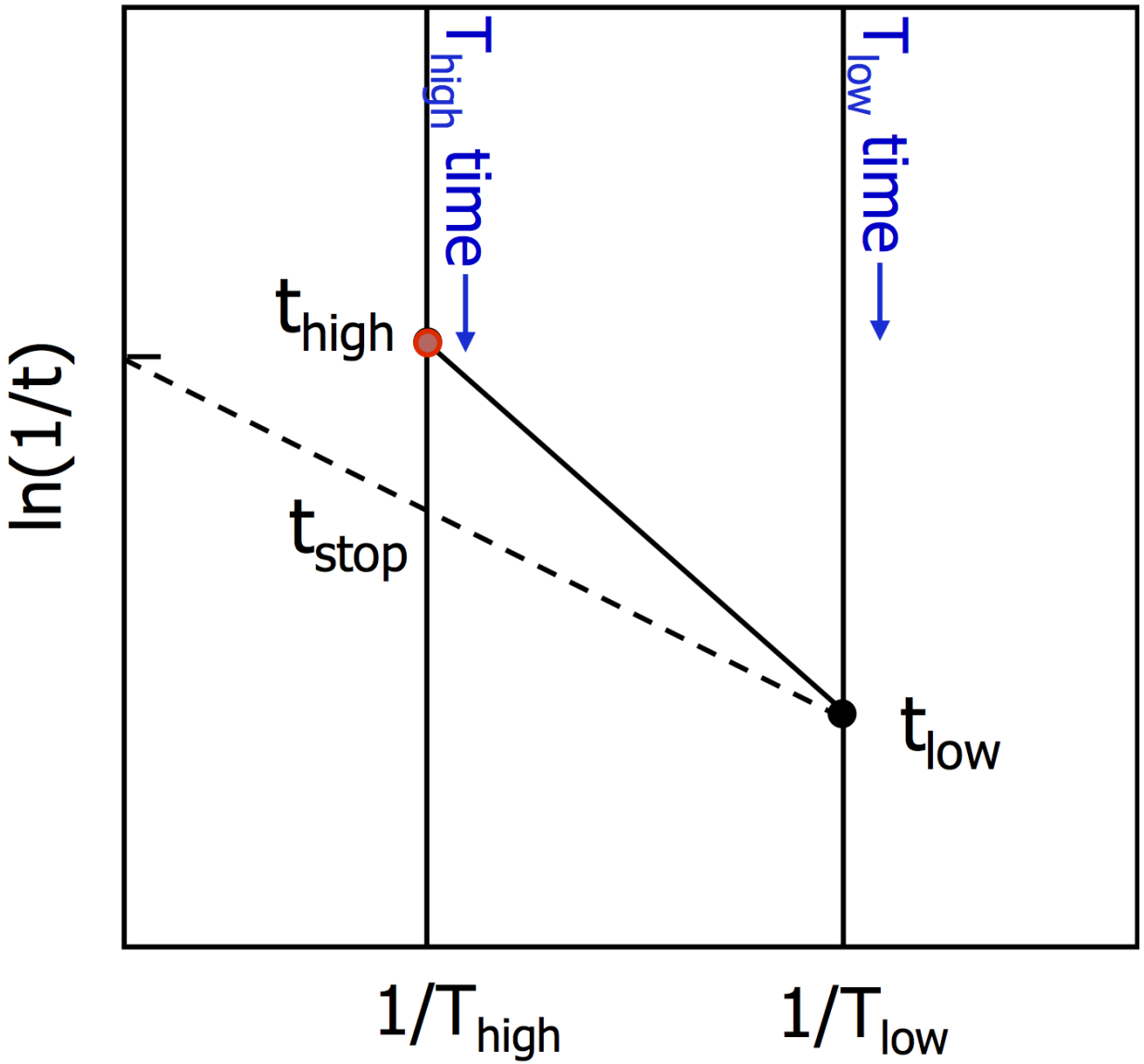
TAD : Illustrations



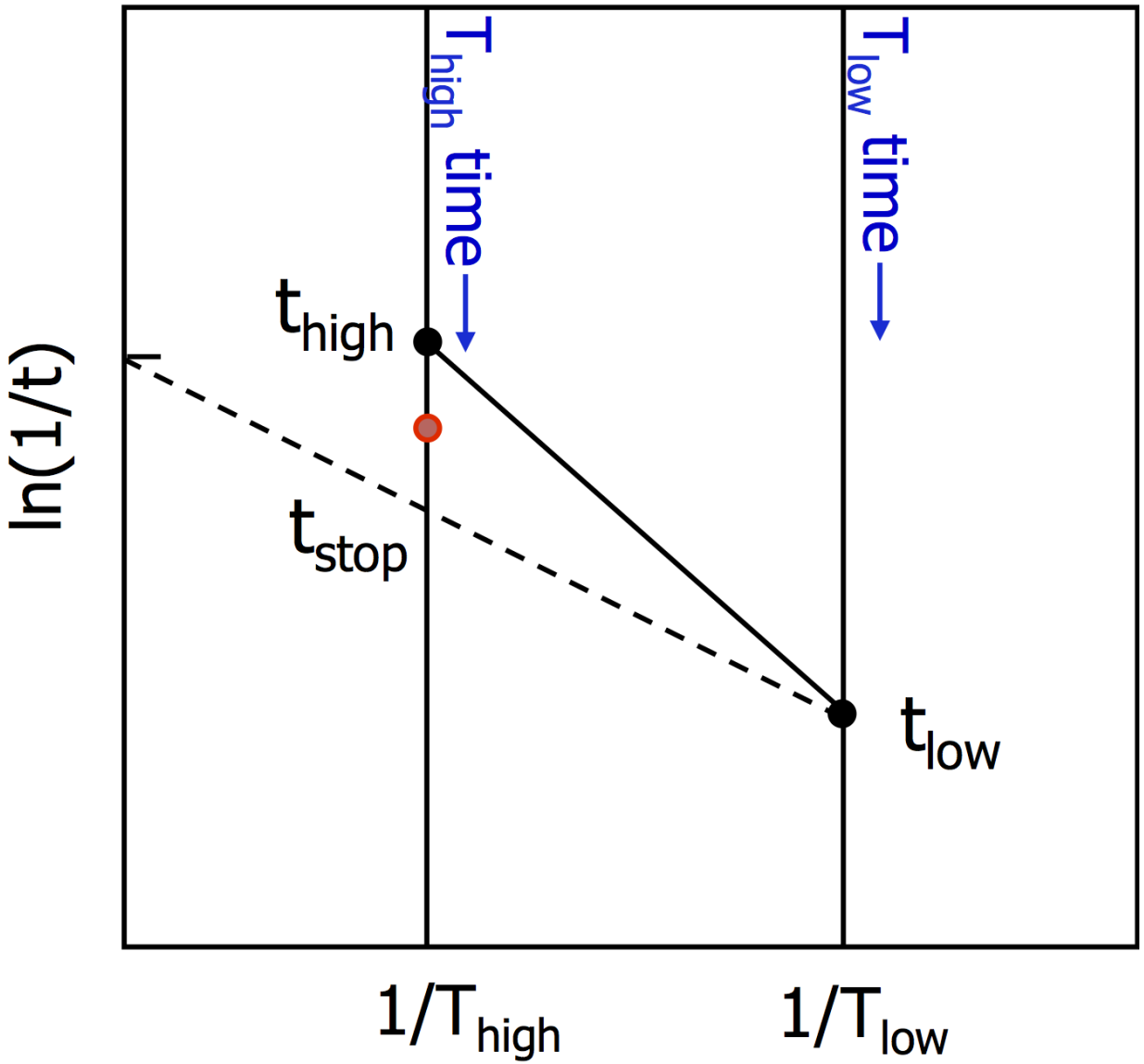
TAD : Illustrations



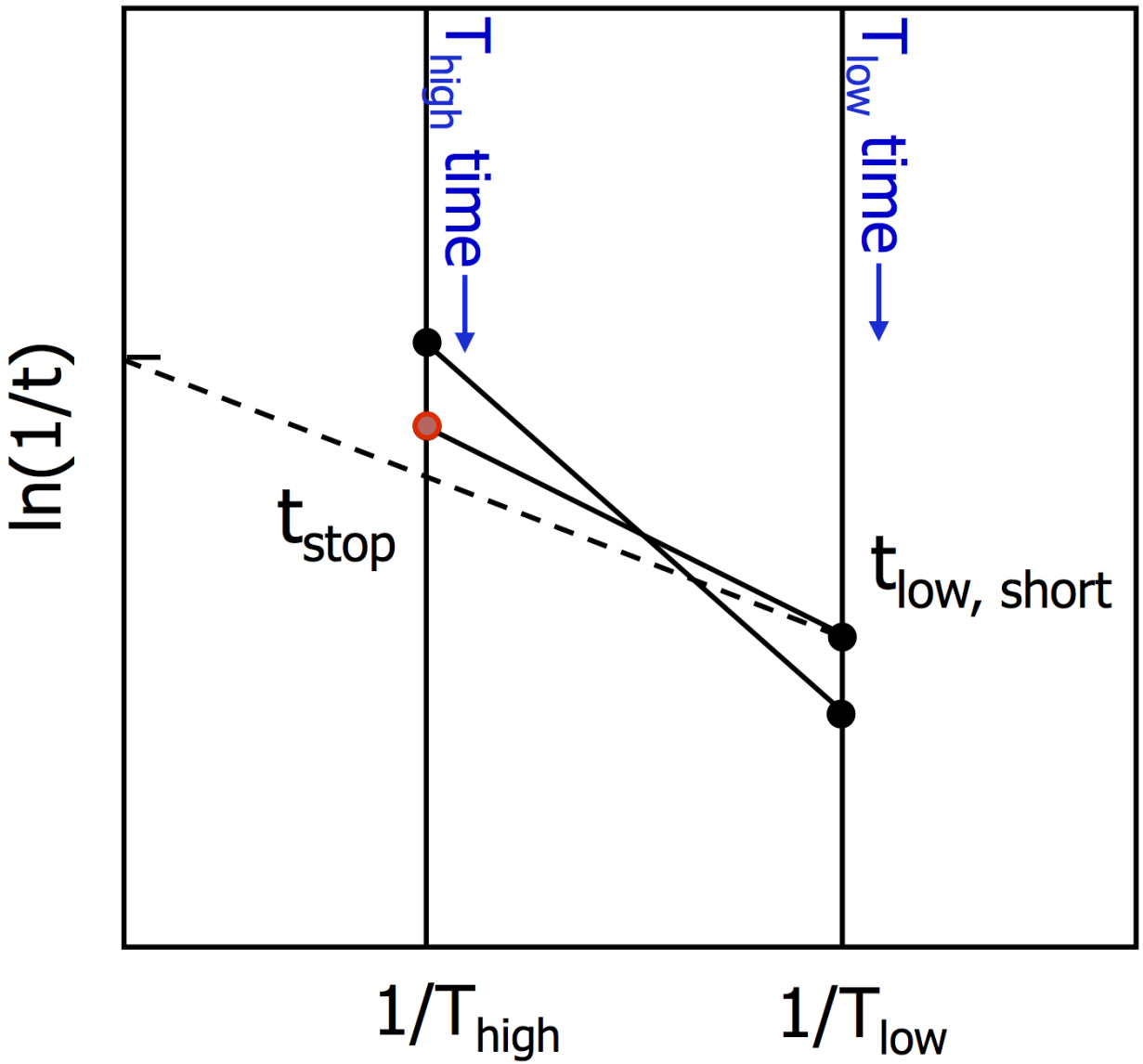
TAD : Illustrations



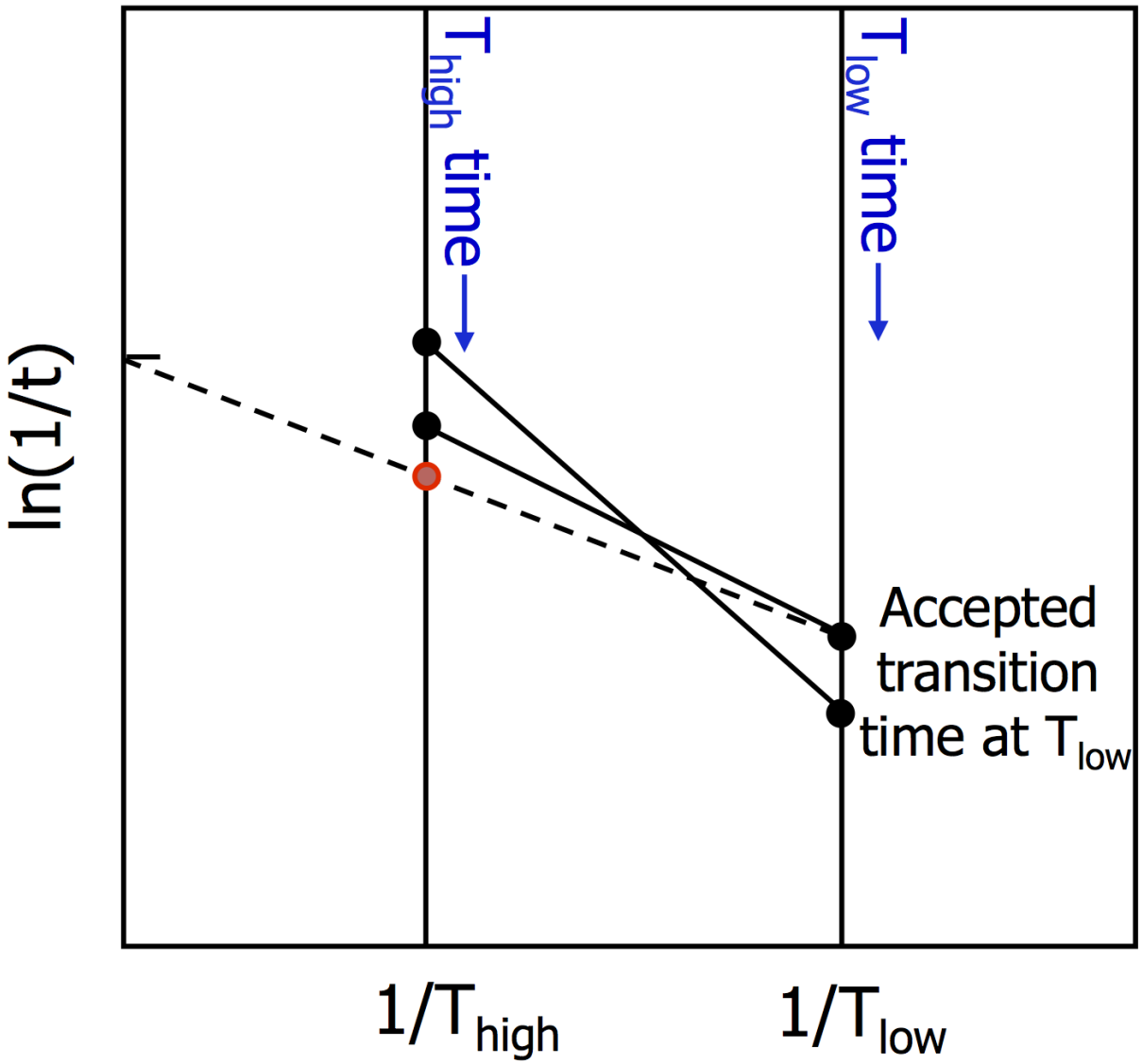
TAD : Illustrations



TAD : Illustrations



TAD : Illustrations



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, 87-110 (2016) for a recent review.

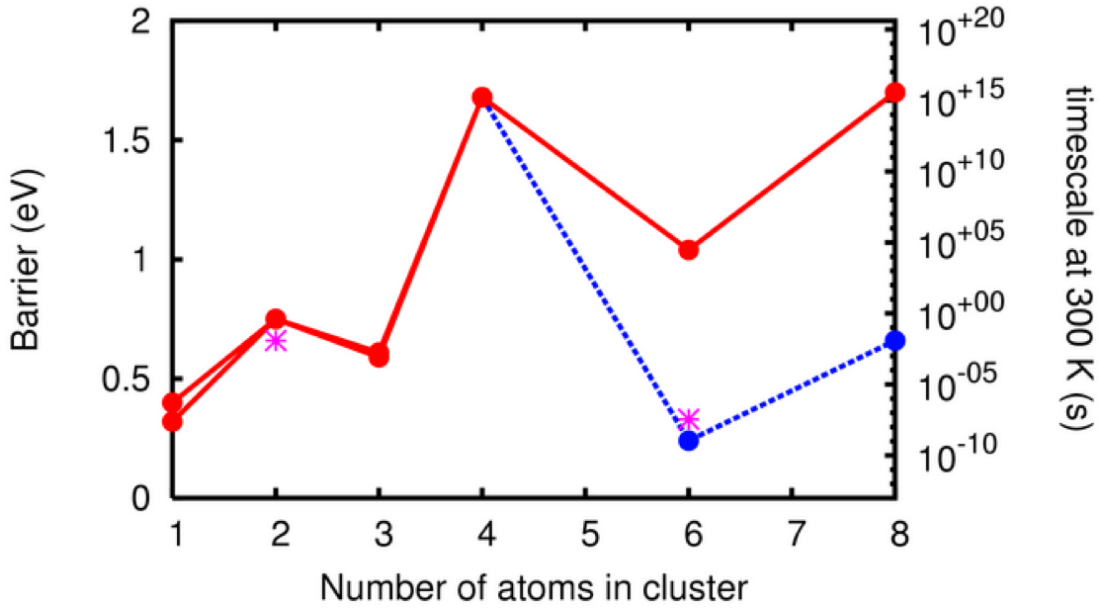
TAD example : Interstitial clusters in MgO

- 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 : $T_{high} = 2000K$, $T_{low} = 300K$

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)

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)].

Illustration of ParRep

Illustration of ParRep

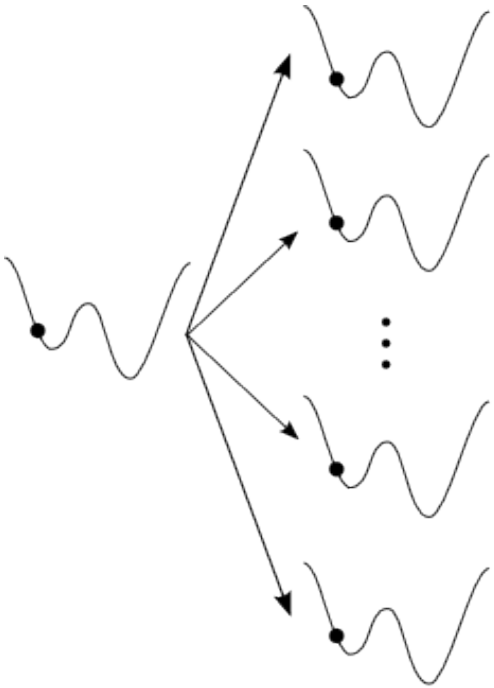
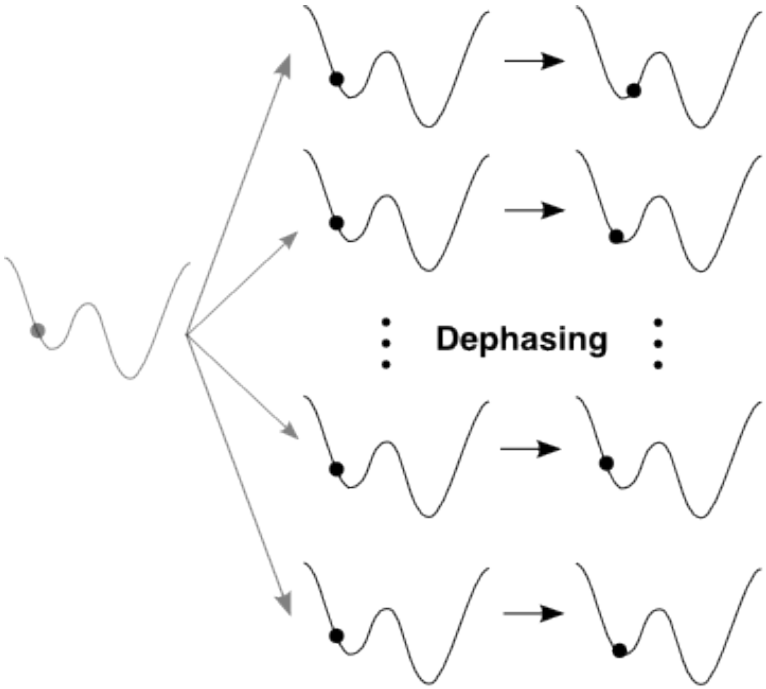
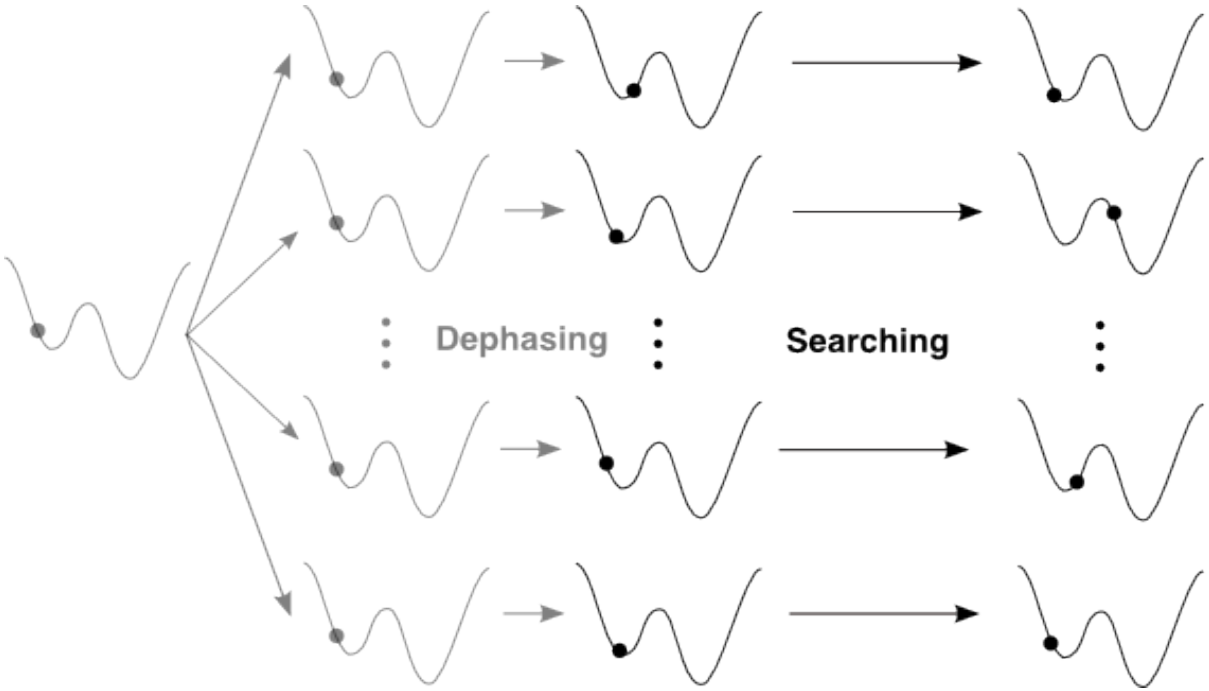


Illustration of ParRep



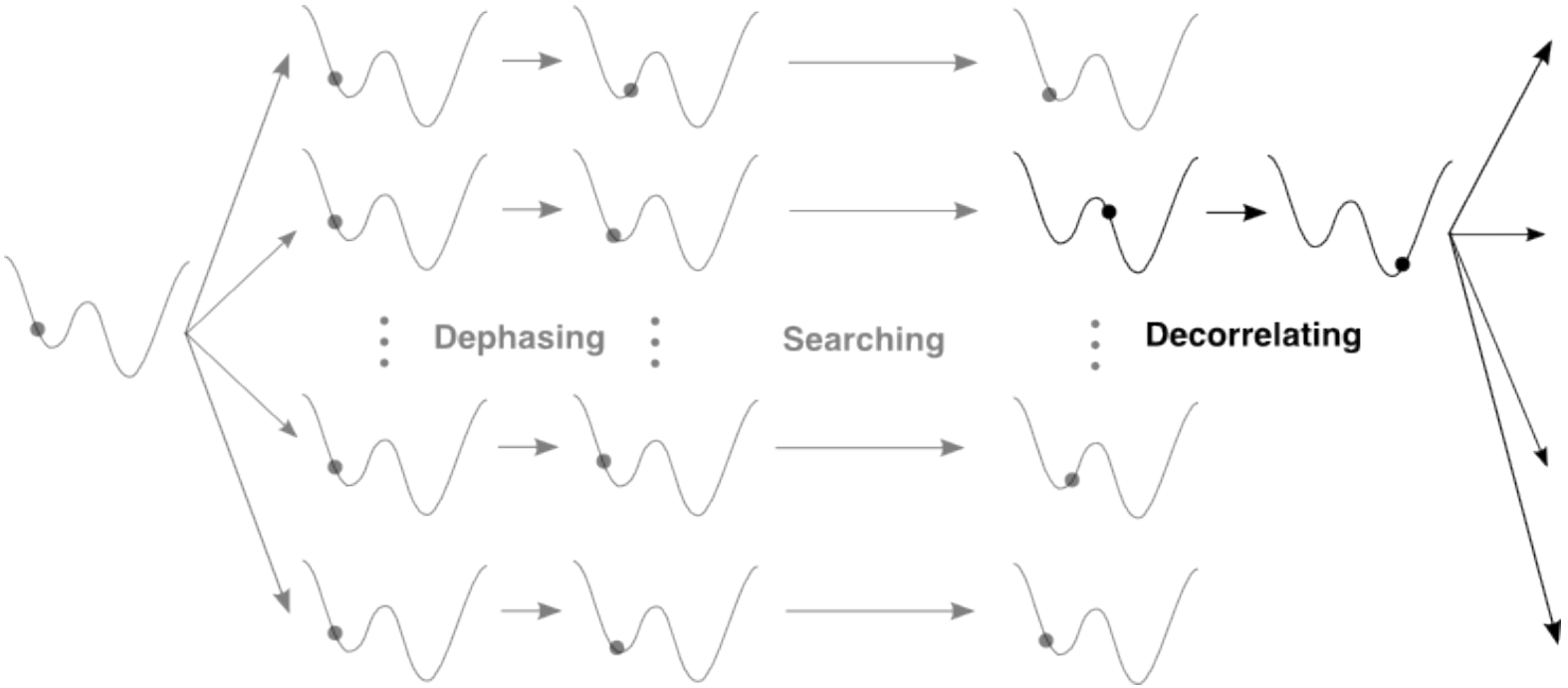
τ_c

Illustration of ParRep



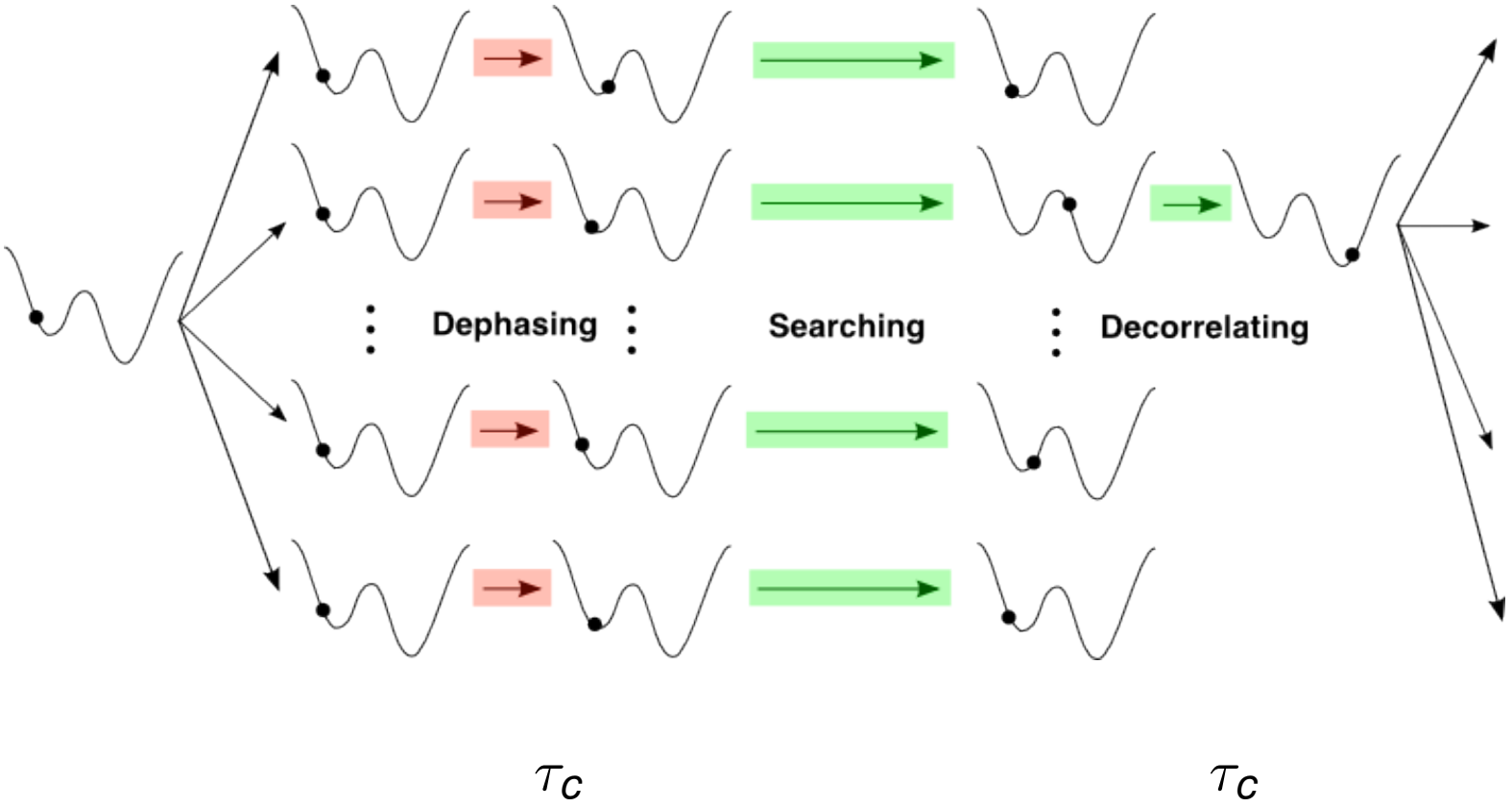
τ_c

Illustration of ParRep



τ_c τ_c

Illustration of ParRep



Justification of ParRep

Assume the the state-to-state dynamics are Markovian, i.e., memory-less. Then ParRep is statistically **exact**.

Even if kinetics are not purely Markovian, ParRep can still be made *arbitrarily accurate* by increasing τ_c.

[Le Bris, Lelievre, Luskin, and DP, Monte Carlo Methods and Applications 18, 119 (2012)]

ParRep in Practice

Pros :

- Very simple
- Arbitrarily accurate
- Flexible in terms of the definition of states
- Can handle driven systems

Cons :

- Requires parallel computers to get some acceleration

See Comp. Mat. Sci. 100, 90 (2015) for a recent review.

Application to Ag Nanowires

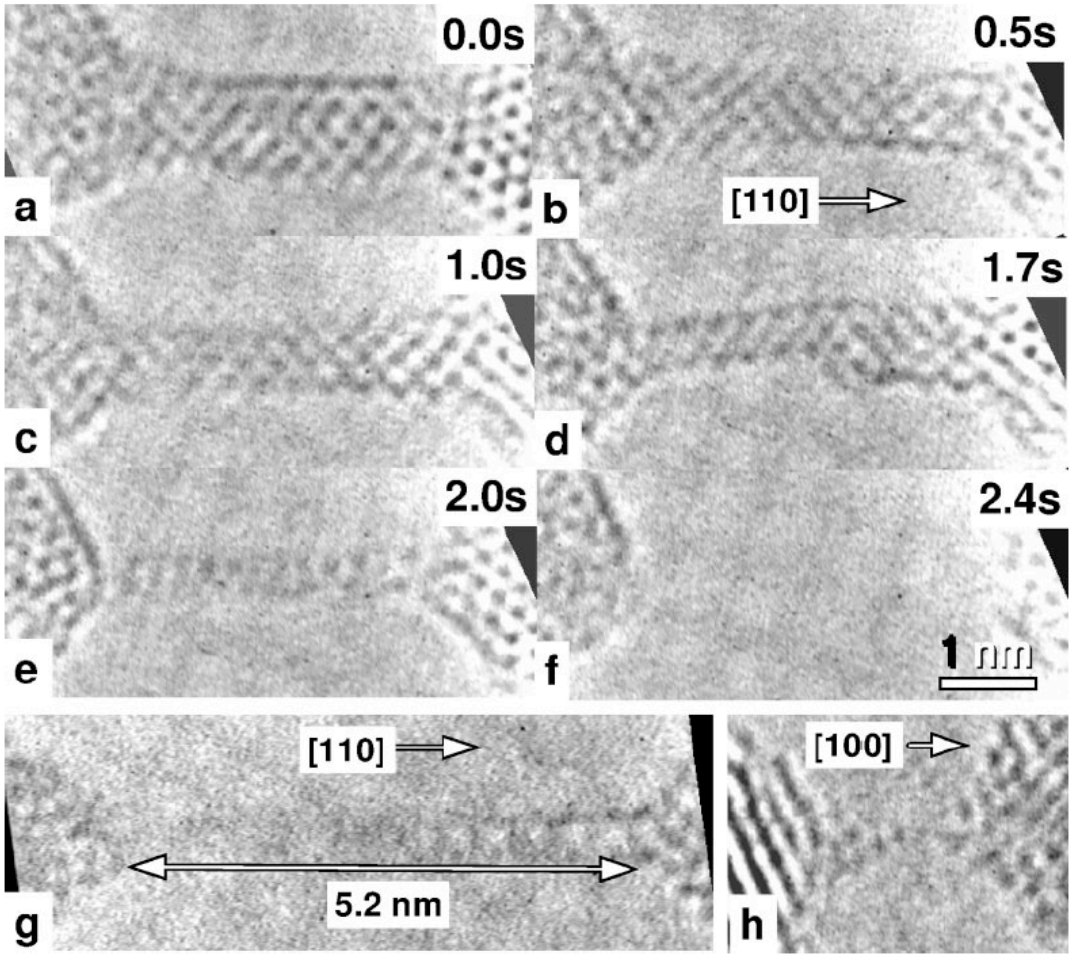
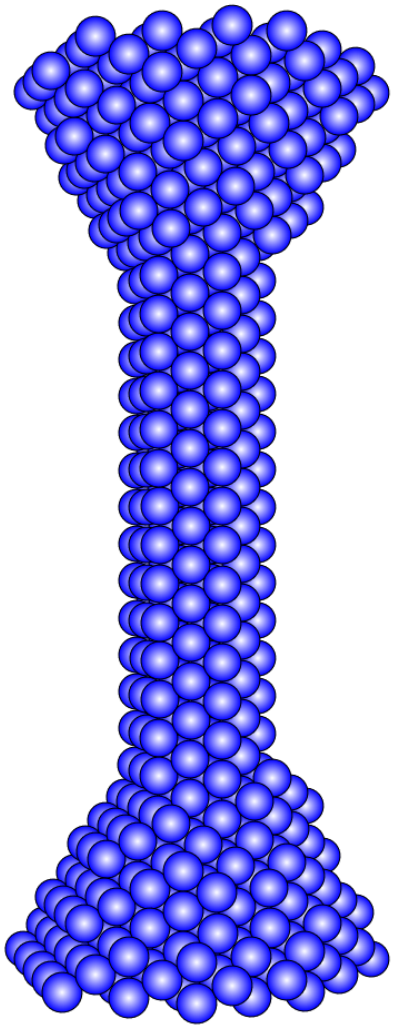


FIGURE – HRTEM imaging of Ag NWs. (Rodrigues *et al.*, PRB 65, 153402)

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

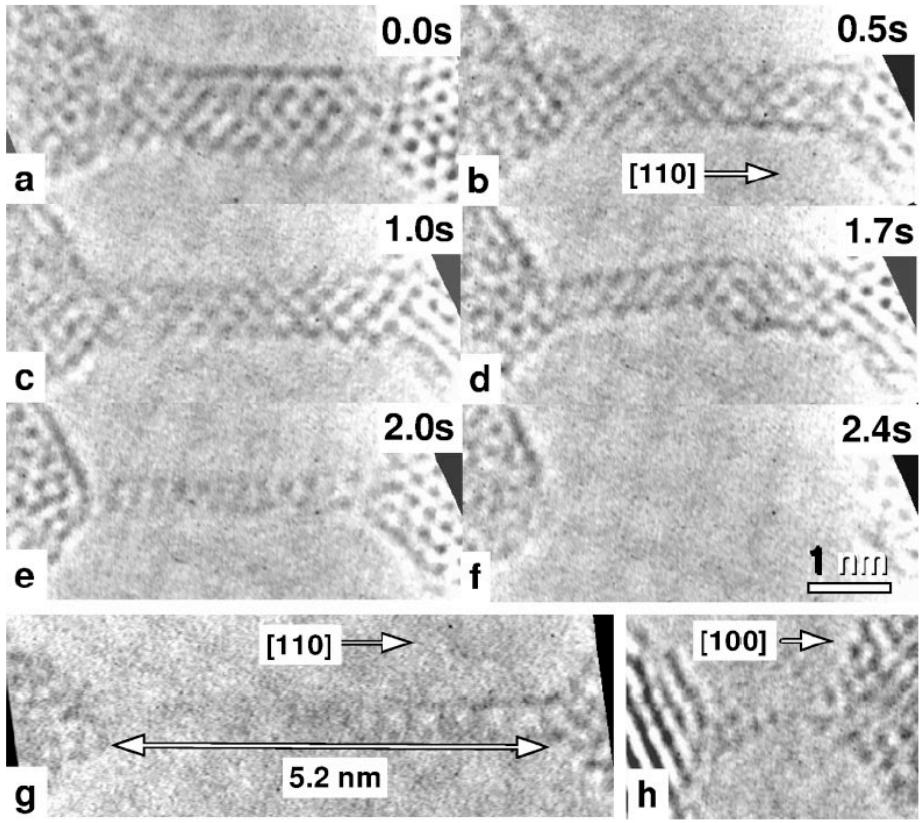


$v = 10^{-5}\mathrm{m/s}, \dot{\epsilon} = 1.5x10^{3}\mathrm{s}^{-1}$

Comparison with Experiments

Qualitative comparison with HRTEM is excellent.

- Uniform thinning (slip/unslip)
- Non-uniform thinning (kinks)
- Postulated “super-elastic” state consistent with our observation of an icosahedral phase that can “unwind” an FCC wire



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
- You are typically interested in paths containing more than one transition.

Practical Considerations : AMD with DFT

Can I use AMD with DFT ? Yes, in principle.... In practice, AMD requires :

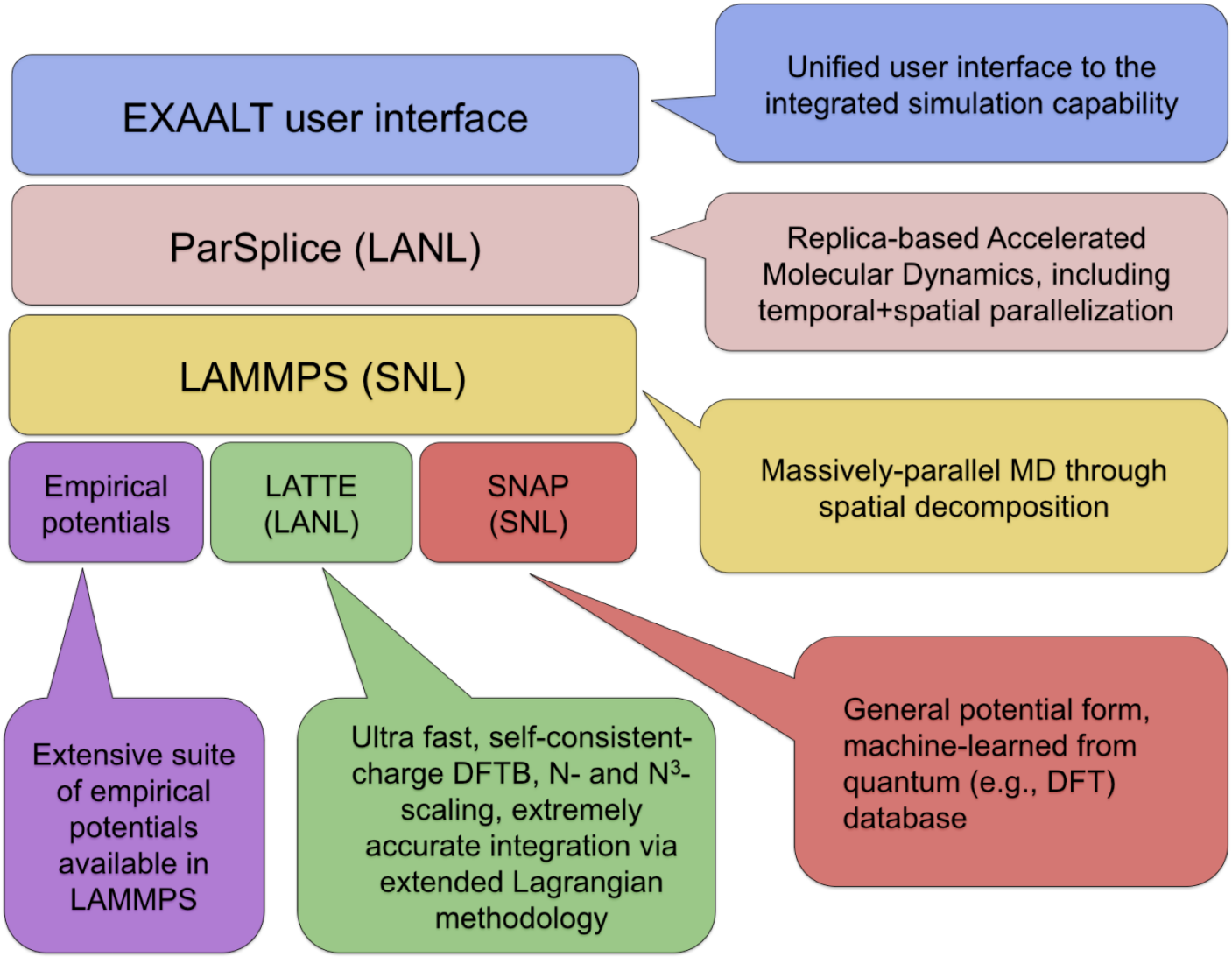
- Running (many) ps of dynamics
- Minimizing the energy
- Computing barriers
- ...

To be practical, you need a force calculation to take (ideally much) less than a minute.

Available tools

- LAMMPS (http ://lammps.sandia.gov/) : ParRep, TAD, HD
- DL POLY (www.ccp5.ac.uk/) : TAD, Hyperdynamics
- EXAALT(http ://gitlab.com/exaalt) (ParSplice + variants)

EXAALT


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 N3-scaling, extremely accurate integration via extended Lagrangian methodology
General potential form, machine-learned from quantum (e.g., DFT) database

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
- Funding : Exascale Computing Project ; US DOE/BES,ASCR ; LANL LDRD

Conclusion

QUESTIONS ?