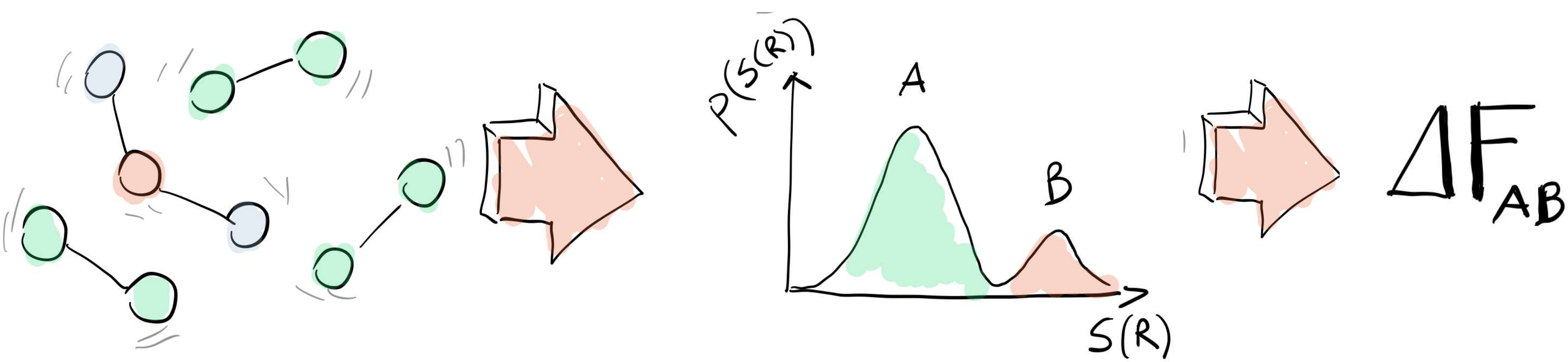
Enhanced Sampling

TYC Materials Modelling

Learning Outcomes

- How can we get thermodynamic information from dynamic trajectories?

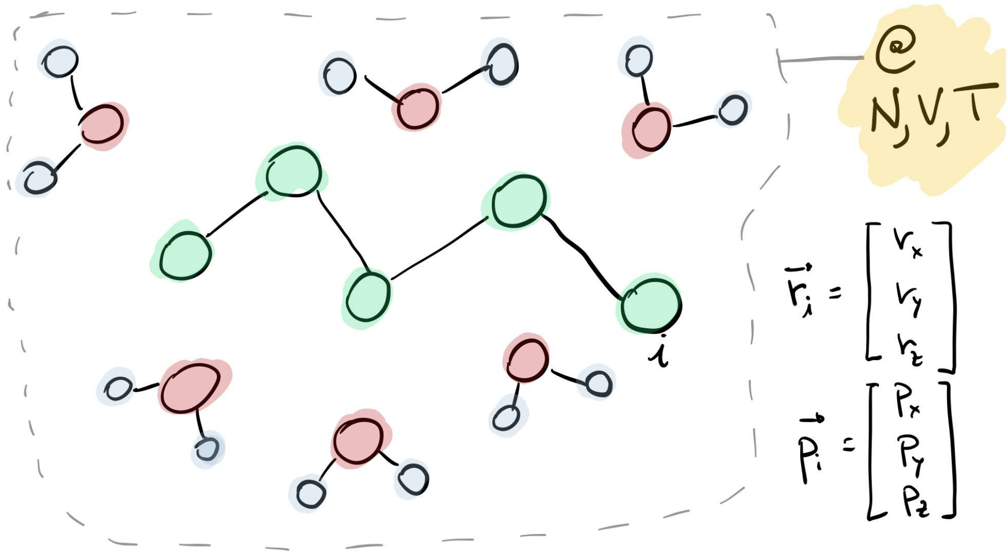


Outline / 1

- Partition Functions, Probability distributions, Free Energy
- Free Energy surfaces: representing free energies in low dimensional collective coordinates spaces.

From trajectories to thermodynamics: a few definitions

- Let's consider a set of interacting particles, representing a molecular system:



The system is characterised by 3N coordinates and 3N momenta, a microscopic state (microstate) of this system is therefore defined by a point in a 6N dimensional Phase Space, denoted with:

$$X = (\underbrace{\vec{p}_1, \vec{p}_2 \dots \vec{p}_N}_{\text{momenta}}, \underbrace{\vec{r}_1, \vec{r}_2, \dots \vec{r}_N}_{\text{COORDINATES}})$$

From trajectories to thermodynamics: a few definitions

- The microstates sampled are members of an ensemble, characterised by a partition function (Q), and a thermodynamic potential (A).
- Considering sampling in the Canonical Ensemble (i.e. @ constant N,V,T) these quantities are:

$$\underbrace{Q(N,V,T)}_{\text{partition function}} = \frac{1}{N!\,h^{3N}} \int dx\, e^{-\beta \underbrace{H(x)}_{\text{Hamitonian}}}$$

$$\underbrace{A(N,V,T)}_{\text{Absolute Free Energy}} = -K_B T \ln \underbrace{Q(N,V,T)}_{\text{Partition Function}}$$

From trajectories to thermodynamics: a few definitions

- The Thermodynamic potential can also be written isolating the contribution of the kinetic and potential energies of the Hamiltonian, and introducing the so-called configurational integral (Z):

$$A(N,V,T) = -k_B T \ln \underbrace{\int dR\, e^{-\frac{U(R)}{k_B T}}}_{Z(N,V,T)} + C$$

configuration integral

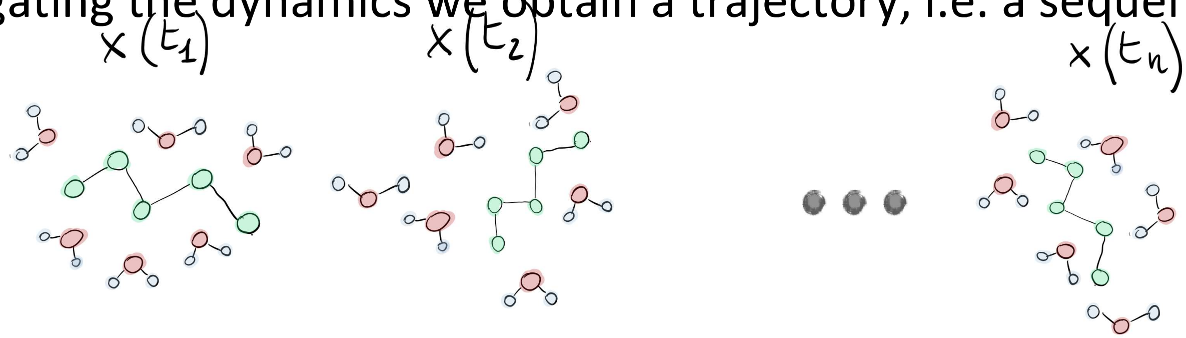
$Z(N,V,T)$

- This form becomes useful when evaluating free energy differences between systems in two different configurations, characterised by the same N, V, and T.

From trajectories to thermodynamics: a few definitions

- By numerically propagating the dynamics we obtain a trajectory, i.e. a sequence of microstates

$x(t)$



THE ERGODIC PRINCIPLE

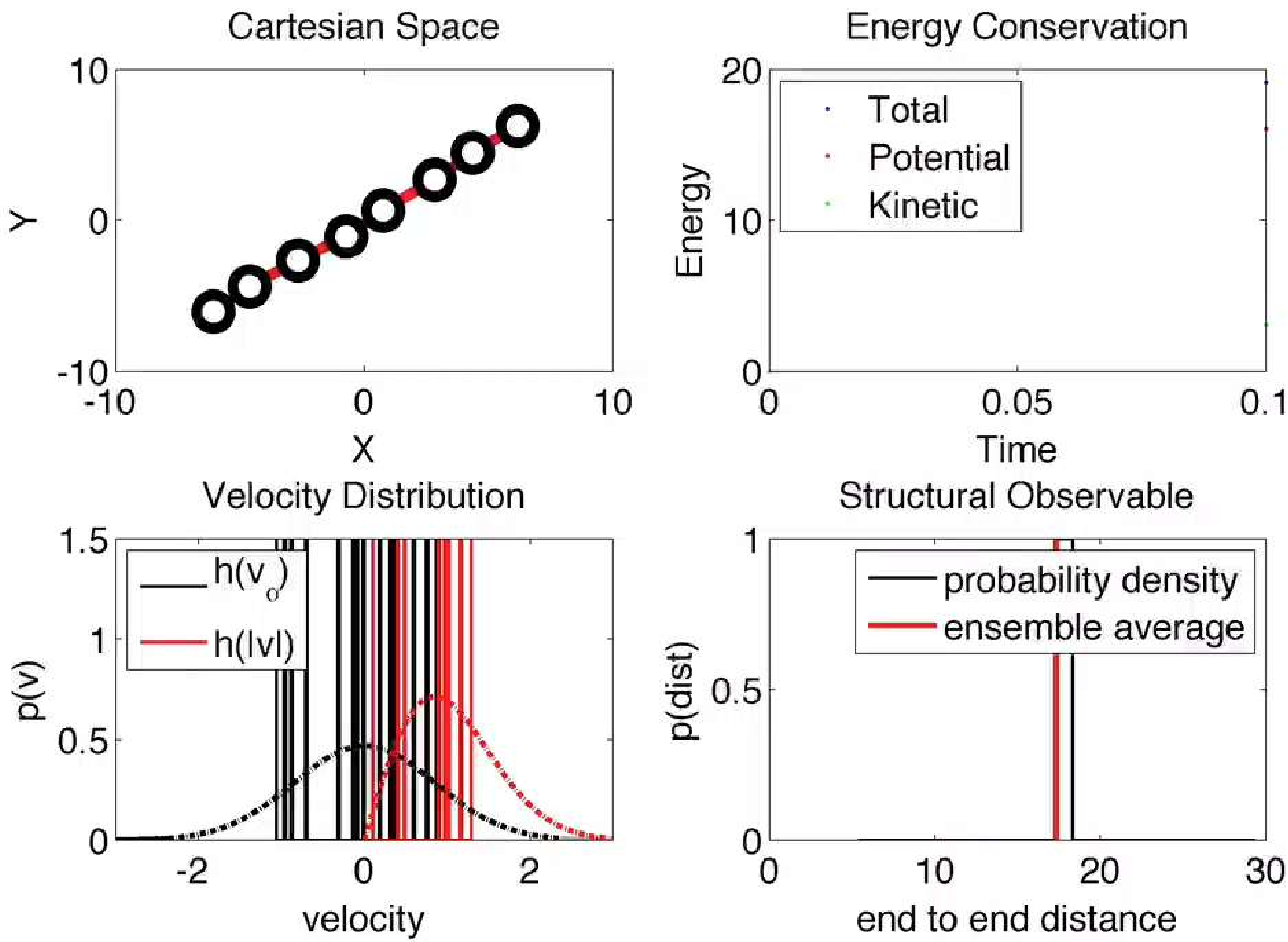
$$O = \langle O \rangle = \lim_{t\to\infty} \underbrace{\frac{1}{t}\int_0^t dt\, \underbrace{O(x(t))}_{\text{Instantaneous Realization of } O(x)}}_{\text{time average}}$$

!!!

macroscopic observable

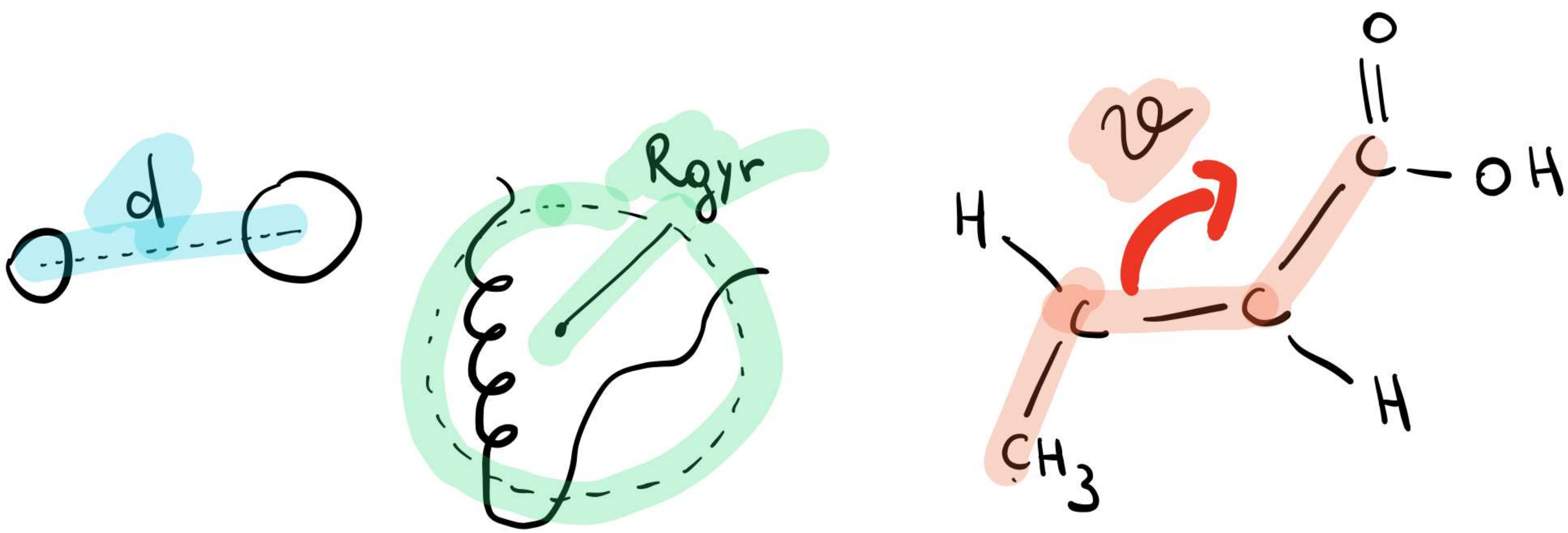
Ensemble average

The Ergodic principle at play



Representing configurational ensembles in human-readable coordinates

- Typically we cannot interpret directly trajectories in phase space (6N dimensional!).
- In order to analyse trajectories, understand mechanisms, develop theories, etc … we need to represent the ensemble of configurations in a lower-dimensional, human readable space.
- Examples of low dimensional variables can be the radius of gyration of a peptide, the distance between two ions forming an ion pair, a dihedral angle in a conformationally flexible molecule, an order parameter in a molecular crystal:



From probability density to free energy profile

- All these quantities share one characteristic: they all function of the coordinates of the system.
- We shall name these quantities "collective variables" or CVs. CVs are indicated with s(R) in the following

- Given a CV s(R) we can define an equilibrium probability density in CV space p(S) as:

$$p(s) = \int dR \left[\delta(s(R) - s) \right] \cdot p(R)$$

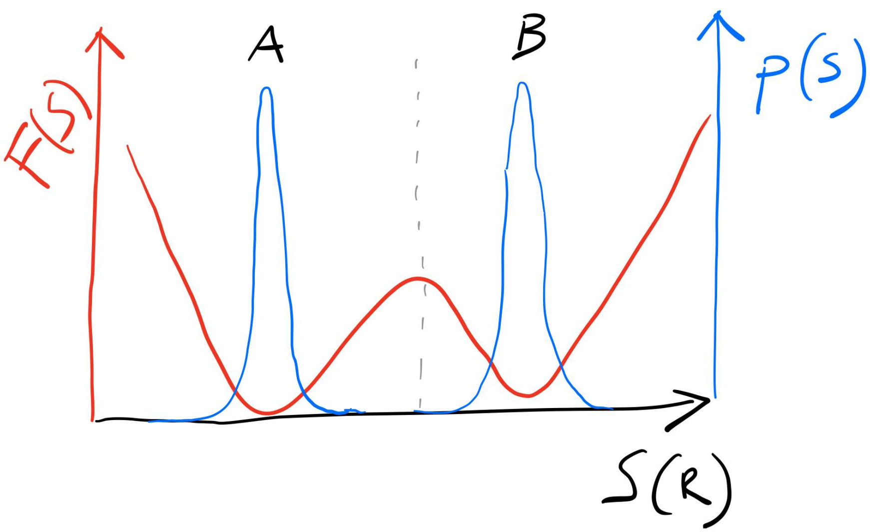
$$p(R) = \frac{e^{-\beta U(R)}}{\underbrace{\int e^{-\beta U(R)} dR}_{\text{configurational integral } Z}}$$

- At this point we can compute a free energ profile, function of the collective variable S

$$F(s) = -KT \ln p(s)$$

From free energy profiles to free energy differences

- If the CV allows to clearly identify highly populated states as local minima, one can readily compute free energy differences from a free energy profile.



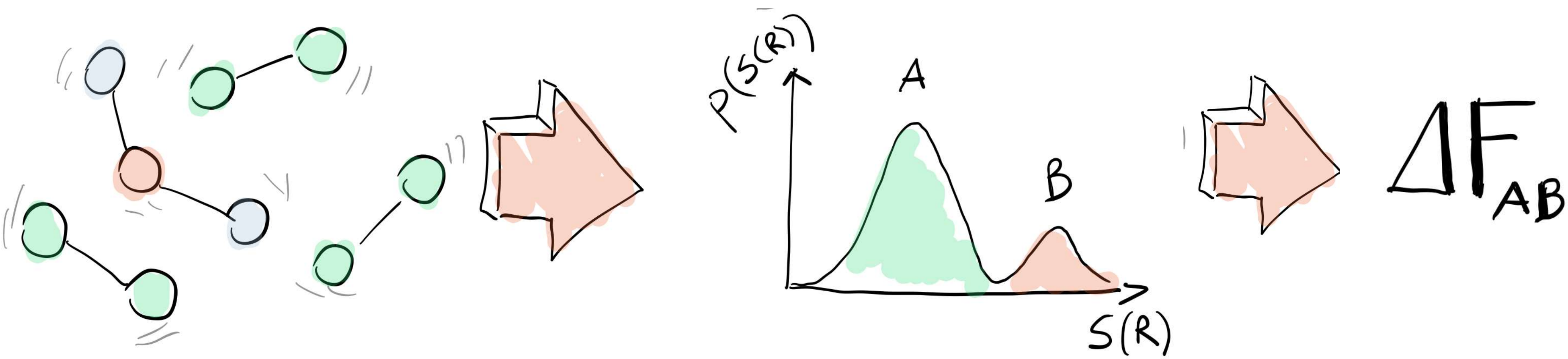
$$\Delta F_{AB} = -kT \ln \frac{\int_B p(s)\, ds}{\int_A p(s)\, ds}$$

- Combining these ideas with the ergodic principle provides a method for the calculation of free energy differences between states sampled in MD simulations

$$\Delta F_{AB} = -kT \ln \frac{\sum_{i \in B} H_i(t \to \infty)}{\sum_{i \in A} H_i(t \to \infty)}$$

Learning Outcomes

- How can we get thermodynamic information from dynamic trajectories?



Outline / 2

- The Sampling Problem
- Introducing a bias potential to overcome sampling limitations
- Unbiased probability distributions from biased sampling

A practical example: 1D Langevin dynamics on a double well potential

plumed.dat

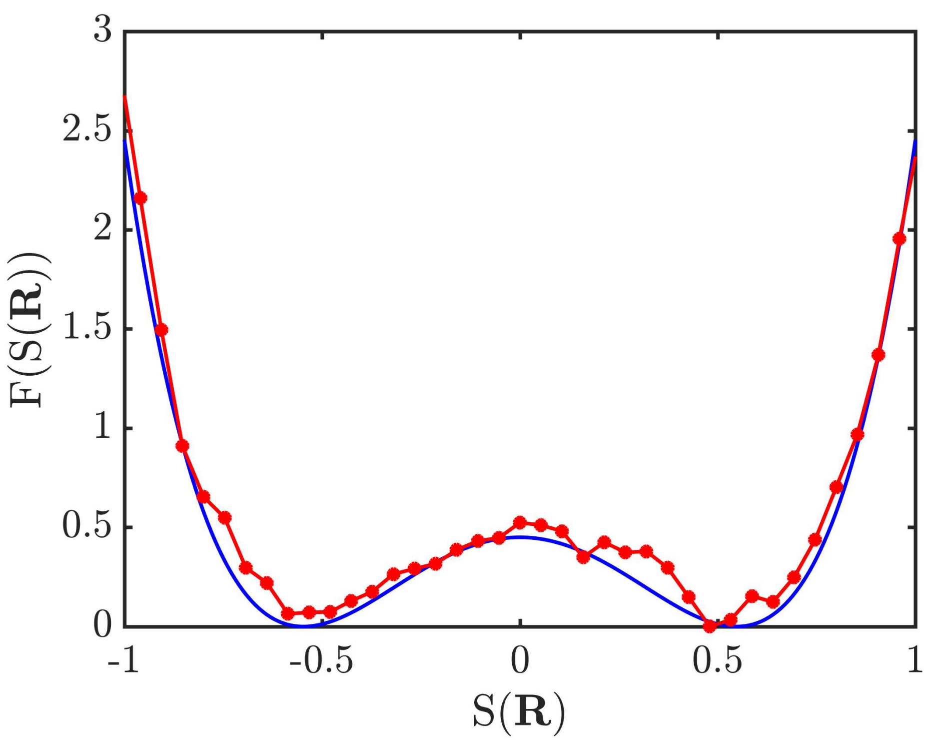
```
p: DISTANCE ATOMS=1,2 COMPONENTS
ff: MATHEVAL ARG=p.x PERIODIC=NO FUNC=(-3*x^2+5*x^4)
bb: BIASVALUE ARG=ff

HISTOGRAM ...
  ARG=p.x
  KERNEL=DISCRETE
  GRID_MIN=-4
  GRID_MAX=4
  GRID_BIN=150
  LABEL=hh
... HISTOGRAM

DUMPGRID GRID=hh FILE=histo

PRINT FILE=position ARG=p.x
```

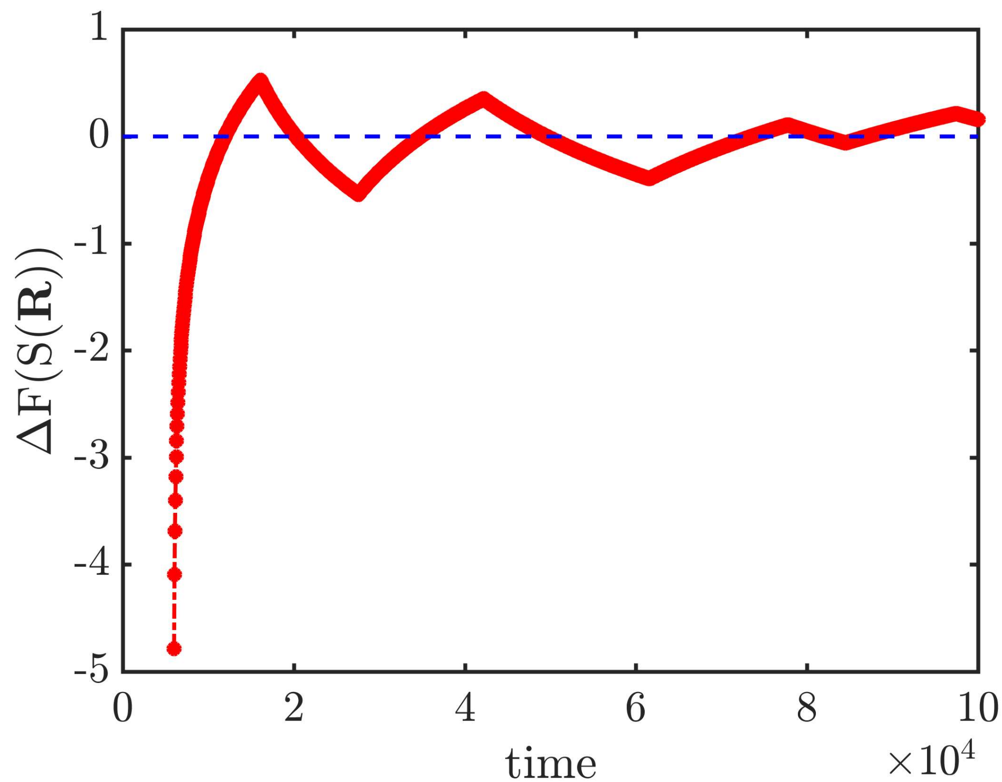
input

```
temperature 1
tstep 0.005
friction 1
dimension 1
nstep 200000
ipos -1.0
periodic false
```

```
matteo@matteo-2 [03:13:56]> plumed pesmd < input
```

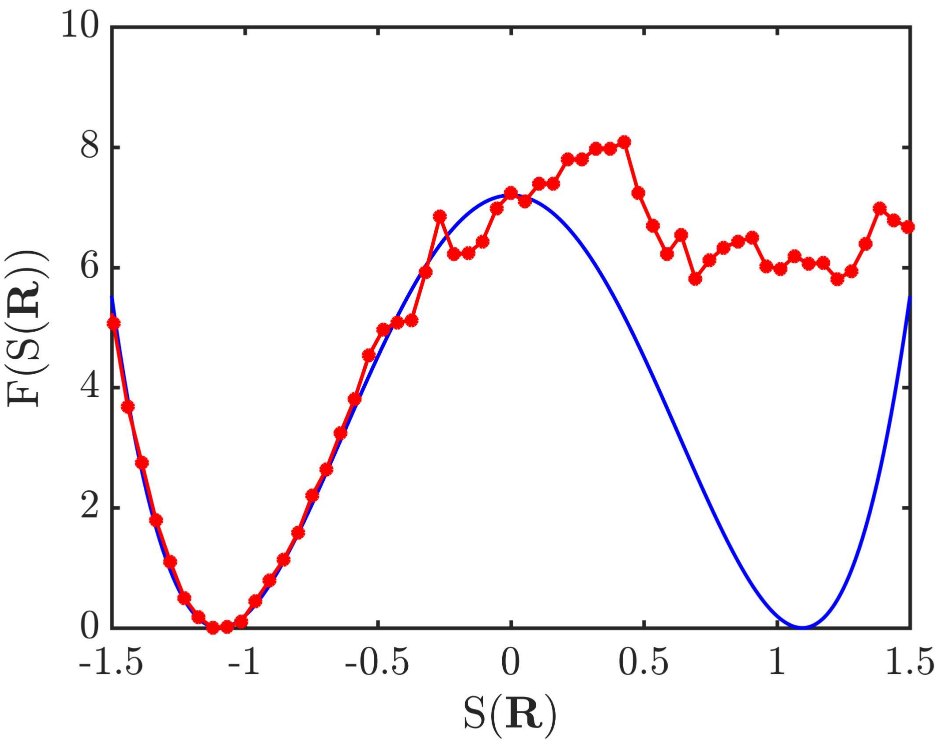
A practical example: 1D Langevin dynamics on a double well potential

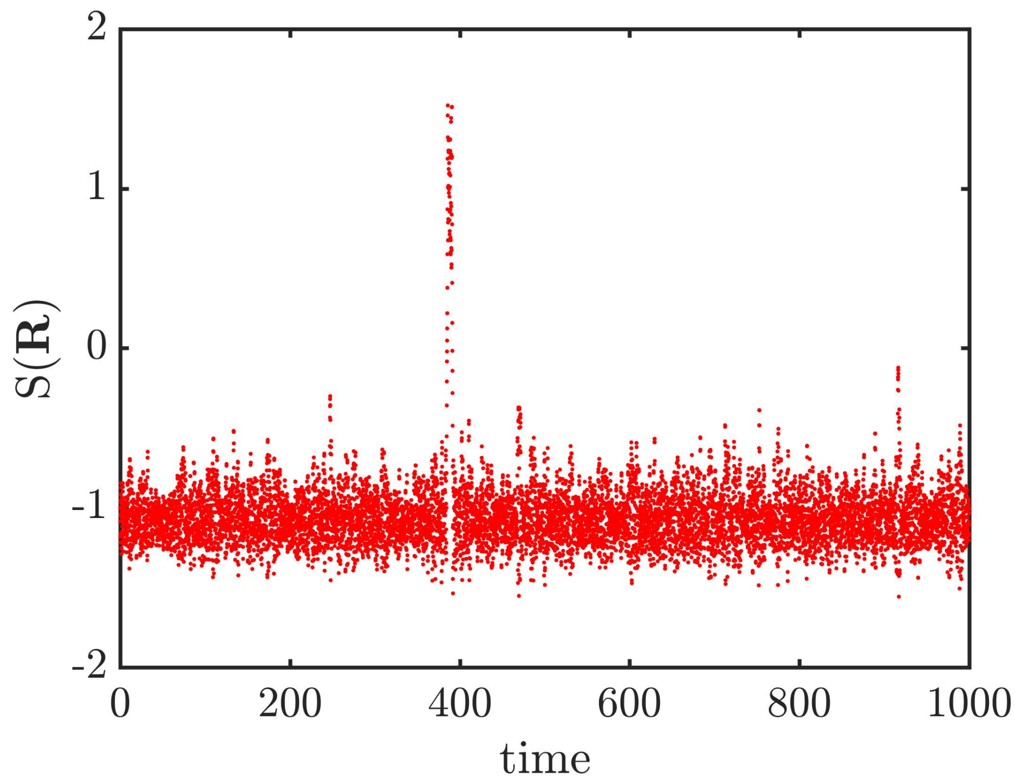




A practical example: 1D Langevin dynamics on a double well potential

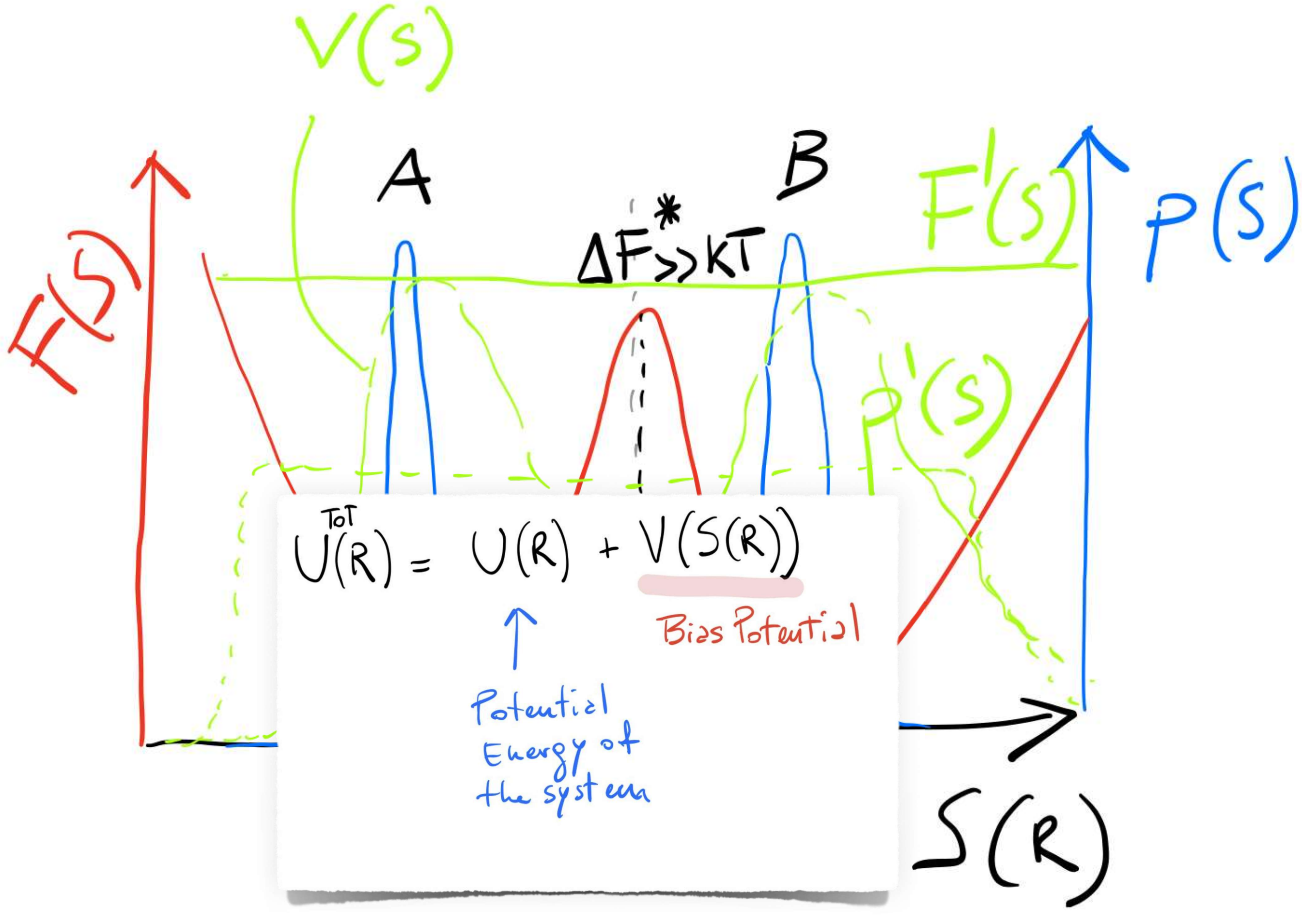
- Is this approach applicable to any transition? ergodicity in practice!





How can we sample more efficiently?

- Q: What is a desirable distribution to sample the entire S(R) space?



Free Energy Perturbation

A: "reference"

B: "perturbed"

$$U_B = U_A + \Delta U_{BA}$$

perturbation

$$Z_A = \int e^{-U_A(\vec{r})/k_BT} \, d\vec{r}$$

$$Z_B = \int e^{-U_B(\vec{r})/k_BT} \, d\vec{r} = \int e^{-\beta U_A(\vec{r})} e^{-\beta \Delta U_{BA}(\vec{r})} \, d\vec{r}$$

$$Z_B = \frac{Z_B \cdot Z_A}{Z_A} = Z_A \cdot \int \underbrace{\frac{e^{-\beta U_A(\vec{r})}}{\int e^{-\beta U_A(\vec{r})} d\vec{r}}}_{P_A(\vec{r})} \cdot e^{-\Delta U_{BA}(\vec{r})} \, d\vec{r}$$

$$Z_B = Z_A \cdot \int P_A(\vec{r}) \, e^{-\Delta U_{BA}(\vec{r})} \, d\vec{r} = Z_A \left\langle e^{-\Delta U_{BA}(\vec{r})} \right\rangle_A$$

$$\Delta F_{BA} = -kT \ln \frac{Z_B}{Z_A} = -kT \ln \left\langle \exp^{-\Delta U_{BA}(\vec{r})} \right\rangle_A$$

How can we get the correct statistics from biased sampling?

reference → unbiased system : $U_0(\vec{r})$

perturbation → bias potential $V(S(\vec{r}))$

perturbed system → $U_b(\vec{r}) = U_0(\vec{r}) + V(S(\vec{r}))$

$$Z_b = Z_0 \cdot \langle e^{-\beta V(S(\vec{r}))} \rangle_0 \qquad S = f(\vec{r})$$

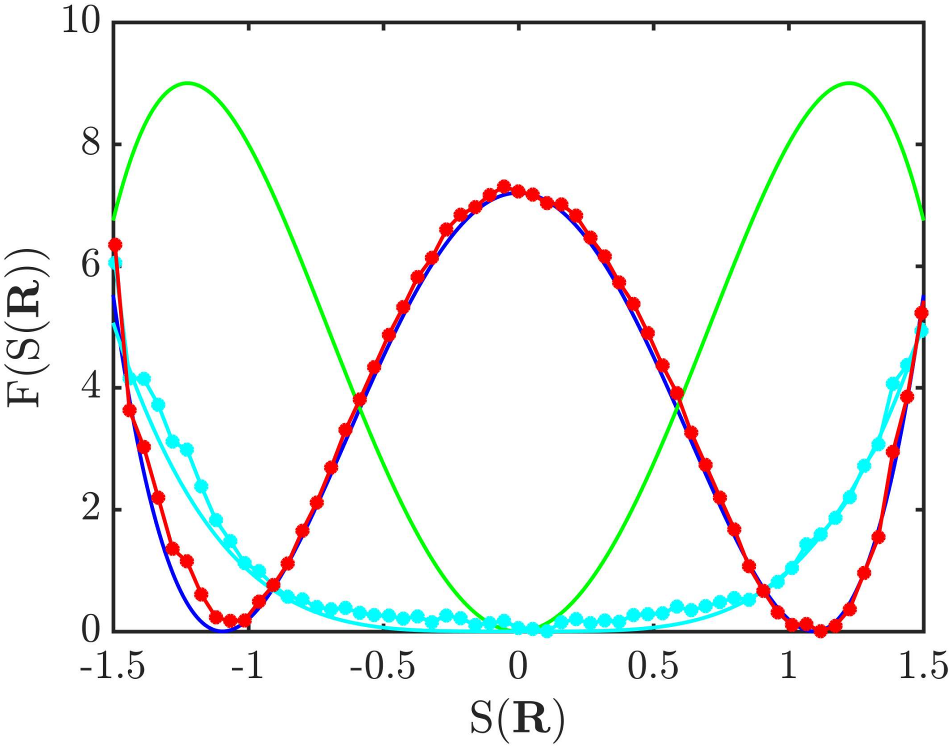
$$p^b(S) = \frac{1}{Z_b} \cdot \int \delta(f(\vec{r}) - S) \cdot e^{-\beta U_0(\vec{r})} \cdot e^{-\beta V(S(\vec{r}))} d\vec{r}$$

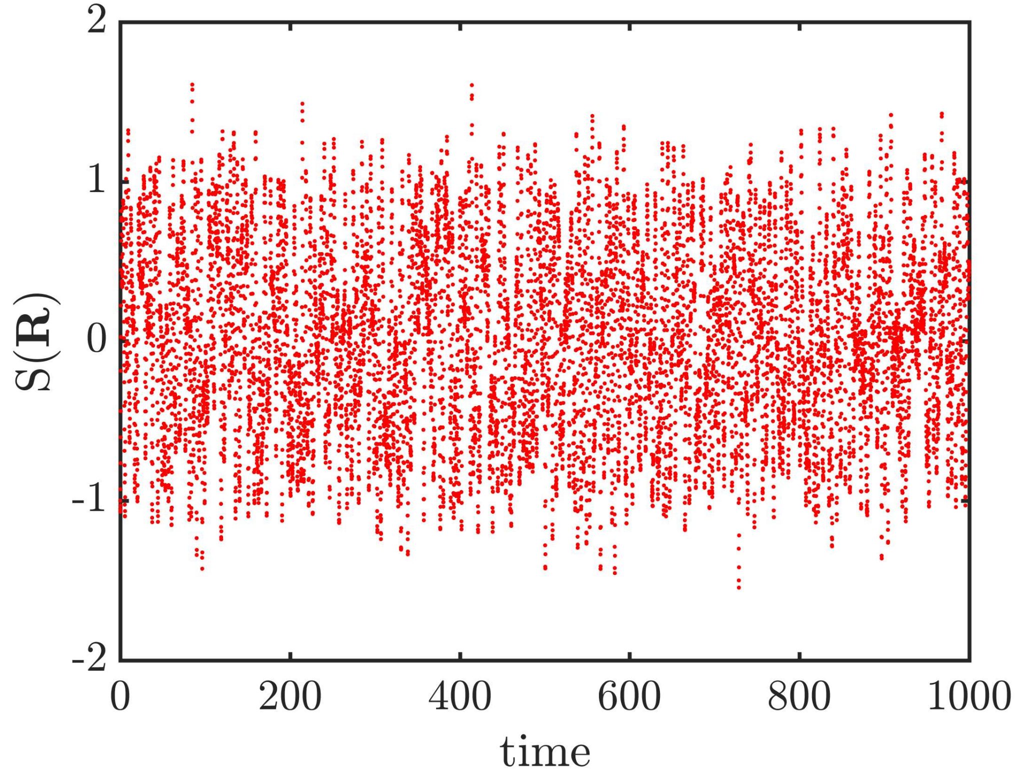
$$= \frac{\int \delta(f(\vec{r}) - S) \cdot e^{-\beta U_0(\vec{r})} \cdot e^{-\beta V(S(\vec{r}))} d\vec{r}}{Z_0 \langle e^{-\beta V(S(\vec{r}))} \rangle_0}$$

$$= p_0(s) e^{-\beta V(s)} \cdot \langle e^{-\beta V(s)} \rangle_0^{-1}$$

$$p_0(s) = p_b(s) \cdot e^{\beta V(s)} \cdot \underbrace{\langle e^{-\beta V(s)} \rangle_0}_{\text{constant}}$$

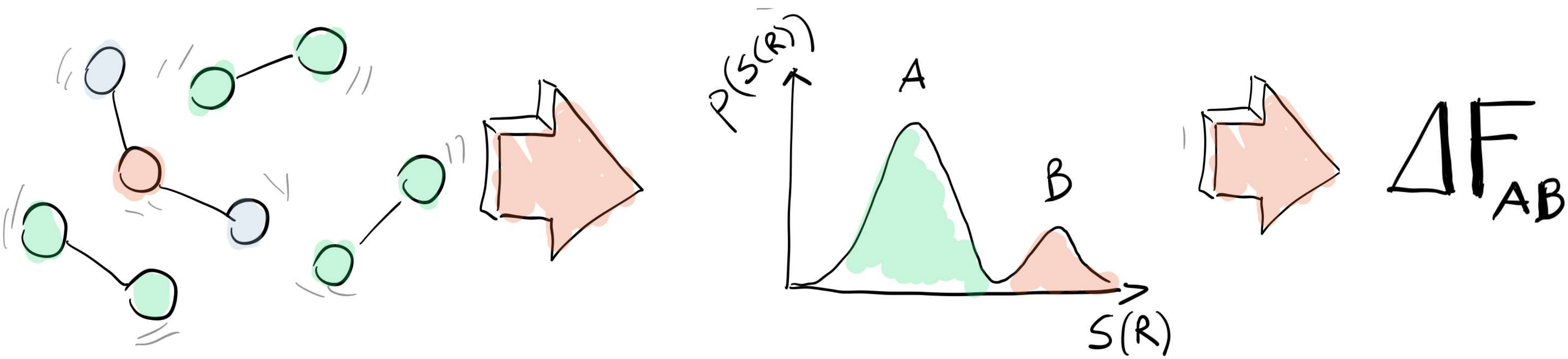
Unbiased statistics from biased sampling: an example





Learning Outcomes

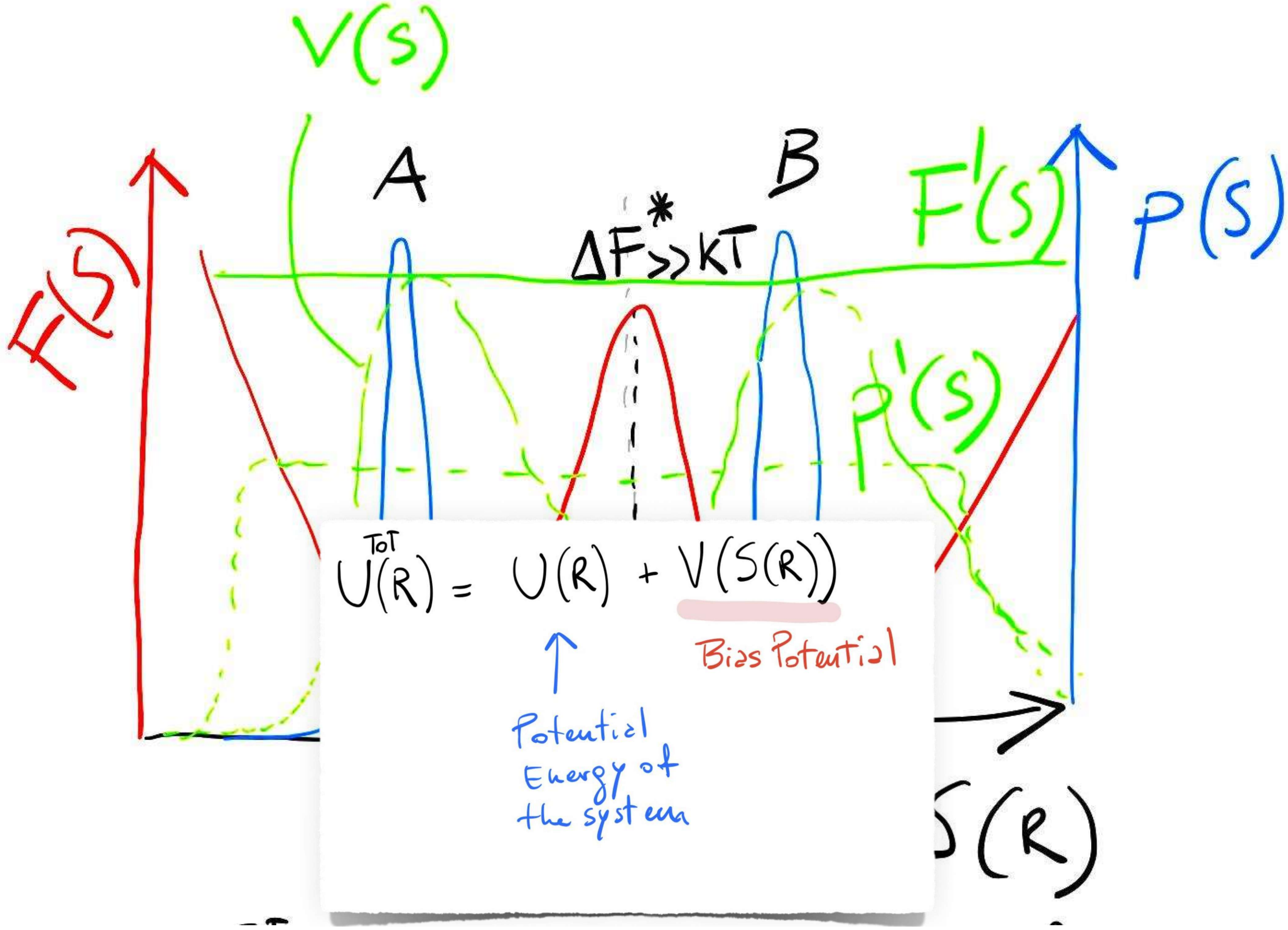
- What is the sampling problem?
- What is biased sampling?
- How can we recover equilibrium probabilities from biased sampling?



Outline / 3

- Biased sampling with Umbrella Sampling + weighted histogram analysis method (WHAM)
- Adaptive biasing with metadynamics

Sampling configurations in the presence of a bias potential



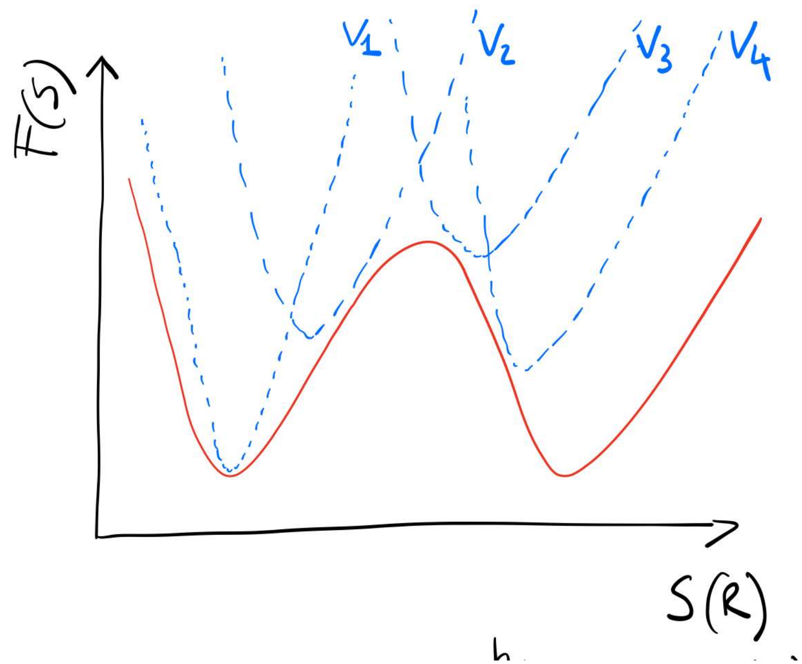
$$\overset{TOT}{U}(R) = U(R) + V(S(R))$$

Potential Energy of the system

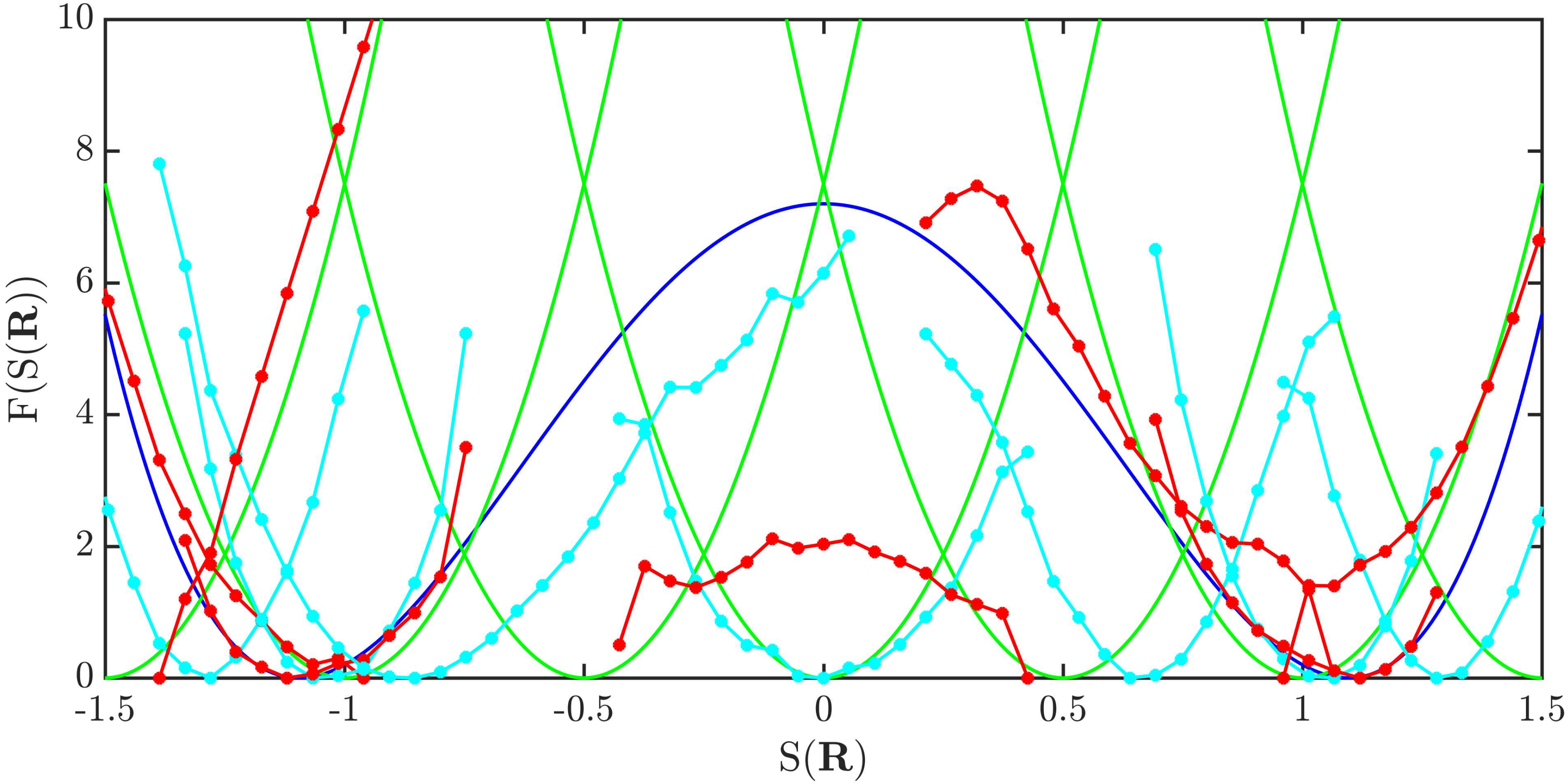
Bias Potential

Umbrella Sampling

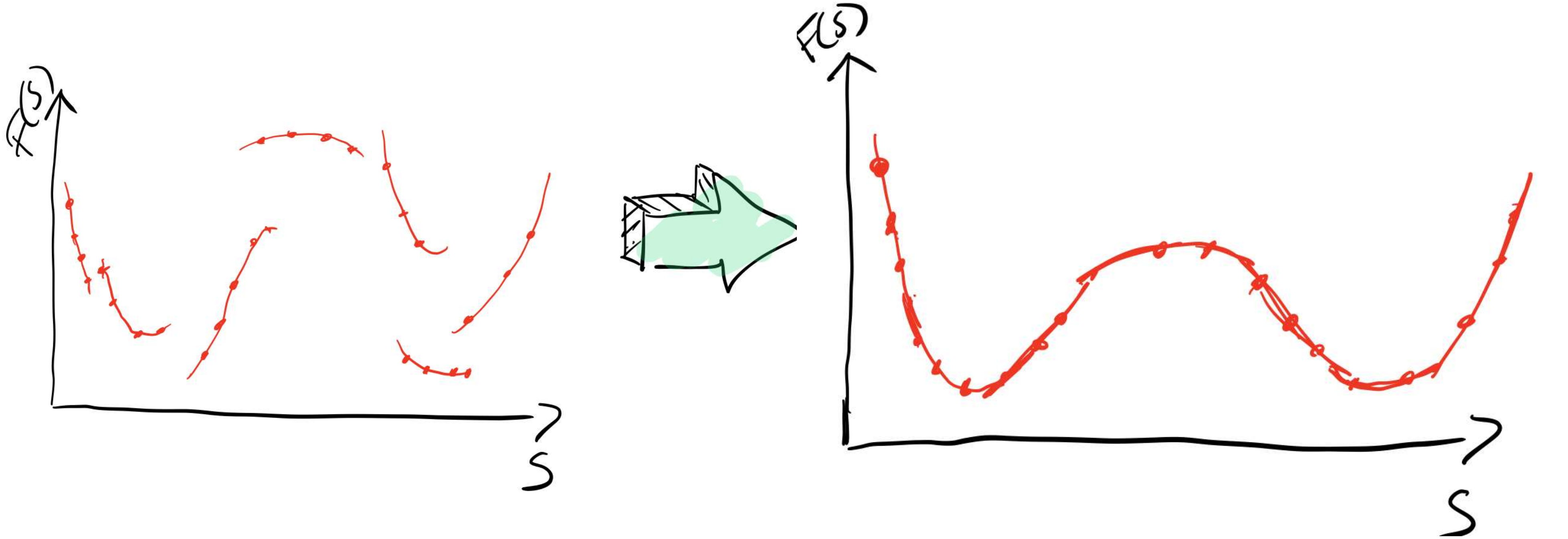
- Hypothesis: we have a reasonable idea of an accessible pathway to connect states two metastable states of interest.
- We design a series of biased simulations that allow to sample extensively configurations along that pathway



Umbrella Sampling: an example

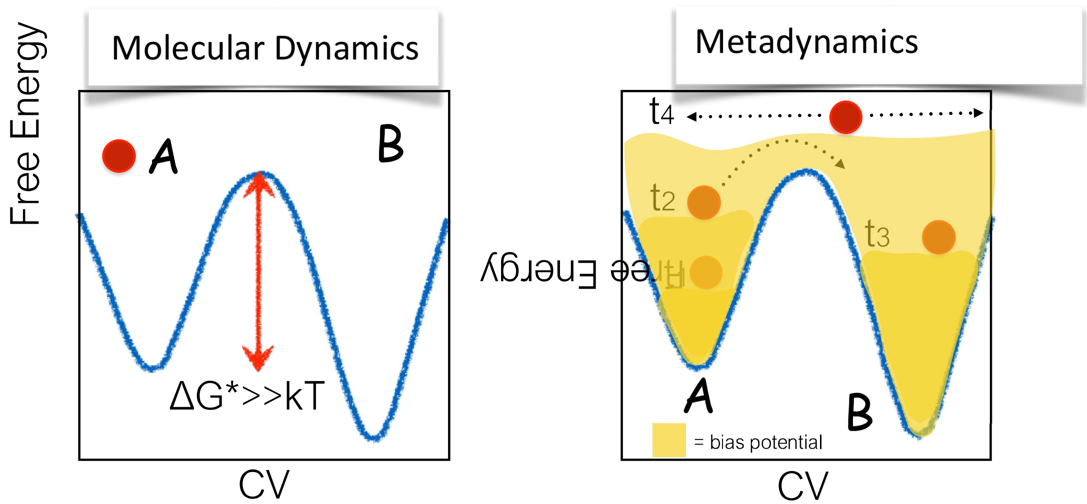


A global free energy profile from partially overlapping simulations

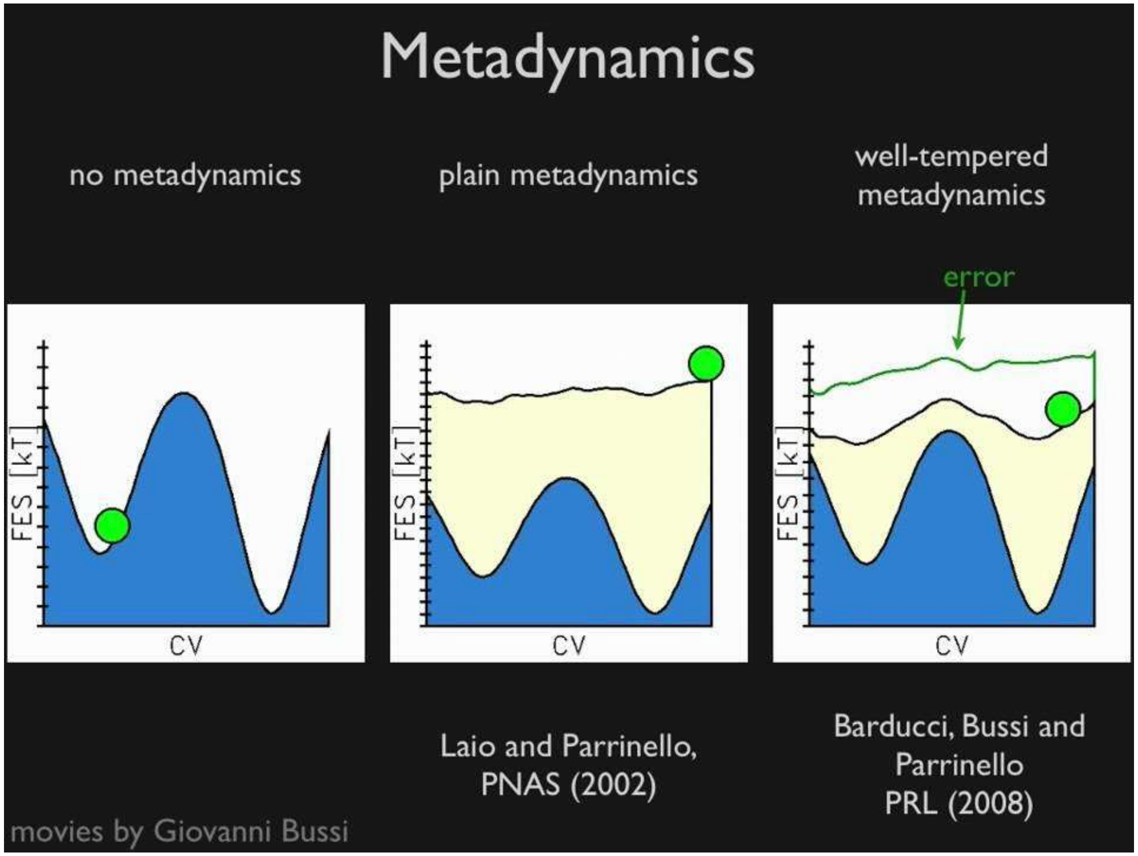


A different approach: a blind search for the optimal bias potential

- What if we want to gain information of low energy pathways between two metastable states?
- We need to build the bias potential adaptively!



MD vs. Metadynamics vs WT MetaD



Outline / 3

- Biased sampling with Umbrella Sampling + weighted histogram analysis method (WHAM)
- Adaptive biasing with metadynamics

Outline / 4

- Applying mean force integration ideas to metadynamics
- Kinetics from metadynamics
- Applications

Free Energy surfaces from biased sampling

$$F(\mathbf{s}) = -\beta^{-1}\ln p^b(\mathbf{s}) - V(\mathbf{s}) - \langle V(\mathbf{s})\rangle_u$$

biased probability density

bias

ensemble average of the bias potential

$$\langle V(\mathbf{s})\rangle_u = \beta^{-1}\ln \frac{\int_\Omega e^{-\beta F(\mathbf{s})-\beta V(\mathbf{s})}d\mathbf{s}}{\int_\Omega e^{-\beta F(\mathbf{s})}d\mathbf{s}}$$

$\langle V(\mathbf{s})\rangle_u$

- is a constant which can be evaluated numerically using self-consistent iterative methods (WHAM)
- In metadynamics it is a function of time: i.e. *c(t)* [Tiwary and Parrinello JPCB 2015]

"Getting around" $\langle V(\mathbf{s})\rangle_u$ through Mean Force Integration I

inindependent

$$F(\mathbf{s}) = -\beta^{-1}\ln p^b(\mathbf{s}) - V(\mathbf{s}) - \langle V(\mathbf{s})\rangle_u$$

constant with respect to **s**

$$\frac{dF(\mathbf{s})}{d\mathbf{s}}$$

mean force in CV space

[Umbrella Integration Kastner and Thiel, JCP 2005]

Unperturbed mean force in **s**

$$\frac{dF(\mathbf{s})}{d\mathbf{s}} = -\frac{d\beta^{-1}\ln p^b(\mathbf{s})}{d\mathbf{s}} - \frac{dV(\mathbf{s})}{d\mathbf{s}}$$

Bias force

Perturbed mean force in **s**

“Getting around” $\langle V(\mathbf{s})\rangle_u$ through Mean Force Integration II

the metadynamics potential is iteratively updated in discrete steps
[t; t + τ]

$$\frac{dF_t(\mathbf{s})}{d\mathbf{s}}$$

mean force in CV space sampled in [t; t + τ]

Bias force

Unperturbed mean force in **s**

$$\frac{dF_t(\mathbf{s})}{d\mathbf{s}} = -\frac{d\beta^{-1}\ln p_t^b(\mathbf{s})}{d\mathbf{s}} - \frac{dV_t(\mathbf{s})}{d\mathbf{s}}$$

subscript *t* indicates estimates sampled during the iteration [t; t + τ]

Perturbed mean force in **s**

An analytical expression for mean force in CV space from metadynamics

$$\left\langle \frac{dF_t(\mathbf{s})}{ds} \right\rangle_t = \frac{\sum_{t'=1}^{t} p^b_{t'}(\mathbf{s}) \frac{dF_{t'}(\mathbf{s})}{ds}}{\sum_{t'=1}^{t} p^b_{t'}(\mathbf{s})}$$

localised weight in CV space

Derivative of the bias potential

$$\frac{dV_t(\mathbf{s})}{ds} = \sum_{t'=1}^{t} \frac{w_0(s - s_{t'})}{{\sigma_M}^2} \exp\left[-\frac{1}{2}\frac{(s - s_{t'})^2}{{\sigma_M}^2}\right]$$

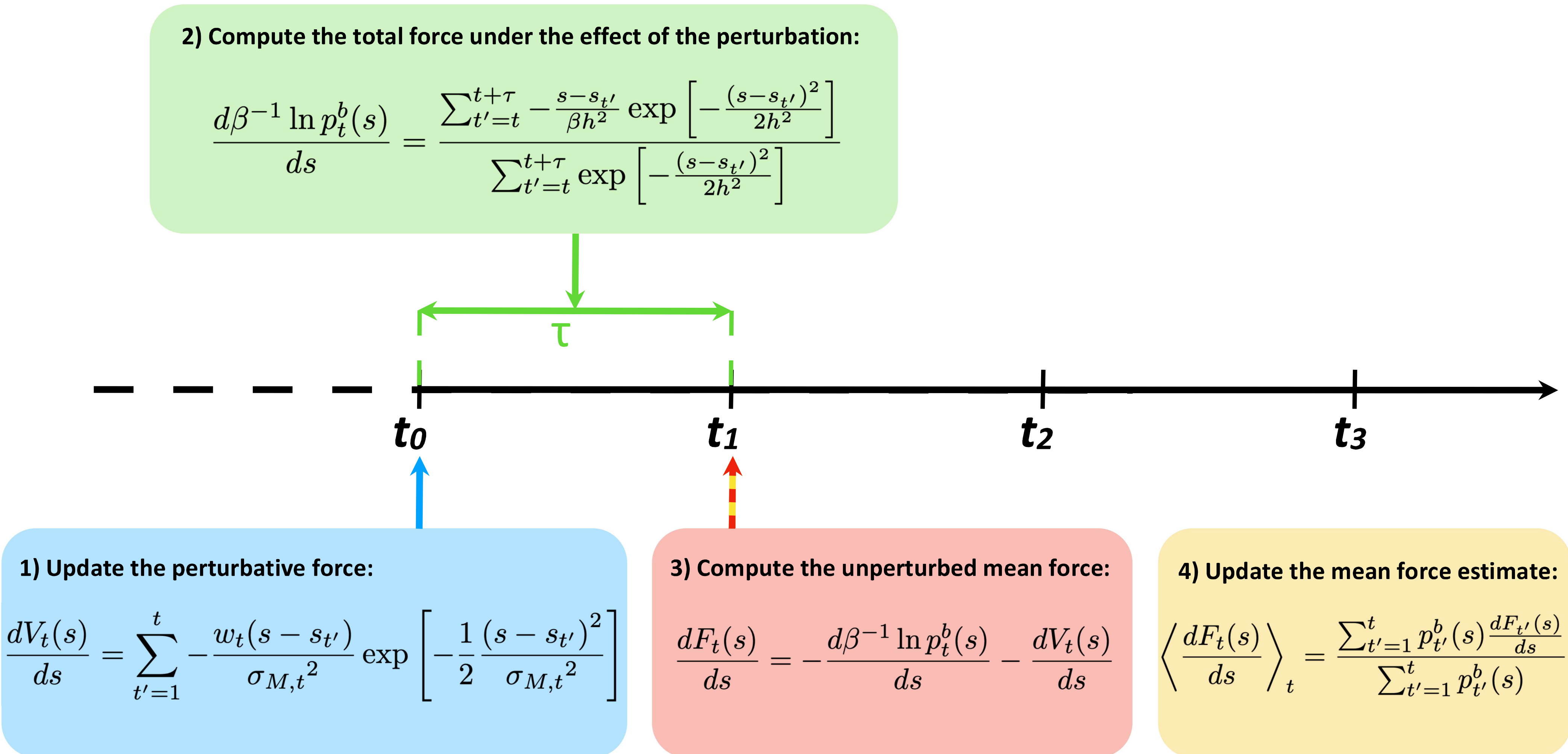
$$\frac{dF_t(\mathbf{s})}{ds} = -\frac{d\beta^{-1}\ln p^b_t(\mathbf{s})}{ds} - \frac{dV_t(\mathbf{s})}{ds}$$

Kernel density approximation of pb

$$p^b_t(\mathbf{s}) = \frac{1}{n_\tau h\sqrt{2\pi}} \sum_{t'=t}^{t+\tau} \exp\left[-\frac{(s - s_{t'})^2}{2h^2}\right]$$

$$\frac{d\beta^{-1}\ln p^b_t(\mathbf{s})}{ds} = \frac{\sum_{t'=t}^{t+\tau} \frac{s-s_{t'}}{h^2} \exp\left[-\frac{(s-s_{t'})^2}{2h^2}\right]}{\sum_{t'=t}^{t+\tau} \exp\left[-\frac{(s-s_{t'})^2}{2h^2}\right]}$$

$$\left\langle \frac{dF_t(\mathbf{s})}{ds} \right\rangle_t = \frac{\sum_{t'=1}^{t} \sum_{t''=t'}^{t'+\tau} \frac{s-s_{t''}}{\beta\, n_\tau h^3\sqrt{2\pi}} \exp\left[-\frac{(s-s_{t''})^2}{2h^2}\right]}{\sum_{t'=1}^{t} p^b_{t'}(\mathbf{s})} + \frac{\sum_{t'=1}^{t} p^b_{t'}(\mathbf{s}) \frac{dV_{t'}(\mathbf{s})}{ds}}{\sum_{t'=1}^{t} p^b_{t'}(\mathbf{s})}$$

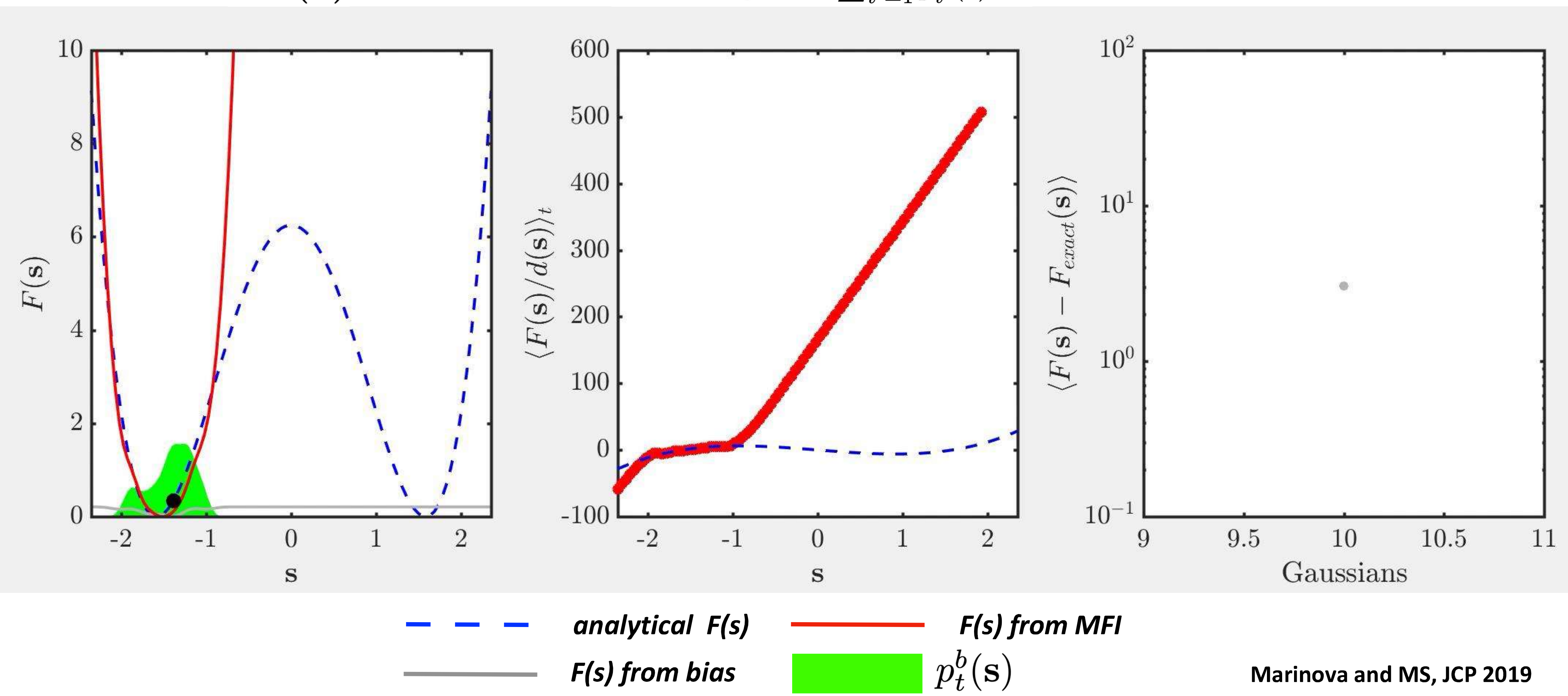

2) Compute the total force under the effect of the perturbation:
$\frac{d\beta^{-1}\ln p_t^b(s)}{ds} = \frac{\sum_{t'=t}^{t+\tau} -\frac{s-s_{t'}}{\beta h^2}\exp\left[-\frac{(s-s_{t'})^2}{2h^2}\right]}{\sum_{t'=t}^{t+\tau}\exp\left[-\frac{(s-s_{t'})^2}{2h^2}\right]}$
τ
t_0
t_1
t_2
t_3
1) Update the perturbative force:
$\frac{dV_t(s)}{ds} = \sum_{t'=1}^{t} -\frac{w_t(s-s_{t'})}{{\sigma_{M,t}}^2}\exp\left[-\frac{1}{2}\frac{(s-s_{t'})^2}{{\sigma_{M,t}}^2}\right]$
3) Compute the unperturbed mean force:
$\frac{dF_t(s)}{ds} = -\frac{d\beta^{-1}\ln p_t^b(s)}{ds} - \frac{dV_t(s)}{ds}$
4) Update the mean force estimate:
$\left\langle \frac{dF_t(s)}{ds} \right\rangle_t = \frac{\sum_{t'=1}^{t} p_{t'}^b(s)\frac{dF_{t'}(s)}{ds}}{\sum_{t'=1}^{t} p_{t'}^b(s)}$

MFI in action

$F(\mathbf{s})$

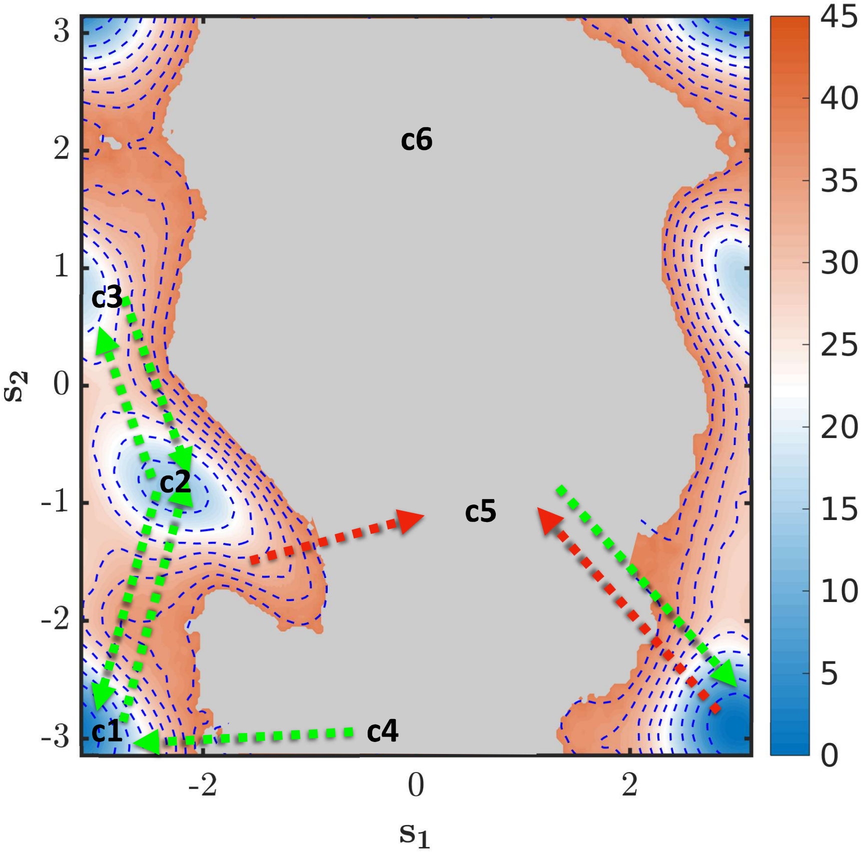
$$\left\langle \frac{dF_t(\mathbf{s})}{d\mathbf{s}} \right\rangle_t = \frac{\sum_{t'=1}^{t} p_{t'}^{b}(\mathbf{s}) \frac{dF_{t'}(\mathbf{s})}{d\mathbf{s}}}{\sum_{t'=1}^{t} p_{t'}^{b}(\mathbf{s})}$$

Error



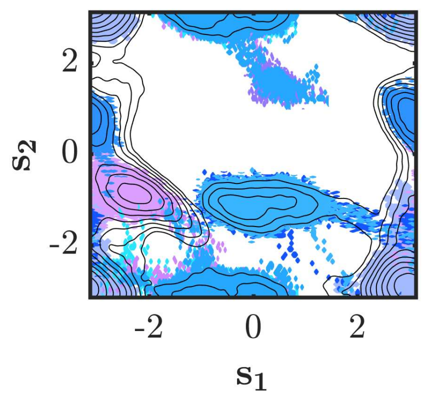
Marinova and MS, JCP 2019

Ibuprofen free energy surface in the crystal bulk from a swarm independent short metadynamics simulations



Set	Initial state	Final State	Average Length [ns]
1	c1	any other	1.7
2	c2	any other	0.05
3	c3	any other	0.013
4	c4	any other	0.015
5	c5	any other	0.05
6	c6	any other	0.0008

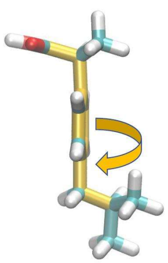
Set 1-6



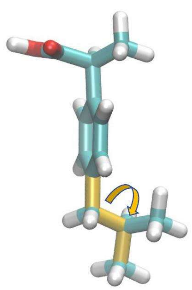
Set 1-10

Ibuprofen in the crystal bulk

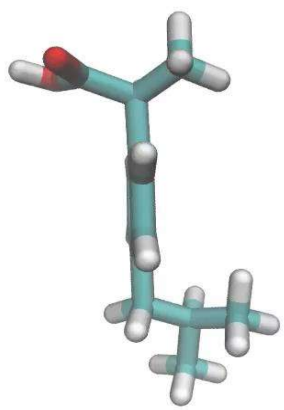
- Two internal torsional angles can describe ibuprofen conformational flexibility

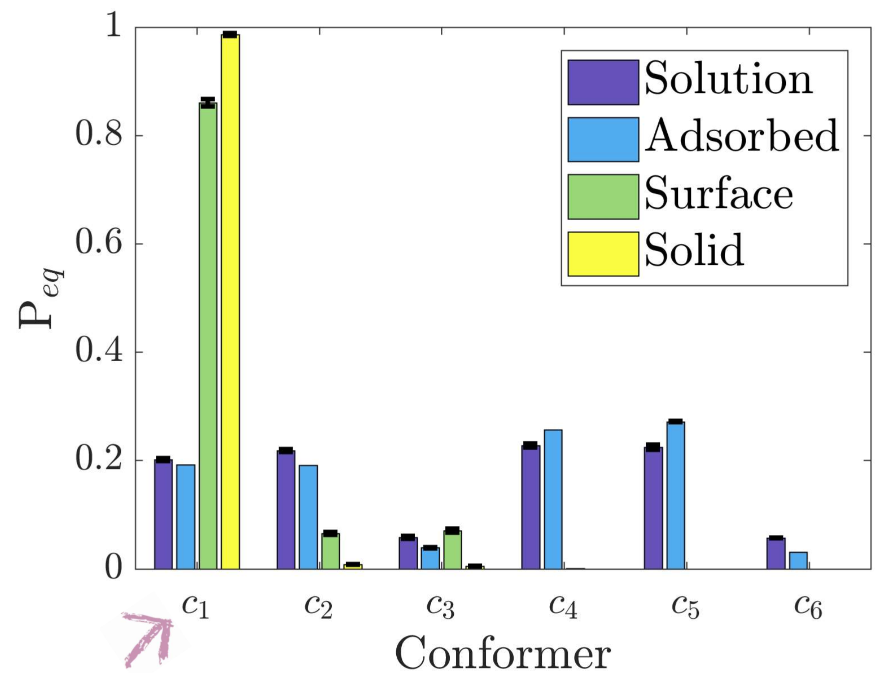


global torsional angle



local torsional angle

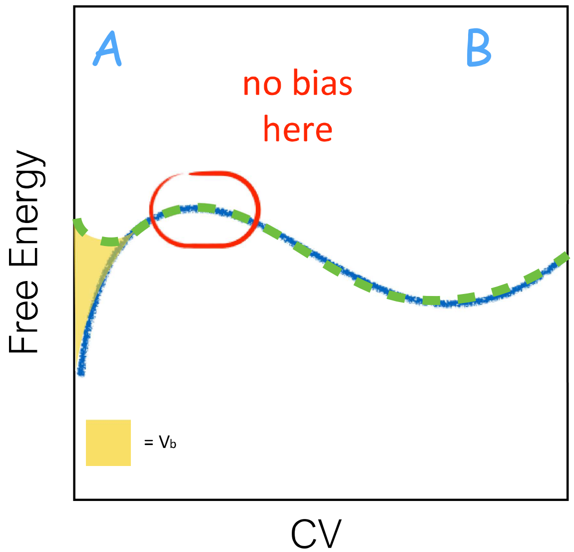




conformer in the crystal

V Marinova, GPF Wood, I Marziano, M Salvalaglio, JCTC, 2018

Kinetics from meta-dynamics



$$k_{A\to B} = \kappa \frac{Z_{TS}}{Z_A}$$

$$k^*_{A\to B} = \kappa^* \frac{Z^*_{TS}}{Z^*_A}$$

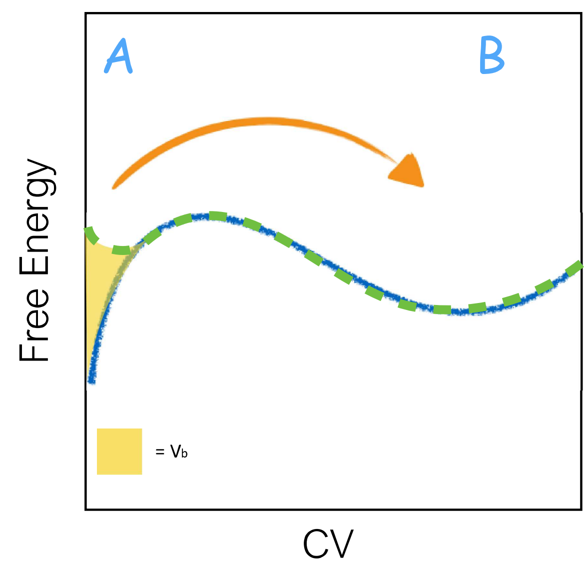
$$\frac{k^*_{A\to B}}{k_{A\to B}} = \frac{Z_A}{Z^*_A}\frac{\kappa^* Z^*_{TS}}{\kappa Z_{TS}}$$

$$k_{A\to B} = k^*_{A\to B}\langle e^{-\beta V_b}\rangle$$

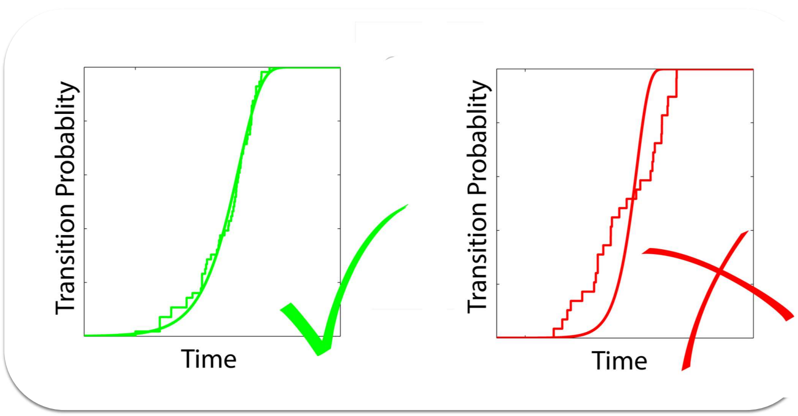
UNBIASED FREQUENCY — BIASED FREQUENCY" — BIAS

$$\frac{k^*_{A\to B}}{k_{A\to B}} = \frac{Z_A}{Z^*_A} = \langle e^{\beta V_b}\rangle$$

Kinetics from meta-dynamics



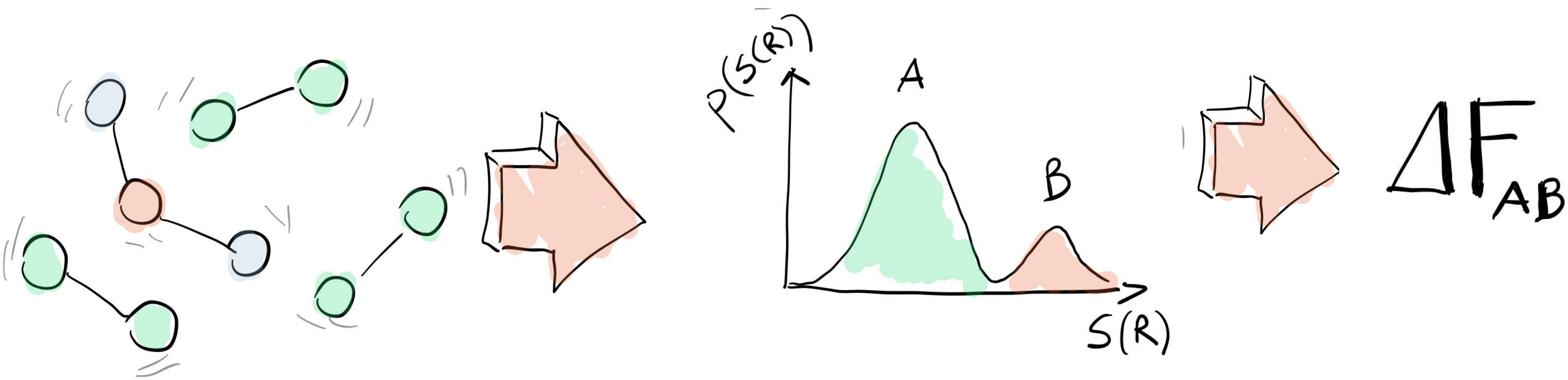
- The **transition** between basins A and B is a **rare event.**
- The **escape** from a metastable state can be interpreted as a **Poisson process**.
- The **survival probability** is **exponentially distributed**.
- This **feature** has to be **preserved** in the calculation of the transition times distribution from **biased simulations**



measuring the **compatibility** of the transition probability **with** that of a **Poisson process reveals** the **corruption** of the transition state

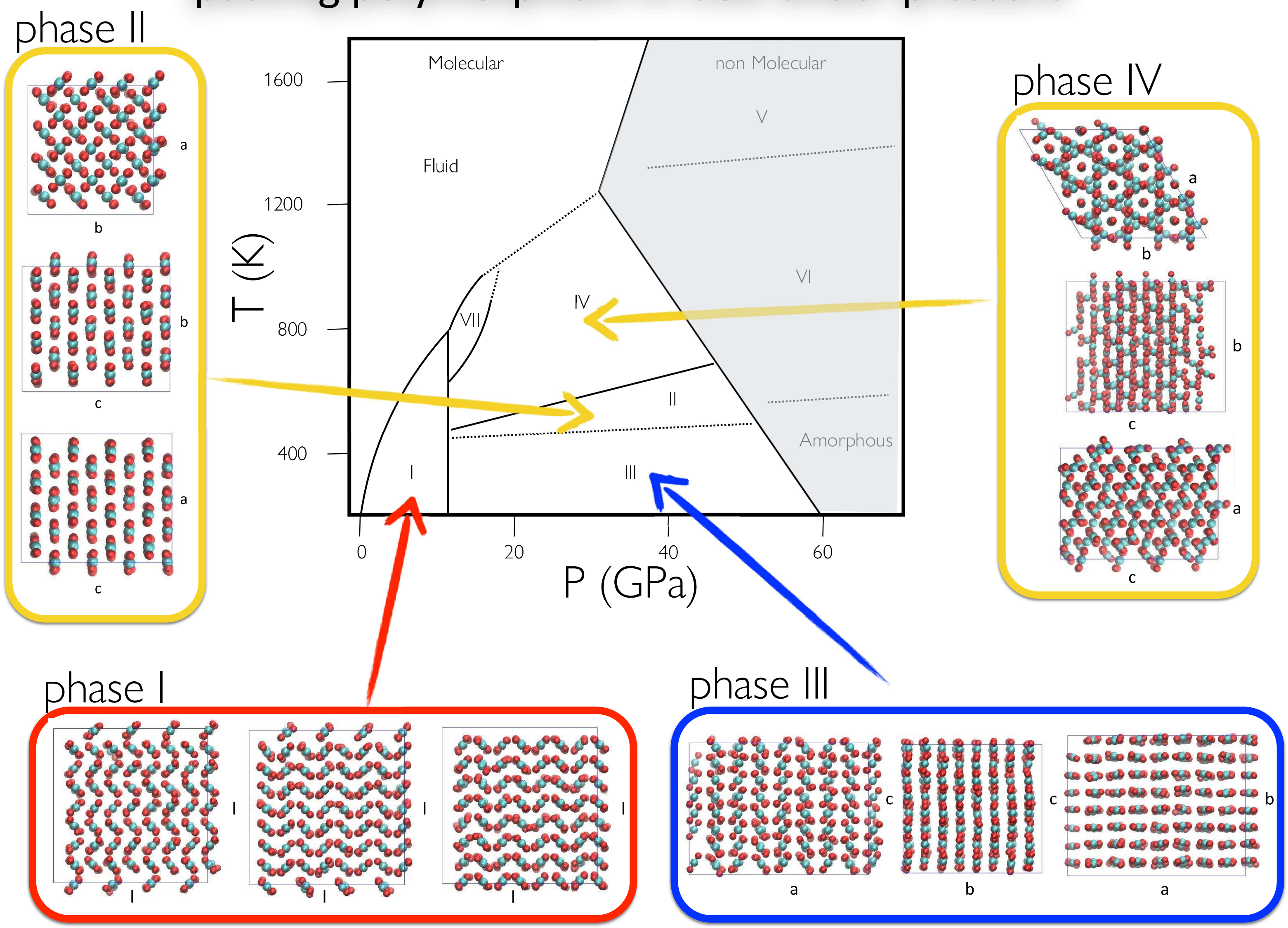
Learning Outcomes - RECAP

- Unbiased statistics from biased sampling
- Umbrella Sampling
- Metadynamics

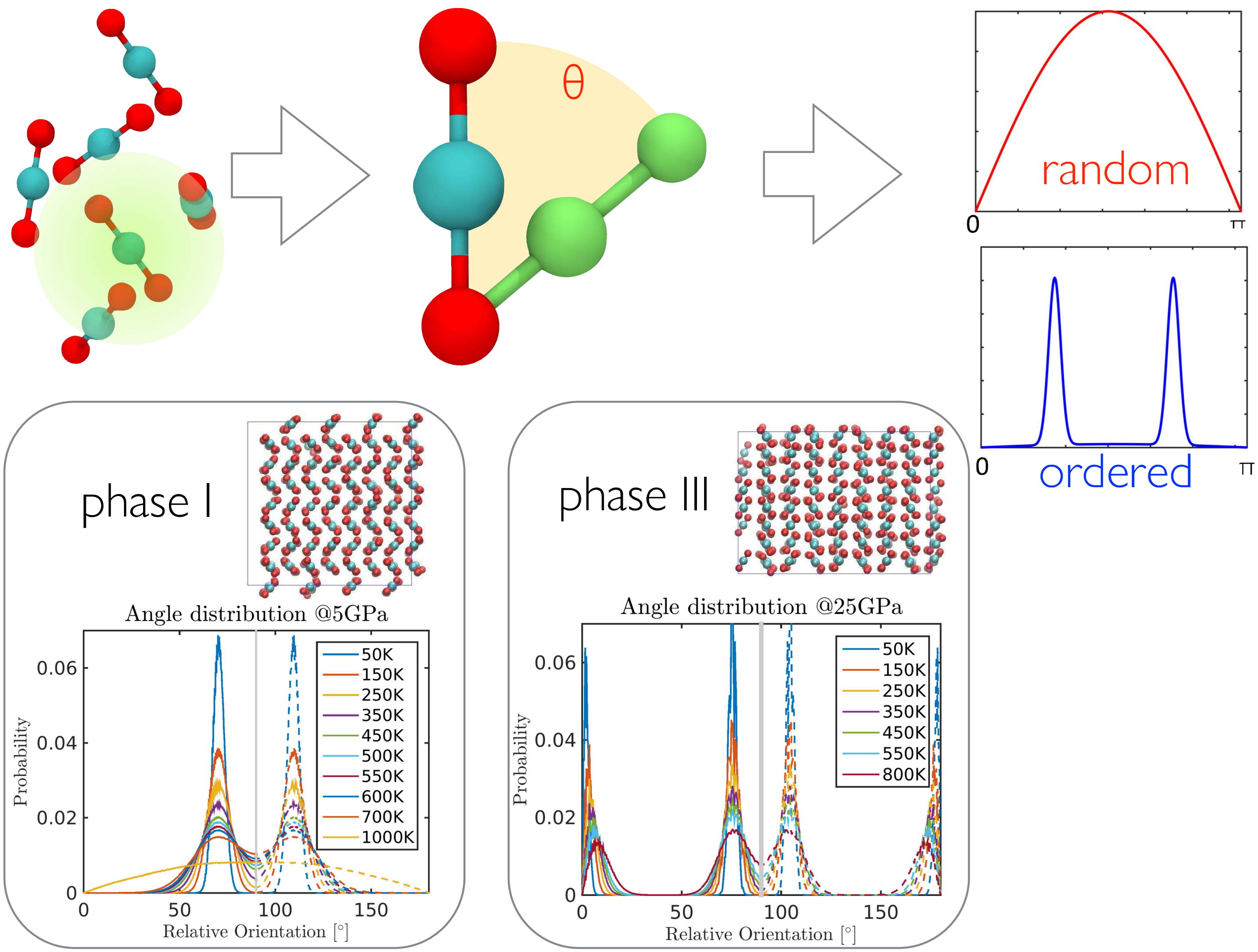


- A few examples follow

packing polymorphism in CO2 under pressure

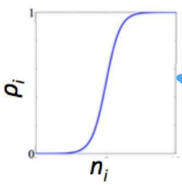


Characterising CO_2 packing with orientation distributions

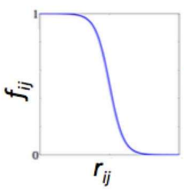


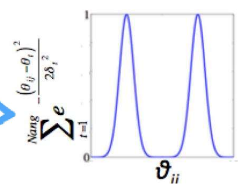
from angle distributions to collective variables

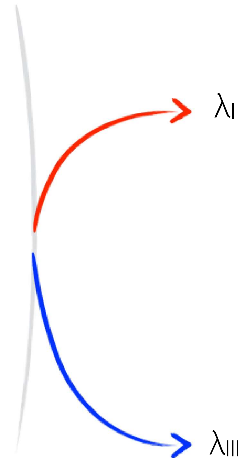
$$\lambda = \frac{1}{N}\sum_{i=1}^{N}\Gamma_i$$

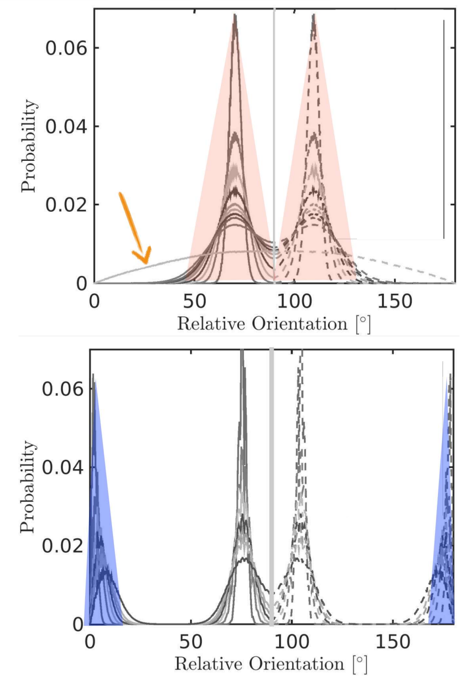


$$\Gamma_i = \frac{\rho_i}{n_i}\sum_{j=1}^{N} f_{ij}\sum_{t=1}^{Nang} e^{-\frac{(\theta_{ij}-\theta_t)^2}{2\delta_t^2}}$$

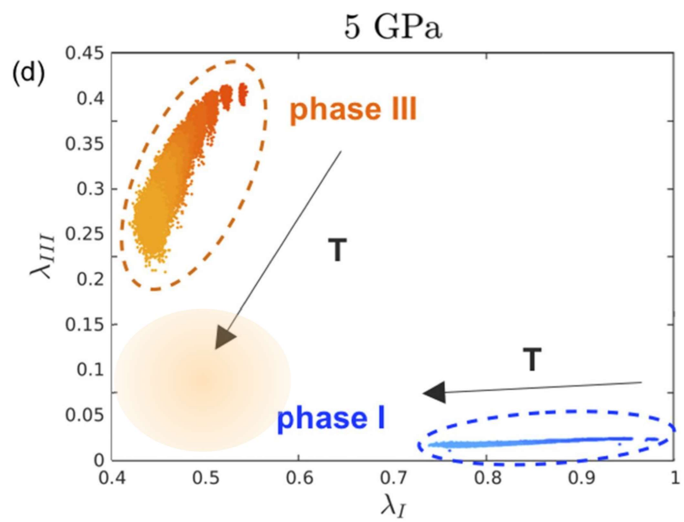




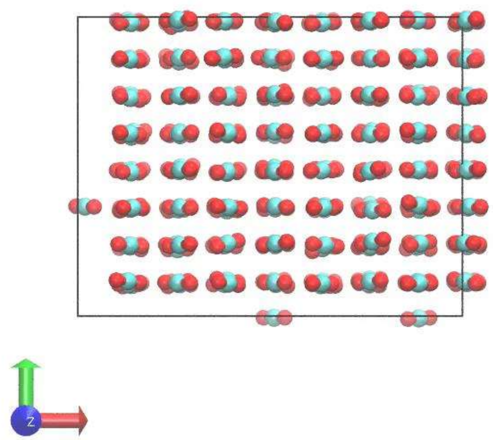


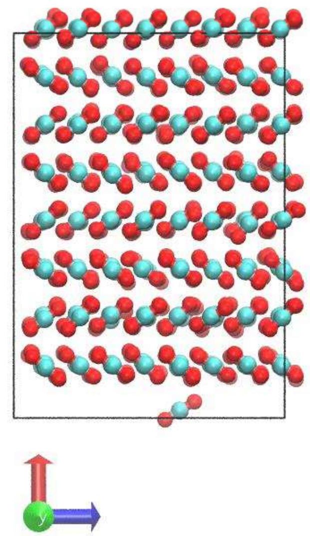


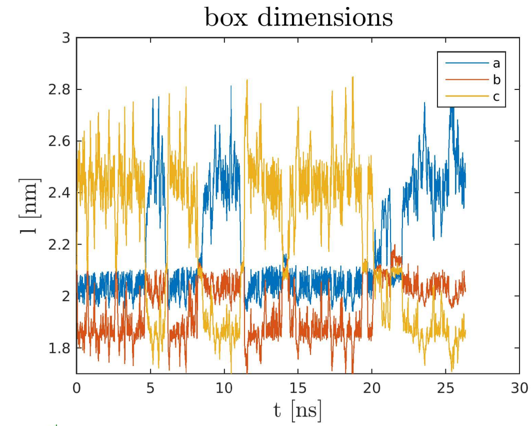
Kernel positioned in order to maximise the difference between two phases

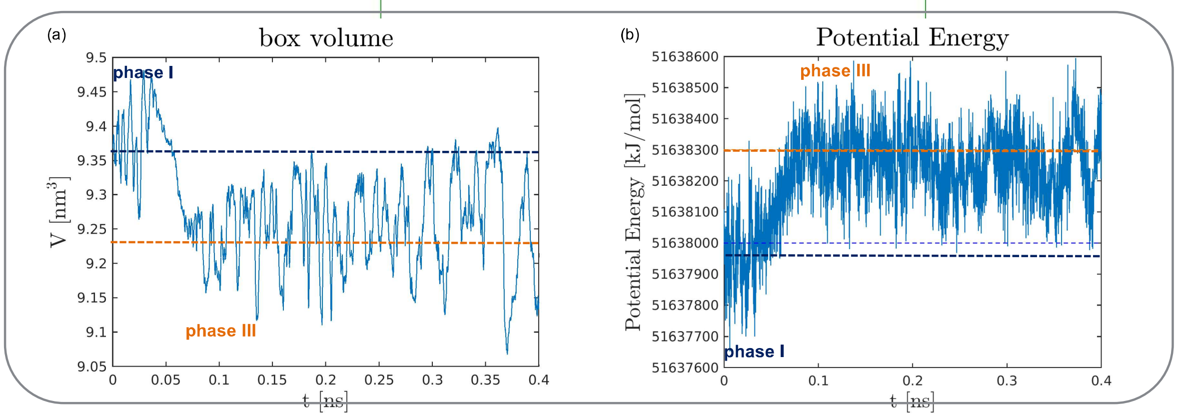


Sampling the I-III polymorphic transtion: a typical trajectory

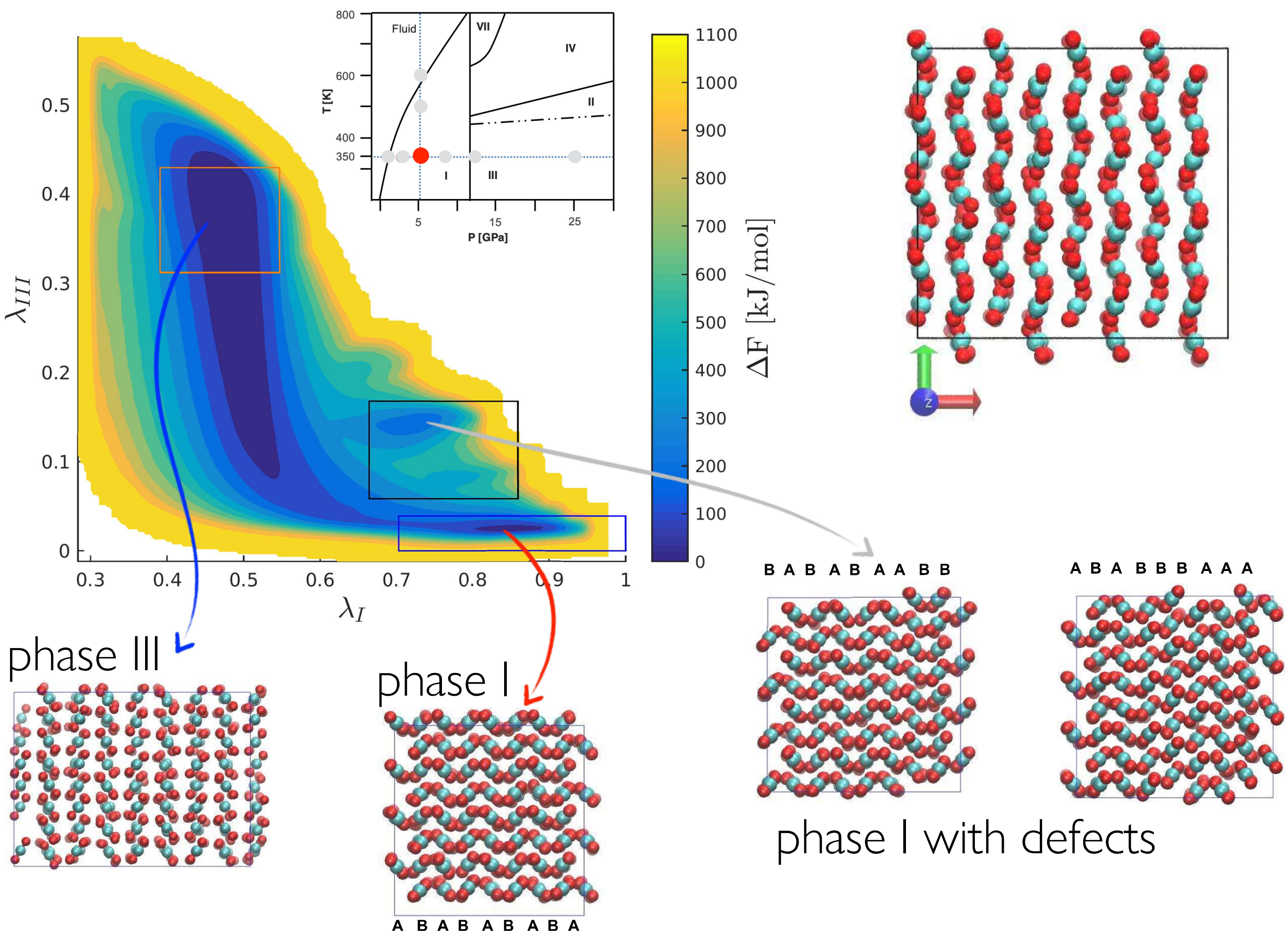




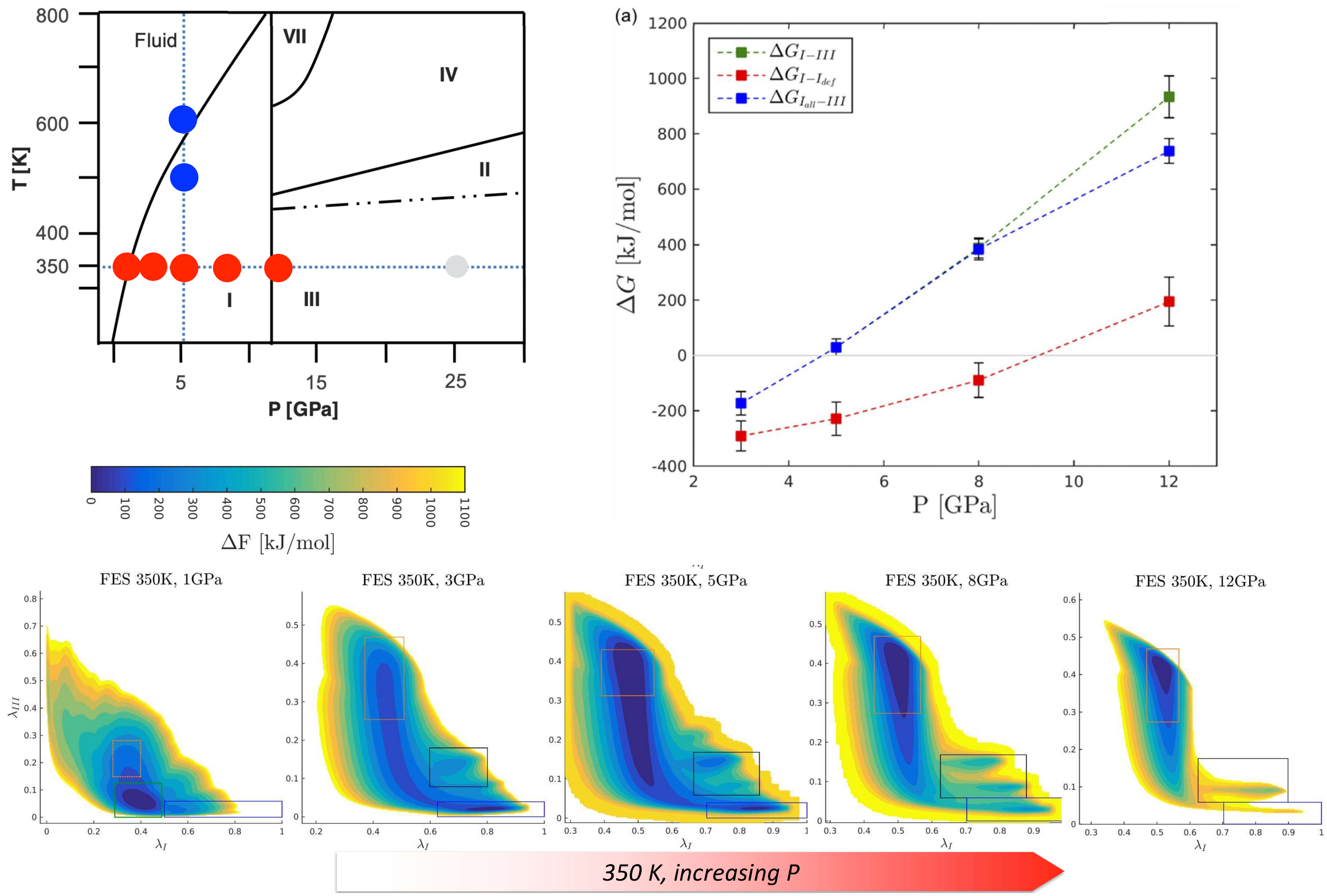




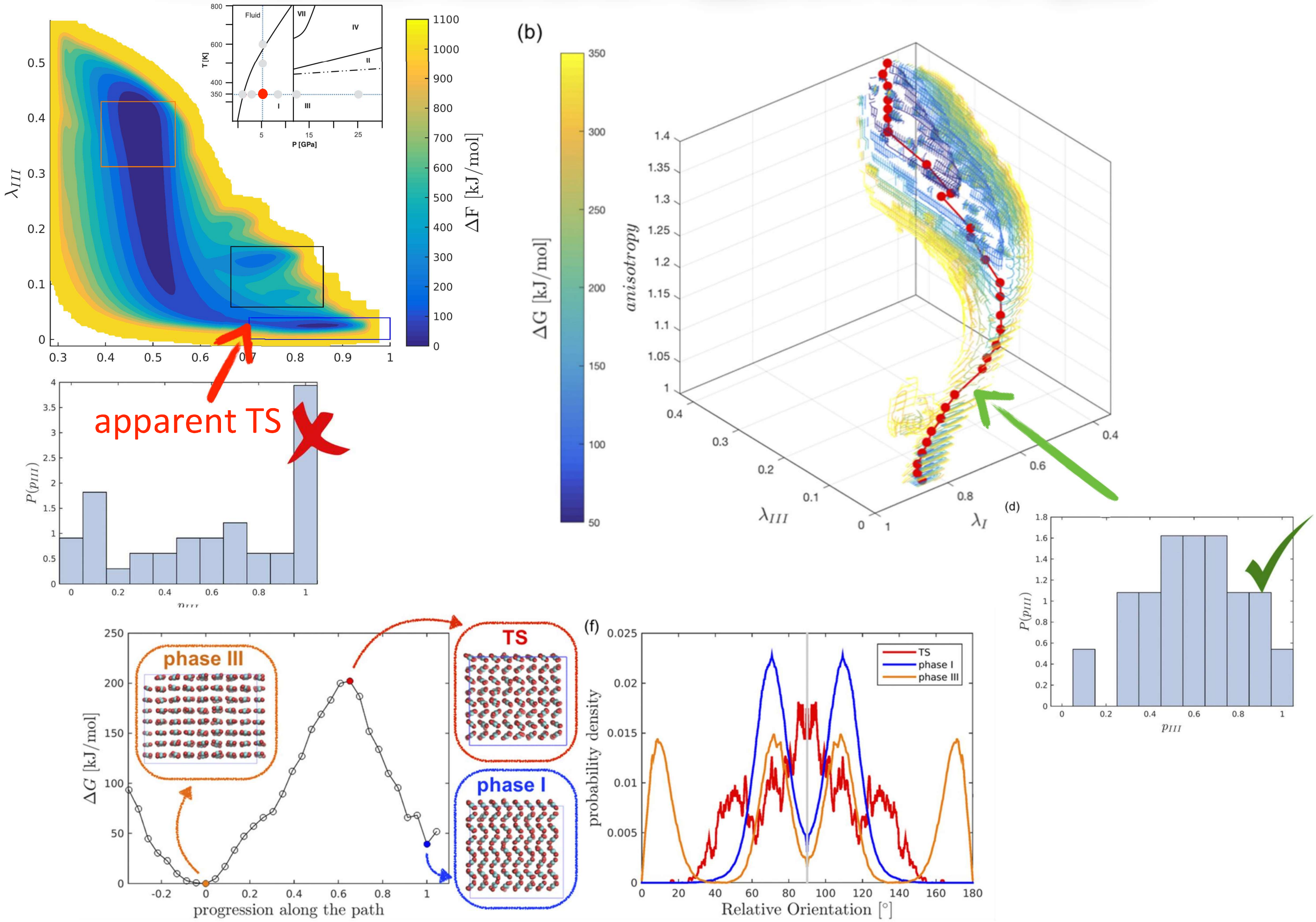
a typical Free Energy Surface



Free Energy Surfaces & Phase Diagram



I-III Transition Mechanism: apparent TS in CV space



Nucleation rates from metadynamics

Characteristic time

$$\tau = \frac{1}{J(S,T)\,V}$$

System Volume

Nucleation rate

a *@T, S=const*

$$\tau \propto \frac{1}{V}$$

b *@V=const*

$$\tau \;\propto\; \frac{1}{J(S,T)} \;\propto\; S^{-1}\exp\left(\beta\Delta F^{*}\right) = S^{-1}\exp\left(\frac{4(\beta\sigma a)^{3}}{27\left(\ln S^{2}\right)}\right)$$

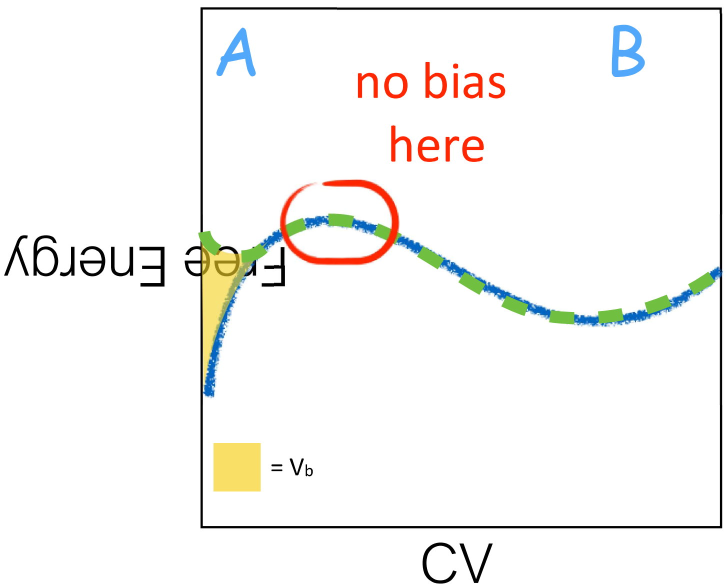
for S approaching 1, the characteristic time diverges

To compute J for S approaching 1
very large systems *(1E6 atoms)*
and/or
extremely long timescales *(hours)*
are required

finite size $\quad \tau_N \propto S^{-1}\exp\left(\beta\Delta F_N^{*}\right)$

macroscopic $\quad \tau \propto S^{-1}\exp\left(\beta\Delta F^{*}\right)$

Nucleation rates from metadynamics



$$k_{A\to B} = \kappa \frac{Z_{TS}}{Z_A}$$

$$k^*_{A\to B} = \kappa^* \frac{Z^*_{TS}}{Z^*_A}$$

$$\frac{k^*_{A\to B}}{k_{A\to B}} = \frac{Z_A}{Z^*_A} \frac{\kappa^* Z^*_{TS}}{\kappa Z_{TS}}$$

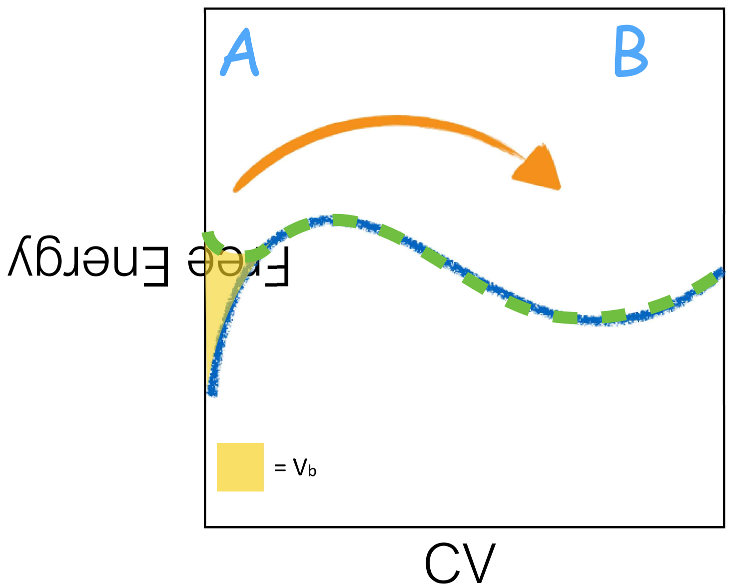
$$k_{A\to B} = k^*_{A\to B} \langle e^{-\beta V_b} \rangle$$

UNBIASED FREQUENCY — BIASED FREQUENCY" — BIAS

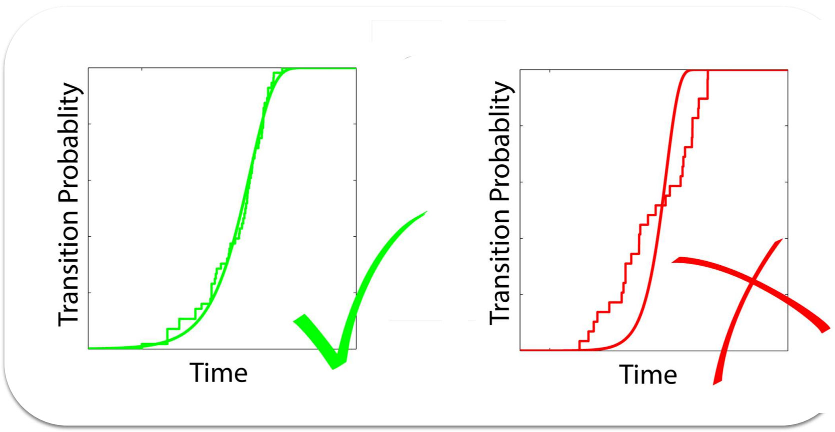
$$\frac{k^*_{A\to B}}{k_{A\to B}} = \frac{Z_A}{Z^*_A} = \langle e^{\beta V_b} \rangle$$

Tiwary and Parrinello, PRL, 2013

How to check that the TS is not affected by bias deposition?



- The **transition** between basins A and B is a **rare event.**
- The **escape** from a metastable state can be interpreted as a **Poisson process**.
- The **survival probability** is **exponentially distributed**.
- This **feature** has to be **preserved** in the calculation of the transition times distribution from **biased simulations**



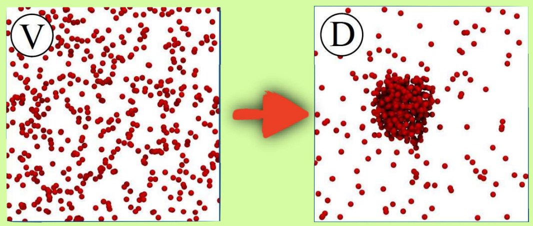
measuring the **compatibility** of the transition probability **with** that of a **Poisson process** **reveals** the **corruption** of the transition state

Salvalaglio, Tiwary and Parrinello, JCTC, 2014

Estimating nucleation timescales

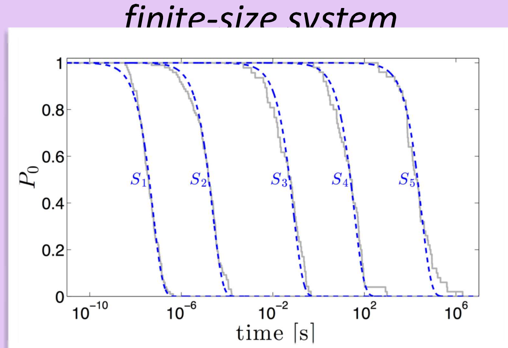
1 *Setup initial conditions:* ***T,S,V***

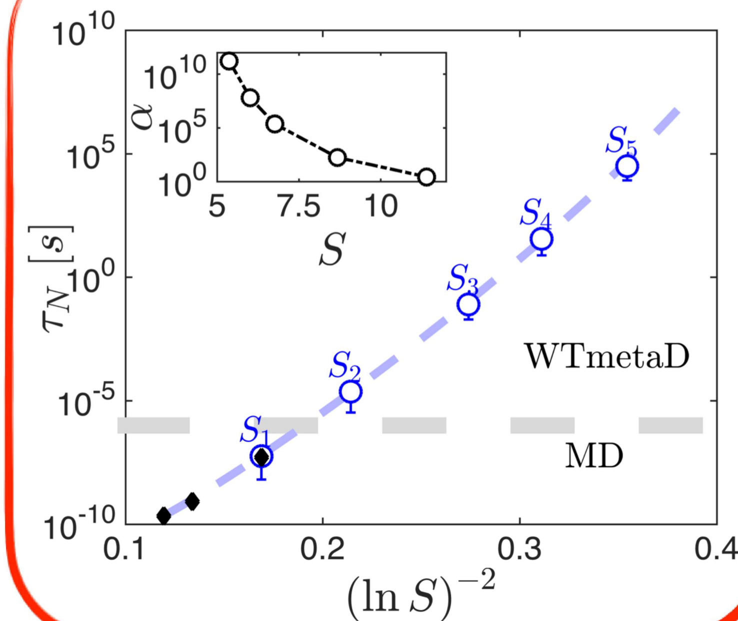
2 *Run a series of WTmetaD simulations*



each simulation provides one realisation of the transition time

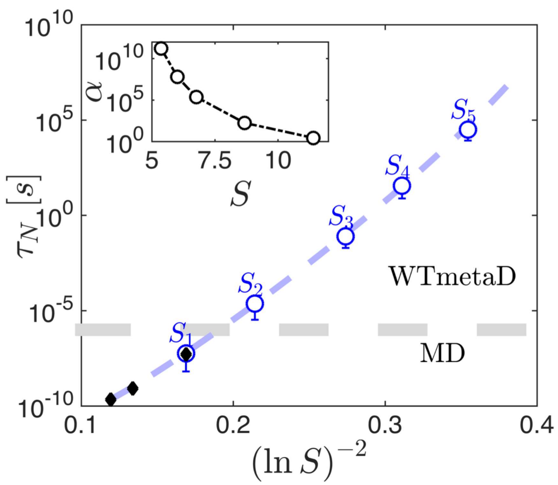
3 *Build survival probability distributions and compute the average nucleation time in the finite-size system*





Timescale ✓

Correcting for finite size effects

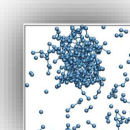


$$J_N = 1/\tau_N V$$

$$J = \phi J_N \simeq J_N \exp\left(\beta\left(\Delta F_N^* - \Delta F^*\right)\right)$$

$$\tau_N \propto S^{-1} \exp\left(\beta \Delta F_N^*\right)$$

- *Free energy barrier in the finite-sized system*
- *function of T,S,V and the* ***surface tension***



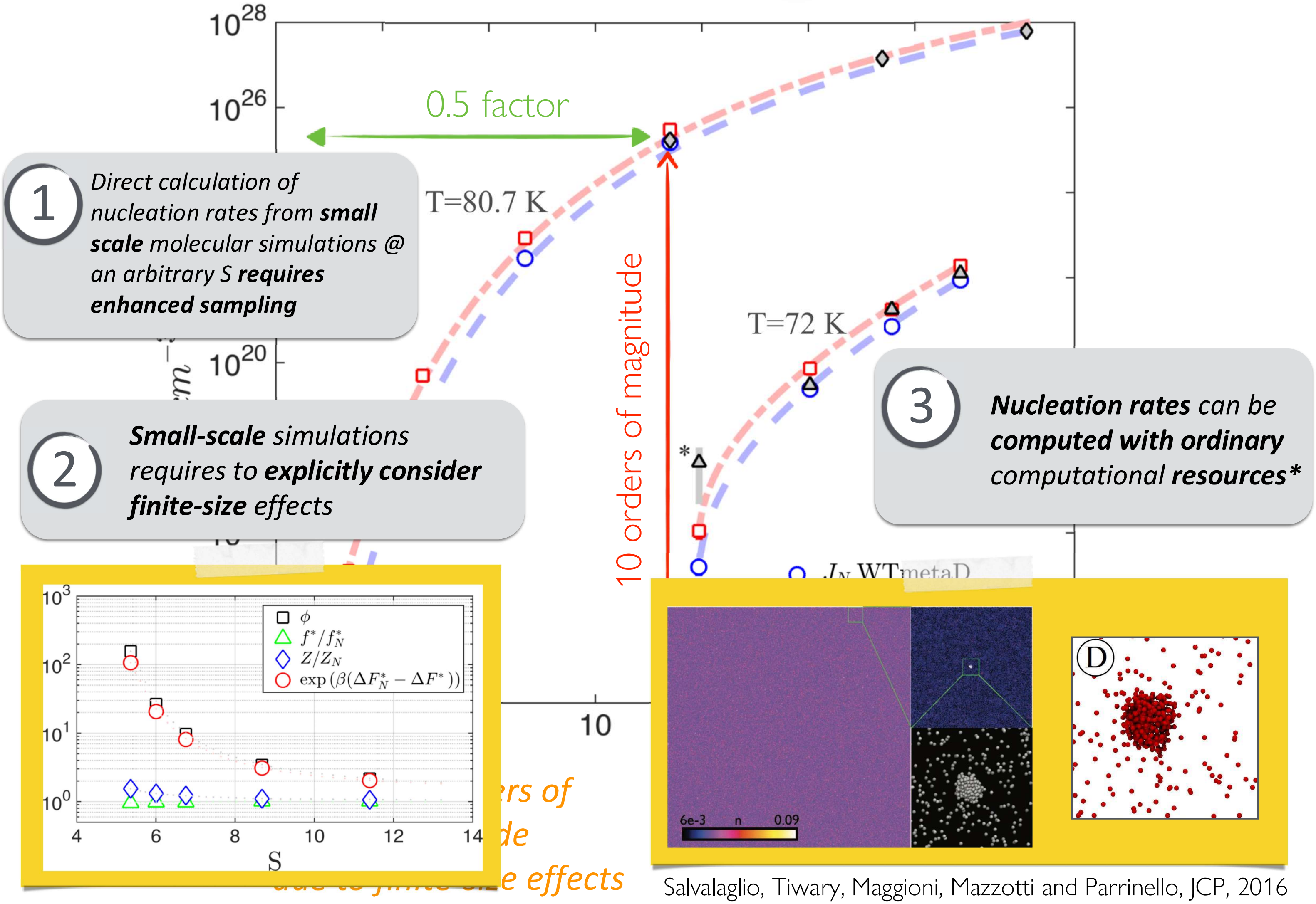
Finite Size

take home messages

1. *Direct calculation of nucleation rates from* ***small scale*** *molecular simulations @ an arbitrary S* ***requires enhanced sampling***

2. ***Small-scale*** *simulations requires to* ***explicitly consider finite-size*** *effects*

3. ***Nucleation rates*** *can be* ***computed with ordinary*** *computational* ***resources****



Salvalaglio, Tiwary, Maggioni, Mazzotti and Parrinello, JCP, 2016