https://github.com/SSAGESproject/SSAGES
https://github.com/SSAGESLabs/PySAGES

Email: jwhitme1@nd.edu

Advanced Sampling Using SSAGES/PySAGES: Background

Jonathan K. Whitmer

University of Notre Dame

Argonne National Laboratory Hands-On Workshop
10/13/22

What is advanced sampling?

Credit: B. Sikora.

Accelerated Dynamics

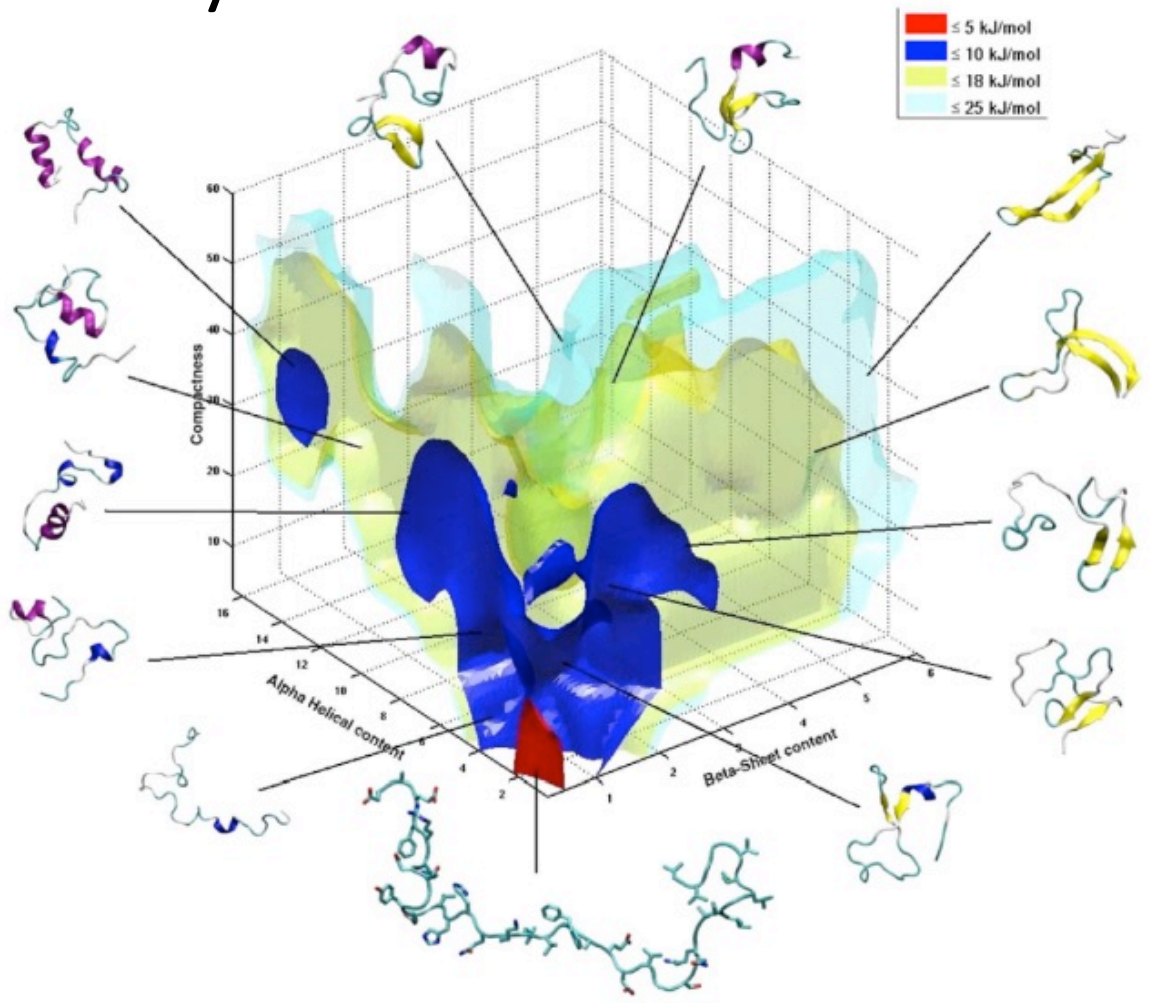
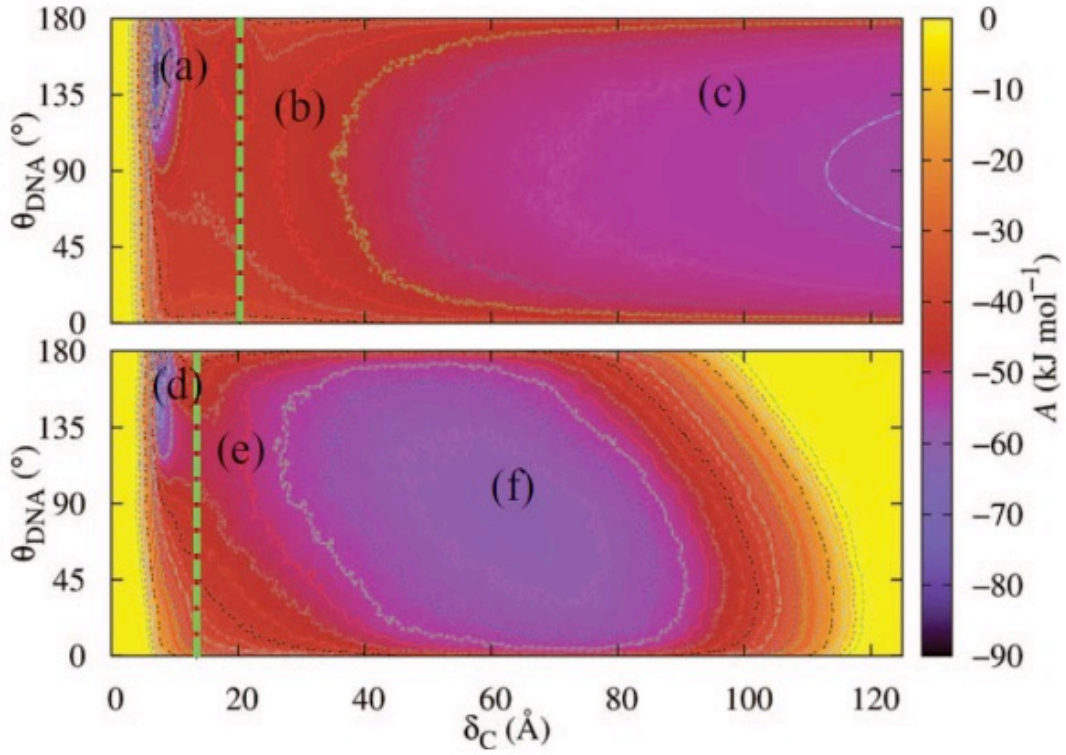
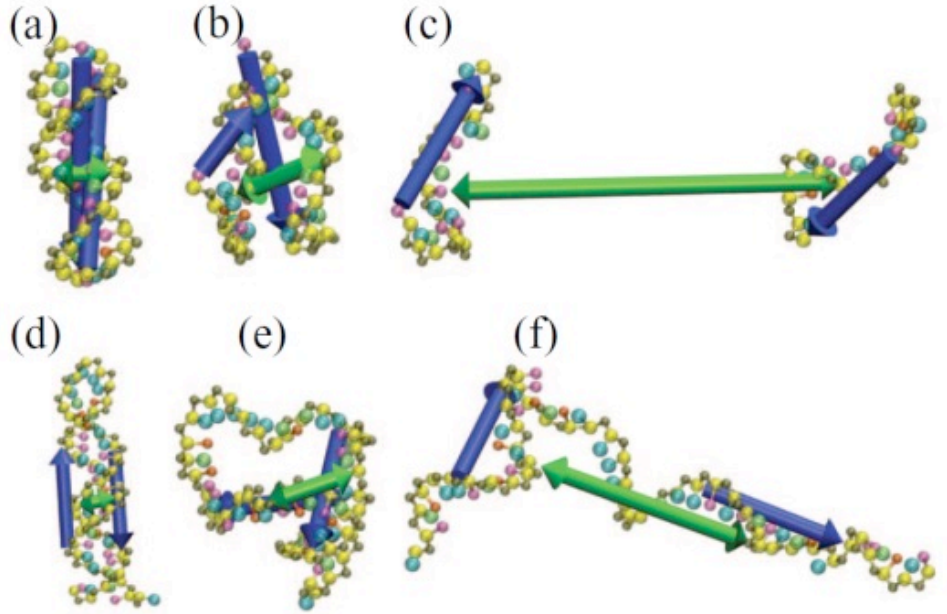


Figure 1. Free energy landscape of the Aβ40 peptide. The free energy landscape is shown as a function of three collective variables used in the NMR-guided metadynamics simulations: anti-parallel β-sheet content (X-axis), α-helical content (Y-axis) and number of hydrophobic contacts (or compactness, Z-axis). Isosurfaces are shown at 5 (red), 10 (blue), 18 (yellow) and 25 kJ/mol (cyan); white regions are not visited as they have higher free energies. Representative structures sampled during the simulation are also shown.

Granata, D. *et al.* The inverted free energy landscape of an intrinsically disordered peptide by simulations and experiments. *Sci. Rep.* **5,** 1–15 (2015).

Accelerated Dynamics





Hinckley, D. M., Freeman, G. S., Whitmer, J. K. & de Pablo, J. J. An experimentally-informed coarse-grained 3-site-per-nucleotide model of DNA: Structure, thermodynamics, and dynamics of hybridization. *J. Chem. Phys.* **139,** 144903 (2013).

Molecular Simulation and the Partition Function

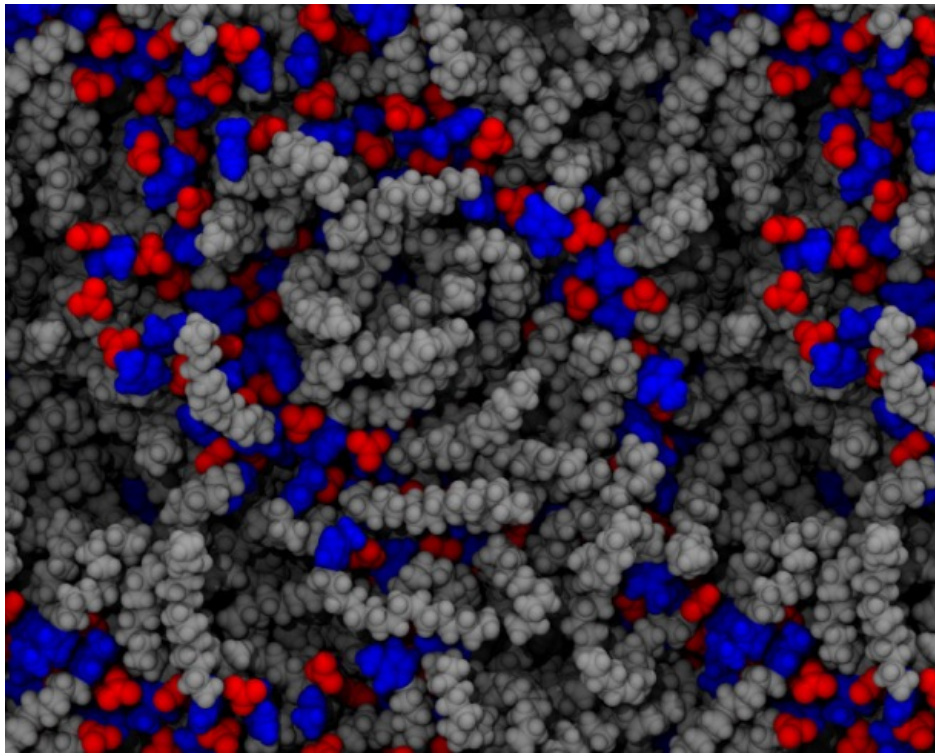
The state vector or *system point* in a constant-volume system* is defined as

$$\boldsymbol{\Xi} := (\mathbf{X}; \mathbf{P}) = (\mathbf{x}_1, \ldots, \mathbf{x}_N; \mathbf{p}_1, \ldots, \mathbf{p}_N)$$

If $\boldsymbol{\Xi}$ is known at any time, it is *completely determined* for all other times (past and present) through the classical equations of motion (NVE).

A point $\boldsymbol{\Xi}$ in phase space Ω is a *microstate* of the system.

Ergodic trajectories in molecular dynamics (MD) simulations evolve to each microstate commensurate with the *ensemble* chosen

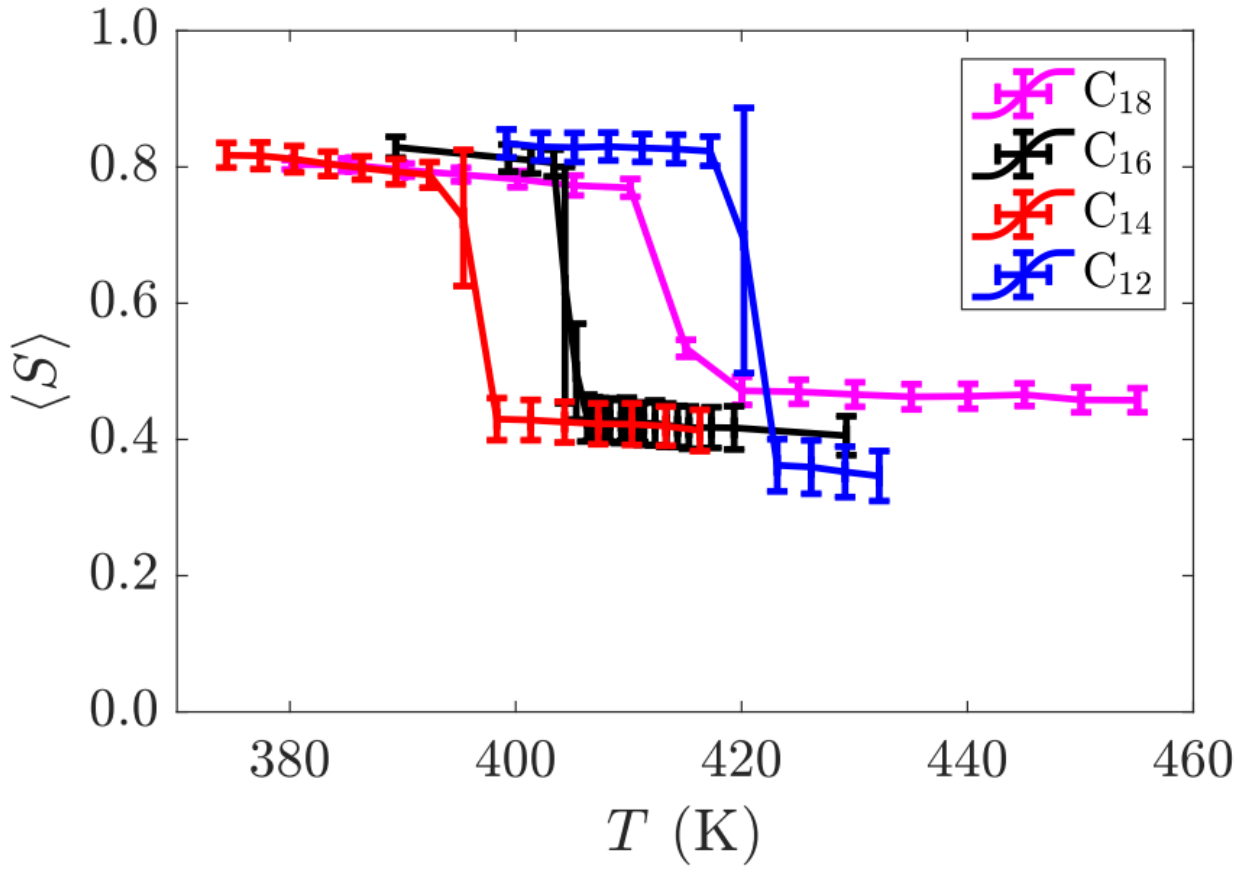


M. Quevillon and J. Whitmer, *Materials* (2018)
M. Quevillon, A. Panteleev and J. Whitmer, *In Prep* (2018)

Molecular Simulation and the Partition Function

- We are interested in collective properties "ξ" of the system: structural ***order parameters***, ***reaction coordinates*** denoting extent of transformation, ***collective variables*** labelling similar states.
- Can be thermodynamic quantities (e.g. $\langle U \rangle$) but are not necessarily. Liquid crystal order parameters $\langle S \rangle$, elastic deformations, numbers of hydrogen bonding contacts can also be of interest
- These are measured as a histogram or time-average over the system state.

$$\langle A \rangle = \frac{\int_\Omega d^{3N}\mathbf{x}\; d^{3N}\mathbf{p}\; A(\mathbf{x}) e^{-\beta U(\mathbf{x})}}{\int_\Omega d^{3N}\mathbf{x}\; d^{3N}\mathbf{p}\; e^{-\beta U(\mathbf{x})}}$$



M. Quevillon and J. Whitmer, *Materials* (2018)
M. Quevillon, A. Panteleev and J. Whitmer, *In Prep* (2018)

Evaluation of the partition function…

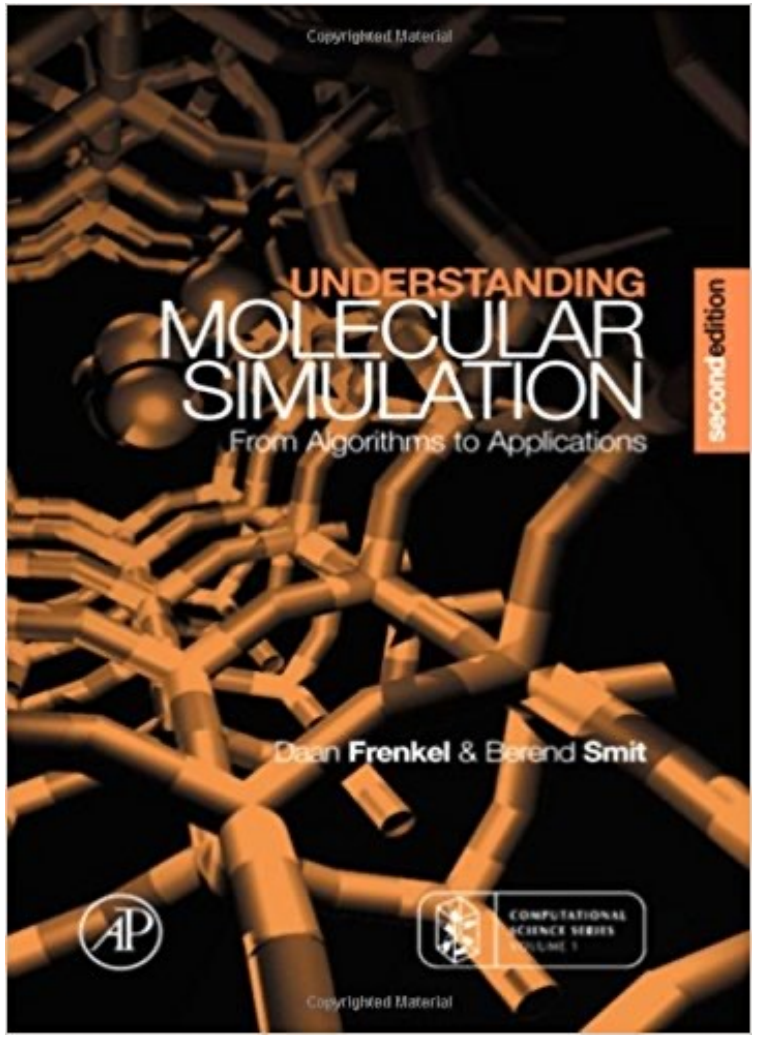
…is effectively impossible:

"…the number of quantum states that contribute to the average is so astronomically large…that a numerical evaluation of all expectation values is unfeasible."

$$\langle A \rangle = \frac{\sum_i A(i) e^{-\beta E_i}}{\sum_i e^{-\beta E_i}}$$

$$\langle A \rangle = \frac{\int_\Omega d^{3N}\mathbf{x}\; d^{3N}\mathbf{p}\; A(\mathbf{x}) e^{-\beta U(\mathbf{x})}}{\int_\Omega d^{3N}\mathbf{x}\; d^{3N}\mathbf{p}\; e^{-\beta U(\mathbf{x})}}$$

While we would ideally like to know the partition function, it's not realistic to evaluate all microstates of the system.



D. Frenkel and B. Smit. *Understanding Molecular Simualtion, 2nd Ed.* Academic Press, 2001.

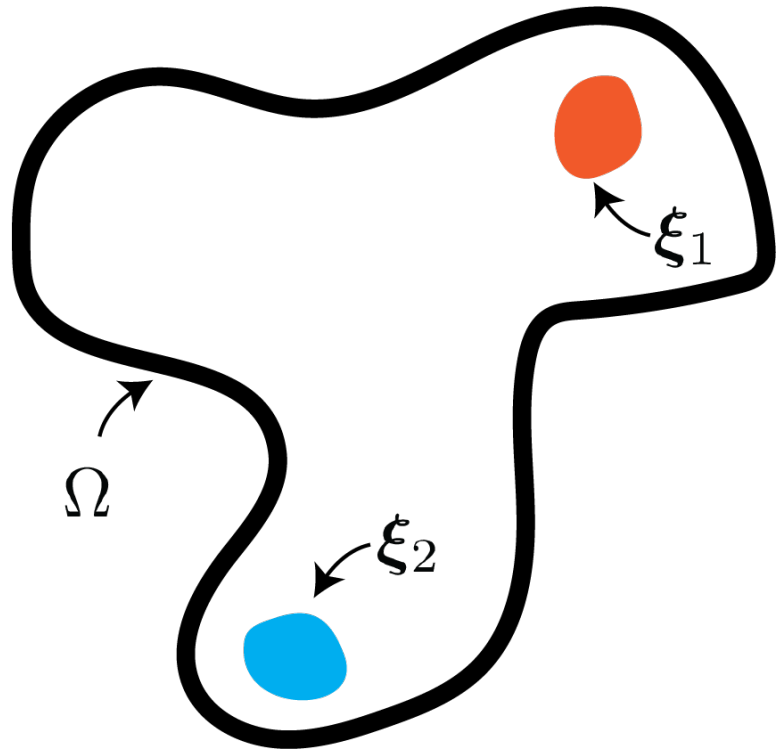
…but we *can* do somewhat better.

Labelling states by small numbers of collective properties permits partial evaluation of the partition function.

$$Z(\boldsymbol{\xi}) = \int_{\Omega} d^{3N}\mathbf{x}\; d^{3N}\mathbf{p}\; \delta(\mathbf{s}(\mathbf{x}) - \boldsymbol{\xi}) e^{-\beta U(\mathbf{x})}$$

Thus, we can efficiently obtain information about average values **and** relative free energies.

$$P(\boldsymbol{\xi}) = \frac{Z(\boldsymbol{\xi})}{Z} := e^{-\beta F(\boldsymbol{\xi})}$$



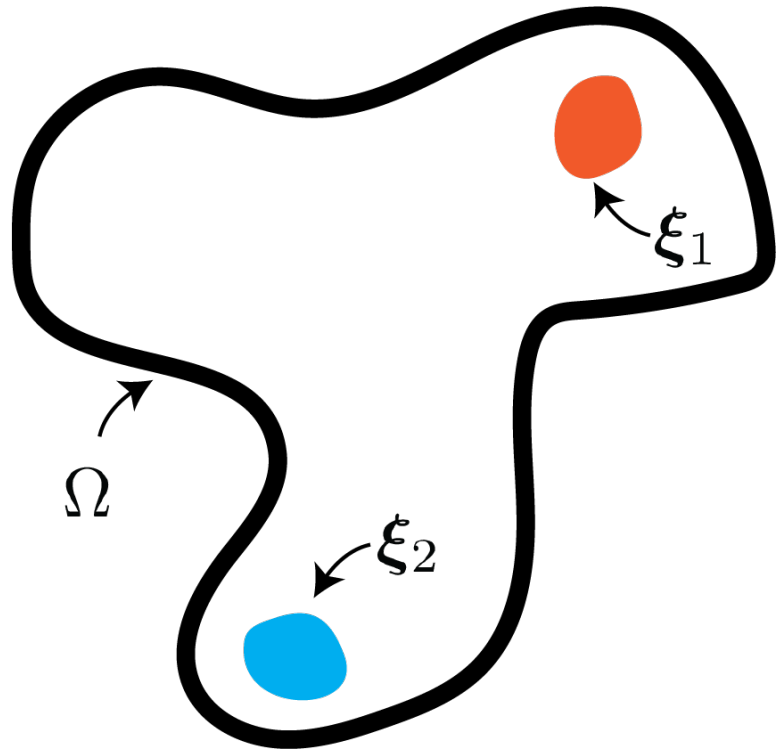
…but we *can* do somewhat better.

Labelling states by small numbers of collective properties permits partial evaluation of the partition function.

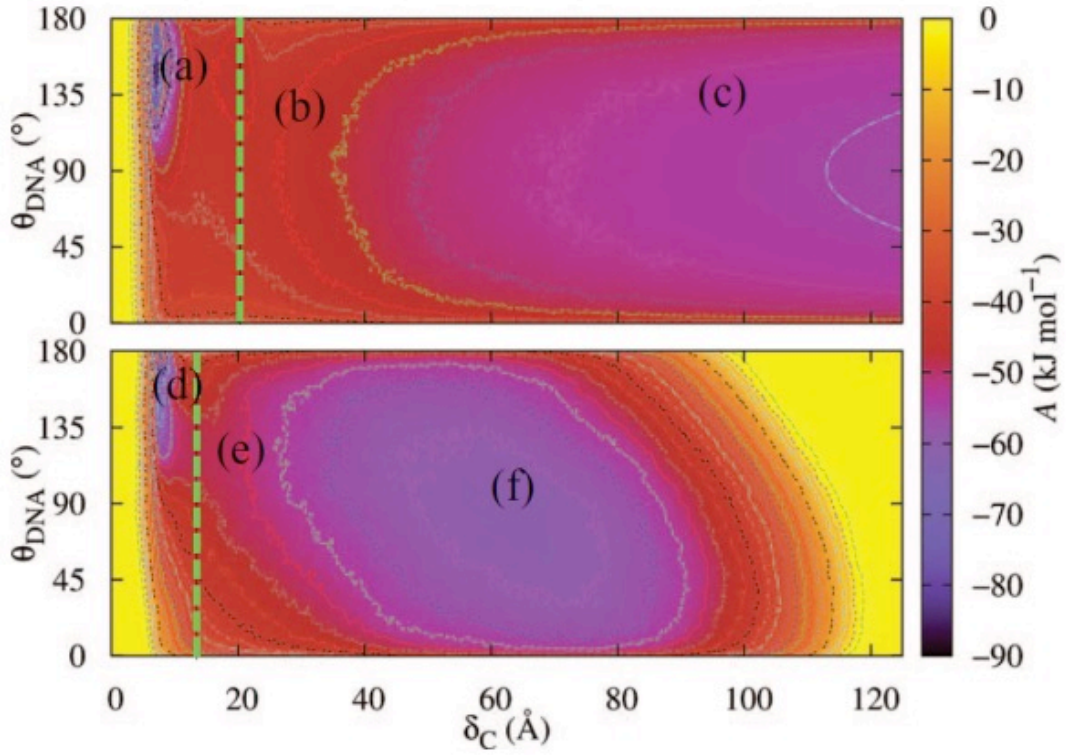
$$Z(\boldsymbol{\xi}) = \int_{\Omega} d^{3N}\mathbf{x}\; d^{3N}\mathbf{p}\; \delta(\mathbf{s}(\mathbf{x}) - \boldsymbol{\xi}) e^{-\beta U(\mathbf{x})}$$

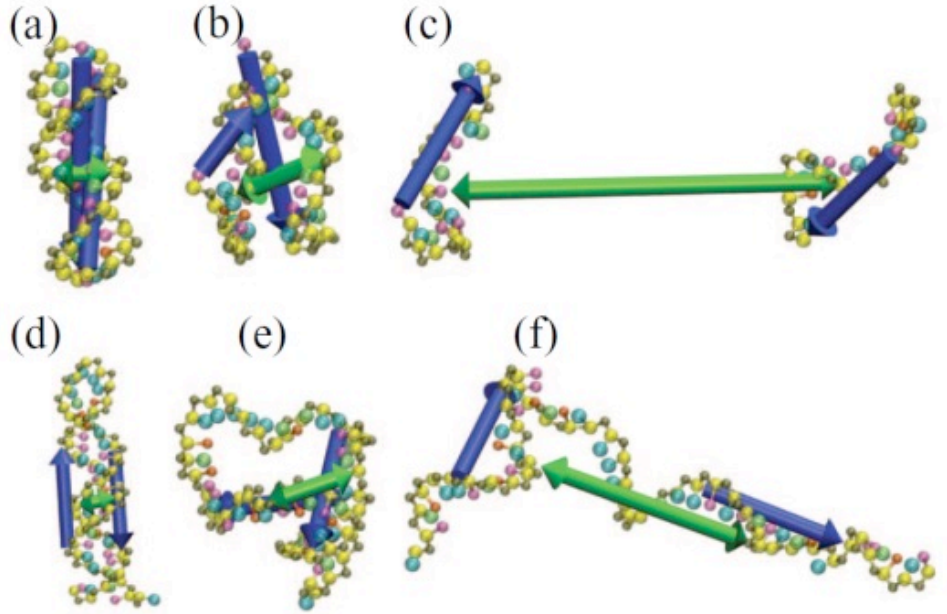
Thus, we can efficiently obtain information about average values **and** relative free energies.

$$P(\boldsymbol{\xi}) = \frac{Z(\boldsymbol{\xi})}{Z} := e^{-\beta F(\boldsymbol{\xi})}$$



Accelerated Dynamics





Hinckley, D. M., Freeman, G. S., Whitmer, J. K. & de Pablo, J. J. An experimentally-informed coarse-grained 3-site-per-nucleotide model of DNA: Structure, thermodynamics, and dynamics of hybridization. *J. Chem. Phys.* **139,** 144903 (2013).

Collective Variables

$$Z = \sum_s g(s) e^{-\beta u(s)}$$

Collective Variables

$$Z = \sum_s g(s) e^{-\beta U(s)} = \int_\Omega d^{3N}x \, e^{-\beta U(x)}$$

Collective Variables

$$Z = \sum_s g(s)\, e^{-\beta U(s)} = \int_\Omega d^{3N}x\, e^{-\beta U(x)}$$

$g(s)$: density-of-states

Ω: explorable domain $\Omega \subseteq \mathbb{R}^{3N}$

Collective Variables

$$Z = \sum_s g(s)\, e^{-\beta U(s)} = \int_\Omega d^{3N}x\, e^{-\beta U(x)}$$

density-of-states

explorable domain $\Omega \subseteq \mathbb{R}^{3N}$

Partition Function of NVT ensemble*

* WLOG. This example is canonical.

Collective Variables

$$Z(\underline{\xi}) = \int_{\Omega} \delta(\underline{s}(\underline{x}) - \underline{\xi})\, d^{3N}\underline{x}\, e^{-\beta U(\underline{x})}$$

Partition Function of $NVT\underline{\xi}$ ensemble

Collective Variables

$$Z(\underline{\xi}) = \int_{\Omega} \delta(\underline{s}(\underline{x}) - \underline{\xi})\, d^{3N}\underline{x}\, e^{-\beta U(\underline{x})}$$

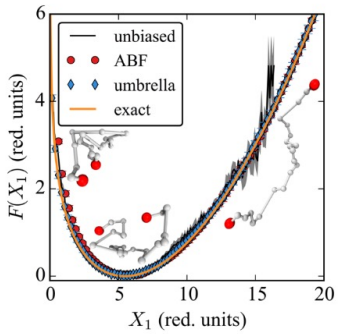
Partition Function of $NVT\underline{\xi}$ ensemble

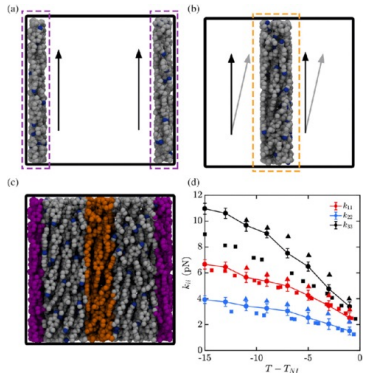
$$P(\underline{\xi}) = \frac{Z(\underline{\xi})}{Z} := e^{-\beta F(\underline{\xi})}$$

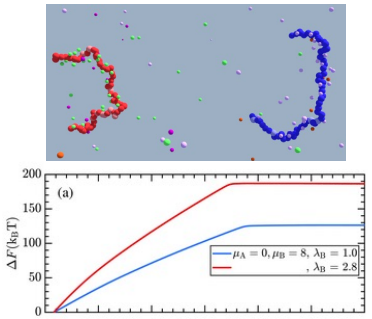
Collective Variables

Types of CVs

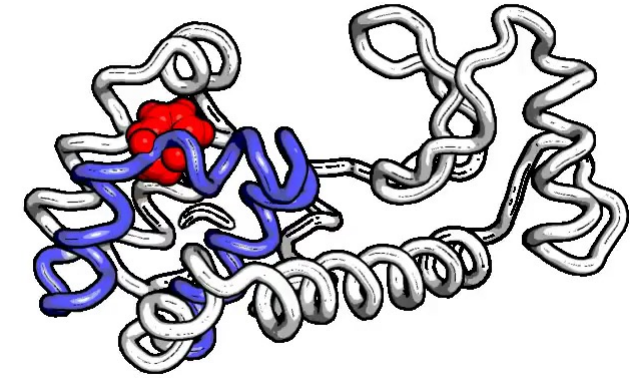
Pairwise distance	Pairwise Clustering
Box volume	Gyration Tensor
Angle	Torsional Angle
Alpha Helix	Parallel Beta Sheet
Anti-Parallel Beta Sheet	Rouse Modes
Particle Positions	Neural Network
Elastic Deformations	Path-Based CV
Orientational Order Parameters	Bespoke/User-generated

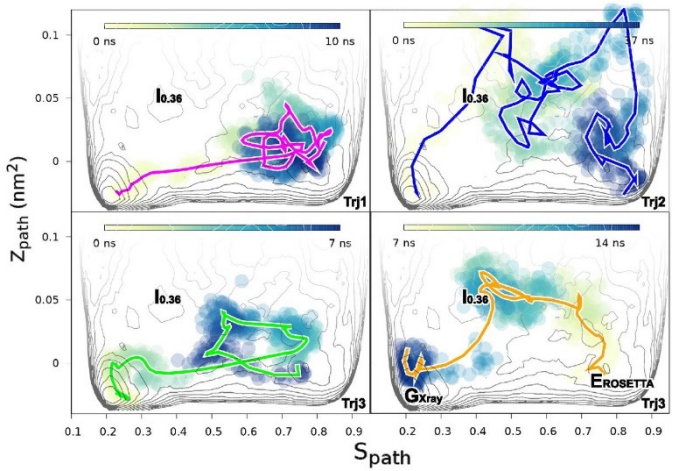






Sidky, et al. *J. Chem. Phys.* **148**, 044104 (2018)
Sidky, et al, PRL **120**, 107801 (2018)
Rathee, et al. *JACS,* **ASAP**, (2018)

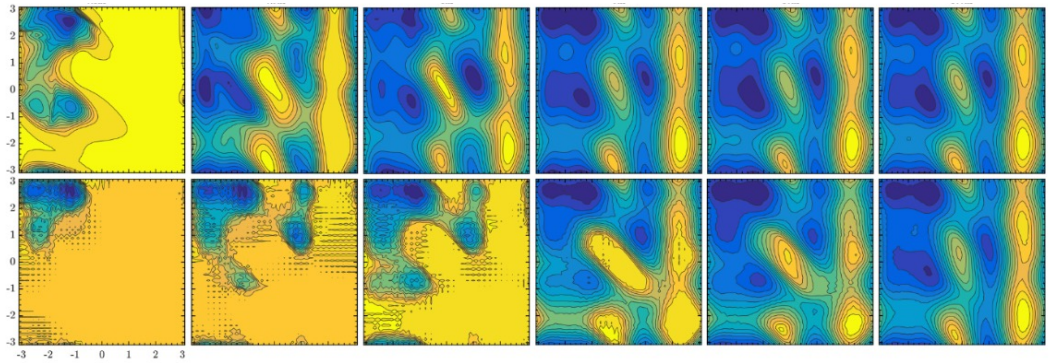




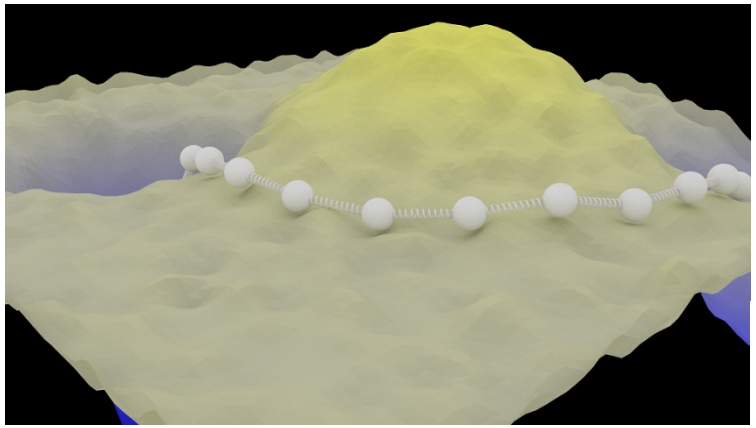
Wang, Y., Papaleo, E. & Lindorff-Larsen, K. Mapping transiently formed and sparsely populated conformations on a complex energy landscape. *Elife* **5,** (2016).

Classes of Methods

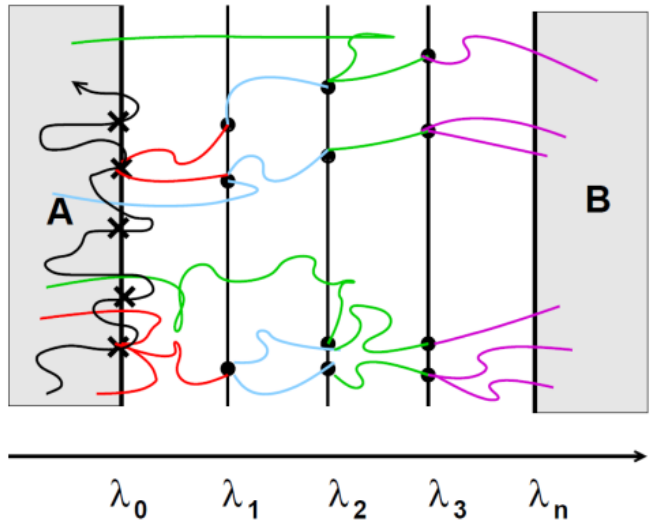
- Flat Histogram Methods



- Minimum FEP Methods



- Reactive Path Methods



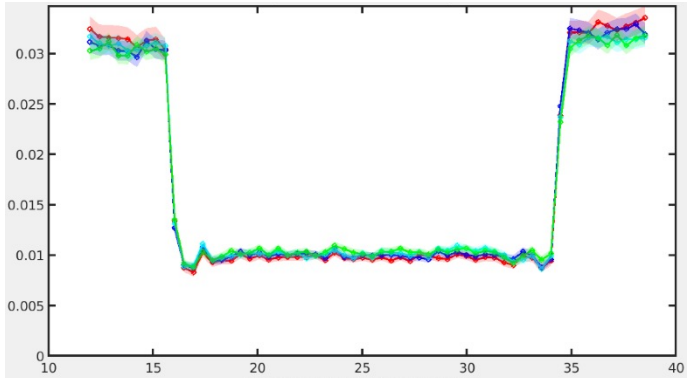
Flat Histogram Methods

Progress Through State Labelling?

Trajectory (MD) or Markov Chain (MC) → Sequence of observed *micro*states

$$\{\boldsymbol{x}_i\}$$

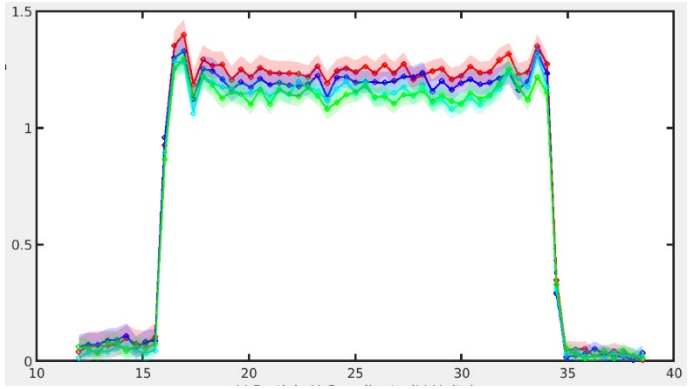
$H(\xi)$



→ Sequence of observed *macro*states

$$\{\boldsymbol{s}(\boldsymbol{x}_i)\}$$

$F(\xi)$



→ Histogram of visited *macro*states

$$H(\xi) = \sum_i \delta(\,\boldsymbol{s}(\boldsymbol{x}_i) - \xi) \sim e^{-\beta F(\xi)}$$

ξ

Images: P. Amorim.

Umbrella Sampling

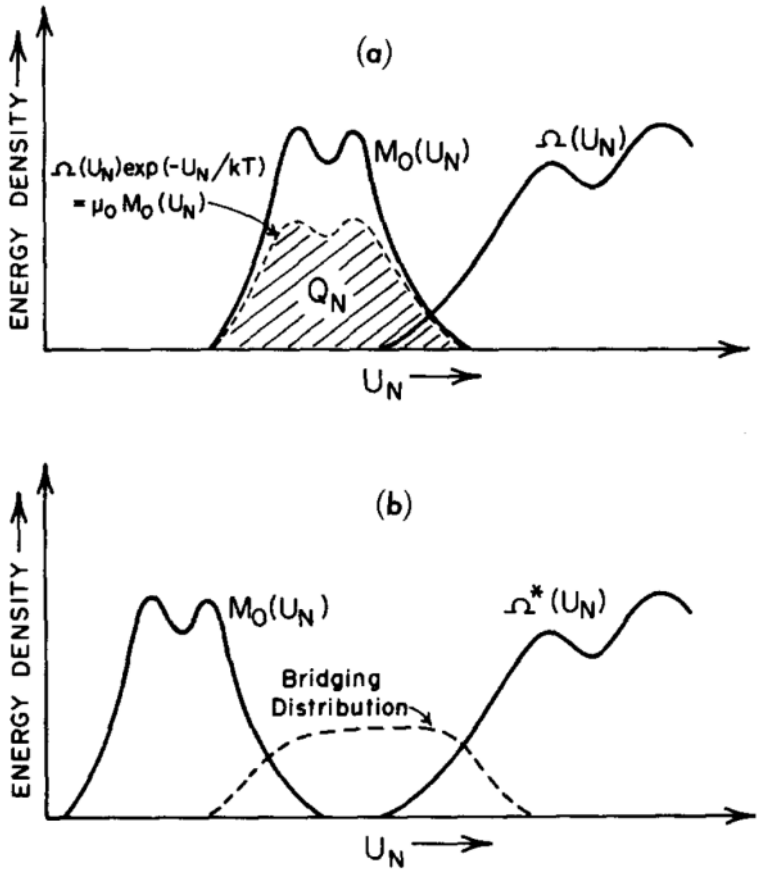


FIG. 1. (a) Schematic diagram of distributions discussed in text. From overlap of $\Omega(U_N)$ [or $\Omega^*(U_N)$] with M_0, one could obtain μ_0 and hence Q_N. (b) Sampled distributions $M_0(U_N)$ and $\Omega^*(U_N)$ will often not overlap; μ_0 can then be obtained by generating a bridging distribution.

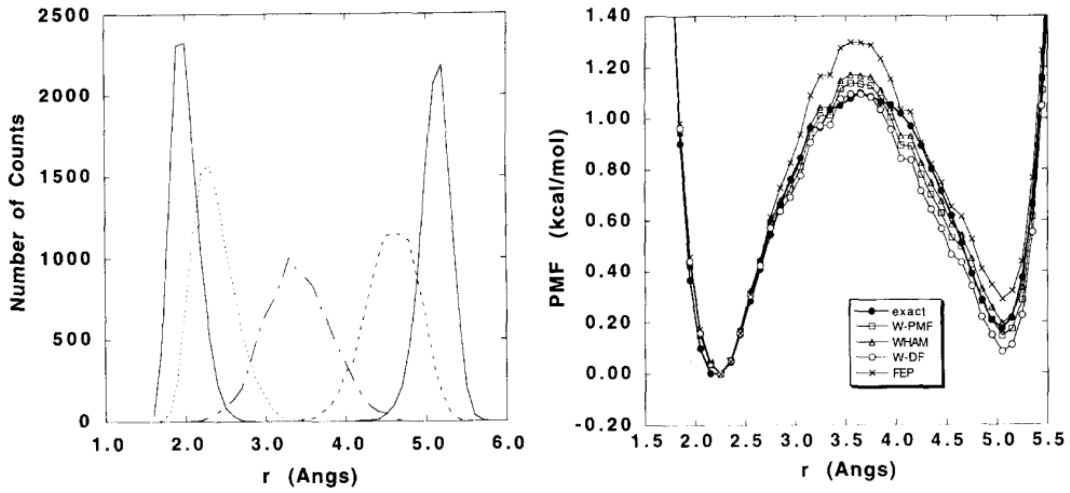
Valleau, J. P. & Card, D. N. Monte Carlo Estimation of the Free Energy by Multistage Sampling. *J. Chem. Phys.* **57,** 5457–5462 (1972).

The Reliable Dinosaur

- Constrain sampling to a region of a state variable through (e.g.) harmonic springs

$$U_{\text{eff}}(\vec{x}) = U(\vec{x}) + k\xi(\vec{x})^2$$

- Removing the bias gives the local free energy surface
- Overlapping regions can be stitched together using the weighted histogram method



It's been around forever, but works provided:

(a) You can generate the initial configruations for each umbrella
(b) You can space umbrellas closely enough to overlap histograms significantly
(c) The free energy surface is "simple", or parallel exchanges are implemented
(d) Diffusion is sufficiently fast to cover the width of the umbrella in accessible timescales

Roux, *Comp. Phys. Comm.* **91** 275 (1995)

Umbrella Sampling

int. energy

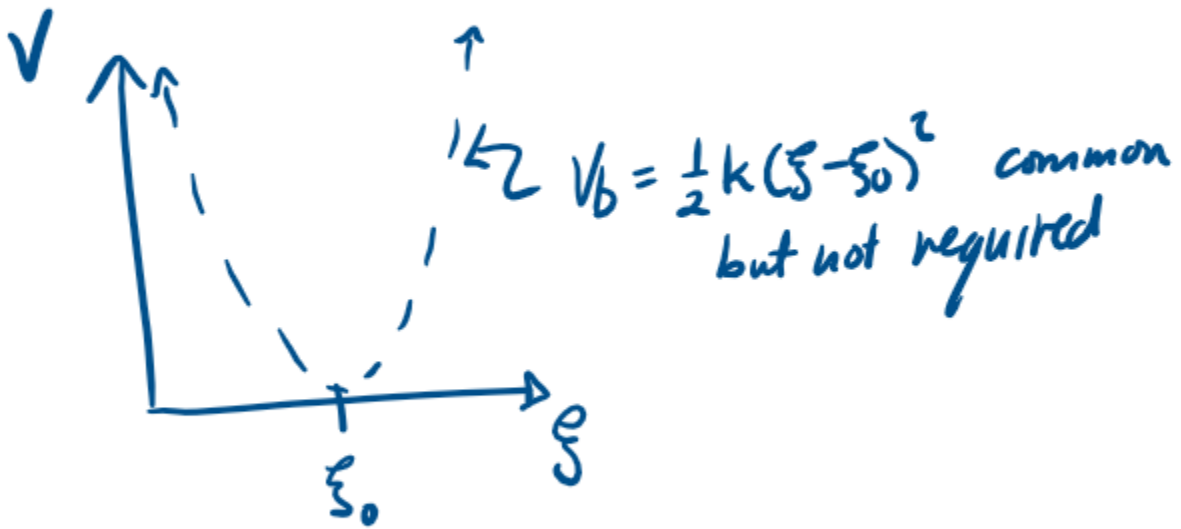
$$U' = U + V(\underline{\xi})$$

bias potential

$$P'(\underline{\xi}) \propto \int d^{3N}\underline{x}\, \delta(\underline{s}(\underline{x}) - \underline{\xi})\, e^{-\beta U(\underline{x}) - \beta V(\underline{\xi})}$$

$$= e^{-\beta V(\underline{\xi})} P(\underline{\xi})$$

$$P(\underline{\xi}) = e^{\beta V(\underline{\xi})} P'(\underline{\xi})$$



Umbrella Sampling

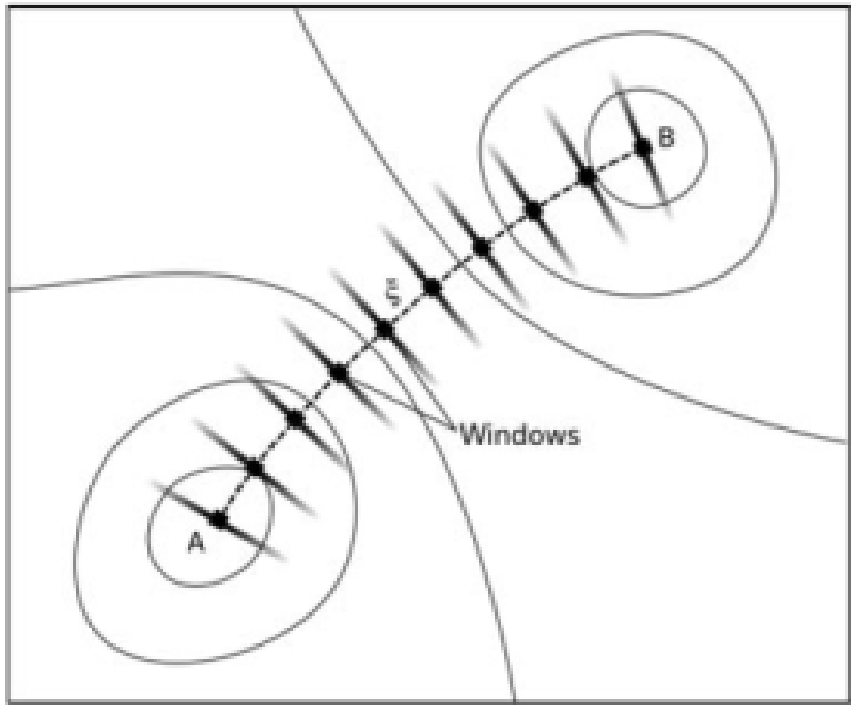


FIGURE 1 | Separation of the reaction coordinate (dashed line) between two states (here represented by two minima on the potential energy surface) into distinct windows. The system is mainly sampled perpendicular to the reaction coordinate in each window.

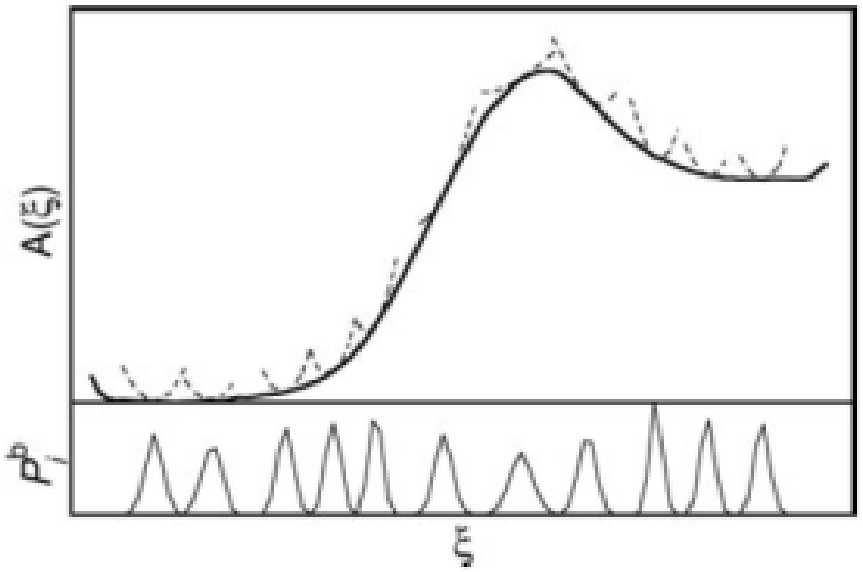
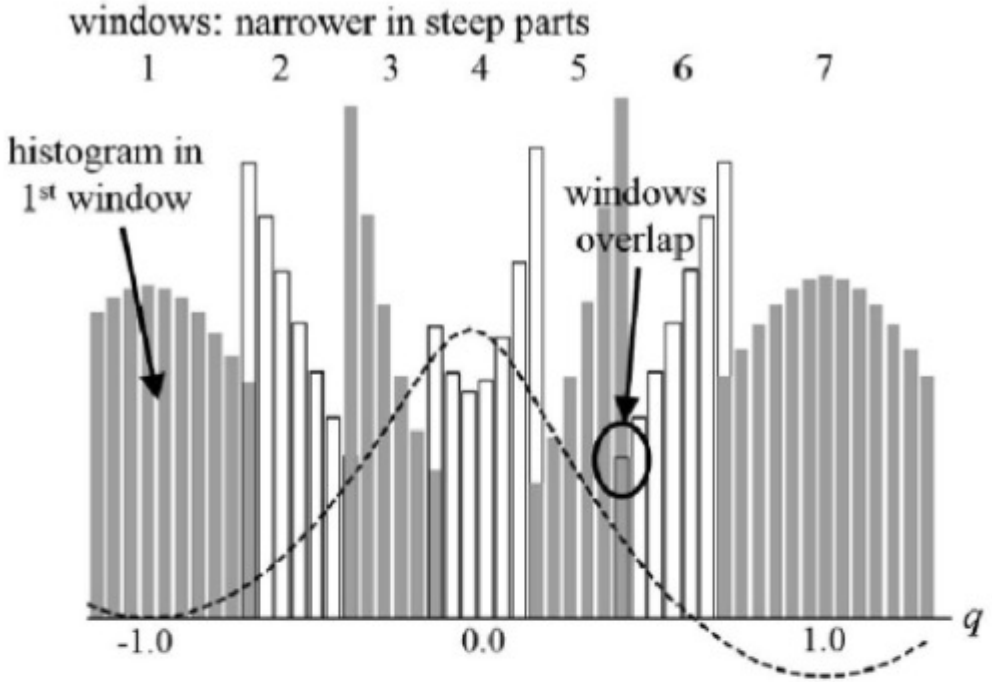
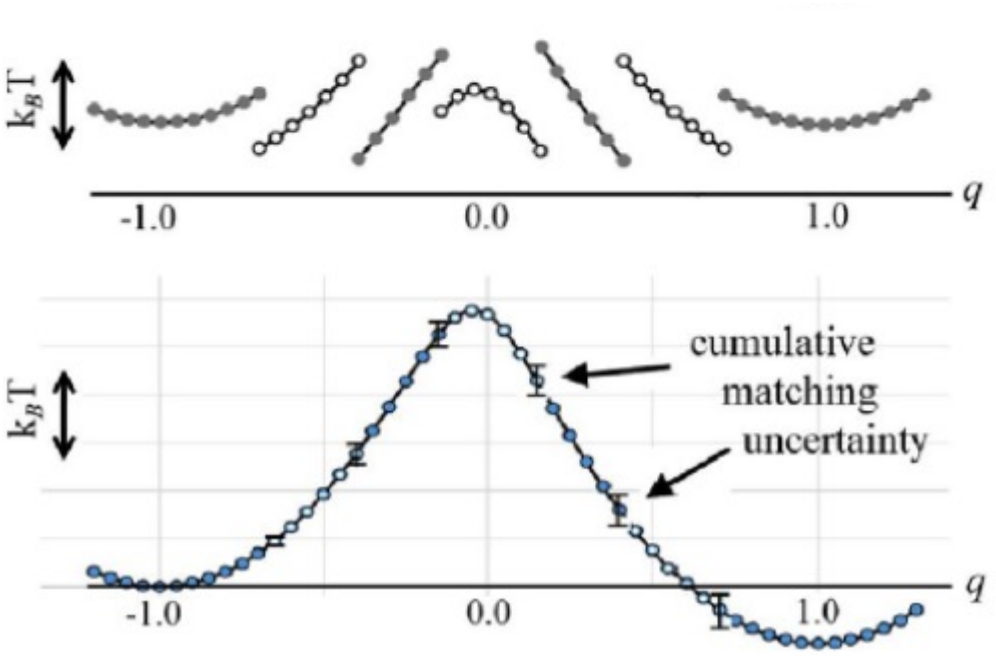


FIGURE 2 | Global free energy (thick solid curve) and the contributions A_i of some of the windows (thin dashed curves). For clarity, only every third window is shown. At the bottom, the biased distributions P_i^b as obtained from the simulation are shown (thin solid curves). Relatively few bins (100) have been used to generate this figure.

Kästner, J. Umbrella sampling. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **1,** 932–942 (2011).

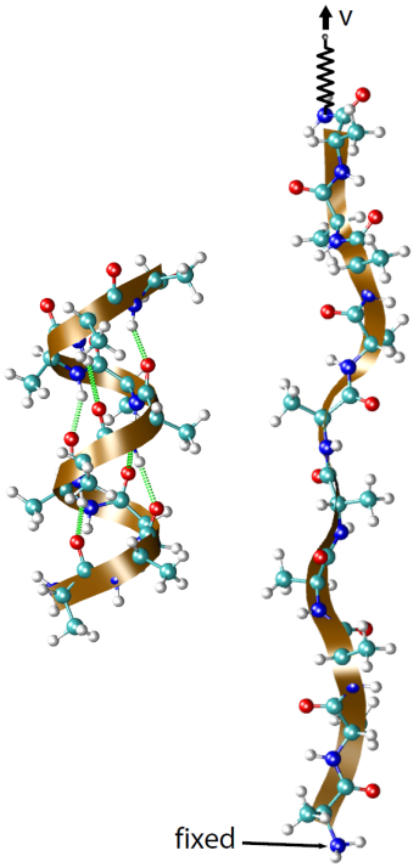
Gluing things together

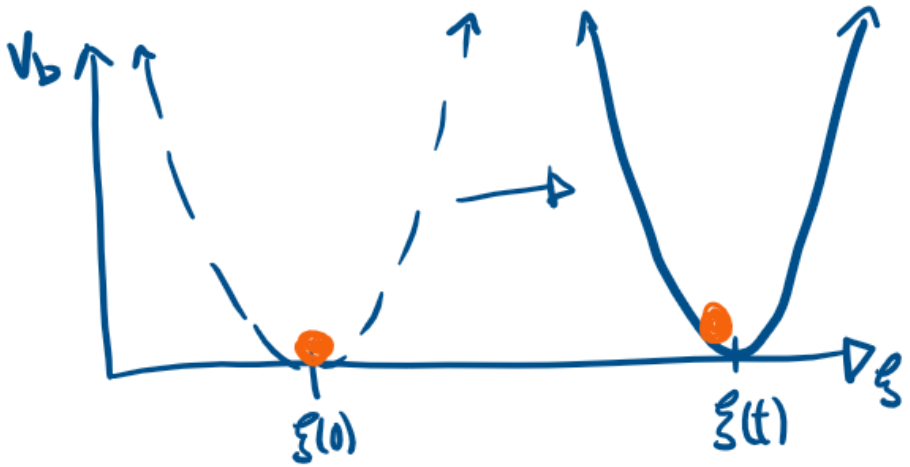




B. Peters, *Reaction Rate Theory and Rare Events*, Elsevier (2017)

Steered Molecular Dynamics





This is a great technique for generating a reactive trajectory (discussed later), but may not sample the ideal rxn coordinates.

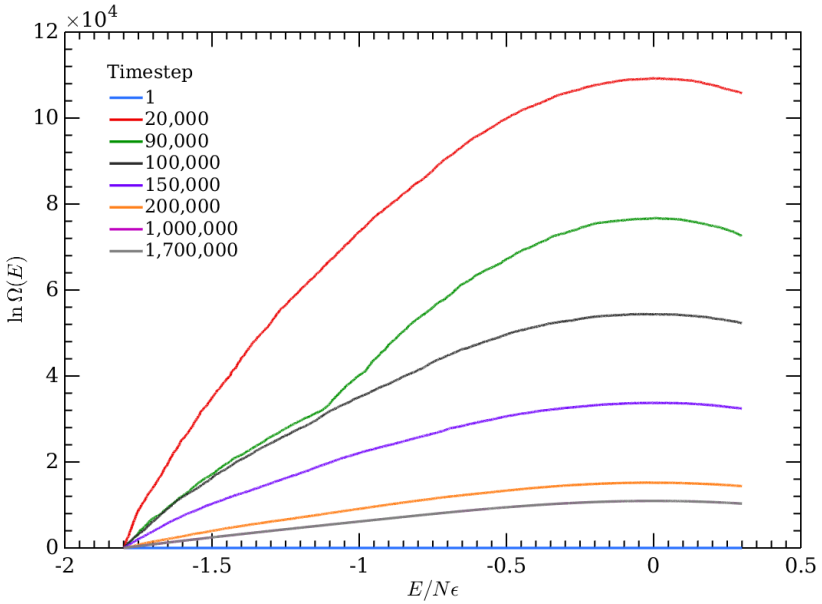
Jarzynski equality:

$$e^{-\beta \Delta F} = \langle e^{-\beta W} \rangle$$

ensemble average of the work performed in pulling protocol. (Practically, it helps to do this slowly)

Park, S., Khalili-Araghi, F., Tajkhorshid, E. & Schulten, K. Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. *J. Chem. Phys.* **119,** 3559 (2003).

Multicanonical (Wang-Landau) Sampling



Metropolis MC

$$\alpha(\underline{x}\to\underline{x}')\,P(\underline{x})\,w(\underline{x}\to\underline{x}') = \alpha(\underline{x}'\to\underline{x})\,P(\underline{x}')\,w(\underline{x}'\to\underline{x})$$

"Local moves"

$$\alpha(\underline{x}\to\underline{x}') = \min\left(1, \frac{P(\underline{x}')}{P(\underline{x})}\right) = \min\left(1, e^{-\beta\Delta U}\right)^{*}$$

WL is multicanonical ⇒ all microstates equivalent.

$$\alpha(\underline{x}\to\underline{x}') = \min\left(1, \frac{\tilde{g}(U(\underline{x}))}{\tilde{g}(U(\underline{x}'))}\right) \quad \text{then } \hat{g}\to f\hat{g}$$

$$\hat{g}(0)=1,\quad f(0)>1,\quad f\to 1 \text{ as } N_{steps}\to\infty$$

This algorithm converges $\tilde{g}(U)\to g(U)$, the D.O.S.

* WLOG, NVT

Multicanonical/Wang-Landau Method

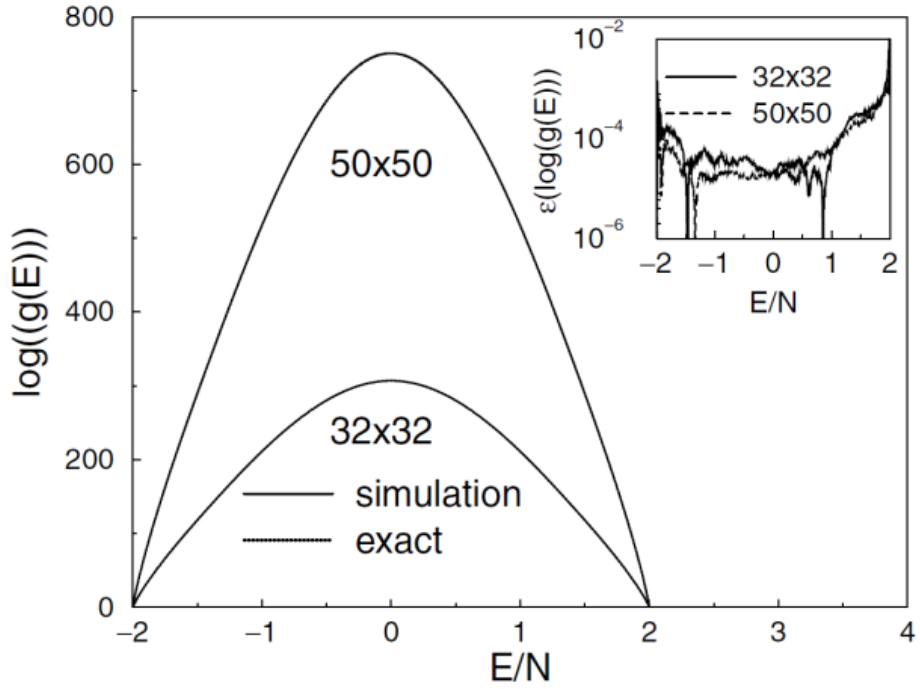


FIG. 1. Comparison of the density of states obtained by our algorithm for 2D Ising model and the exact results calculated by the method in Ref. [13]. Relative errors $\varepsilon(\log(g(E)))$ are shown in the inset.

1. Wang, F. & Landau, D. P. Determining the density of states for classical statistical models: A random walk algorithm to produce a flat histogram. *Phys. Rev. E* **64,** 056101 (2001).
2. Wang, F. & Landau, D. P. Efficient, Multiple-Range Random Walk Algorithm to Calculate the Density of States. *Phys. Rev. Lett.* **86,** 2050–2053 (2001).

Wang-Landau Sampling

MICROCANONICAL PROBABILITY

$$P(E) = 1/\Omega(N, V, E)$$

ACCEPTANCE RULE

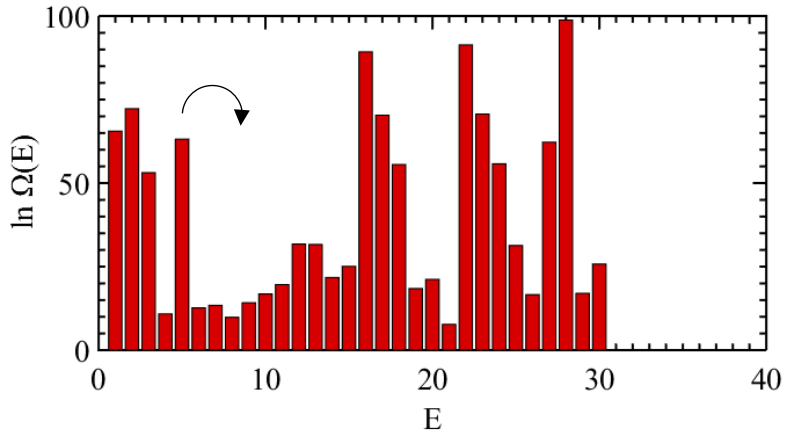
$$P_{acc}(i \to j) = \min\left\{1, \frac{\Omega(E_i)}{\Omega(E_j)}\right\}$$

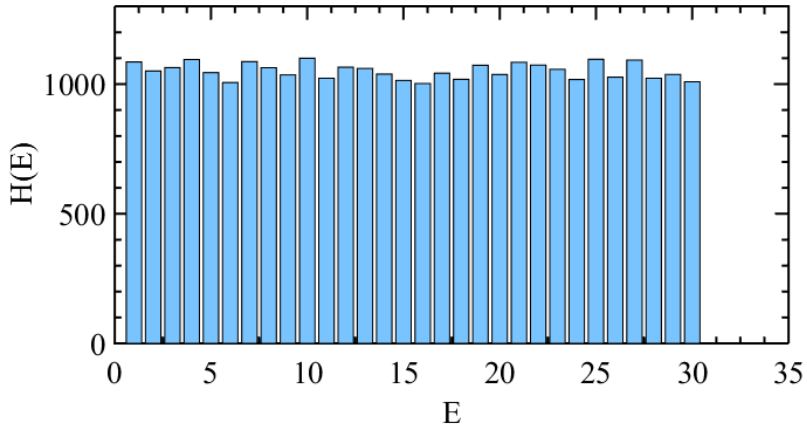
CONSTRUCTING THE ESTIMATOR

$$\mathrm{E} \in [E_{min}, E_{max}] \qquad N = 32$$

$$\Omega(E) = 1 \;\forall\; \mathrm{E} \qquad \mathrm{H}(E) = 0 \;\forall\; \mathrm{E}$$

THE ALGORITHM





RANDOM WALK IN PHASE SPACE

$$\Omega(E_{final}) = f \times \Omega(E_{final})$$

$$\mathrm{H}(E_{final}) = \mathrm{H}(E_{final}) + 1$$

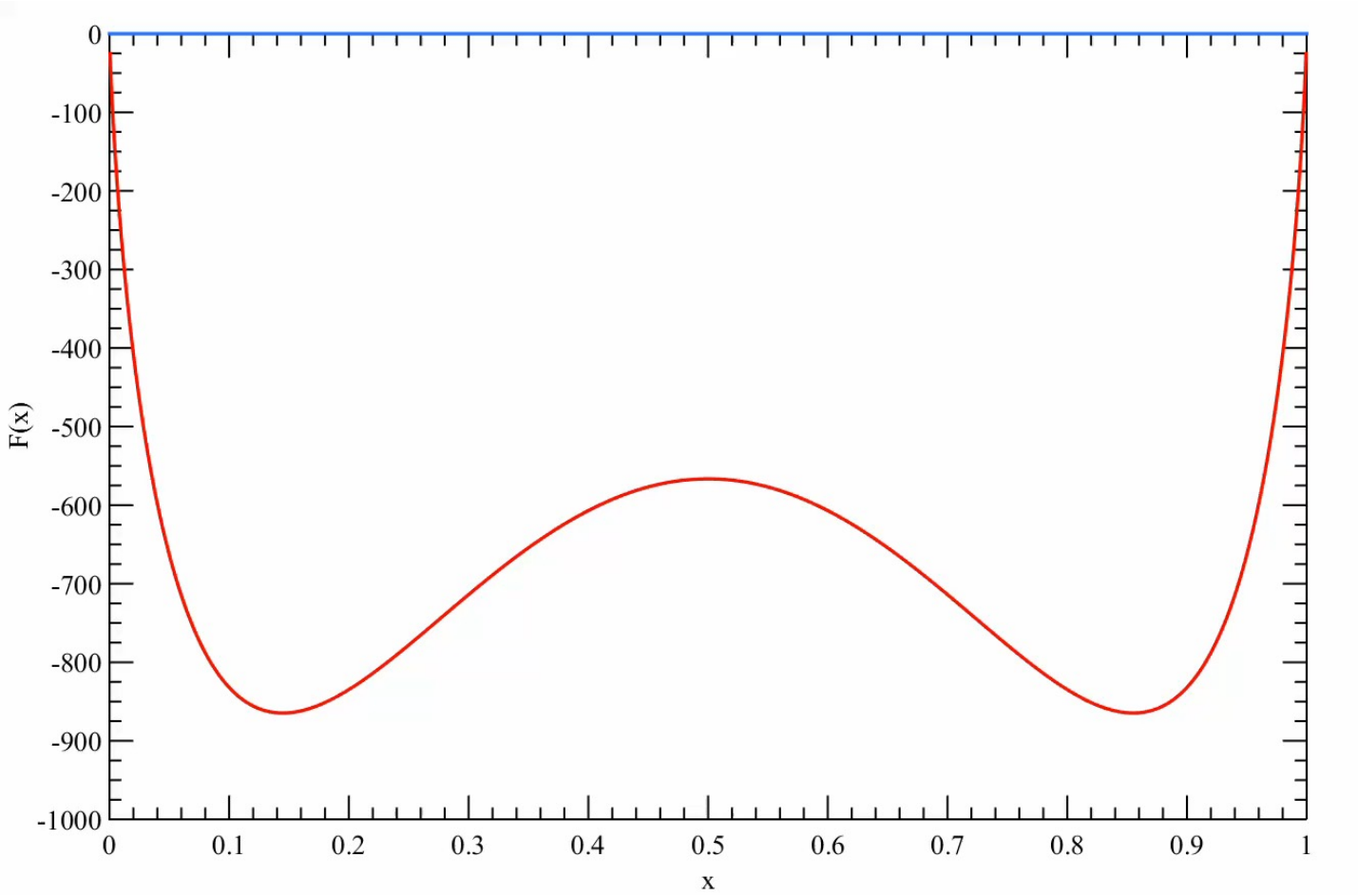
TERMINATION CRITERIA

$$f_{new} = \sqrt{f_{old}}$$

This process is repeated, refining the DOS until the convergence factor is very small (~$1x10^{-8}$)

Wang, F., & Landau, D. P. (2001). *Physical Review Letters*, *86*, 2050–2053.

Expanded Ensemble Density of States



GENERALIZED DOS

$$P_{acc}(i \to j) = \min\left\{1, \frac{\Omega(x_i)}{\Omega(x_j)} e^{-\beta \Delta E}\right\}$$

OBSERVATIONS

As bias accumulates, the walker can overcome free energy barriers

The estimate is discretized over an interval of interest

Most of the time is spent filling in regions of low probability

The estimator converges to the free energy $-\log \Omega(x) \to \beta F(x)$

Metadynamics

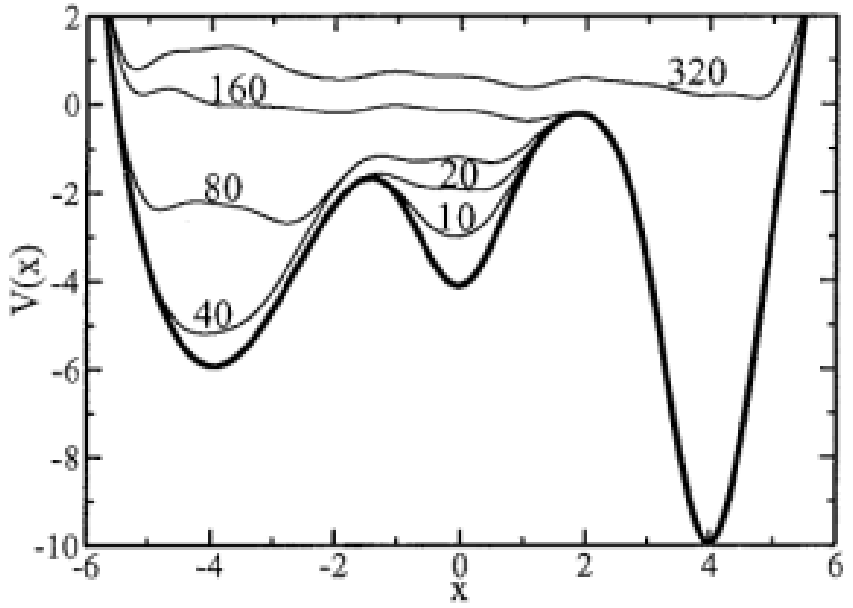
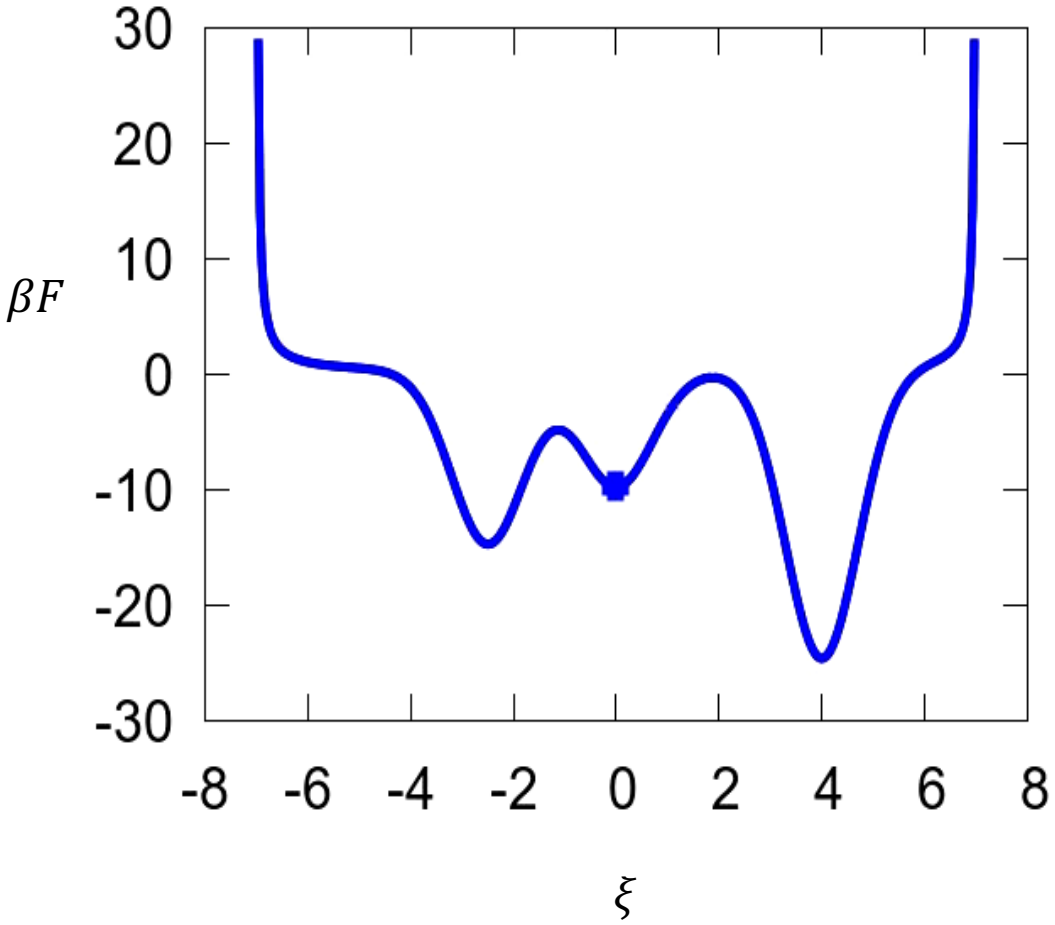


Fig. 1. Time evolution of the sum of a one-dimensional model potential $V(x)$ and the accumulating Gaussian terms of Eq. 2. The dynamic evolution (thin lines) is labeled by the number of dynamical iterations (Eq. 1). The starting potential (thick line) has three minima and the dynamics is initiated in the second minimum.

Laio, A. & Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 12562–12566 (2002).

$$V_{bias}(\xi,t) = \sum_{t_i<t} W(t_i) e^{-\beta \sum_i (s_i(x(t_i)) - \xi_i)^2 / 2\sigma_i^2}$$



Forces evaluated by chain rule: $F_\xi = -\frac{\partial V_{bias}}{\partial \xi} = -\frac{\partial \xi}{\partial x} \cdot \frac{\partial V_{bias}}{\partial \xi}$

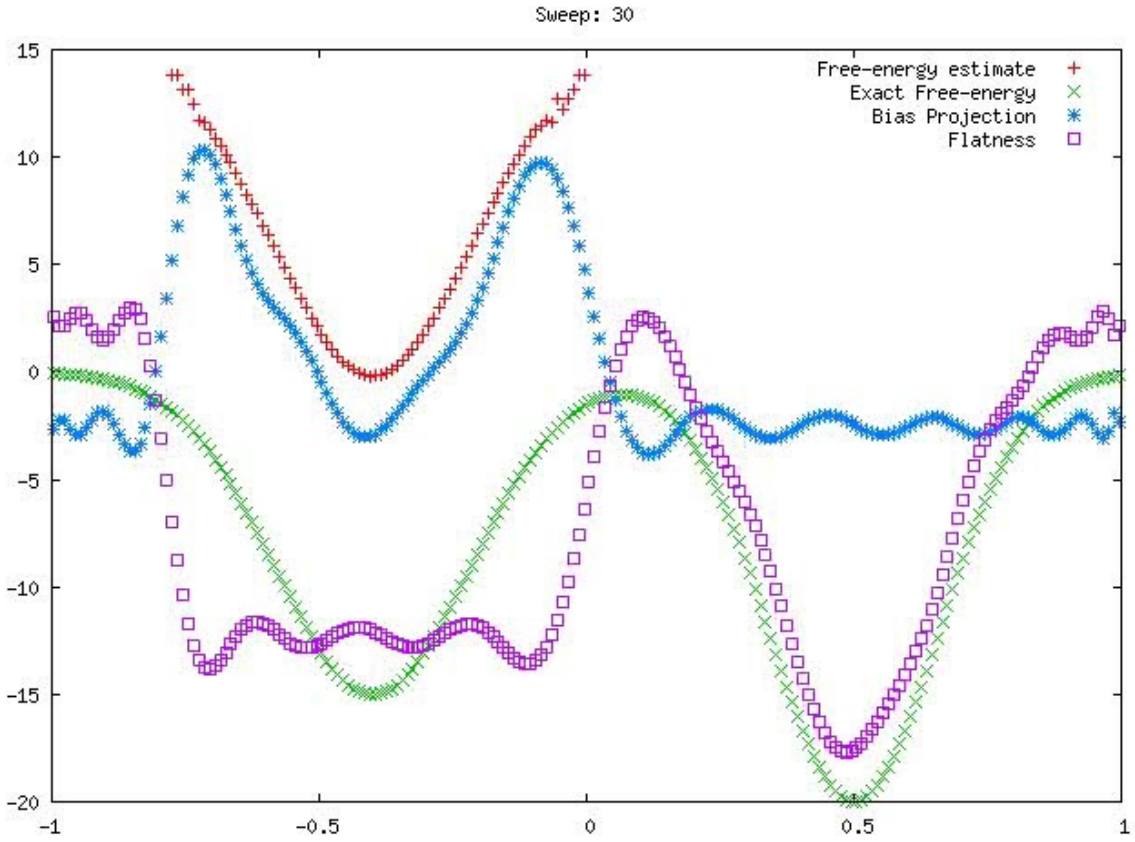
Jacobian

Basis Function Method

- Choose a domain Ξ on which to study the order parameter ξ.
- Choose a set of orthogonal functions $L_i(\xi)$ resolving this domain.
- During simulation, keep a histogram H_i of states visited during interval i
- Transform this into an unbiased histogram via the equation

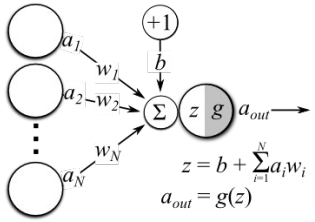
$$\widetilde{H}_i(\xi) = H_i(\xi)e^{\beta V_i(\xi)}$$

- Use this to create a running estimate of the partition function, $Z_{i+1} = \Sigma_{j \le i} W_j \widetilde{H}_j$, where the weight factor W_i is chosen to bias sampling toward times when the free energy surface is well resolved.
- Obtain $V_{i+1} = \Sigma_j \alpha_j L_j(\xi)$ by projection of $\log(Z_{i+1})$ onto basis function expansion



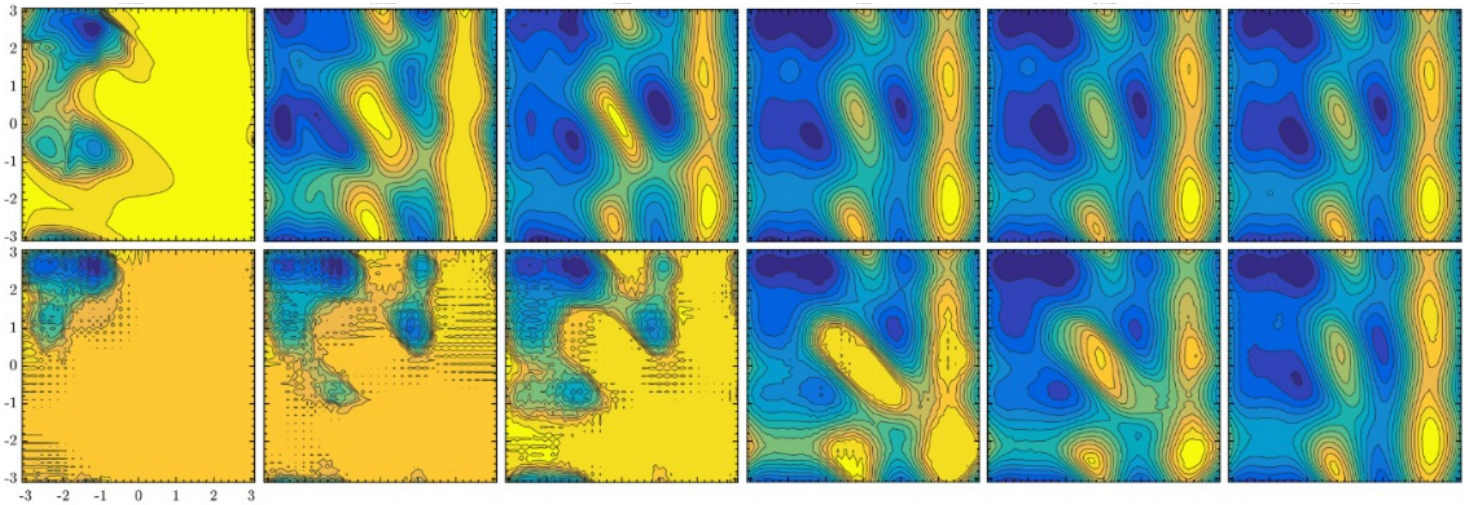
JK Whitmer, C Chiu, AA Joshi, JJ de Pablo, *PRL* **113** (19), 190602 (2014)

Neural Network Methods

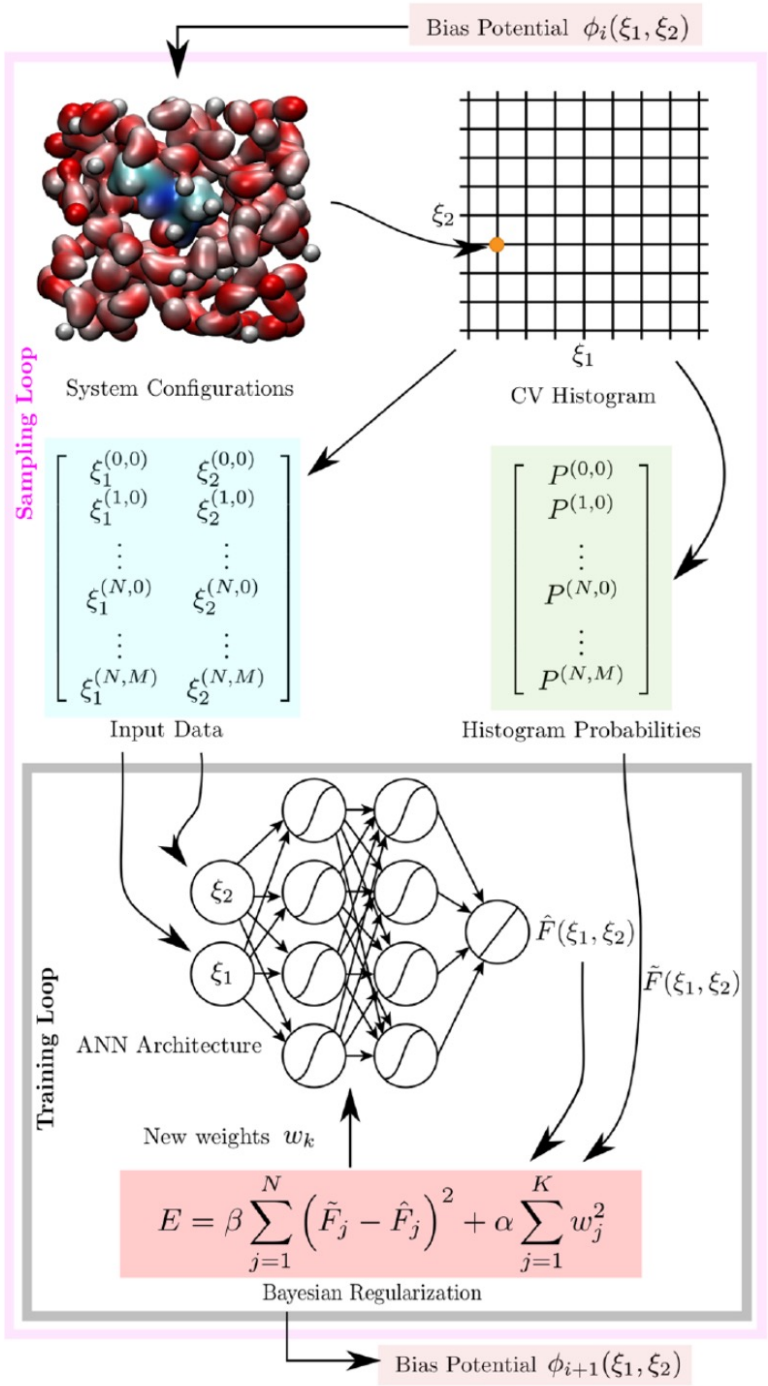


ANN

0.1 ns 0.5 ns 1.0 ns 5.0 ns 10 ns 100 ns



ABF



Sidky, H. & Whitmer, J. K. Learning free energy landscapes using artificial neural networks. *J. Chem. Phys.* **148,** 104111 (2018).

Adaptive Biasing Force

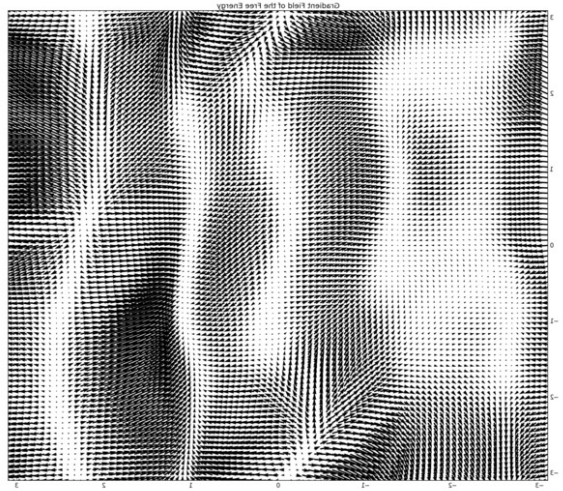
Evaluate the instantaneous force $\underline{F}_{\underline{\xi}}$

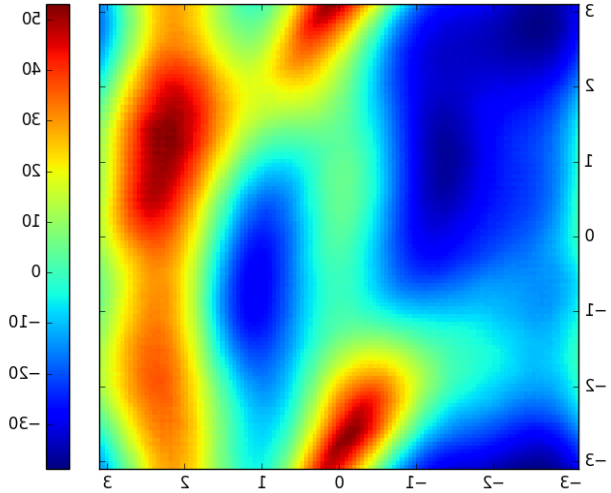
$$\frac{\partial F}{\partial \underline{\xi}} = \left\langle \frac{\partial U}{\partial \underline{\xi}} - \frac{1}{\beta} \frac{\partial (\ln(J))}{\partial \underline{\xi}} \right\rangle_{\xi} = -\langle F_{\xi} \rangle$$

derivative of free energy

"mean force"

Applying $\underline{F}_{bias} = -\langle \underline{F}_{\underline{\xi}} \rangle$ results in an n-dimensional random walk.*

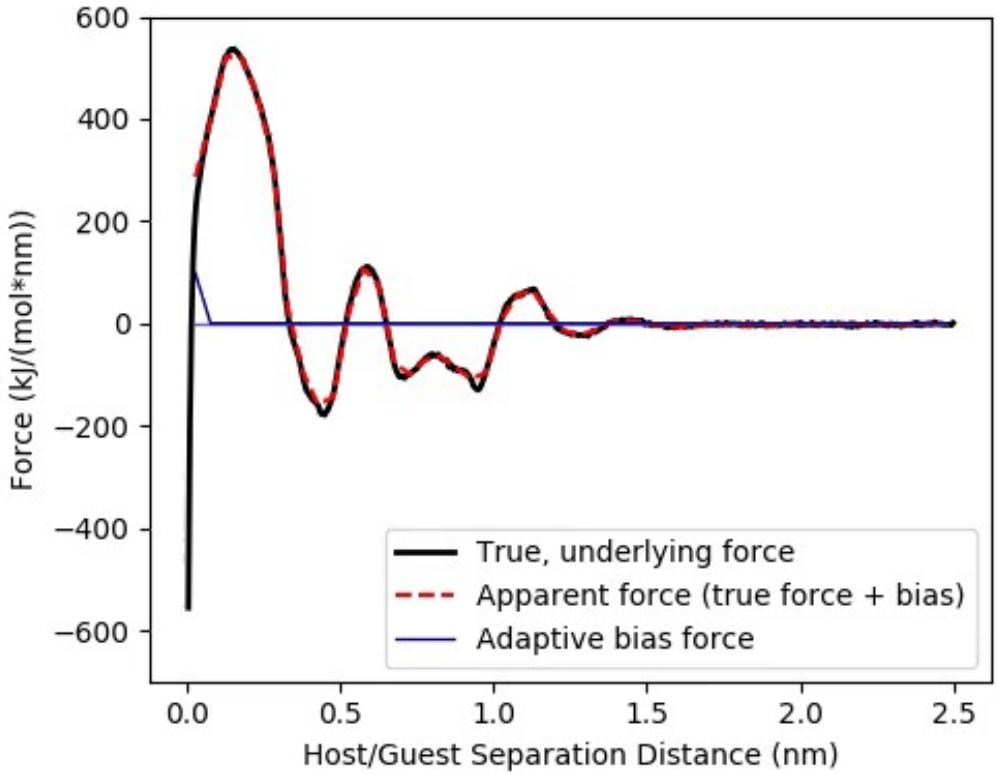


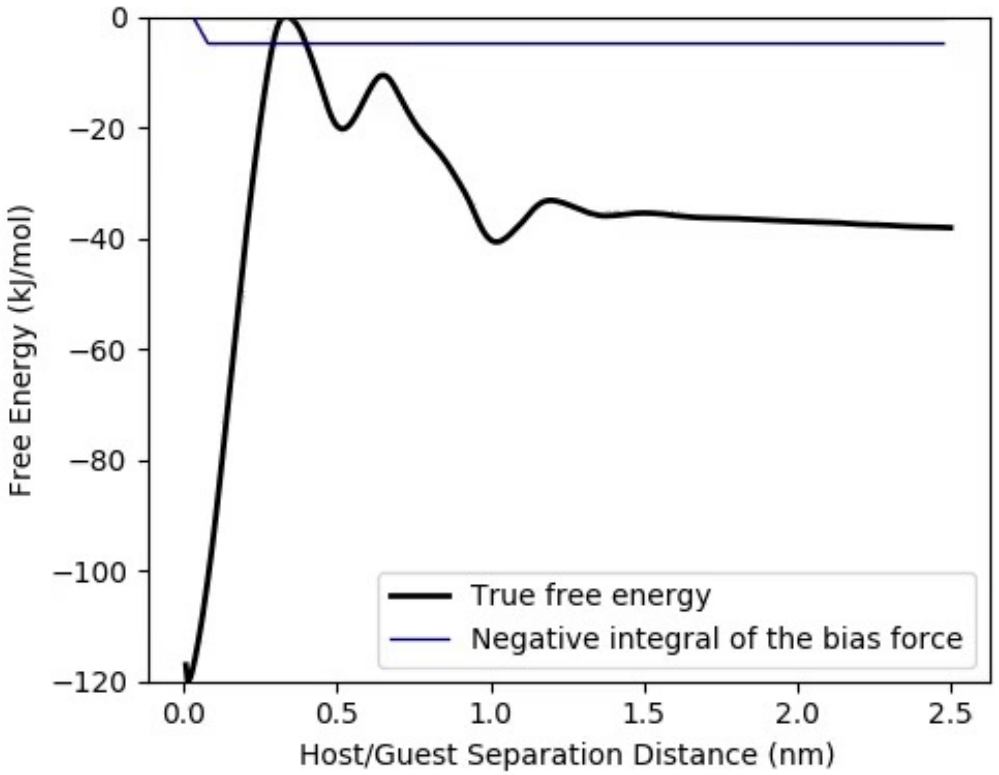


* subject to sufficient statistics**

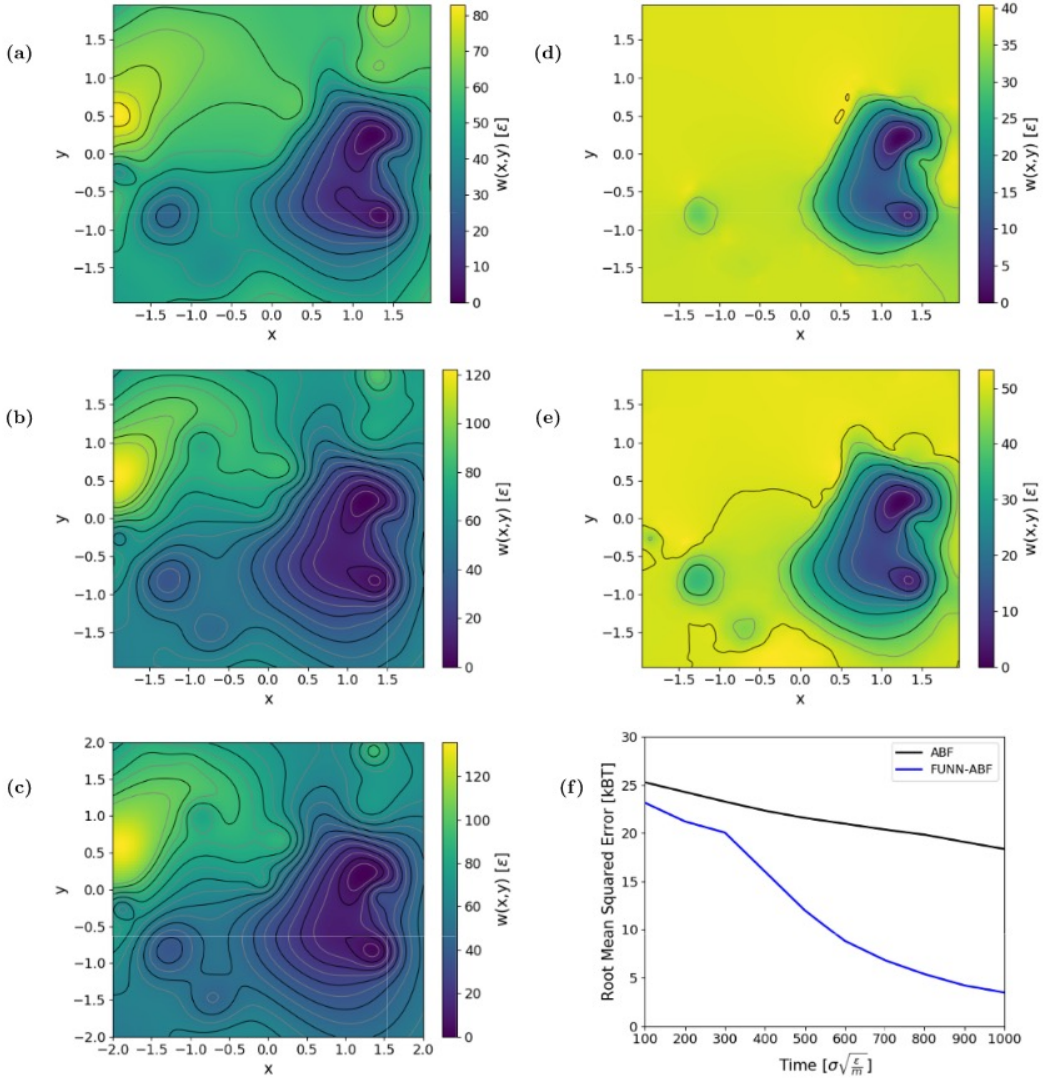
** And pathological cases.

Adaptive Biasing Force





FUNN

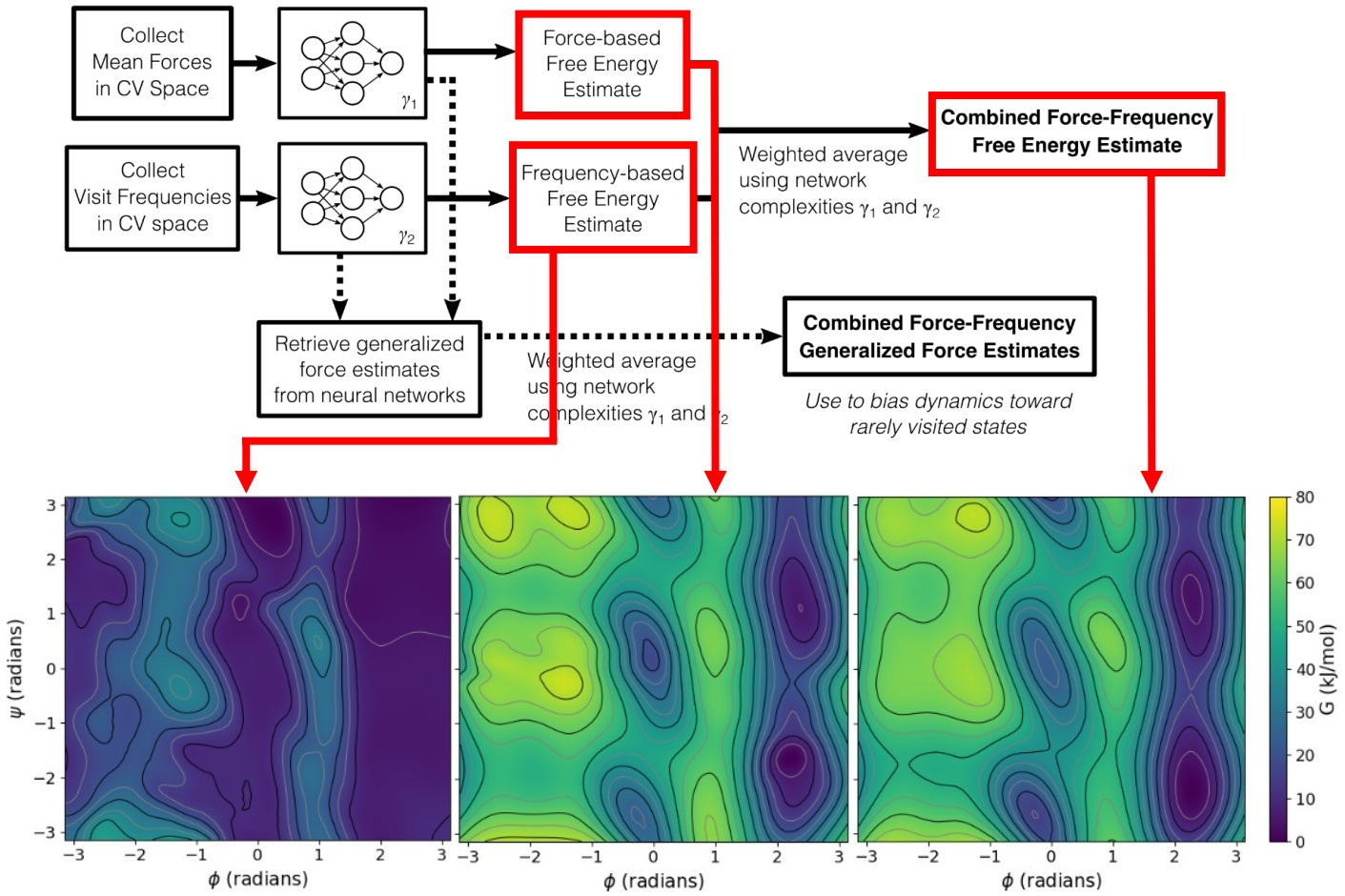


ABF w/ ANNs !

- superior speed from interpolation
- fast scaling peaks from ABF
- Factor of 100x speed improvement on canonical test systems.

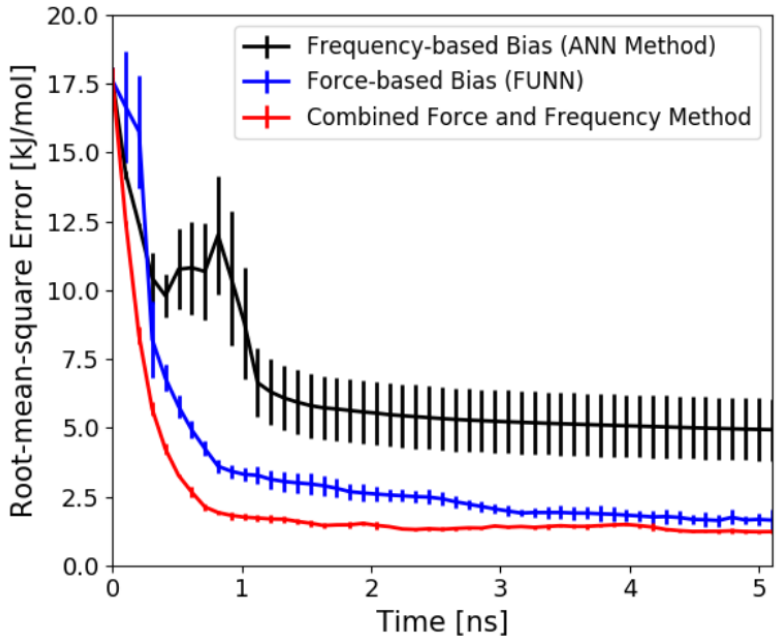
Guo, A. Z. *et al.* Adaptive enhanced sampling by force-biasing using neural networks. *J. Chem. Phys.* **148,** (2018).

The next generation of NN Biases



E.Sevgen, J.K.Whitmer, J.J. de Pablo, et al. *J. Chem. Theory Comput.* **2020**

- Neural Network methods applied to frequency (ANN Method) and estimated mean forces (FUNN)
- Combined Force Frequency (CFF) balances both to obtain the swiftest convergence possible.



Minimum FEP Methods

Transition Path Sampling

have trajectory $\Xi(\Gamma) = \{\underline{x}(0), \underline{x}(\delta t), \ldots, \underline{x}(\Gamma)\}$

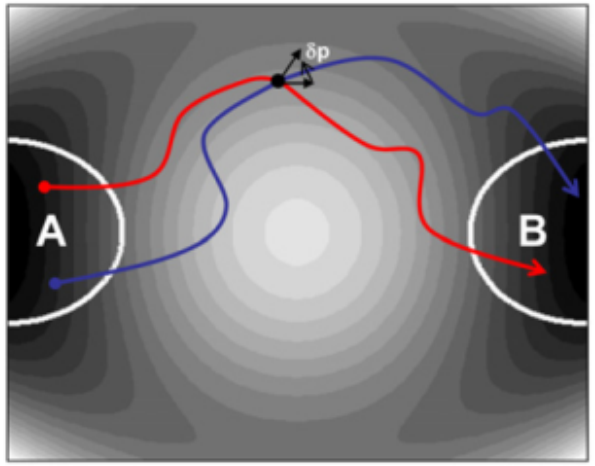
pluck $\underline{x}(t_i)$, perturb.

evolve backward ← → evolve forward

$\Xi'(\Gamma) = \{\underline{x}'(0), \ldots \underline{x}'(t_i), \ldots, \underline{x}'(\Gamma)\}$

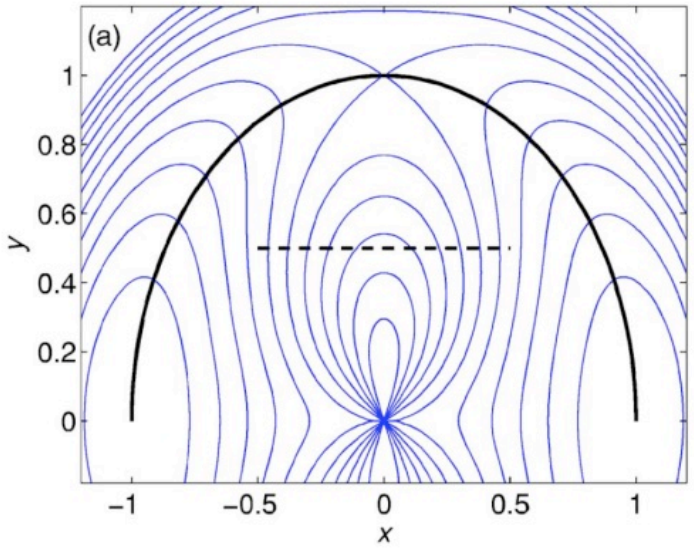
$$P_{acc}(\Xi \to \Xi') = h_A(\underline{x}'(0)) h_B(x'(\Gamma)) \min\left(1, P(\underline{x}'(t_i))/P(\underline{x}(t_i))\right)$$

generates independent paths, reaction rate information from correlation fcns.



Escobedo, F. A., Borrero, E. E. & Araque, J. C. Transition path sampling and forward flux sam
Applications to biological systems. *J. Phys. Condens. Matter* **21,** 333101 (2009).

Zero-Temperature String



Start with a path loosely connecting basins of interest.

$$\{\underline{\xi}\} = \{\underline{\xi}_1, \underline{\xi}_2, \ldots, \underline{\xi}_n\}$$

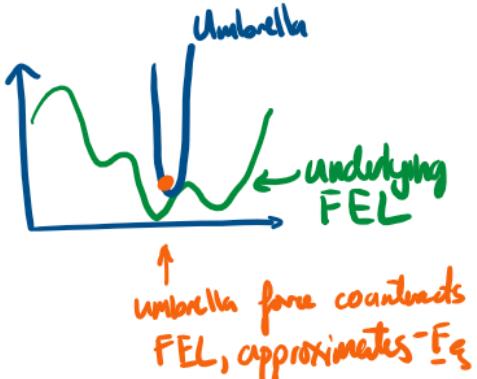
Estimate $\underline{F}_{\underline{\xi}}$ using (e.g.) stiff umbrella sampling

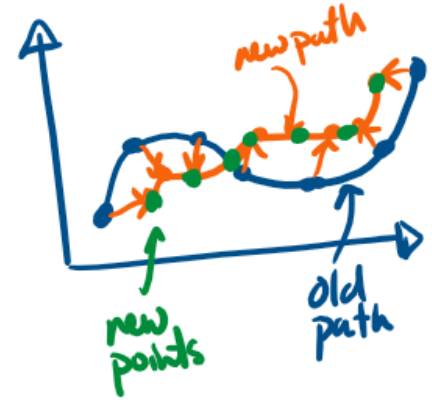
Integrate: $\underline{\xi}_i(t+\delta t) = \underline{\xi}_i(t) + \delta t\, \underline{F}_{\underline{\xi}_i}(t)$

"overdamped"

Note this minimizes the FE by simply following the gradient.

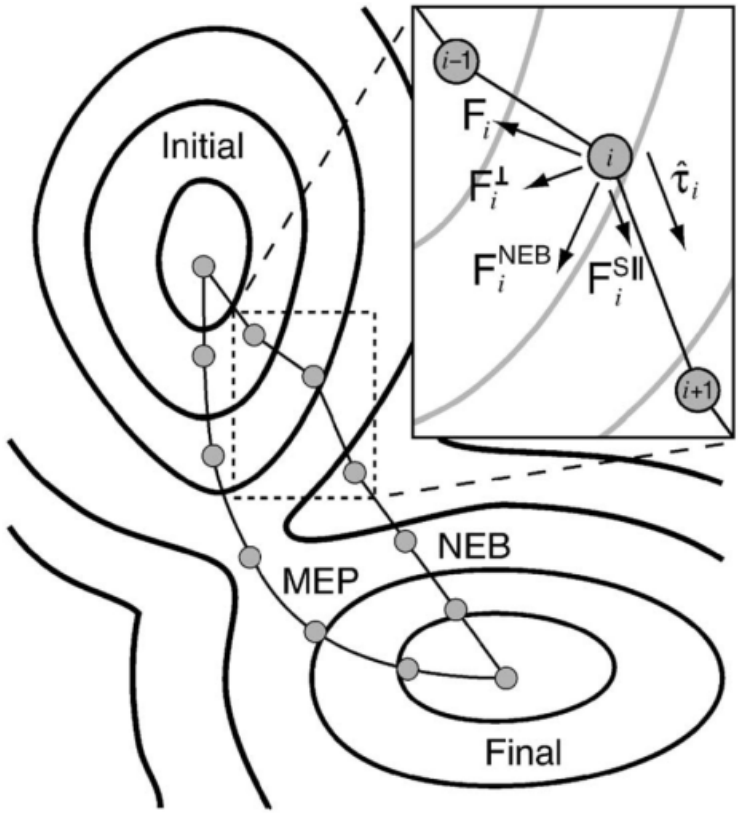
$\underline{\xi}_i$ are commonly redrawn after each step by spline

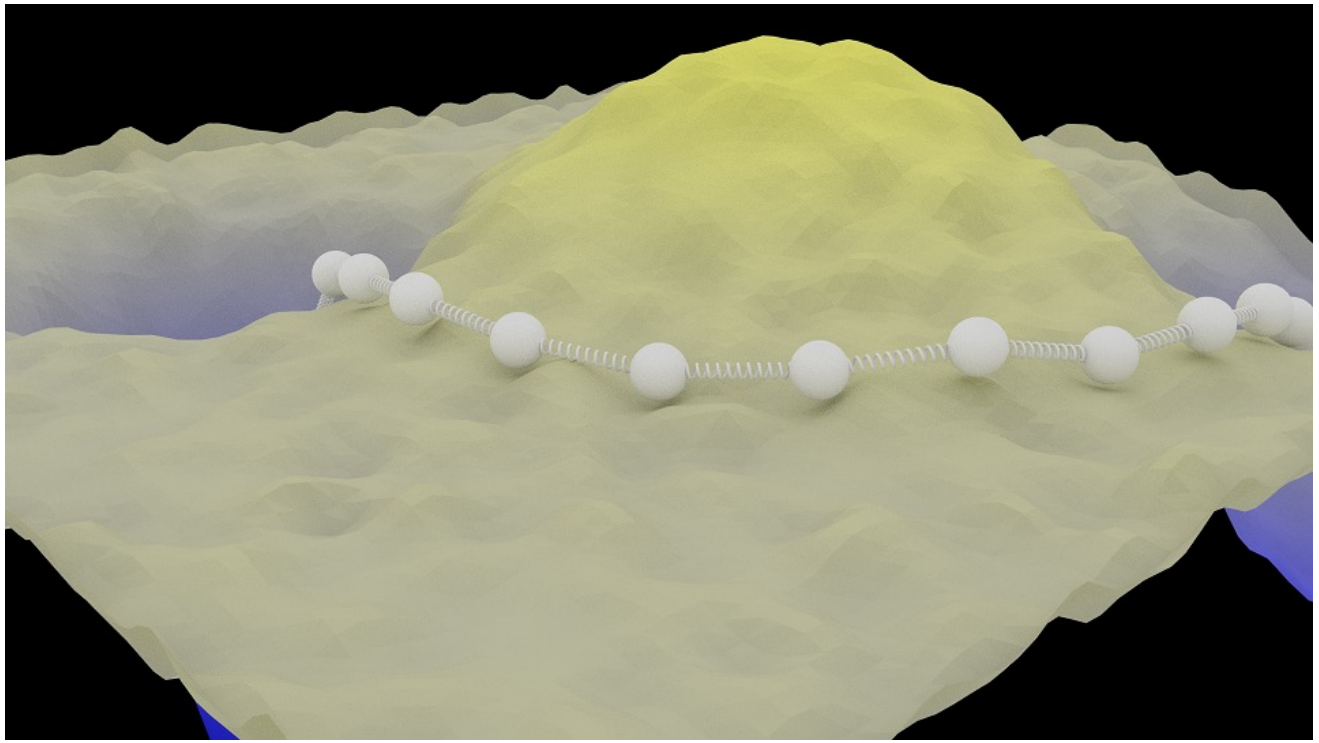




E, W., Ren, W. & Vanden-Eijnden, E. Simplified and improved string method for computing the minimum energy paths in barrier-crossing events. *J. Chem. Phys.* **126,** 164103 (2007).

Elastic Band



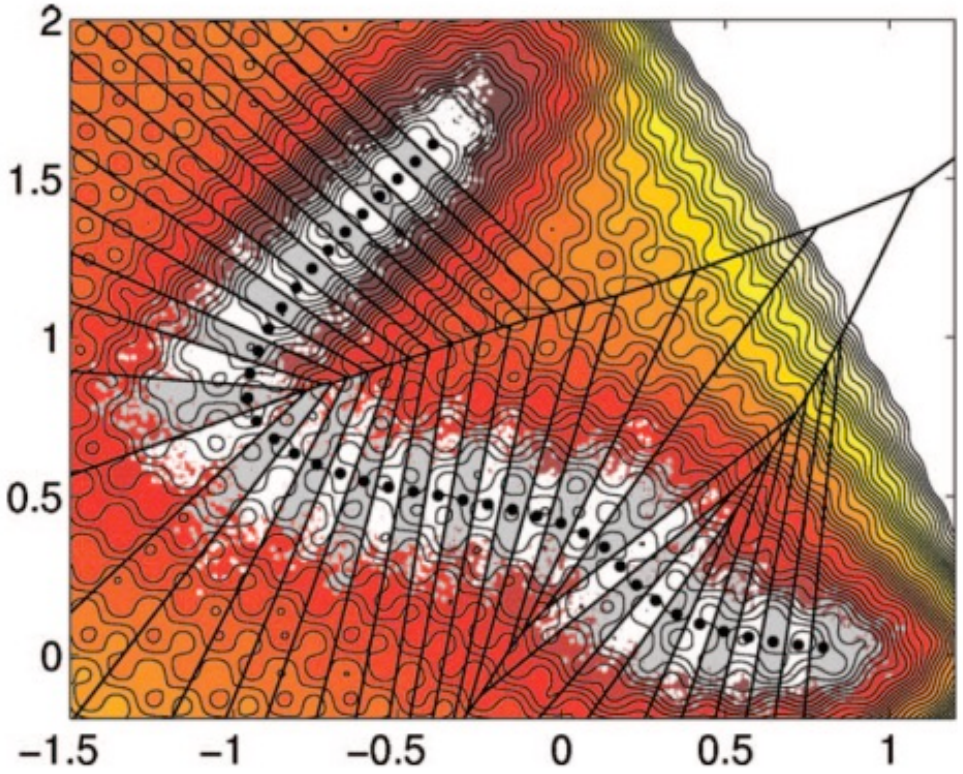


Similar, but $\underline{F}_\xi$ includes spring forces. inhibit separation along the chain.

$$\underline{\xi}_i(t+\delta t) = \underline{\xi}_i(t) + \delta t \left(\underline{F}_{s,\parallel} + \underline{F}_{\xi_i}^{\perp} \right)$$

Sheppard, D., Terrell, R. & Henkelman, G. Optimization methods for finding minimum energy paths. *J. Chem. Phys.* **128,** 134106 (2008).

Finite Temperature String



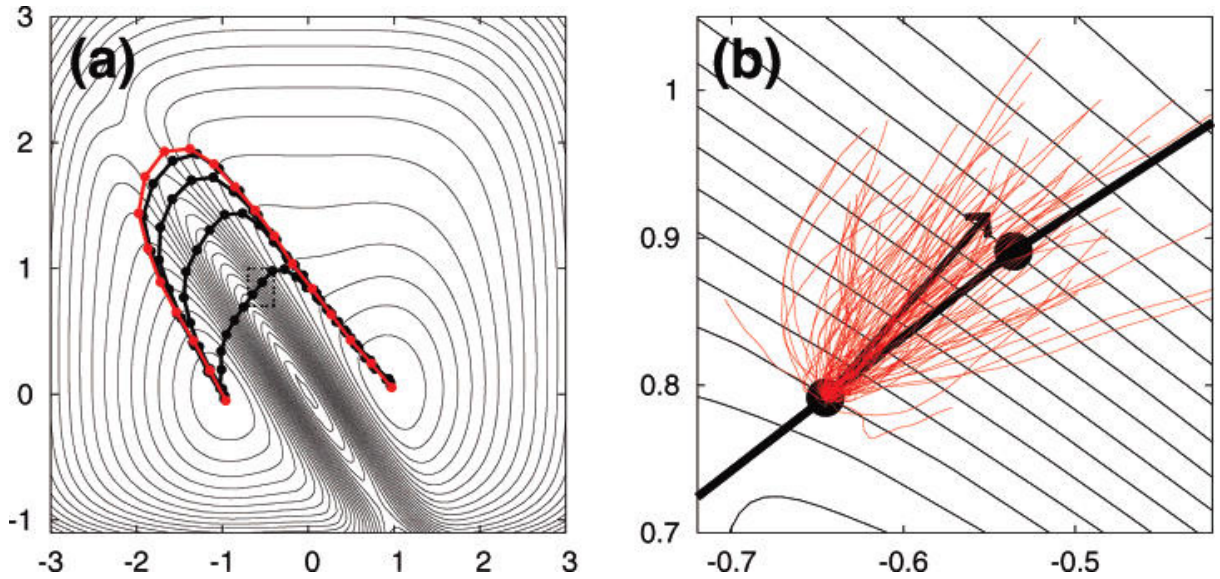
Break up space as Voronoi cells. Compute new center from sampling constrained to that cell.

$$\underline{\xi}(t+\Delta t) = \frac{\delta t}{\Delta t} \sum_n \underline{\xi}_i(t+n\delta t)$$

Redraw Voronoi cells & iterate until convergence.

Vanden-Eijnden, E. & Venturoli, M. Revisiting the finite temperature string method for the calculation of reaction tubes and free energies. *J. Chem. Phys.* **130,** 194103 (2009).

Swarm of Trajectories



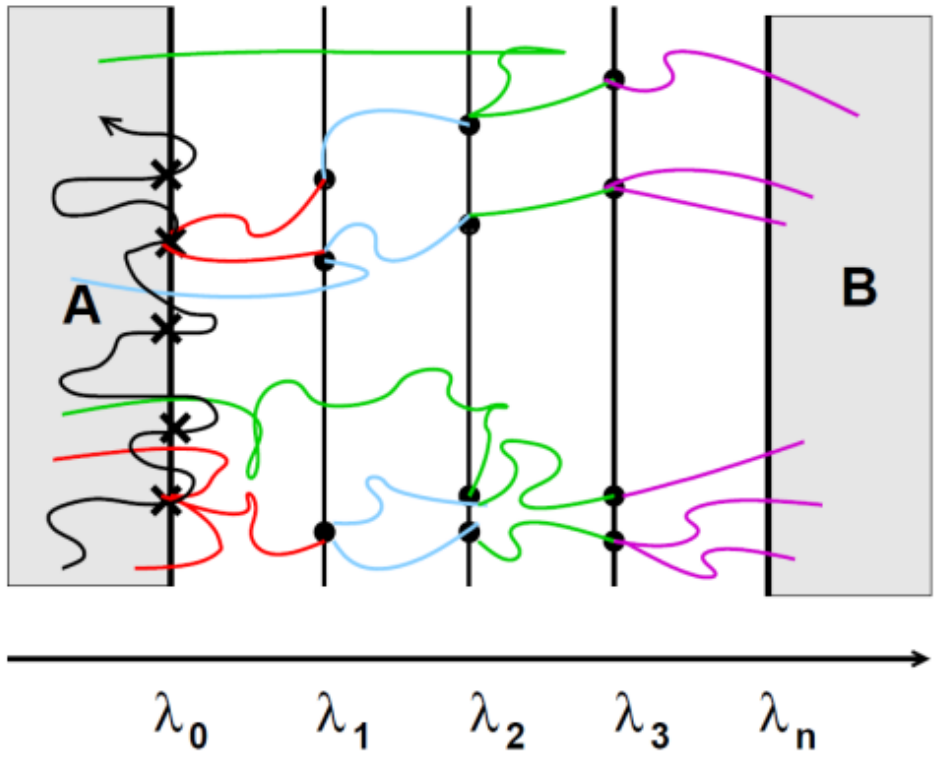
Calculate the average drift in ξ_i from a swarm of short trajectories.

Update the string in the direction of that gradient.

Pan, A. C., Sezer, D. & Roux, B. Finding Transition Pathways Using the String Method with Swarms of Trajectories. *J. Phys. Chem. B* **112,** 3432–3440 (2008).

Reactive Path Methods

Forward Flux



Grow trajectories by breaking space into interfaces, building on past successes.

$$P(\lambda_{i+1} | \lambda_i) = \frac{\#\text{ successes from } i}{\#\text{ tries from } i}$$

Can be used to generate initial paths or calculate reaction rates.

$$k = k_0 \prod_{i=0}^{n-1} P(\lambda_{i+1} | \lambda_i)$$

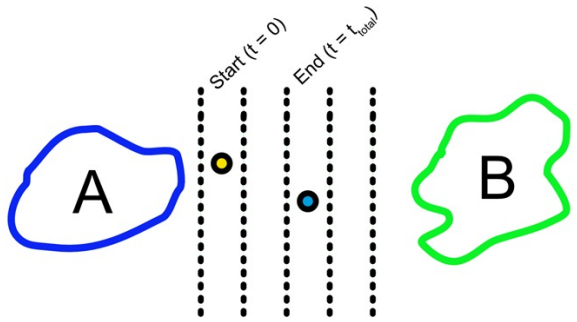
↑ reading flux at 1st interface

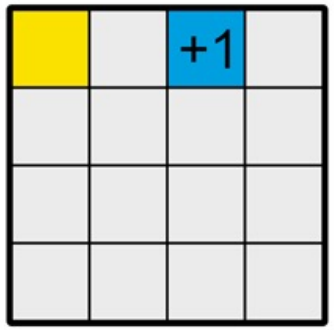
Allen, R. J., Valeriani, C. & Rein Ten Wolde, P. Forward flux sampling for rare event simulations. *J. Physics. Condens. matter* **21,** 463102 (2009).

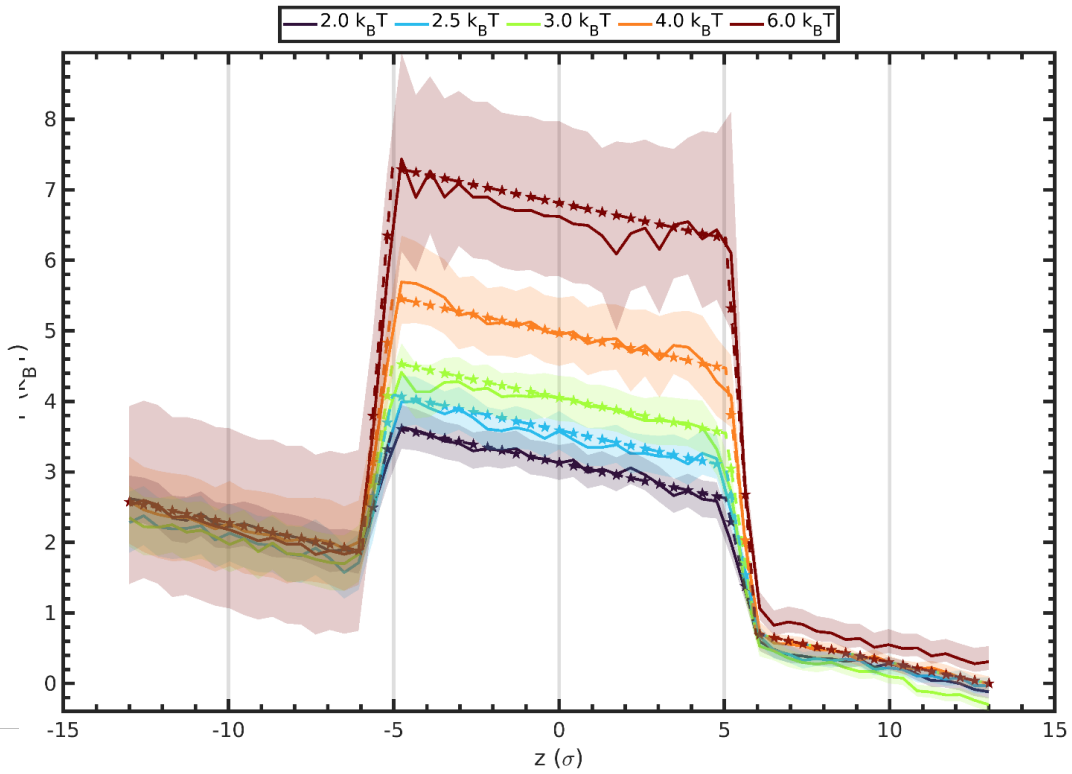
Nonequilibrium Trajectory Sampling

- Short runs are initiated between two basins of interest
- Outcome is recorded in a count matrix that estimates global transition rates.
- Steady states and “free energies” are obtained from the nullspace of the matrix

$$\frac{\partial P_i}{\partial t} = \sum_j k_{ij} \cdot P_j := 0$$

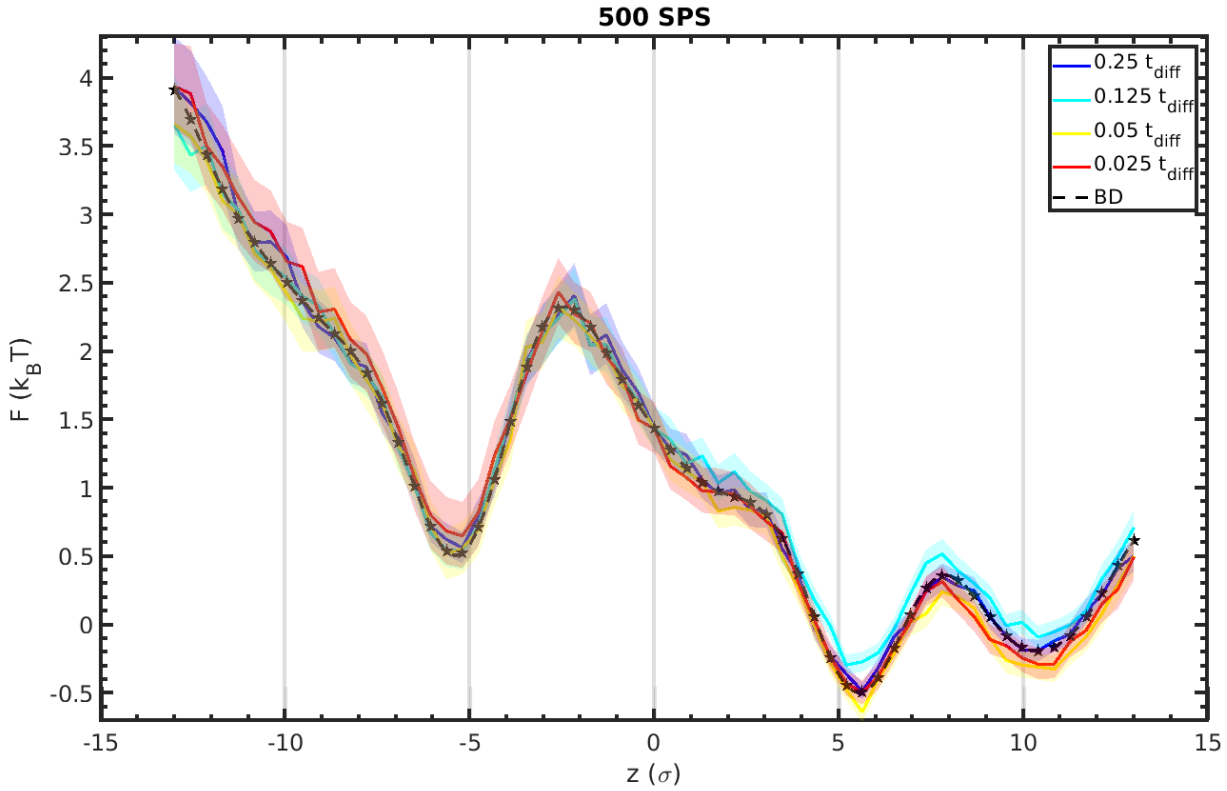


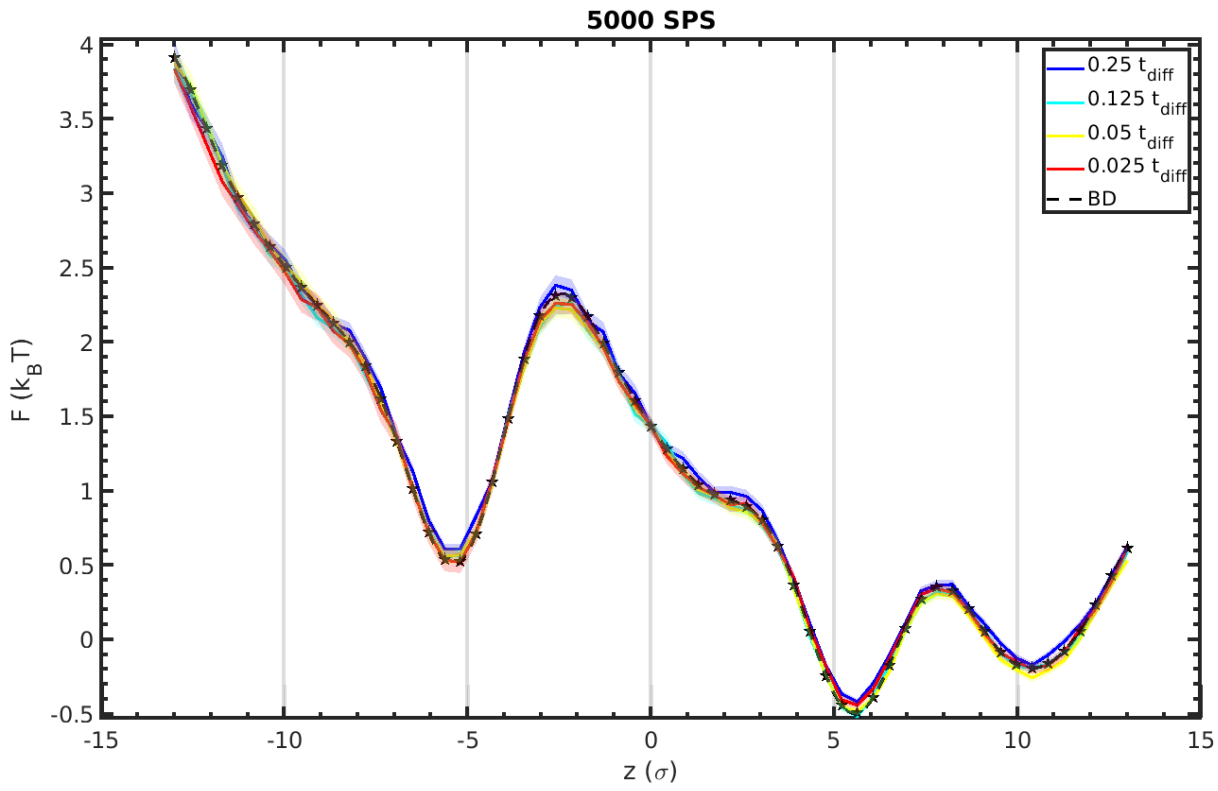




P.H. Amorim Valenca, W.A. Phillip, W. A., J.K. Whitmer, submitted (2022)

NETS is Well-Suited to Exploring New Landscapes





P.H. Amorim Valenca, W.A. Phillip, W. A., J.K. Whitmer, submitted (2021)

SSAGES/PySAGES

Key Capabilities: SSAGES/PySAGES

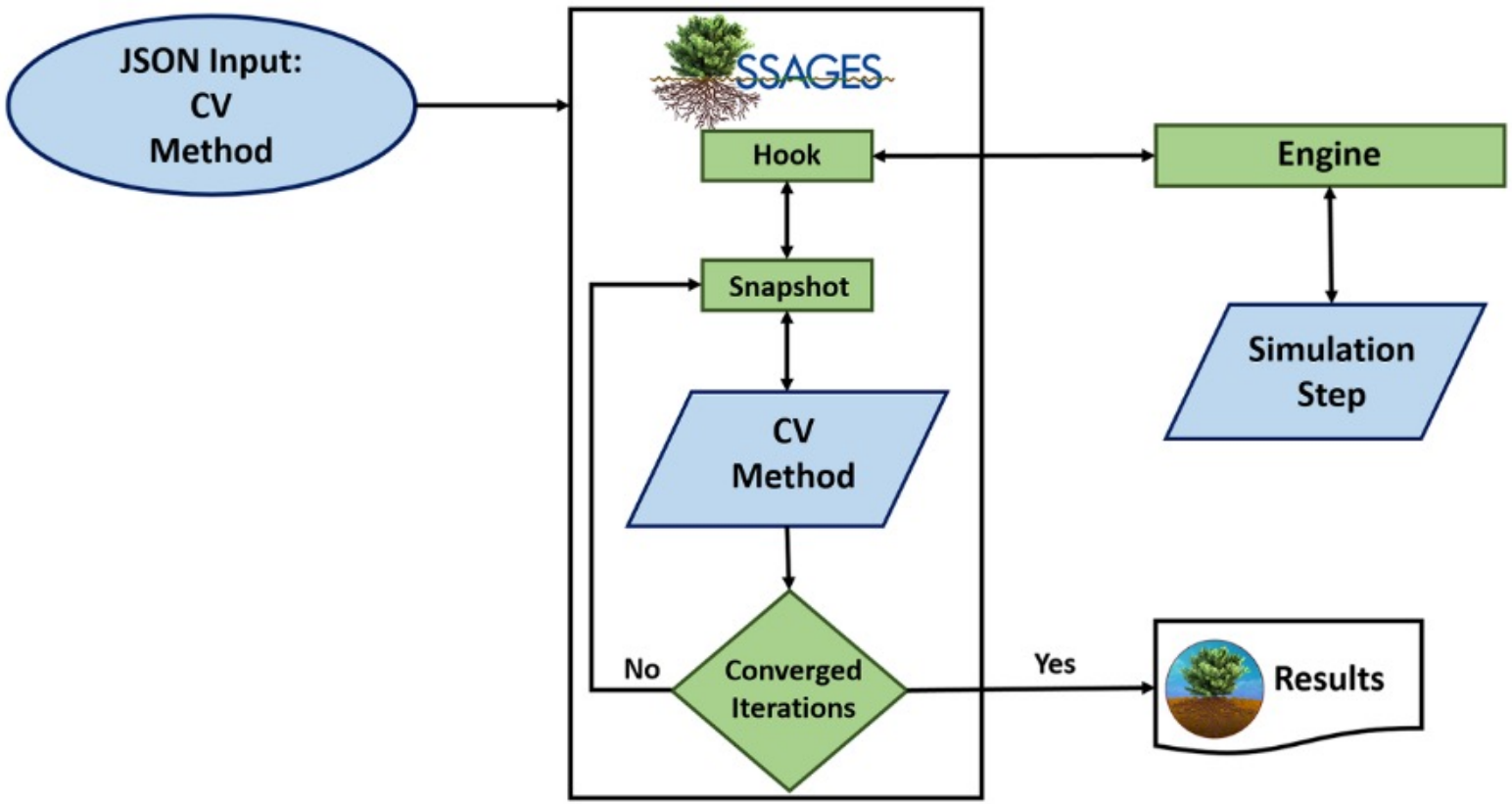
- SSAGES contains a number of methods in several classes to obtain reactive paths and free energy landscapes (FELs).
  - Reactive Pathways (e.g. FTS)
  - "Flat Histogram" (e.g. Metadynamics)
  - Flux Methods (FFS)
- SSAGES is the ***only*** open-source code to implement some methods.
- Sampling along standard collective variables (CVs) is implemented, and methods are extensible to bespoke CVs.
- SSAGES couples to six MD codes,* including:
  - LAMMPS
  - GROMACS
  - OpenMD
  - Qbox
  - HOOMD-blue
- SSAGES would benefit from expanding the number of supported couplings.
  - PySAGES is extending these algorithms (and couplings*) into the GPU-accelerated space.

Method	Software
Adaptive Biasing Force	**S**, Pl, C
ANN/FUNN/CFF/Sobolev Sampling	**S**
Basis Function Sampling/Green's Function Sampling	**S/*S***
Extended Lagrangian	Pl
Forward Flux Sampling	**S**
Maximum Entropy Bias	Pl
Metadynamics	**S***, Pl*, C*, H*
Nudged Elastic Band/String Methods	**S**, *T*
Parallel Bias Metadynamics	***S***, Pl
Replica Exchange	Pl*, L, G, C, Py, O, ***S***
Steered Molecular Dynamics/Moving Restraint Methods	**S**, Pl*, C
Swarm of Trajectories	**S**
Temperature Accelerated/Log Mean Force Dynamics	Pl*
Transition Path Sampling	***S***, O
Transition Interface Sampling	Py, O
Umbrella Sampling	**S,** Pl, C
Variationally Enhanced Sampling	Pl*

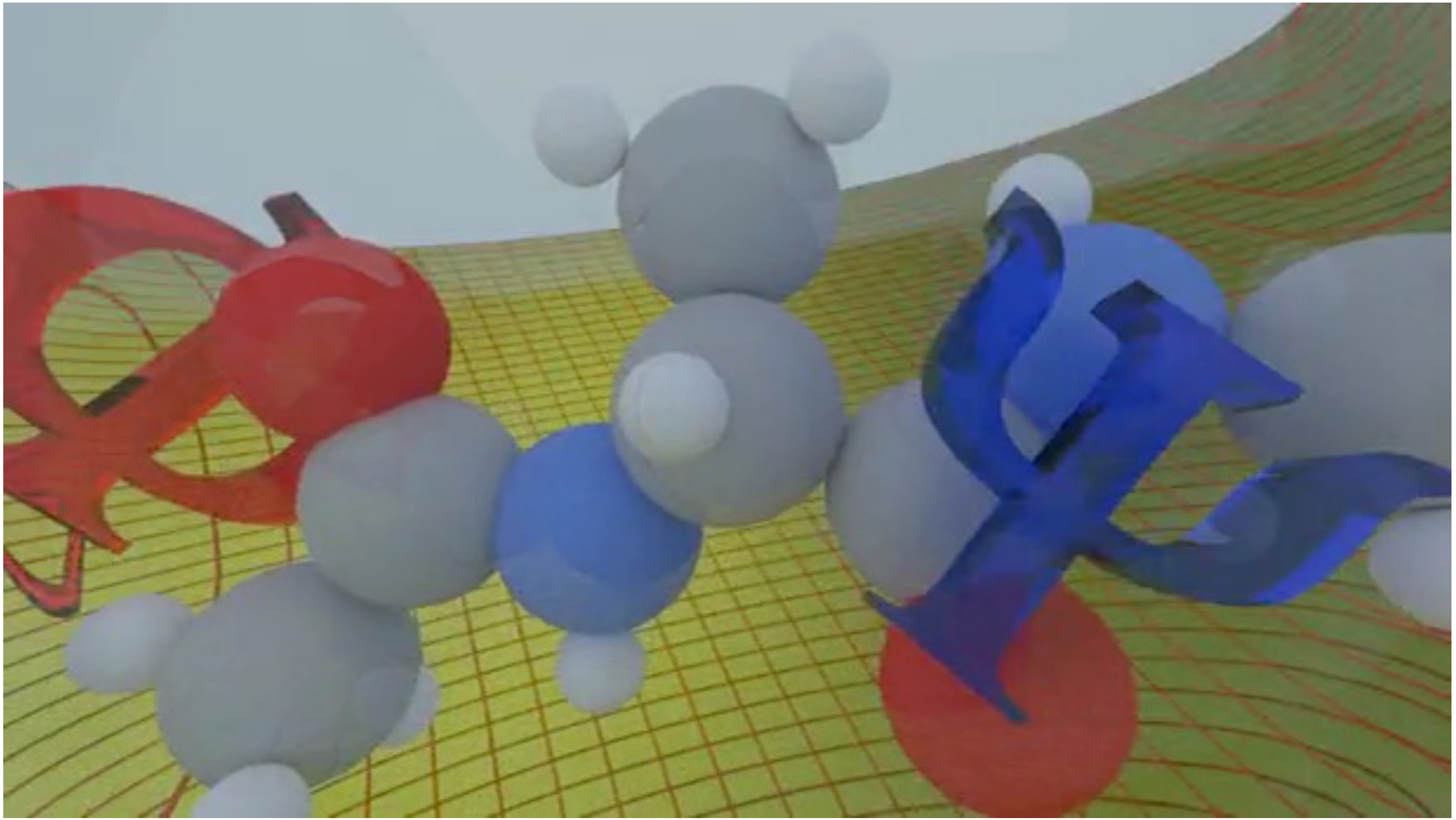
S=SSAGES; Pl=Plumed; L=LAMMPS; G=GROMACS; C=Colvars; Py=Pyretis; O=Open Path Sampling; T=Transition State Tools; H=HOOMD

*Italics for development versions/user plugins only. *Contains multiple variants of the algorithm in release version. Green: Contains a variant with the same essential properties.*

How Does SSAGES Work?

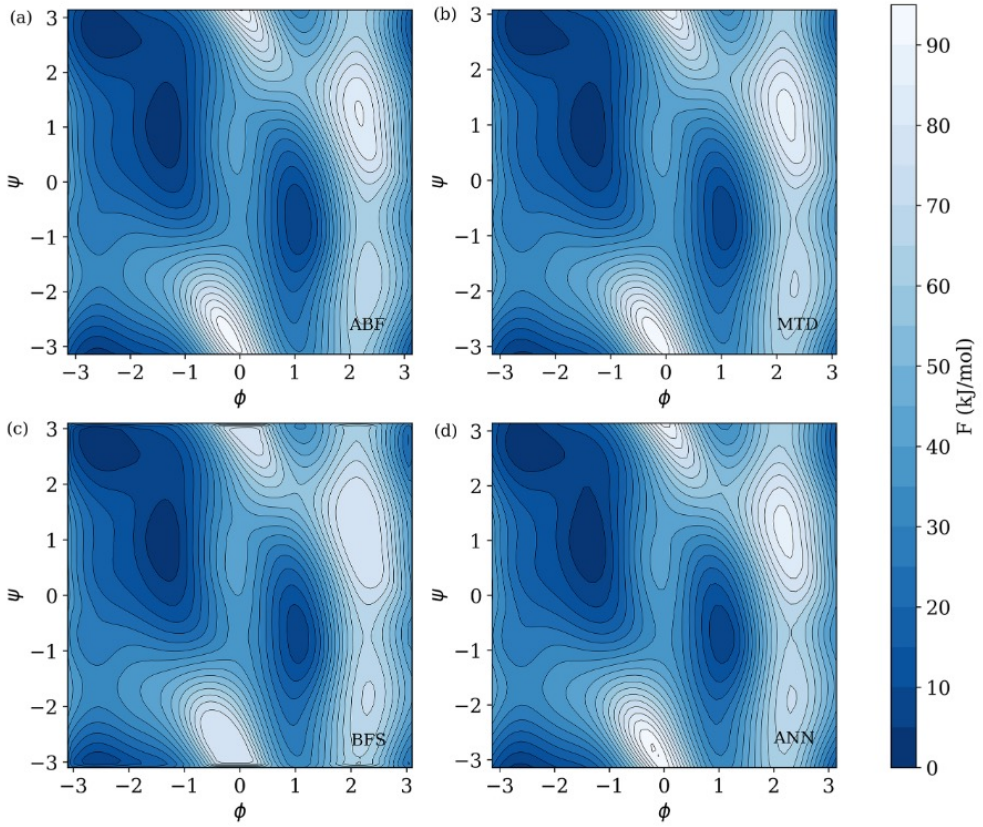


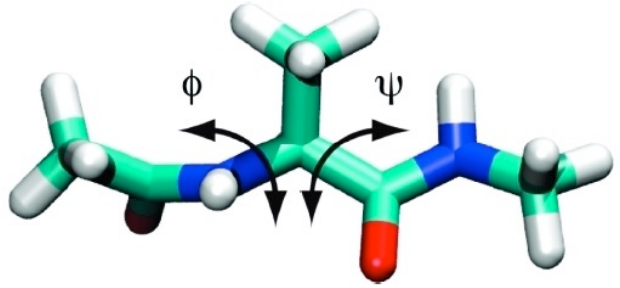
Sidky, H. *et al.* SSAGES: Software Suite for Advanced General Ensemble Simulations. *J. Chem. Phys.* **148,** (2018).

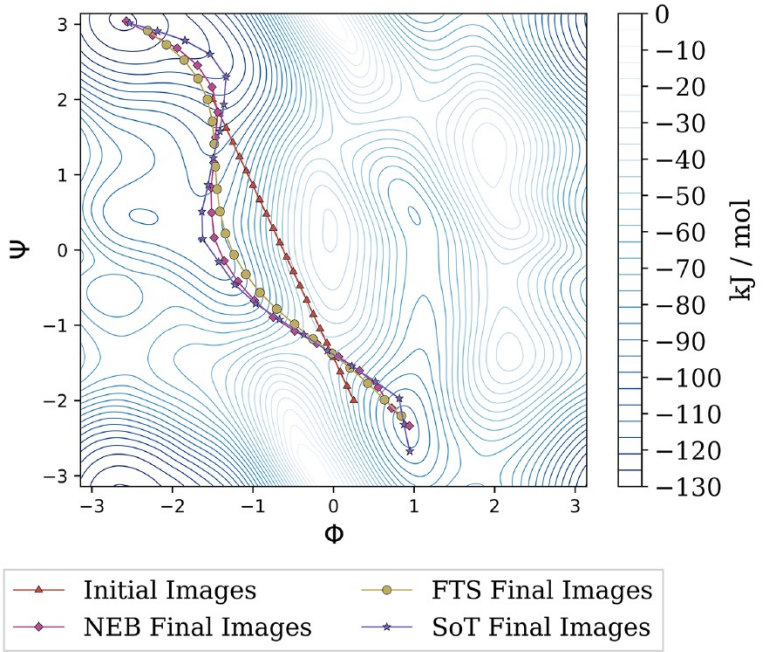


Credit: B. Sikora.

Free Energy Landscape and Transition Pathway

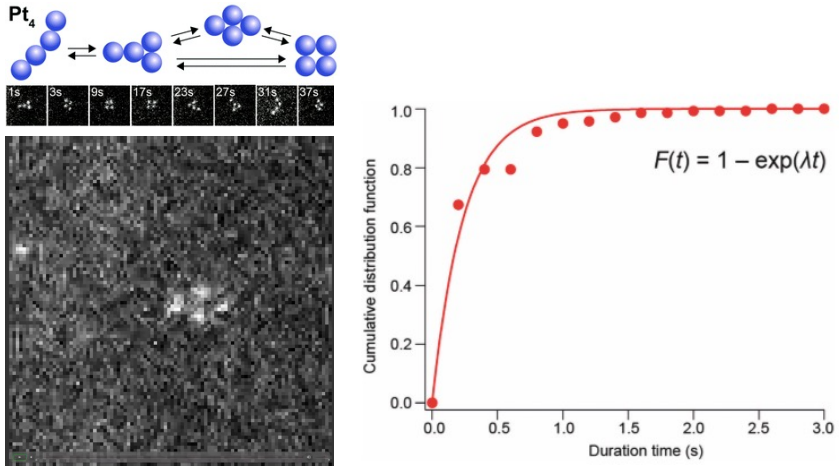




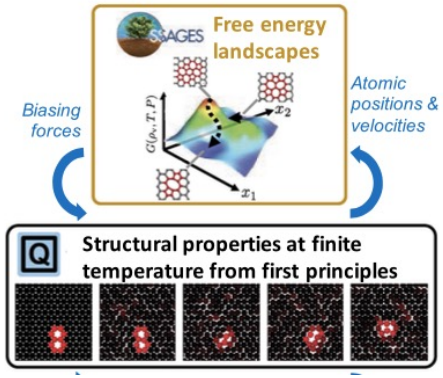


Sidky, H. *et al.* SSAGES: Software Suite for Advanced General Ensemble Simulations. *J. Chem. Phys.* **148,** (2018).

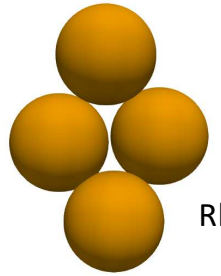
Case Study: Gold Cluster Morphologies



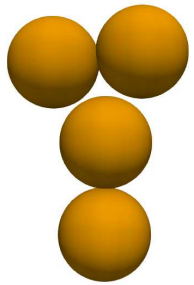
T. Imaoka et al. Chem. Commun., **55**, 4753 (2019)



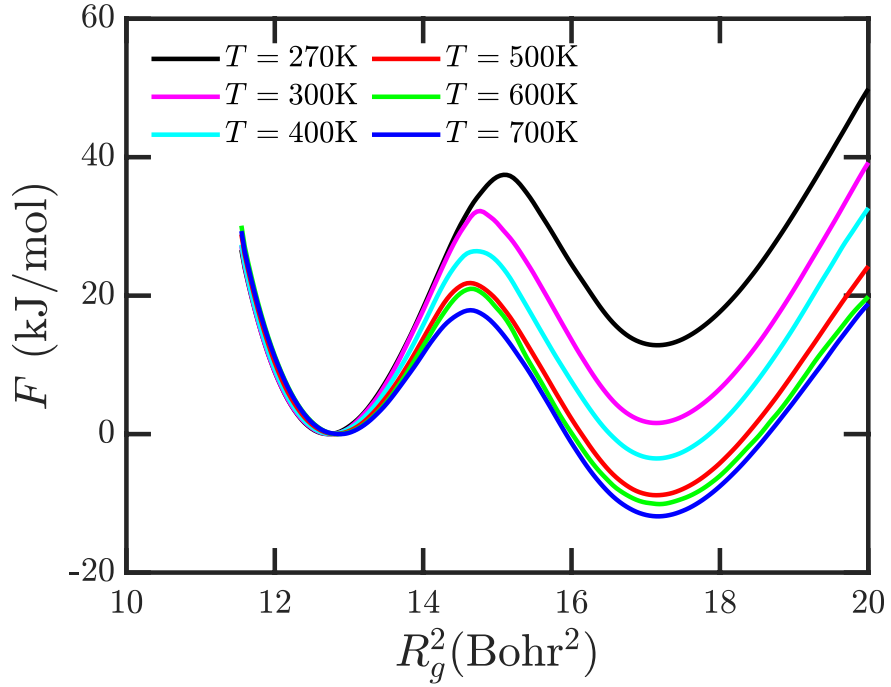
Govoni, Whitmer, de Pablo, Gygi, and Galli.
npj Comput. Mater. 7, 32 (2021))



Rhombus



Y-shaped

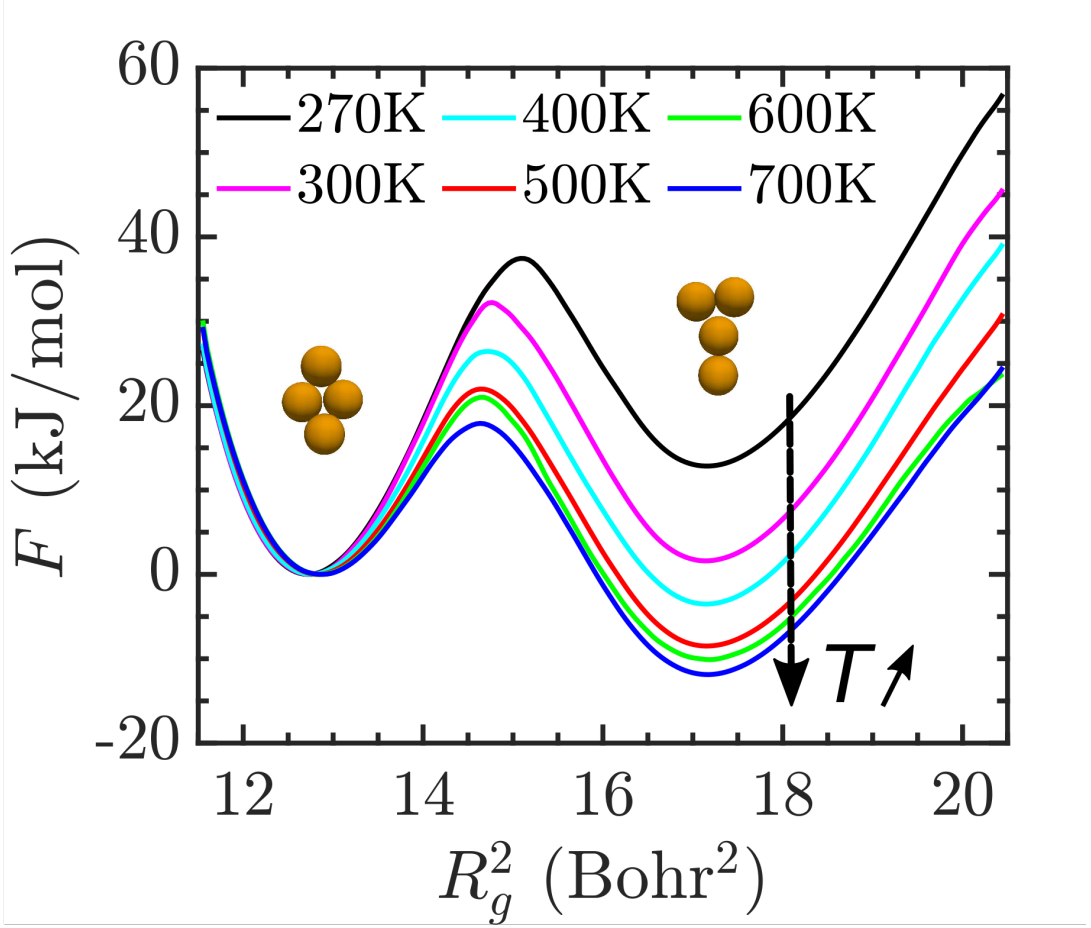


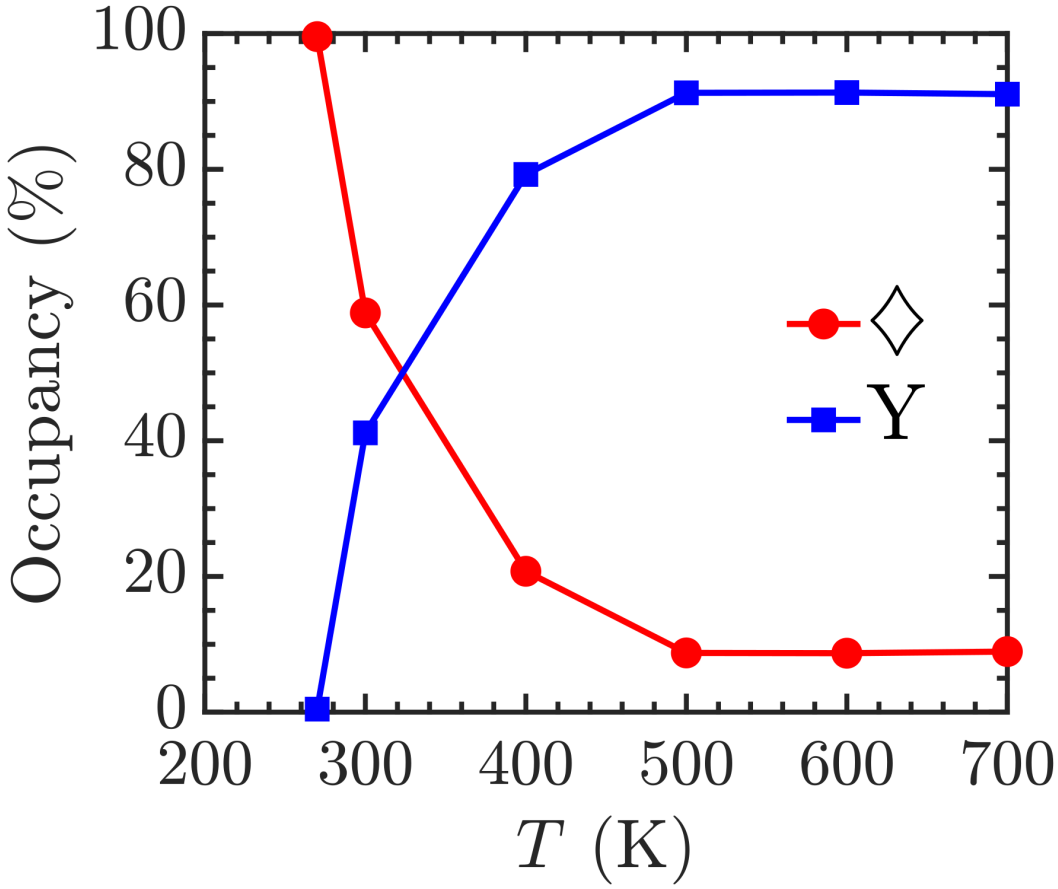
J. Shi, F. Gygi, J. K. Whitmer. *Submitted* (2021).

Free Energy and Occupancy

$$\tilde{P}(i) = \int_{\Xi_i} d\xi e^{-\beta F(\xi)}$$

$$C_Y = \frac{\tilde{P}_Y}{\tilde{P}_R + \tilde{P}_Y} \qquad C_R = \frac{\tilde{P}_R}{\tilde{P}_R + \tilde{P}_Y}$$

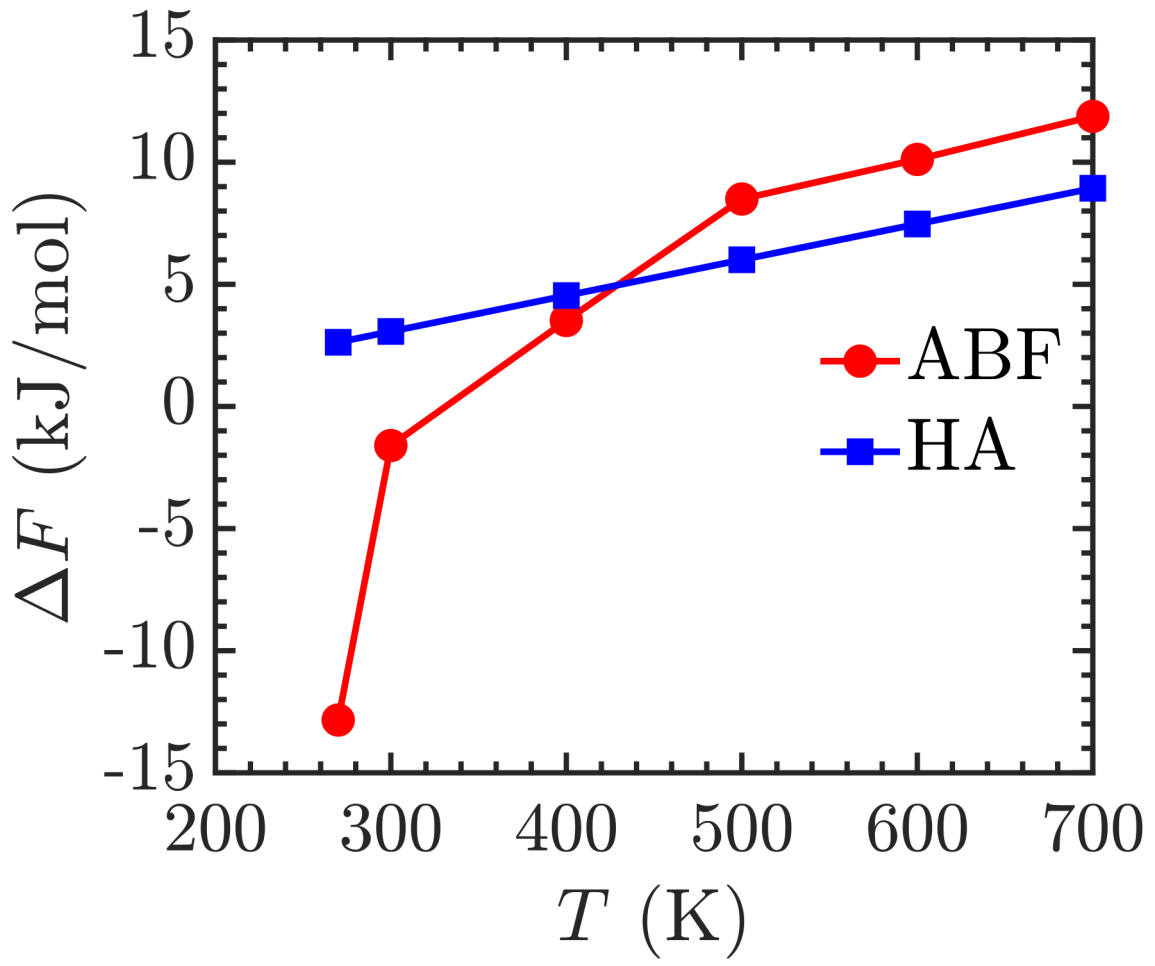




J. Shi, F. Gygi, J. K. Whitmer. *arXiv:2110.05336*.

Comparison with Harmonic Approximation (HA)

- The harmonic approximation reconstructs free energies from 0K DFT by assuming the local dynamics of a set of harmonic oscillators.
- Anharmonic entropy plays a key role in stabilizing clusters thermodynamically.
- It is important to use full free energy landscape sampling in ab initio contexts.



J. Shi, F. Gygi, J. K. Whitmer. *arXiv:2110.05336*.

Summary

- Advanced sampling is a set of tools for enhancing exploration and obtaining thermodynamic information from molecular simulations.

- Three primary classes of method exist: flat histogram, minimum (free) energy path, and reactive path methods

- The codes SSAGES and PySAGES implement many of the most popular flavors of these algorithms; you will learn to use these in the next two hands-on sessions.

Wide-open Problems in Advanced Sampling

- Nonequilibrium systems.
- *Ab initio* free energy computation.
- Collective variable identification/neutralization
- Efficient (and accurate) rate calculations
- Applications (the sky is really the limit…)

Thank you!

SSAGES

MICCoM

U.S. DEPARTMENT OF ENERGY

Office of Science

UNIVERSITY OF NOTRE DAME