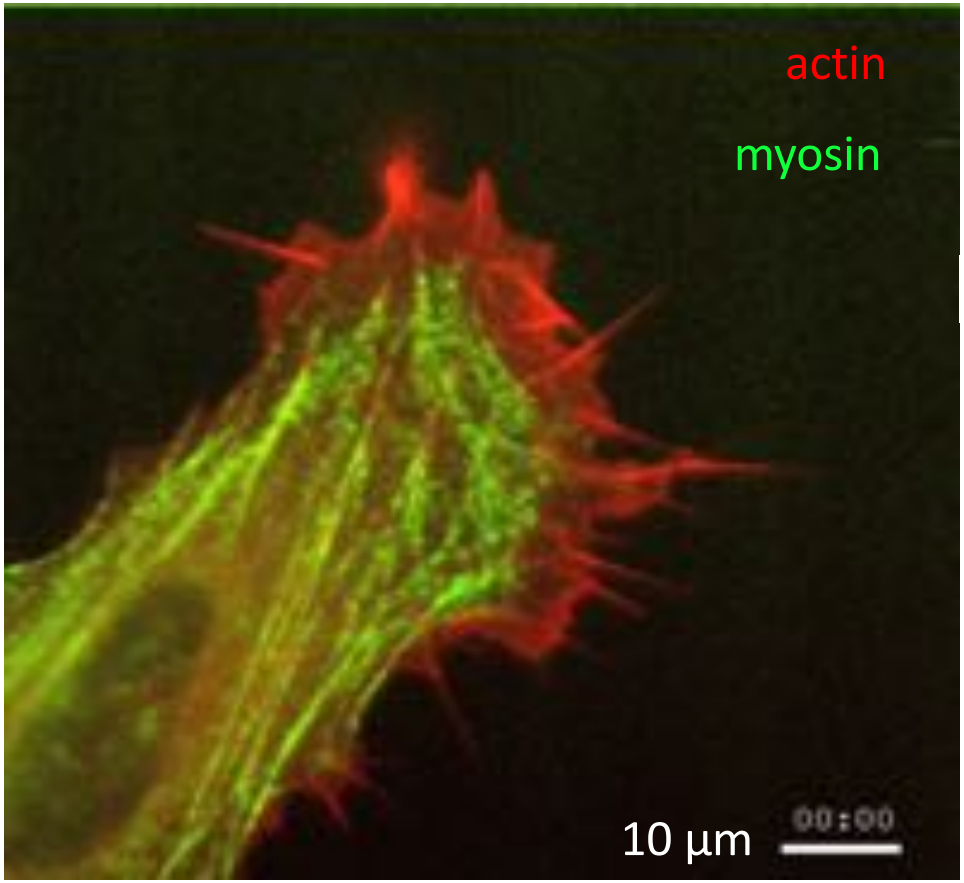
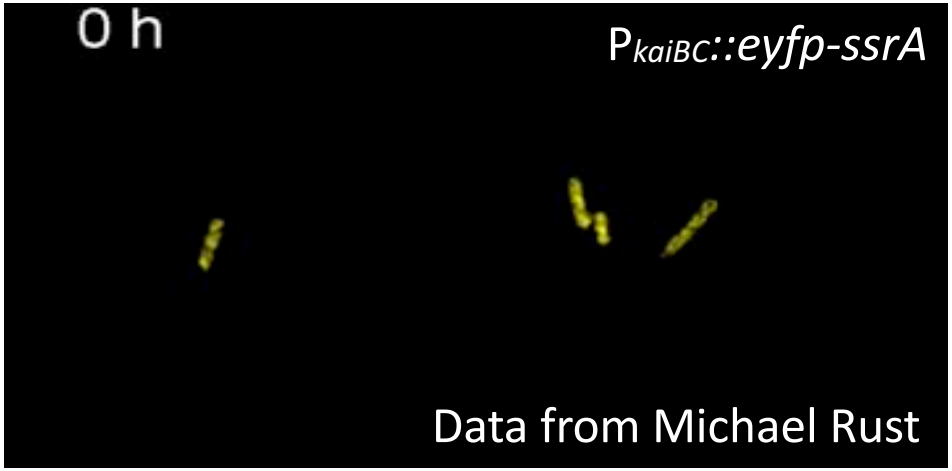
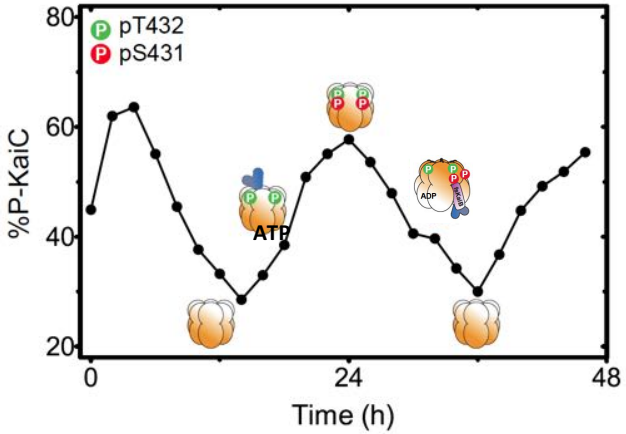
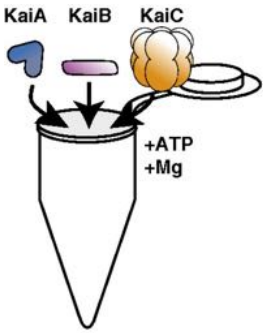
Living systems are far from equilibrium.



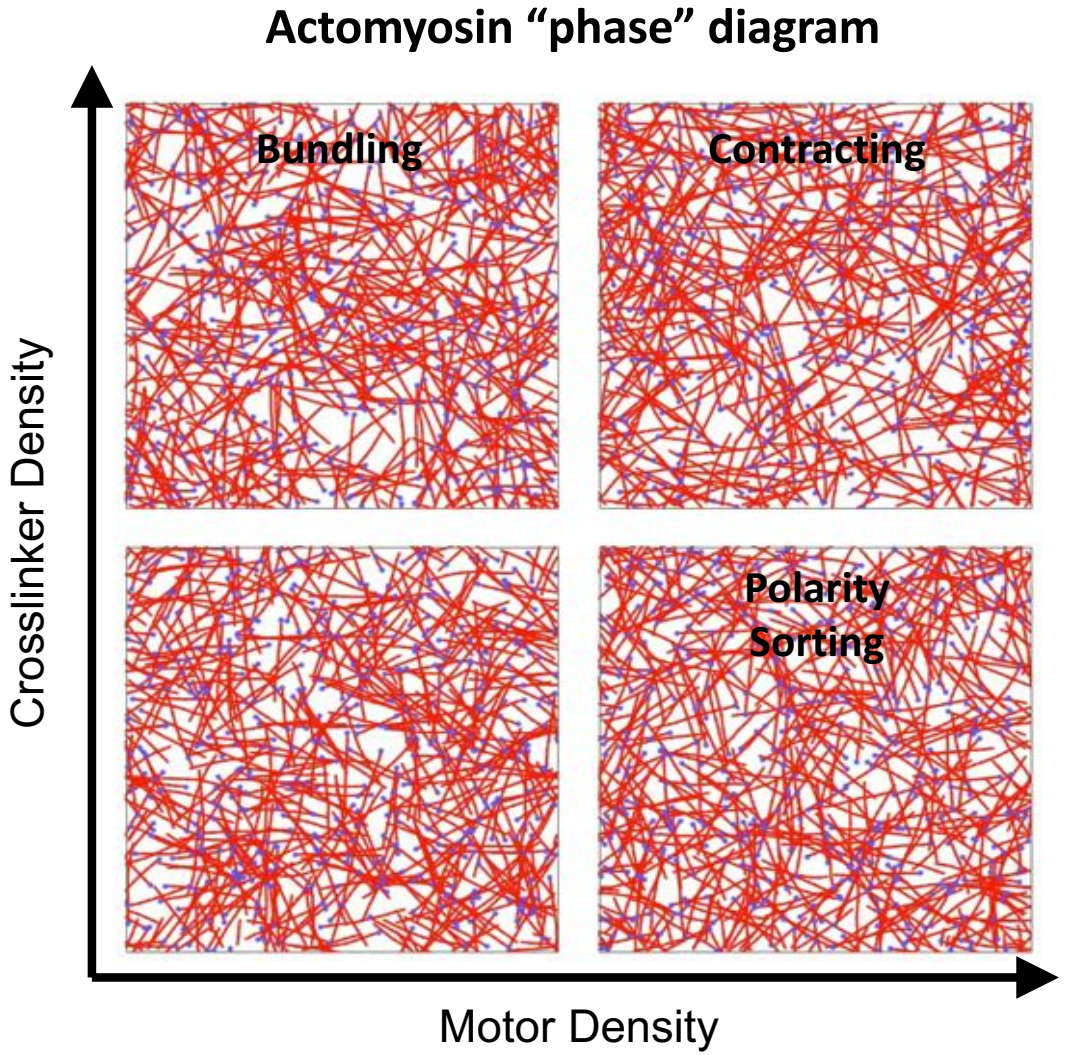
http://cellix.imba.oeaw.ac.at/fileadmin/conferences/Videotour_CellMotility/NEW_2010/Figure_7_400.MOV







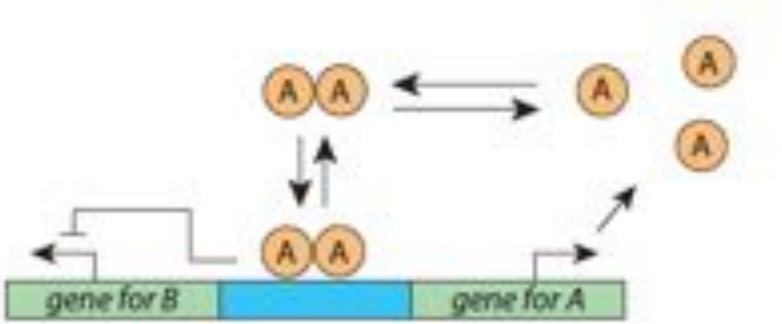
Models of living systems have traditionally been sufficiently simple that they can be simulated directly.

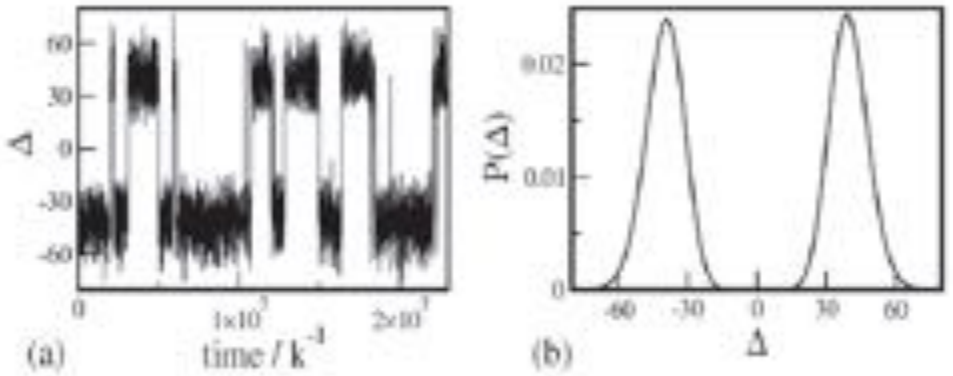


Freedman, Banerjee, Hocky, Dinner, Biophysical Journal (2017)

Examples of models of living systems that are challenging to simulate directly.

Multimodality

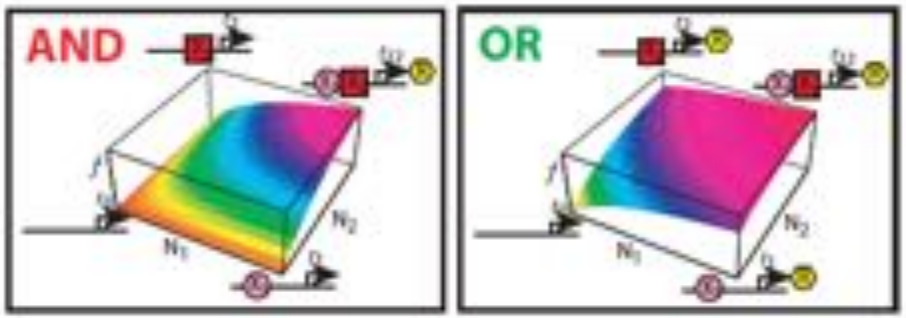




Allen, Warren, and ten Wolde (2005)

Tail statistics

$$\frac{\partial^2}{\partial N_i \partial N_j}\mathrm{CRIF} = K\frac{\langle \delta N_i \delta N_j \delta N_R \rangle}{\langle \delta N_i^2 \rangle \langle \delta N_j^2 \rangle}$$



Warmflash and Dinner (2008)

Nonequilibrium systems present a number of challenges for enhanced sampling methods:

- no *a priori* knowledge of the distribution function;

$$\langle A \rangle = \frac{\int A(q) e^{-\beta E} \, dq}{\int e^{-\beta E} \, dq}$$

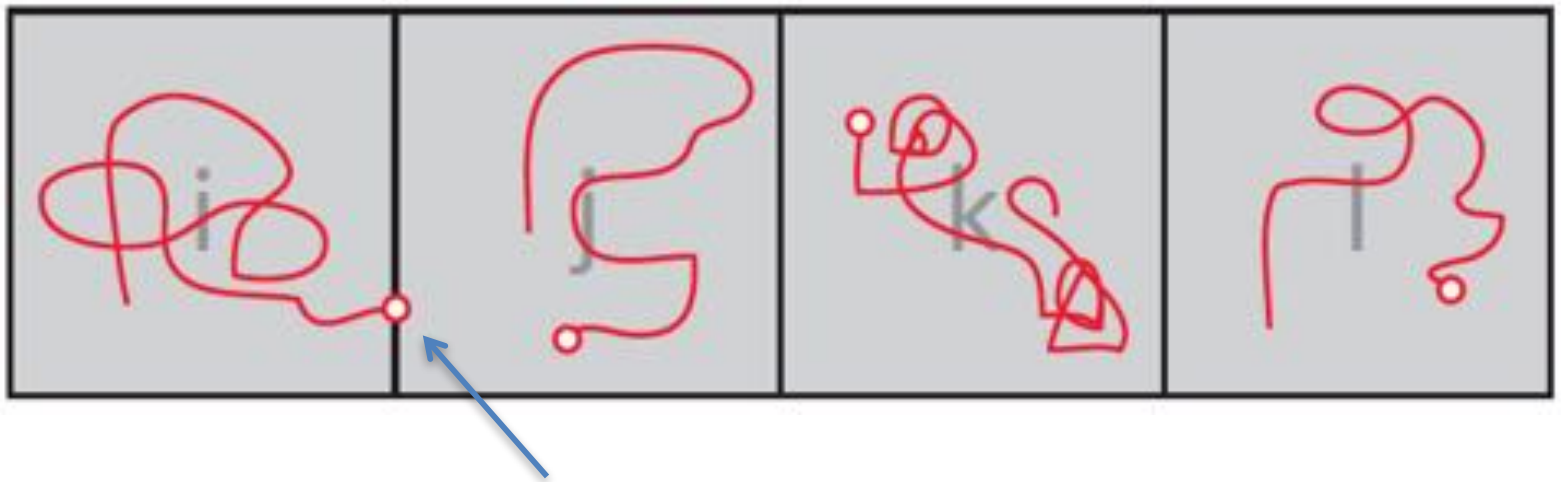
- microscopically irreversible dynamics;

$$P(q)W(q \to q') = P(q')W(q' \to q)$$

- global flows in phase space.

Original Nonequilibrium Umbrella Sampling Algorithm

1) Divide order parameter space
2) Sample each region and estimate fluxes
3) Estimate weights
4) Repeat until weights and fluxes converge



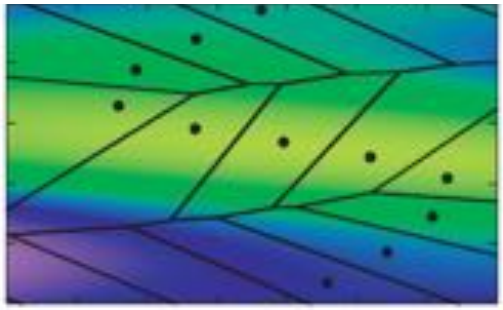
Upon attempted boundary crossing:

- add point to flux list,
- note flux to adjust region weights,
- restart walker, partitioning weight between saved and active copies

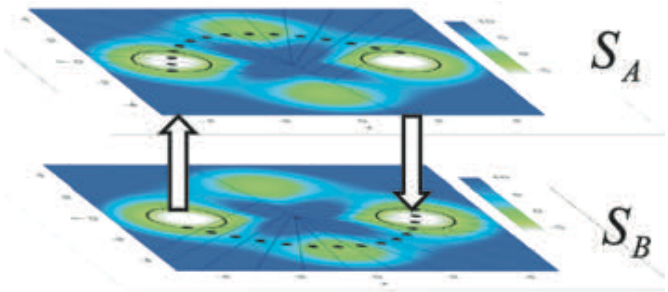
Warmflash, Bhimalapuram, Dinner (2007) J. Chem. Phys. 127, 154112.
Dickson, Dinner (2010) Annu. Rev. Phys. Chem. 61, 441-59.

Further developments of original NEUS

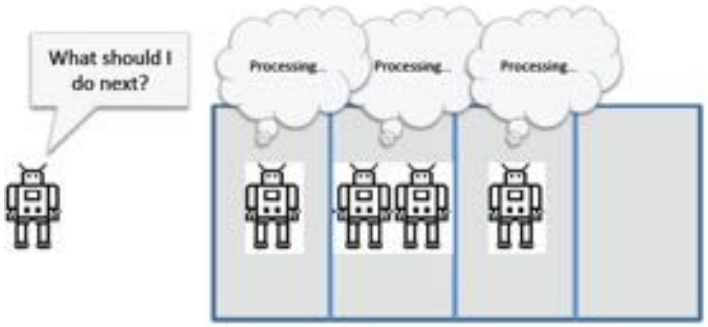
- Reaction path discovery (string method)
  - Dickson et al. JCP 130, 074104 (2009)



- Rates
  - Dickson et al. JCP 131, 154104 (2009)
  - Vanden-Eijnden JCP 131, 044120 (2009)

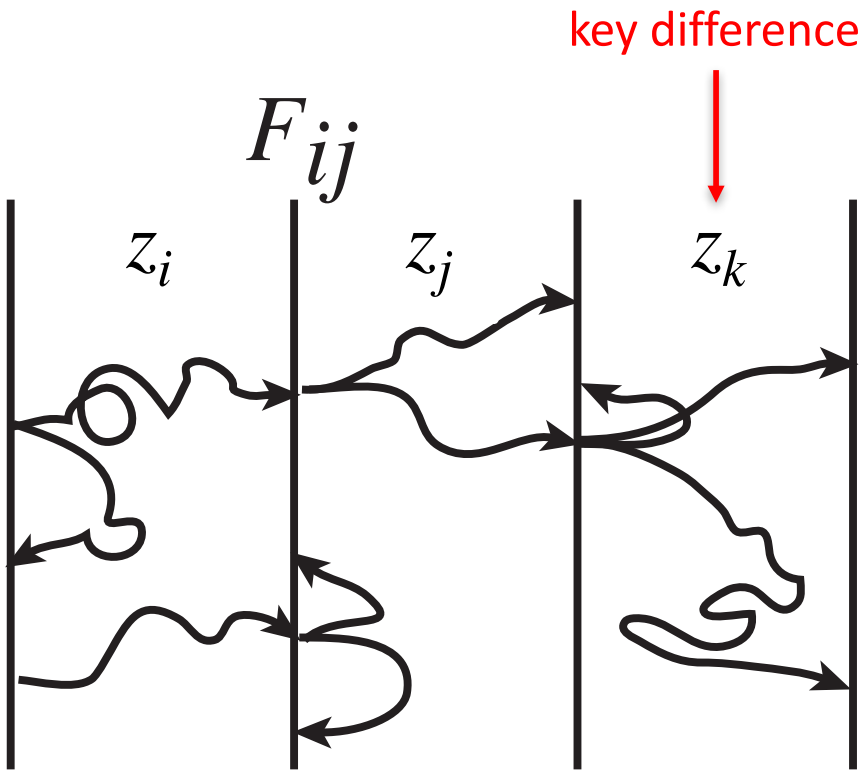


- Parallelism
  - Dickson et al. JCTC 7, 2710-2720 (2011).



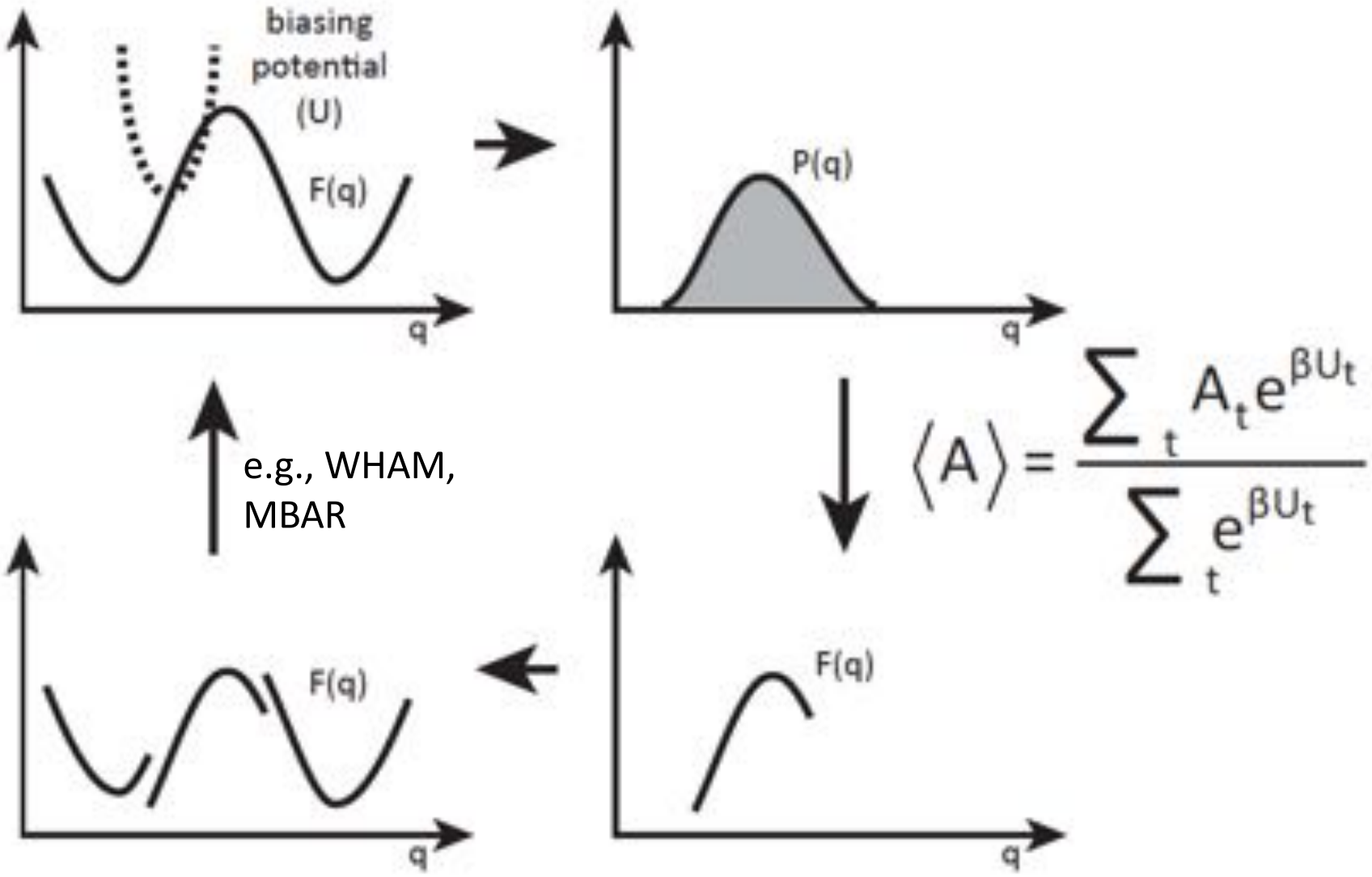
Splitting methods preserve the underlying dynamics but branch and prune the trajectories to focus sampling in specific areas

- Weighted Ensemble (WE)
- Forward Flux Sampling (FFS)
- Steered Transition Path Sampling (STePS)
- Nonequilibrium Umbrella Sampling (NEUS)



NEUS is distinct in that it is also a stratification method.

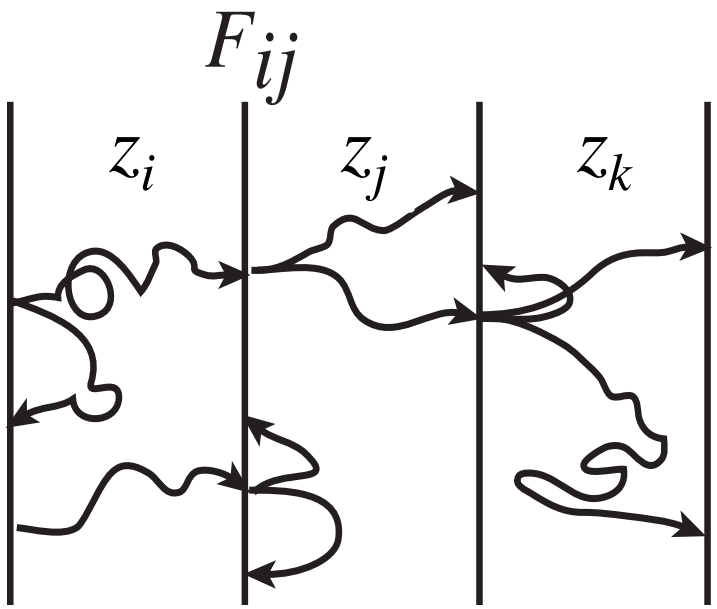
Equilibrium umbrella sampling is a well-known stratification method in molecular simulation.



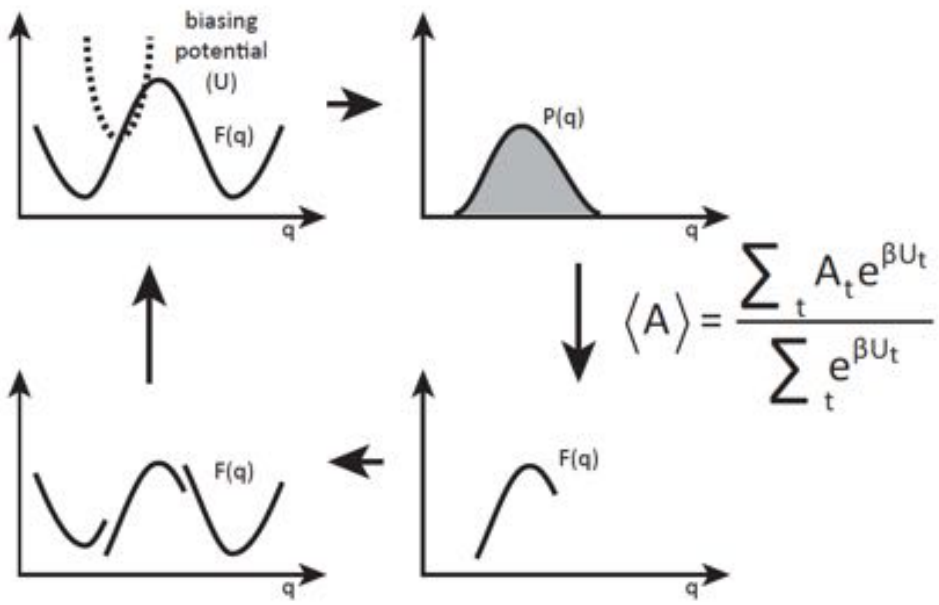
Torrie, Valleau, J Comput Phys 23, 187 (1977)
Pangali, Rao, Berne, J Chem Phys 71, 2975 (1979)

Can we make a framework that encompasses both these methods?

Nonequilibrium US



Equilibrium US



Work with Jeremy Tempkin, Erik Thiede,
Brian Van Koten, and Jonathan Weare

Funding from NIH

Outline

- Motivation

- Eigenvector Method for Umbrella Sampling (EMUS)
  - Formulation
  - Error analysis
  - Examples

  Thiede, Van Koten, Weare, Dinner, J Chem Phys 145, 084115 (2016)
  Dinner, Thiede, Van Koten, Weare, arxiv:1705.08445 (2017)

- Nonequilibrium Umbrella Sampling (NEUS)
  - Formulation
  - Examples

  Dinner, Tempkin, Van Koten, Mattingly, Weare, SIREV; arXiv:1610.09426

Umbrella sampling as an eigenproblem

We want to obtain the free energy, or equivalently the probability:

$$\pi(x) \propto \exp\left(-\beta H_0(x)\right)$$

To this end, we apply a bias, e.g.,

$$\psi_i(x) = \exp\left(-\beta\left(q - q_0^i\right)^2\right)$$

Simulation *i* samples from the probability

$$\pi_i(x) = \frac{\psi_i(x)\pi(x)}{\int \psi_i(x)\pi(x)dx}$$

To align the probabilities from different simulations, we need to determine the normalization, equivalent to the zero of free energy:

$$z_i = \int \psi_i(x)\pi(x)dx$$

Umbrella sampling as an eigenproblem

Once we have the normalization, we can calculate any average from

$$\langle f \rangle = \sum_i z_i \int \frac{f(x)}{\sum_k \psi_k(x)} \pi_i(x) dx$$

In particular,

$$z_j = \langle \psi_j \rangle = \sum_i z_i \int \frac{\psi_j(x)}{\sum_k \psi_k(x)} \pi_i(x) dx$$

or

$$z_j = \sum_i z_i F_{ij}$$

with the stochastic matrix *F* defined as

$$F_{ij} = \int \frac{\psi_j(x)}{\sum_k \psi_k(x)} \pi_i(x) dx$$

EMUS procedure

1. Estimate F and $\langle f \rangle_j$ by sampling from π_j .
2. Solve the eigenvector equation

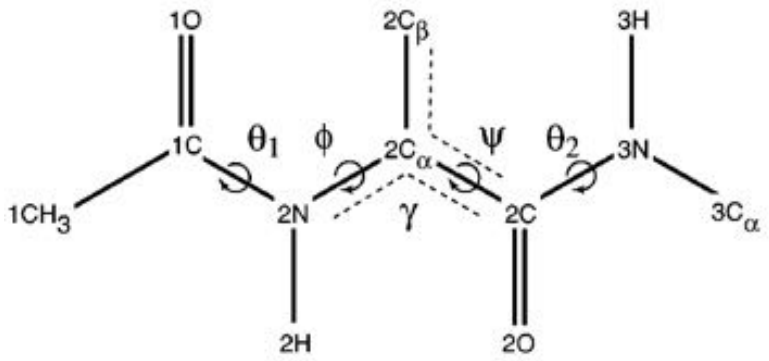
$$z^{\mathrm{T}} F = z^{\mathrm{T}}$$

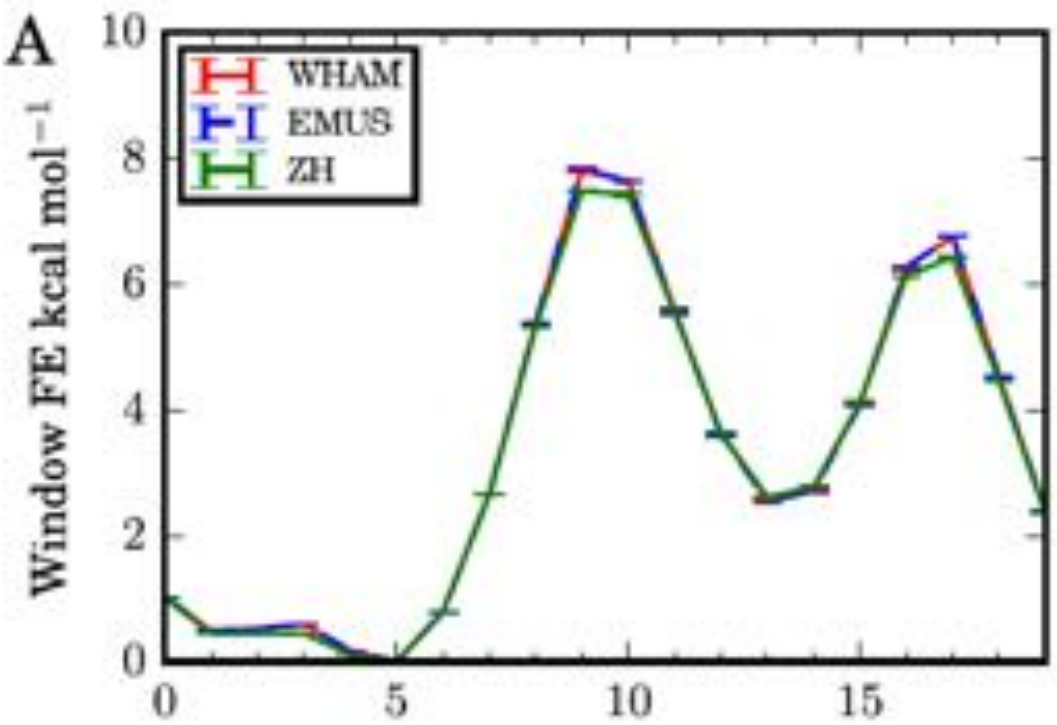
3. Compute the desired expectation as a weighted average

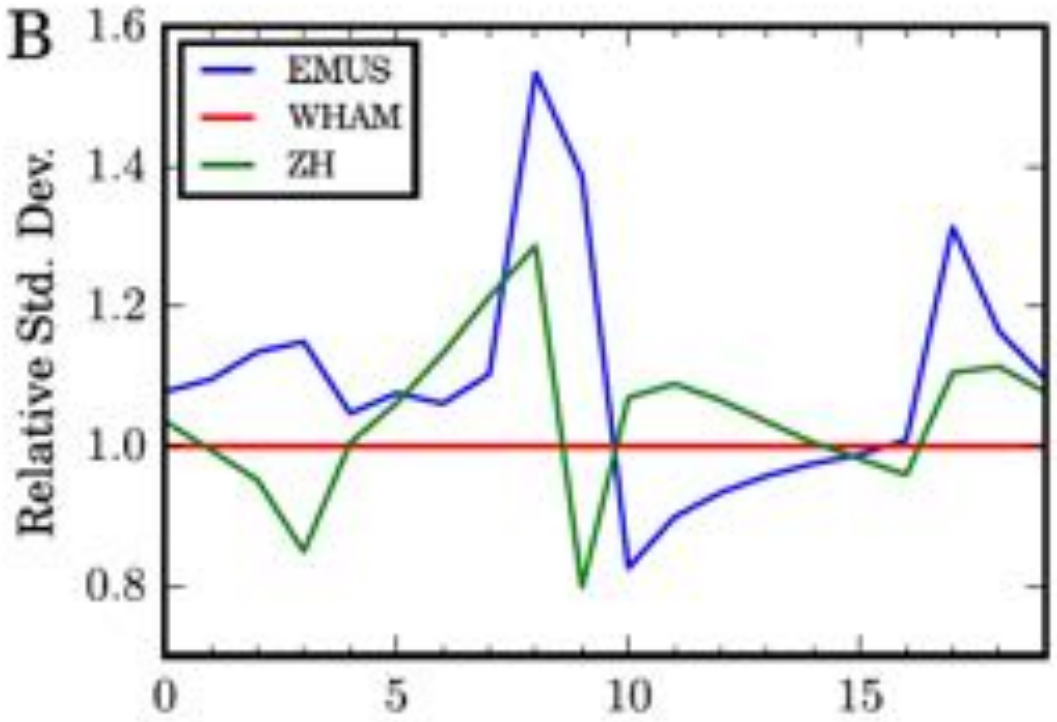
$$\int_{x \in \mathbb{R}^d} f(x)\pi(dx) = \sum_{j=1}^{n} z_j \langle f \rangle_j$$

$$\langle f \rangle_j = \int_{x \in \mathbb{R}^d} f(x)\pi_j(dx).$$

This Eigenvector Method for Umbrella Sampling (EMUS) for obtaining the window normalizations performs comparably to existing methods while facilitating mathematical analysis of the error.







Relation to Multistate Bennett Acceptance Ratio (MBAR)

EMUS satisfies the extended bridge sampling equation

$$z_j \sum_{i=1}^{L} \langle \alpha_{ij}(x)\psi_i(x)\pi(x)\rangle_j = \sum_{i=1}^{L} z_i \langle \alpha_{ij}(x)\psi_j(x)\pi(x)\rangle_i$$

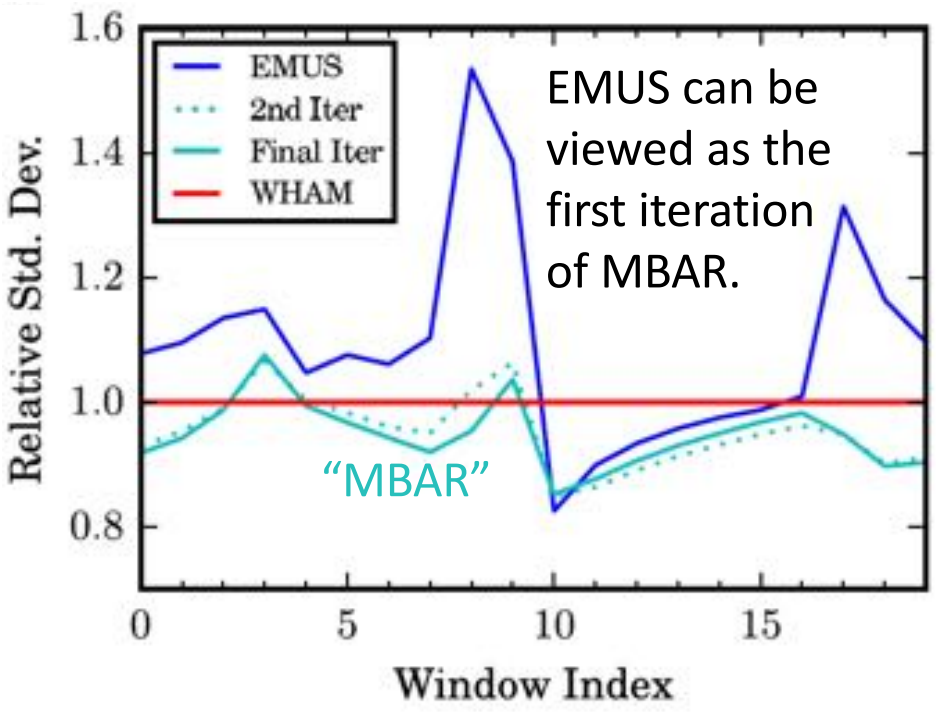
with arbitrary α_{ij}. For MBAR, Shirts and Chodera (2008) chose

$$\alpha_{ij}^{\mathrm{MBAR}}(x) = \frac{n_i/z_i}{\sum_k \psi_k(x)\pi(x)n_k/z_k}$$

while EMUS corresponds to

$$\alpha_{ij}^{\mathrm{EMUS}}(x) = \frac{1}{\sum_k \pi(x)\psi_k(x)}$$

The dependence of α_{ij} on z in MBAR necessitates self-consistent solution and complicates error analysis.



EMUS Error analysis

We can prove a central limit theorem for EMUS. A key result is

$$\mathrm{var}(G_{lm}) \approx \sum_i n_i \mathrm{var}_{\mathrm{i}} \left(\sum_{\mathrm{j}} \frac{\partial G_{lm}}{\partial F_{ij}} \frac{\psi_j(x)}{\sum_k \psi_k(x)} \right)$$

number of independent samples

Delta method

$$\mathrm{var}\big(f(\bar{g})\big) \approx \mathrm{var}(\nabla f(\bar{g}) \cdot g)$$

For practical computation, write

$$\frac{\partial G_{lm}}{\partial F_{ij}} = -\frac{\partial}{\partial F_{ij}} \log\left(\frac{z_l}{z_m}\right) = \frac{1}{z_m}\frac{\partial z_m}{\partial F_{ij}} - \frac{1}{z_l}\frac{\partial z_l}{\partial F_{ij}}$$

EMUS Error analysis

We can go further by exploiting the eigenvector framework for umbrella sampling,

In particular,

$$\frac{\partial z_k}{\partial F_{ij}} = z_i (I - F)^{\#}_{jk}$$

where # denotes the group inverse: $AA^{\#}A = A$, $A^{\#}AA^{\#} = A^{\#}$, $AA^{\#} = A^{\#}A$.

Golub, G; Meyer, C. "Using the QR Factorization and Group Inversion to Compute, Differentiate, and Estimate the Sensitivity of Stationary Probabilities for Markov Chains." *SIAM J. Alg. Disc. Meth.* 7, 12 (1986)

Algorithm for assigning computational effort to minimize the overall error in an average

1. Sample to obtain the entries in *F.*
2. Calculate the group inverse of *I-F*, and in turn the matrix of derivatives $\partial G_{lm}/\partial F_{ij}$.
3. For each window, construct the trajectory

$$v_{ni} = \sum_j \frac{\partial G_{lm}}{\partial F_{ij}} \frac{\psi_j(x_{ni})}{\sum_k \psi_k(x_{ni})}$$

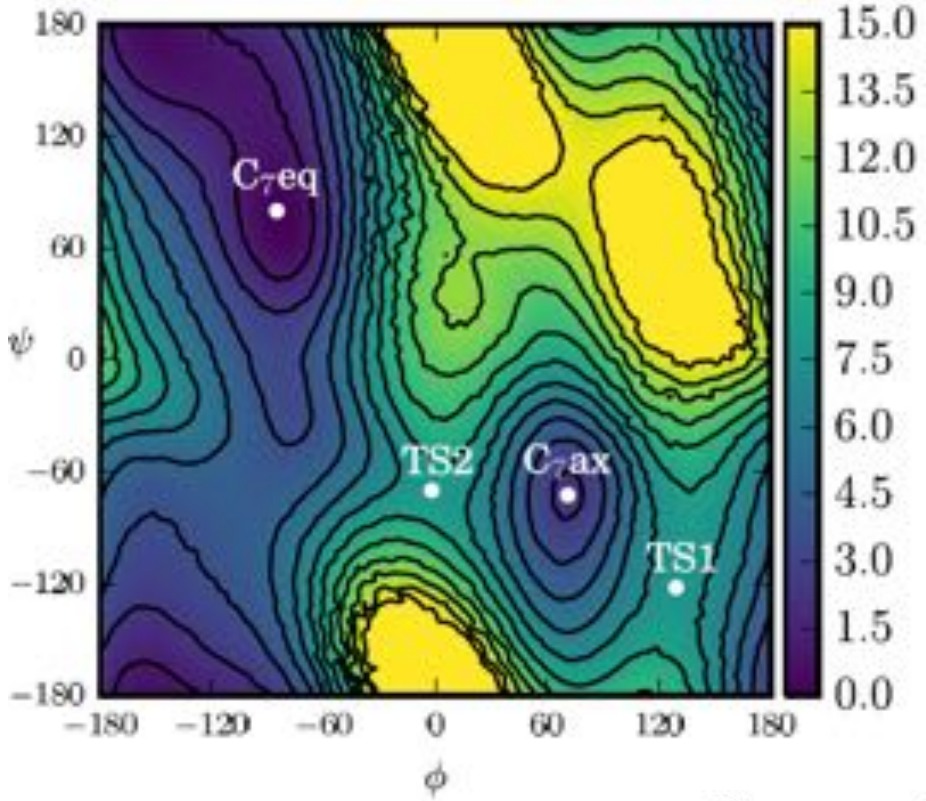
4. Calculate the variance and autocorrelation time of this trajectory to obtain

$$\mathrm{err}_i^2 = \mathrm{var}_i(v_{ni})/n_i$$

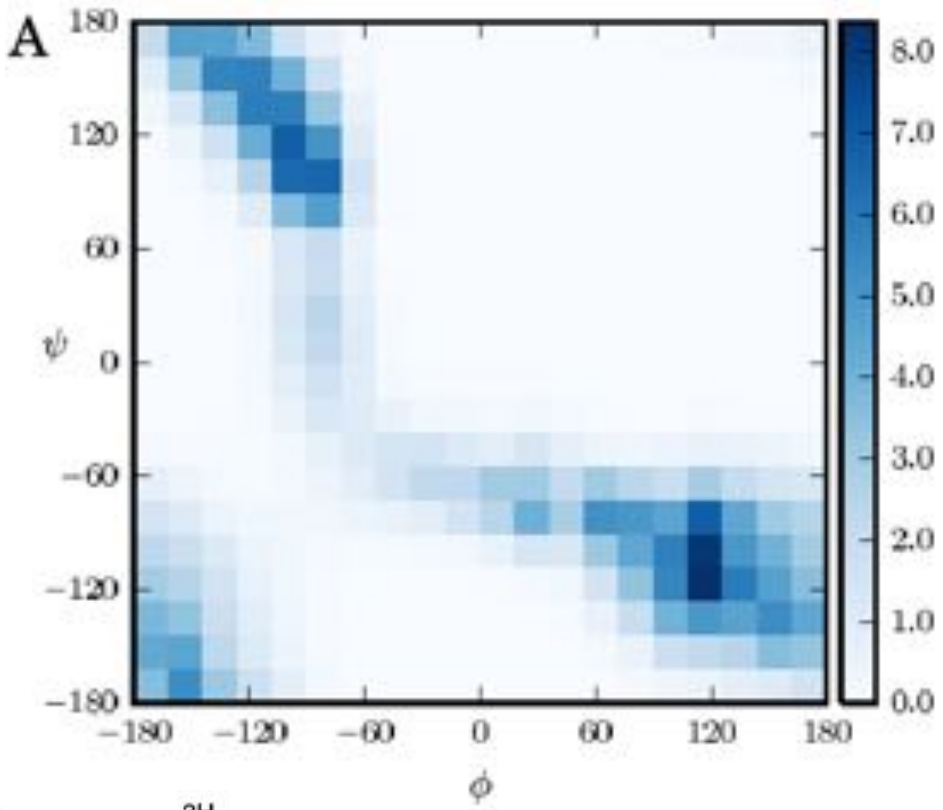
1. Assign computational resources proportional to err_i, which we term the "relative importances".

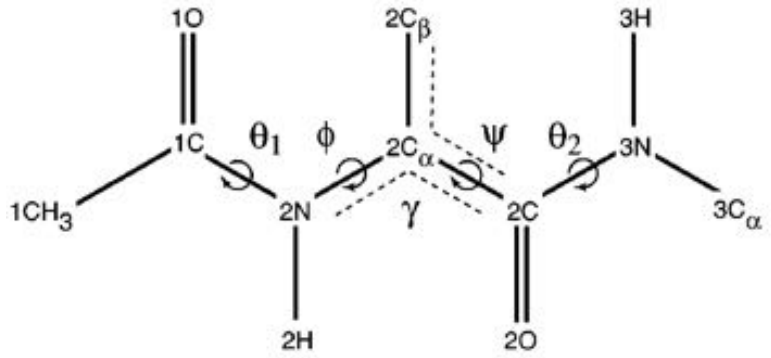
EMUS error contributions to $\Delta G(C_{7eq} - C_{7ax})$ of the alanine dipeptide

Potential of mean force (kcal/mol)



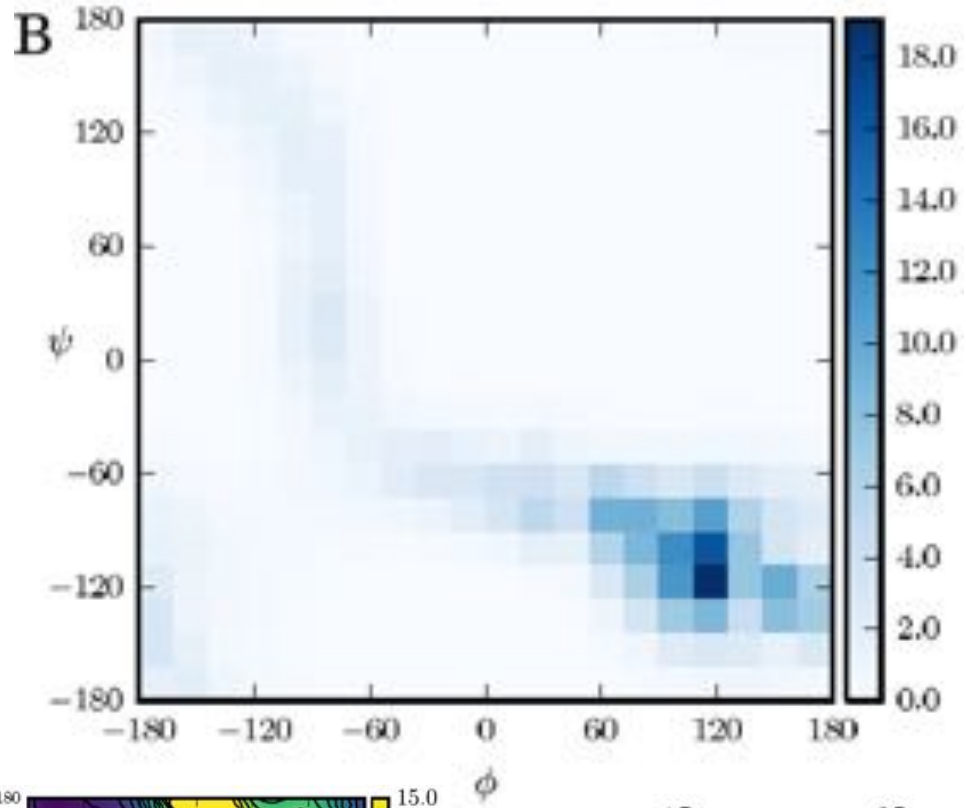
Window importances

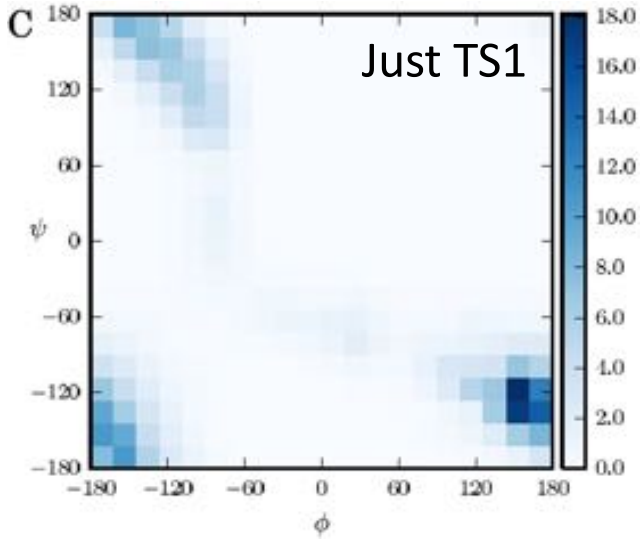


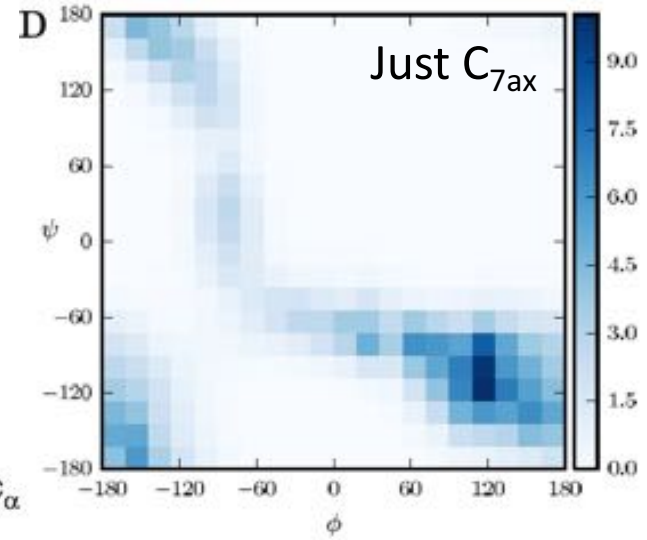


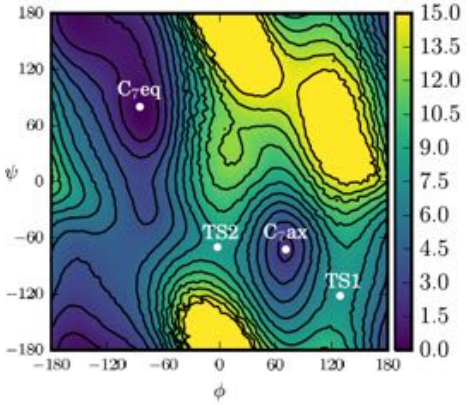
EMUS error contributions to ΔG(TS1 – C_{7ax}) of the alanine dipeptide

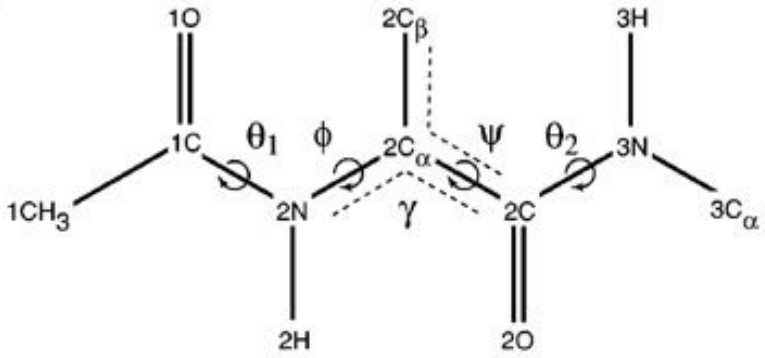
Window importances



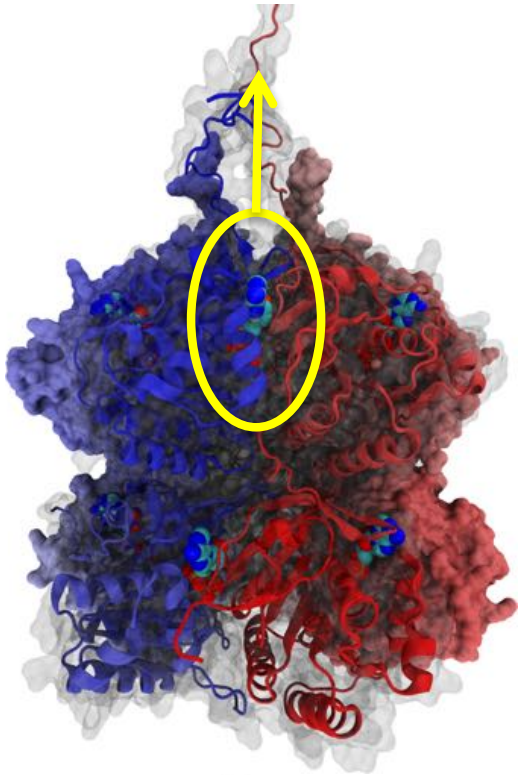


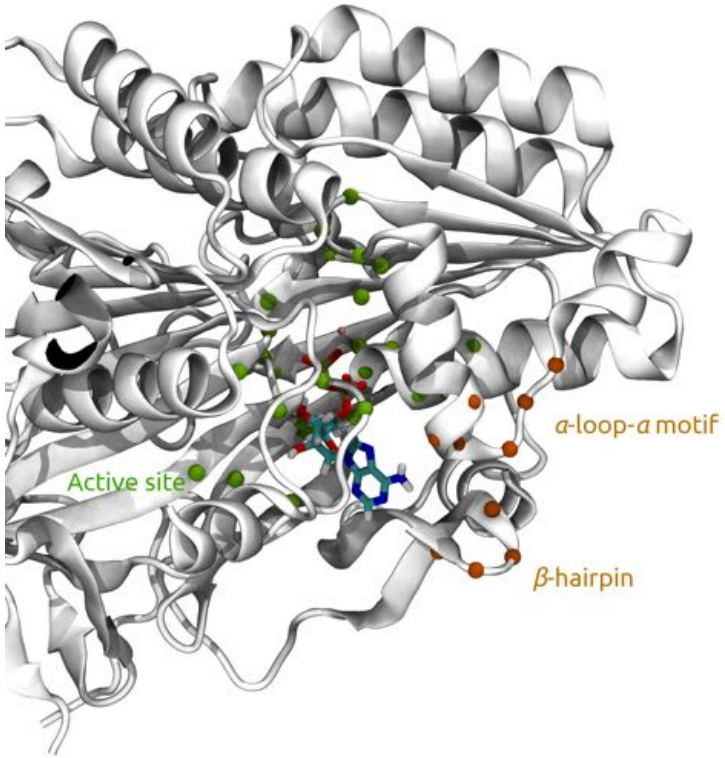


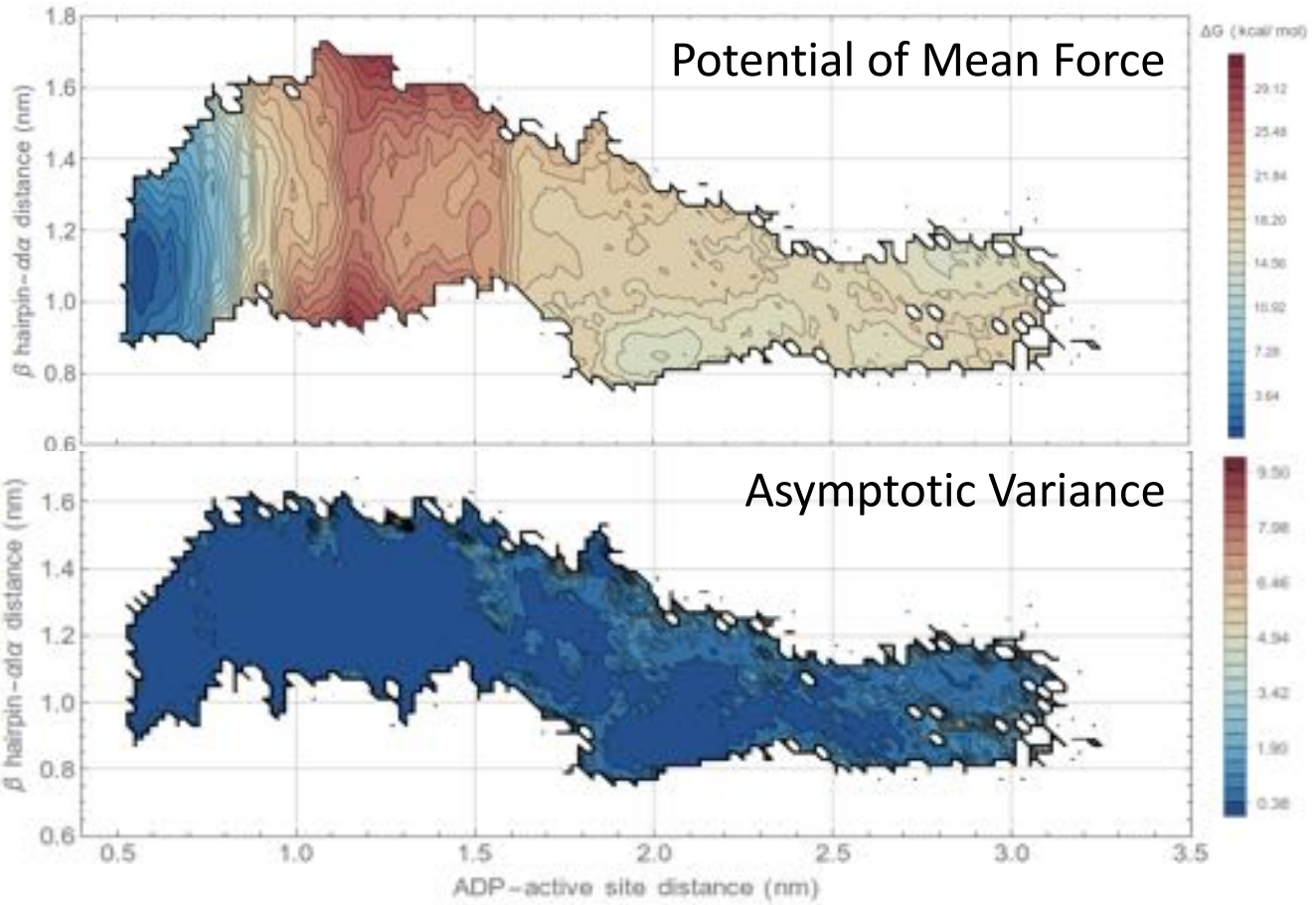




EMUS applied to ADP release from a circadian clock protein (KaiC, AAA+ ATPase)







Simulations by Lu Hong

EMUS Summary

- We cast equilibrium US as an eigenproblem.

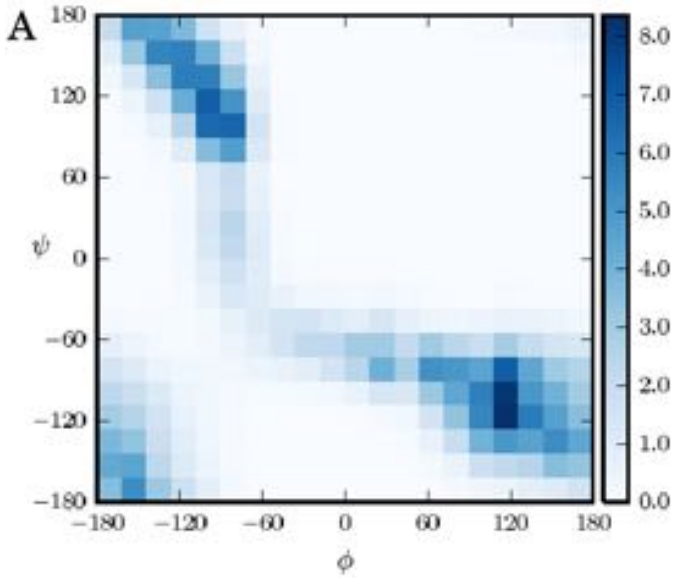
$$z_j = \sum_i z_i F_{ij}$$

$$F_{ij} = \int \frac{\psi_j(x)}{\sum_k \psi_k(x)} \pi_i(x) dx$$

- The non-iterative nature of EMUS facilitates error analysis.

$$\frac{\partial z_k}{\partial F_{ij}} = z_i (I - F)^{\#}_{jk}$$

- Average-specific window contributions can be used to guide sampling.



- https://github.com/ehthiede/EMUS

Outline

- Motivation

- Eigenvector Method for Umbrella Sampling (EMUS)
  - Formulation
  - Error analysis
  - Examples

  Thiede, Van Koten, Weare, Dinner, J Chem Phys 145, 084115 (2016)
  Dinner, Thiede, Van Koten, Weare, arxiv:1705.08445 (2017)

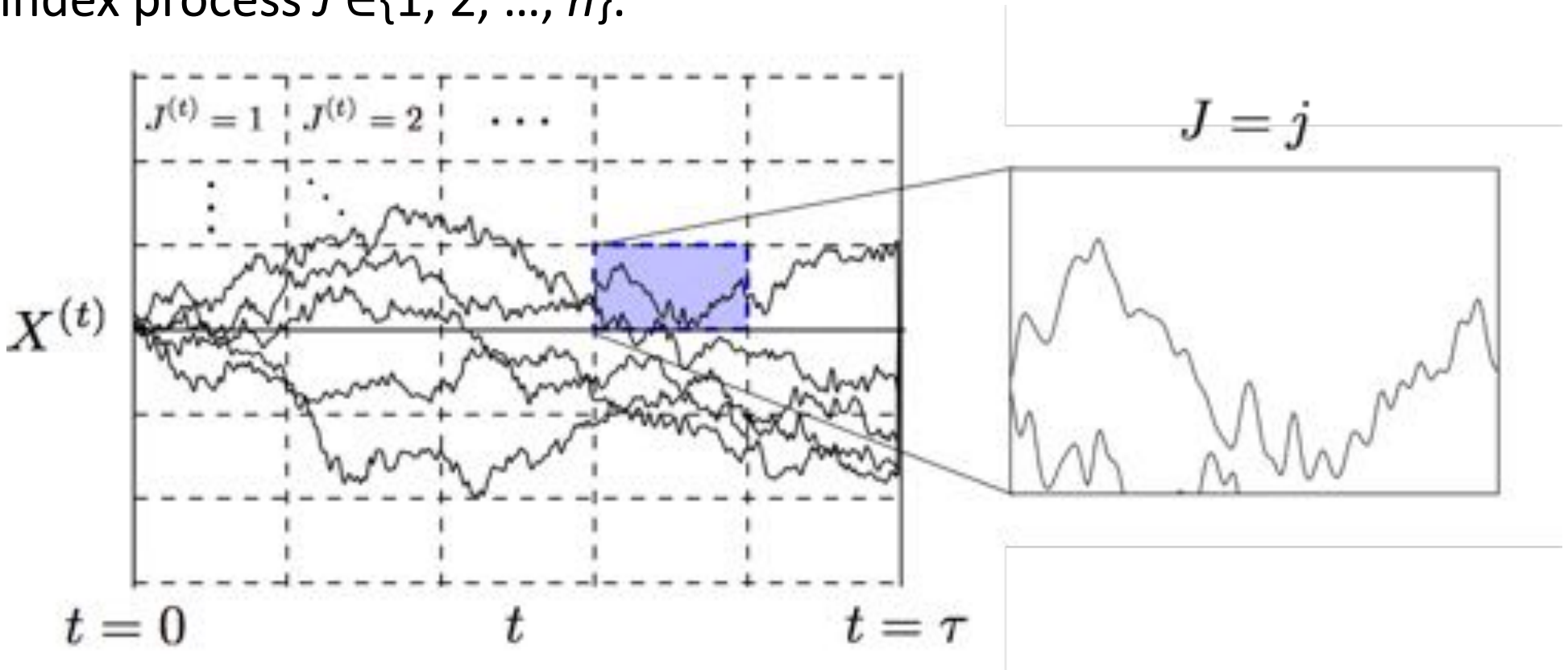
- Nonequilibrium Umbrella Sampling (NEUS)
  - Formulation
  - Examples

  Dinner, Tempkin, Van Koten, Mattingly, Weare, SIREV; arXiv:1610.09426

We seek to calculate nonequilibrium expectations as a sum over windows:

$$\mathbf{E}\left[\sum_{t=0}^{\tau-1} f(t, dx)\right] = \sum_{j=1}^{n} z_j \langle f \rangle_j$$

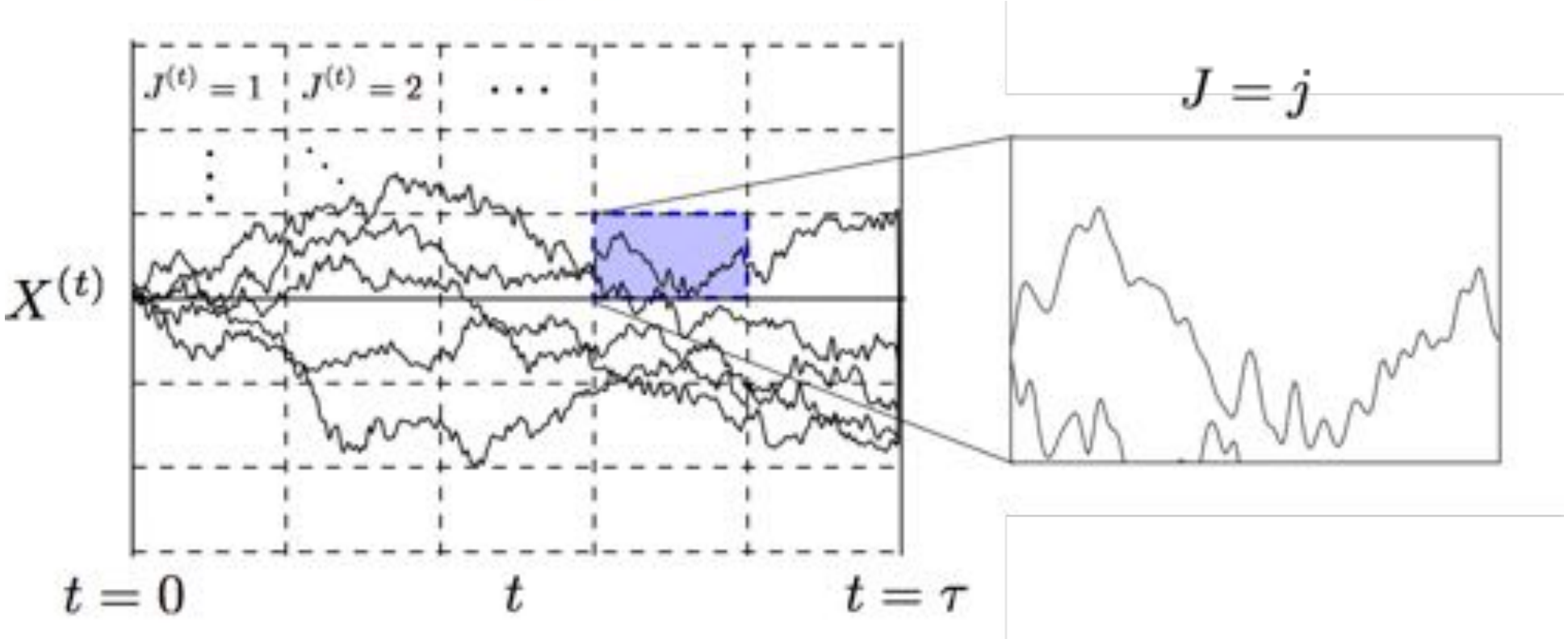
To this end, we define windows in space and time and define an index process $J \in \{1, 2, \ldots, n\}$.



Now,

$$\pi_i(t, dx) = \frac{\mathbf{P}\left[t < \tau, X^{(t)} \in dx, J^{(t)} = i\right]}{z_i}$$

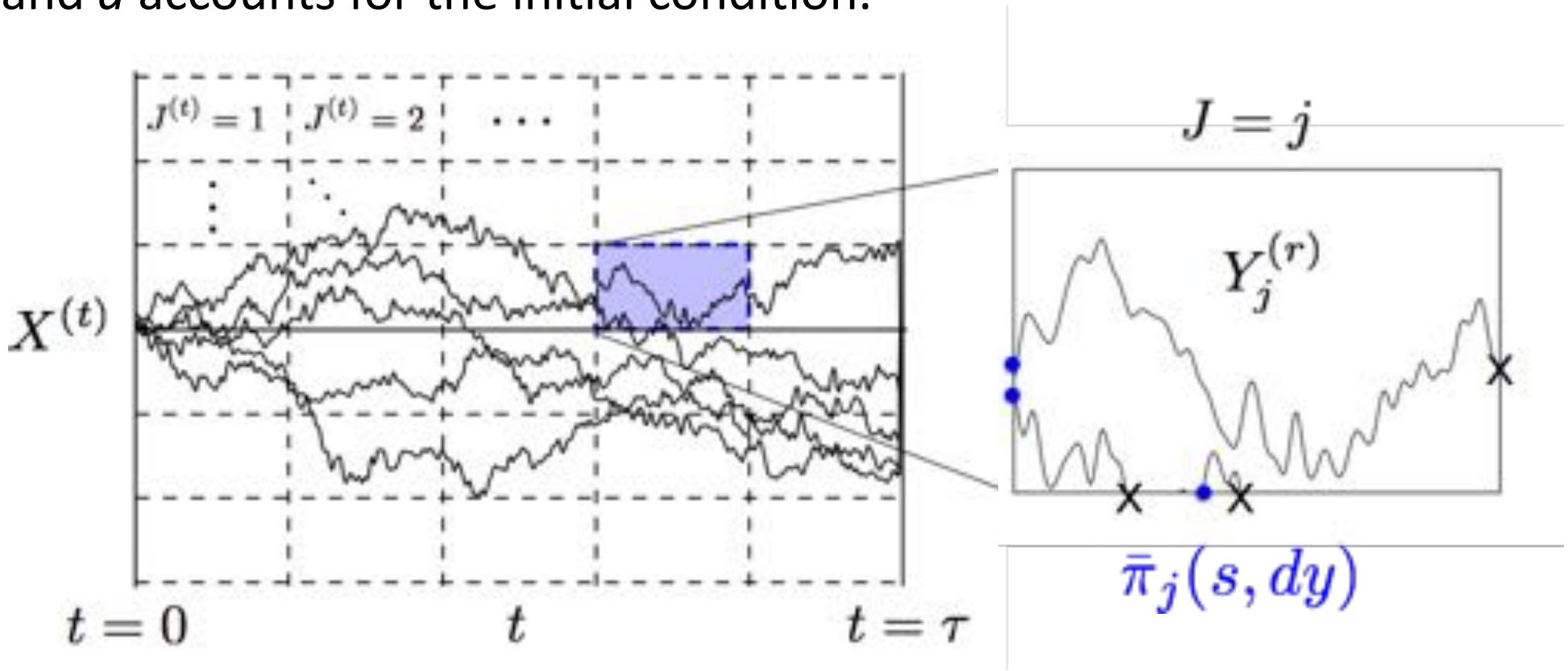
$$z_j = \sum_{t=0}^{\infty} \mathbf{P}\left[t < \tau,\ J^{(t)} = j\right]$$



The key difference from EMUS is that now, to obtain π_j, we sample trajectory segments with initial points drawn from the entry point distribution into window *j*, $\overline{\pi}_j$. If we can do that, the z_i can be shown to satisfy

$$z^{\mathrm{T}} G + a^{\mathrm{T}} = z^{\mathrm{T}}$$

where *G* is the matrix of window-to-window transition probabilities and *a* accounts for the initial condition.



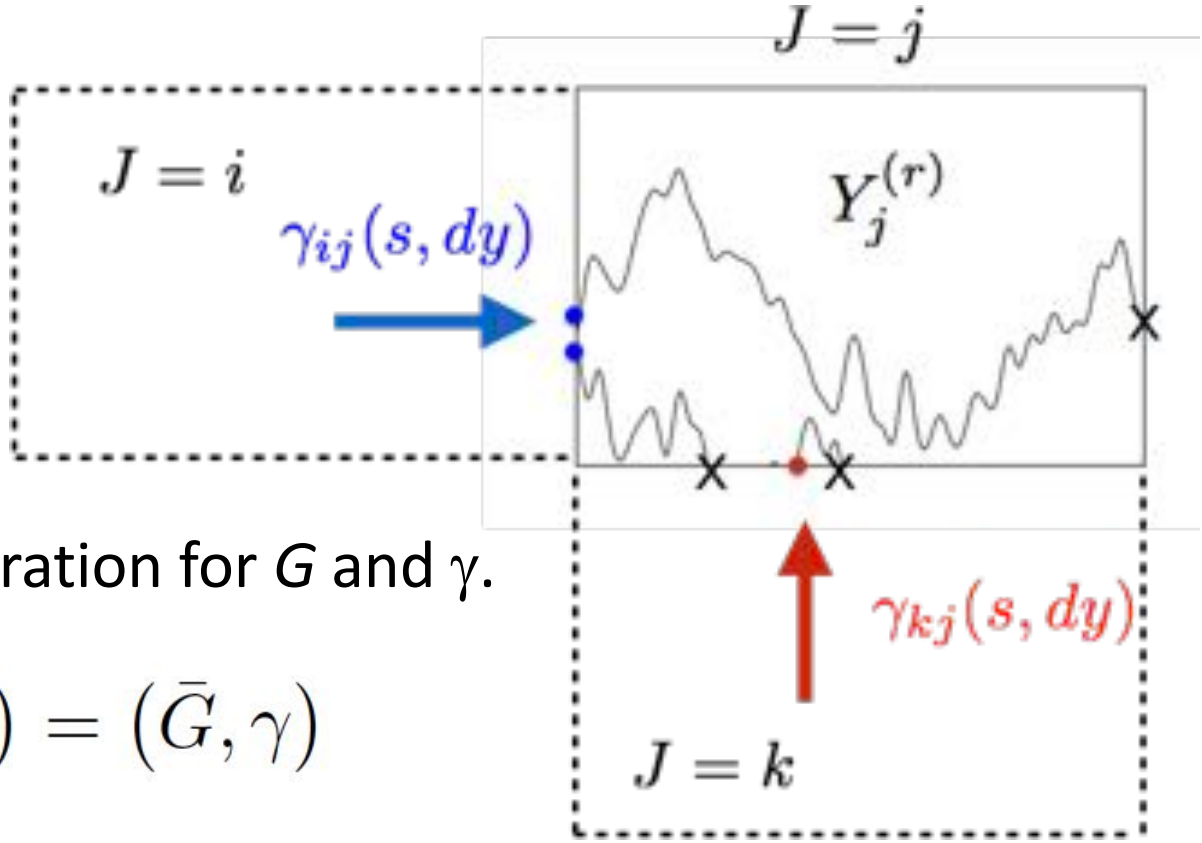
window-to-window transition probabilities (analog of F_{ij} in EMUS)

conditional distribution for a particular neighbor pair

entry-point distribution

$$\bar{\pi}_j(s, dy) = \frac{1}{z_j(1 - G_{jj})} \begin{cases} \sum_{i \neq j} z_i G_{ij} \gamma_{ij}(s, dy), & \text{if } s > 0 \\ a_j \mathbf{P}[X^{(0)} \in dy | J^{(0)} = j] & \text{if } s = 0 \end{cases}$$

We use the statistics of exit points to update estimates of γ_{ij} in neighboring windows.



This defines a fixed-point iteration for G and γ.

$$\left(\mathcal{G}(\bar{G}, \gamma), \Gamma(\bar{G}, \gamma)\right) = (\bar{G}, \gamma)$$

NEUS is a form of stochastic approximation.

NEUS procedure (parallels EMUS)

1. Estimate G and $\langle f \rangle_j$ by sampling from π_j using the latest estimates of the weights z and the entry point distributions.
2. Solve the affine eigenvector equation

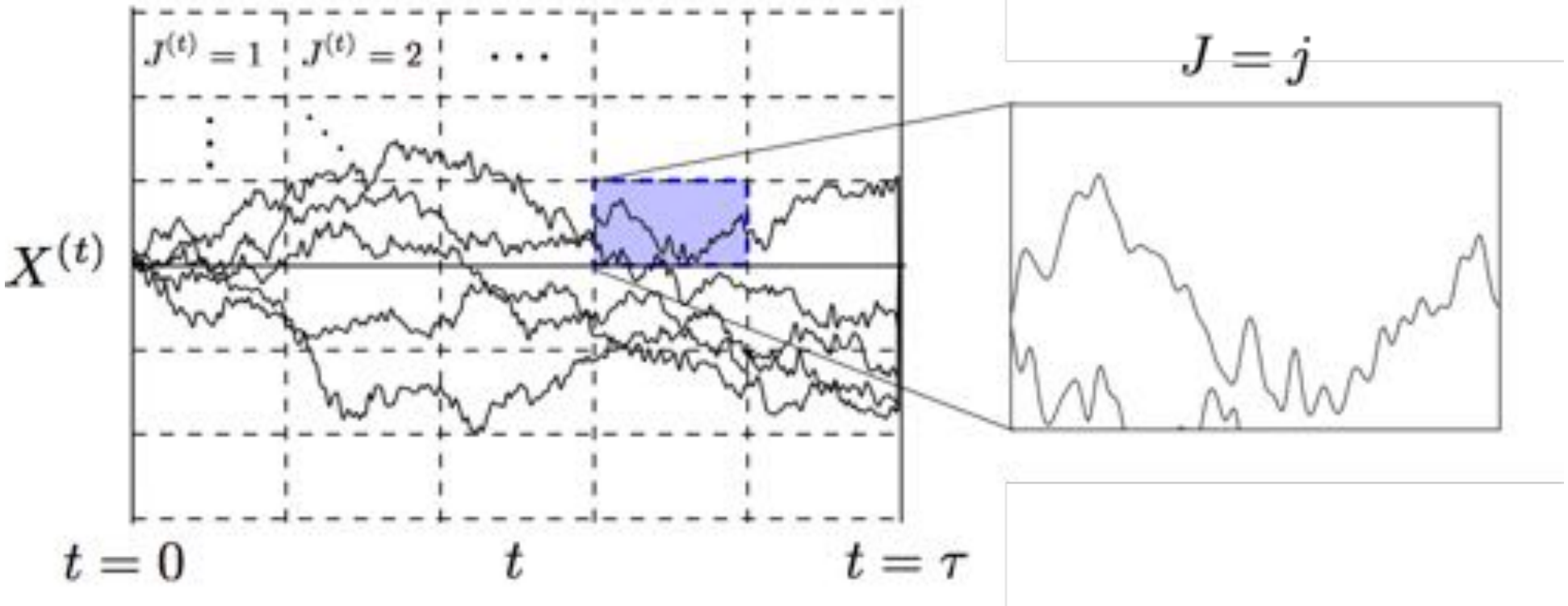
$$z^{\mathrm{T}} G + a^{\mathrm{T}} = z^{\mathrm{T}}$$

3. Compute the desired expectation as a weighted average

$$\mathbf{E}\left[\sum_{t=0}^{\tau-1} f(t, dx)\right] = \sum_{j=1}^{n} z_j \langle f \rangle_j$$

Iterate

In effect, we have obtained a procedure for generating the dynamics of the index process *J* from short trajectory segments.



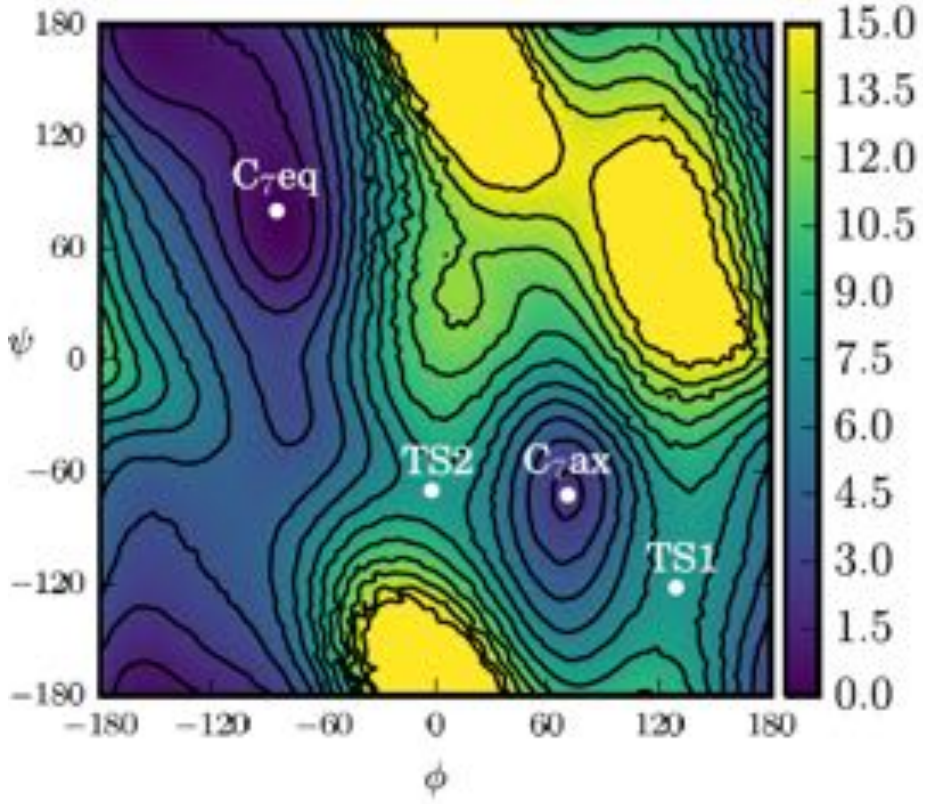
Advantages of new NEUS procedure

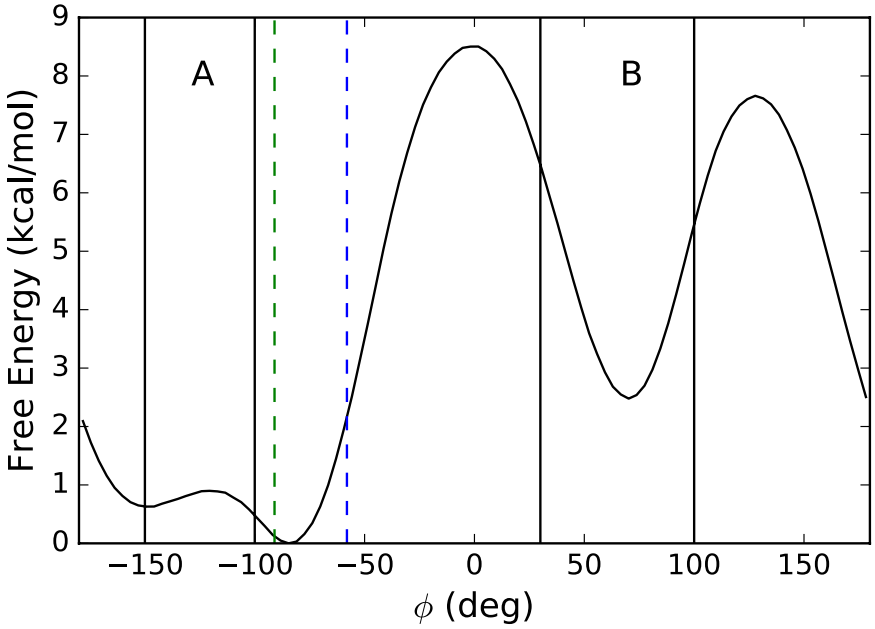
- Unified framework for equilibrium and nonequilibrium US.
- Shows NEUS is a form of stochastic approximation.
- Allows computing expectations over non-stationary processes.
  - Can be used to obtain exact dynamical statistics for microscopically reversible dynamics.
  - Can be used for any process in which the microscopic dynamics defines the distribution function.

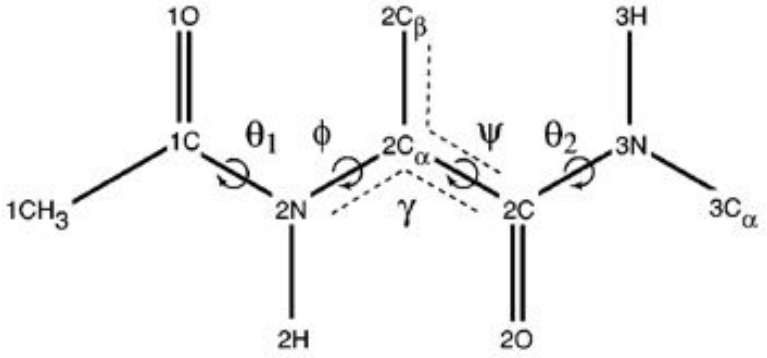
Dinner, Tempkin, Van Koten, Mattingly, Weare, SIREV; arXiv:1610.09426

Example: Compute probability of hitting set B before set A for times less than τ_{max}.

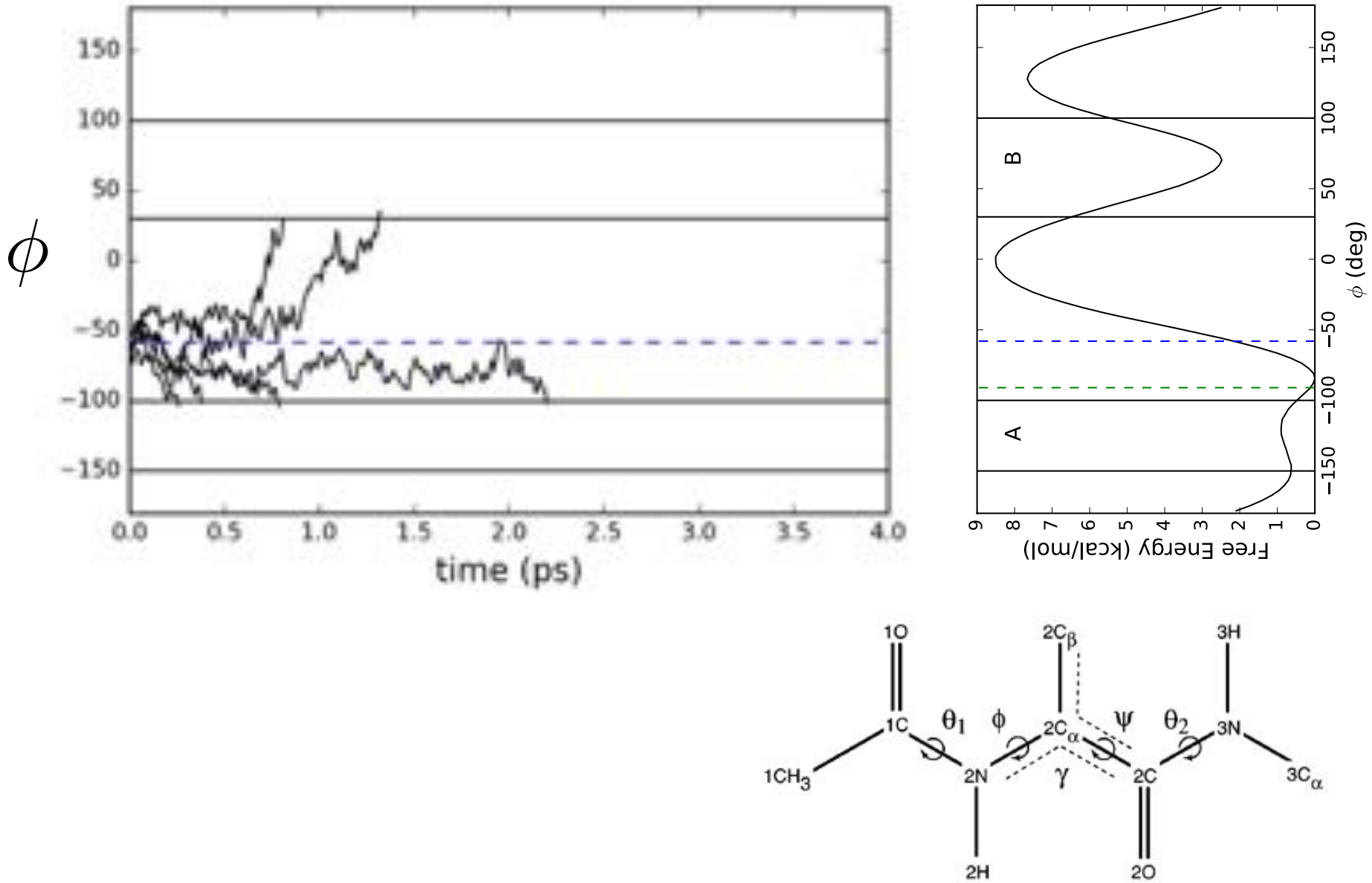
Potential of mean force (kcal/mol)



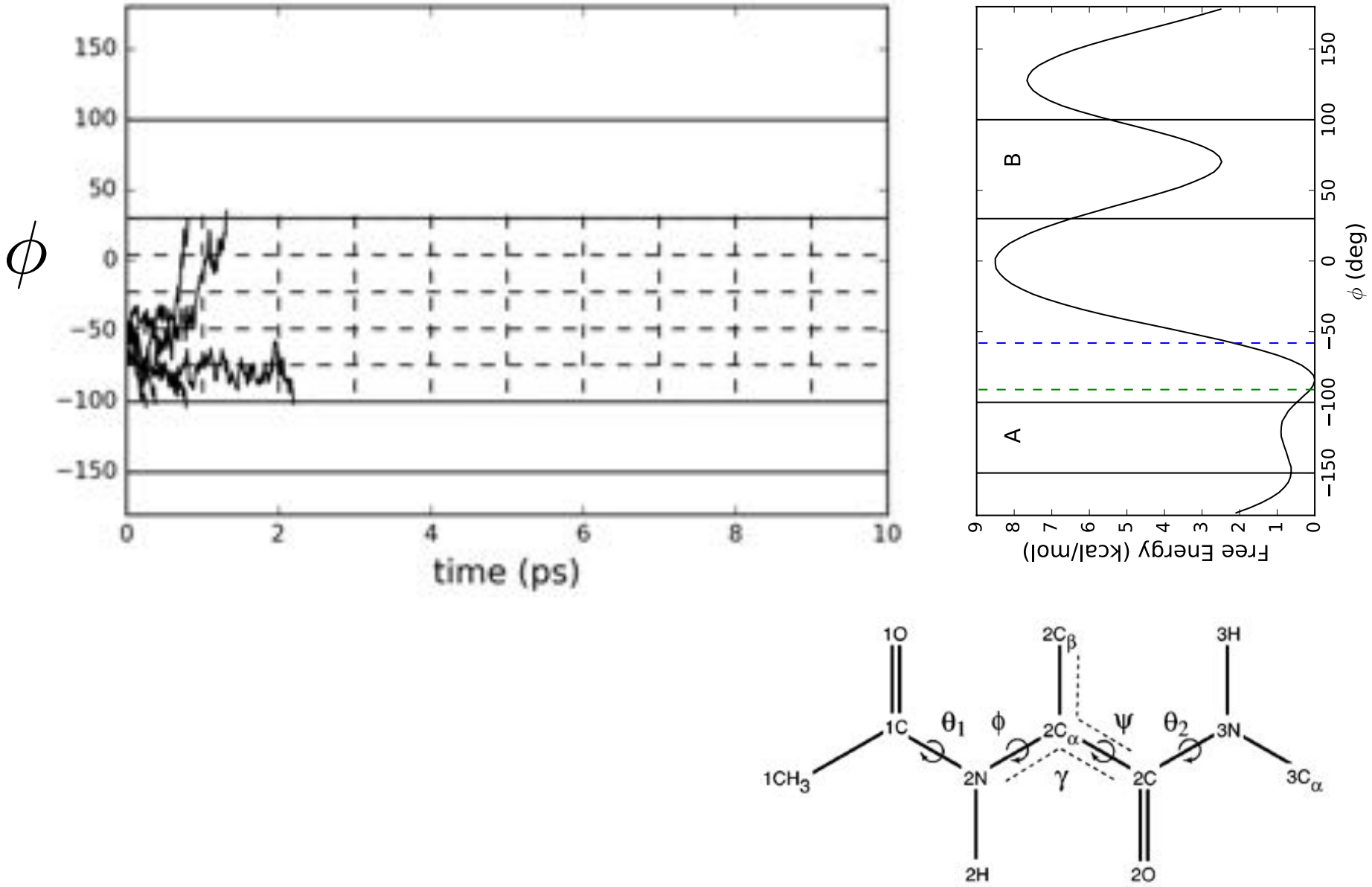




Example: Compute probability of hitting set B before set A for times less than τ_{max} (direct shooting).

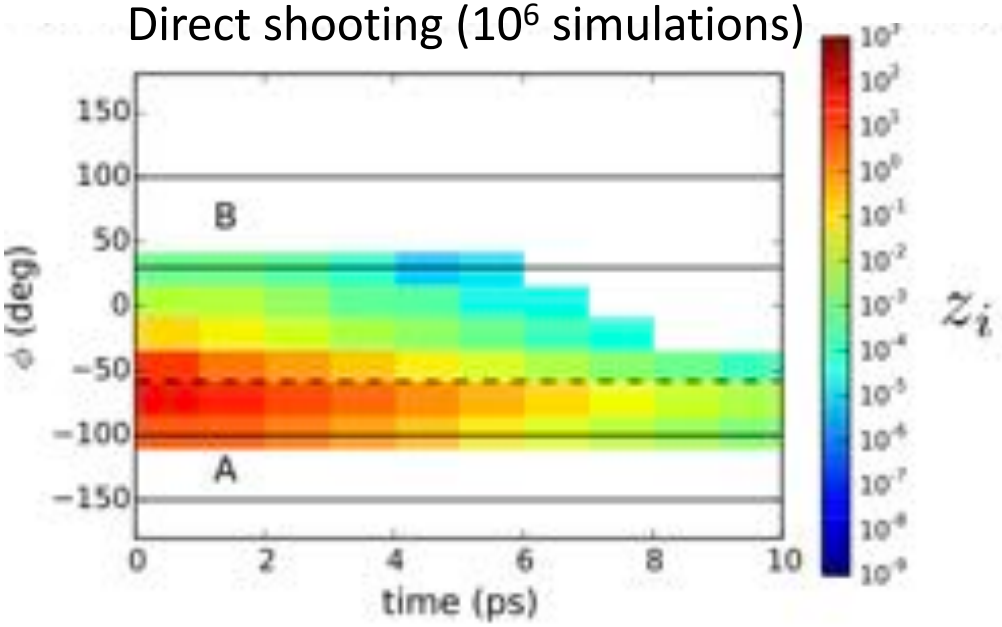


Example: Compute probability of hitting set B before set A for times less than τ_{max} (stratification).

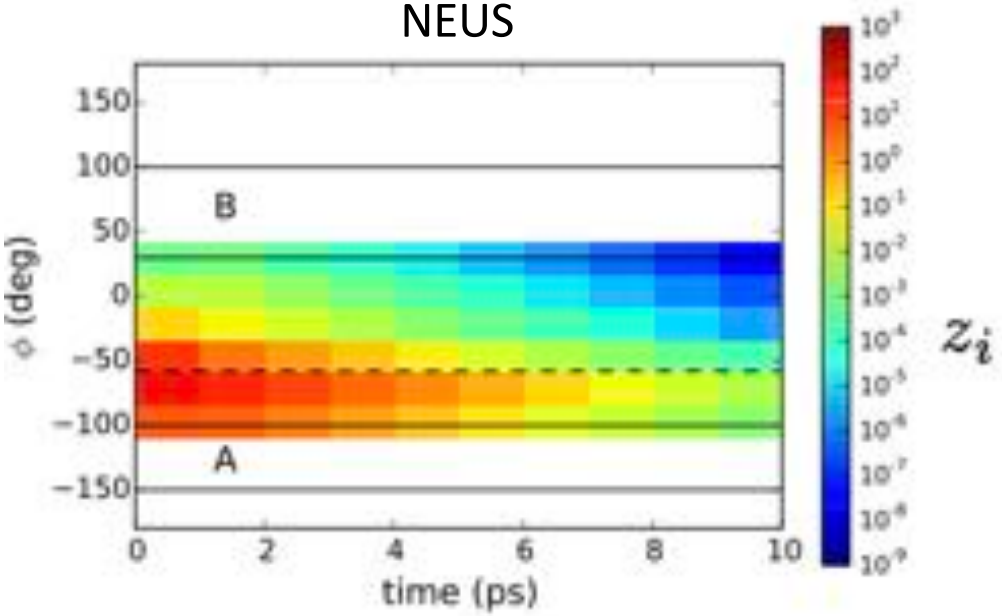


Example: Compute probability of hitting set B before set A for times less than τ_{max}.

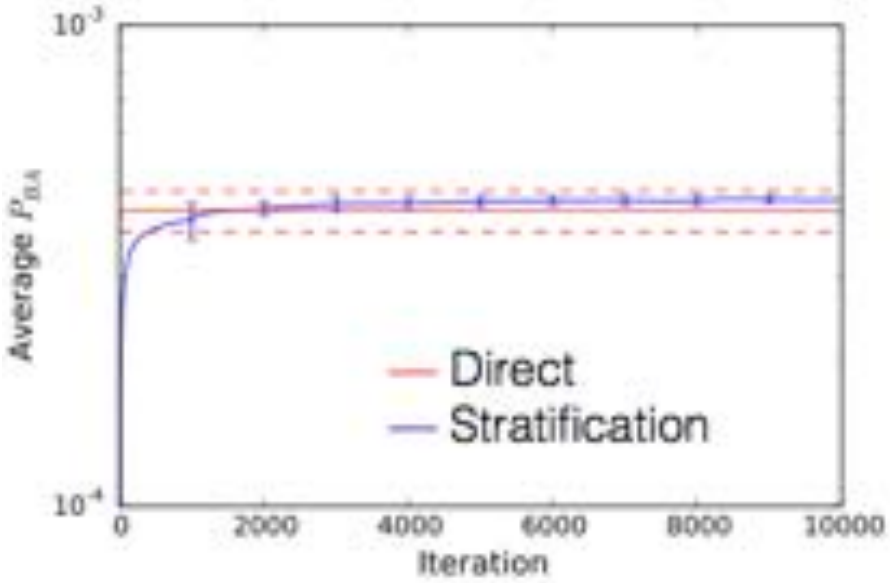
Direct shooting (10^6 simulations)



NEUS

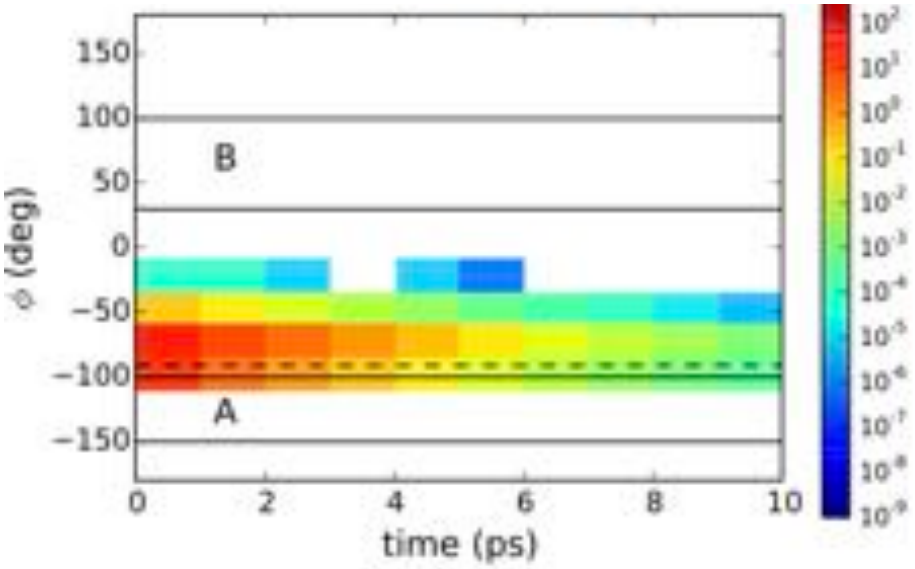


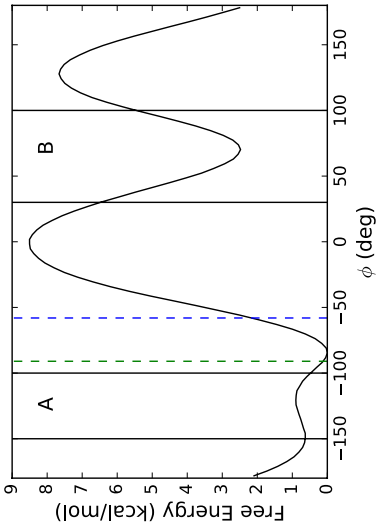
For a sufficiently easy problem, we can verify that we obtain the same result as direct simulation.



Example: Compute probability of hitting set B before set A for times less than τ_{max}.

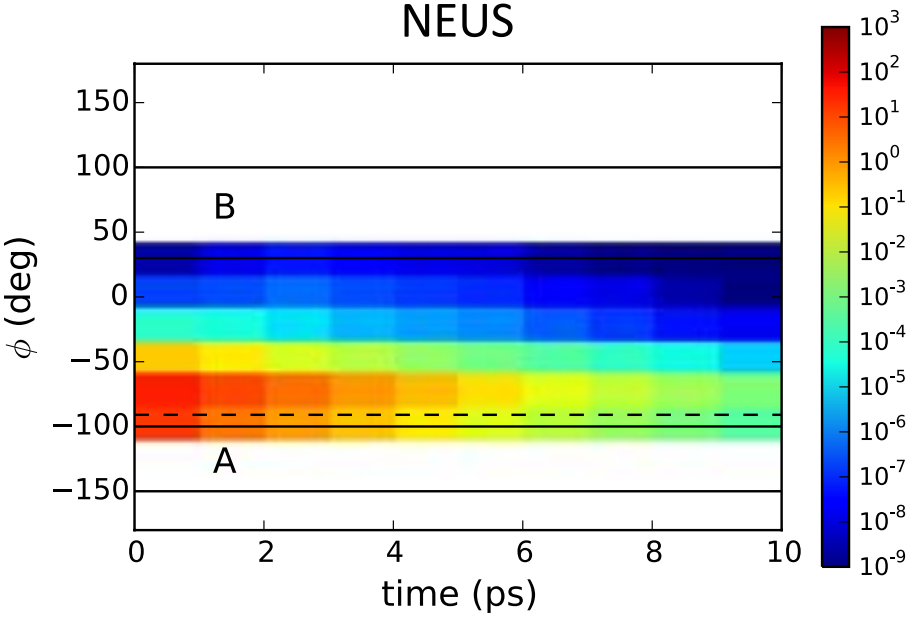
Direct shooting (10^6 simulations)

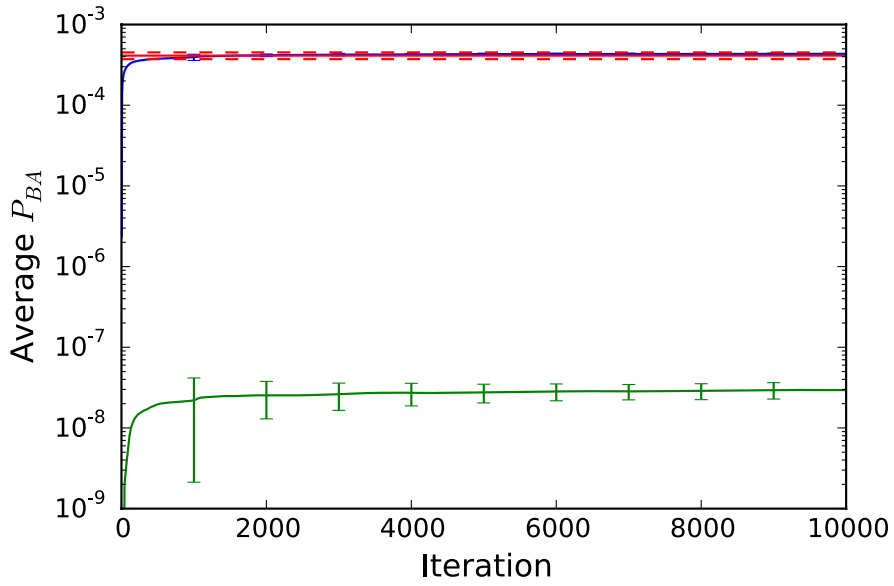




If we shift the initial condition to make the problem harder, direct integration becomes impossible but NEUS costs the same.

NEUS



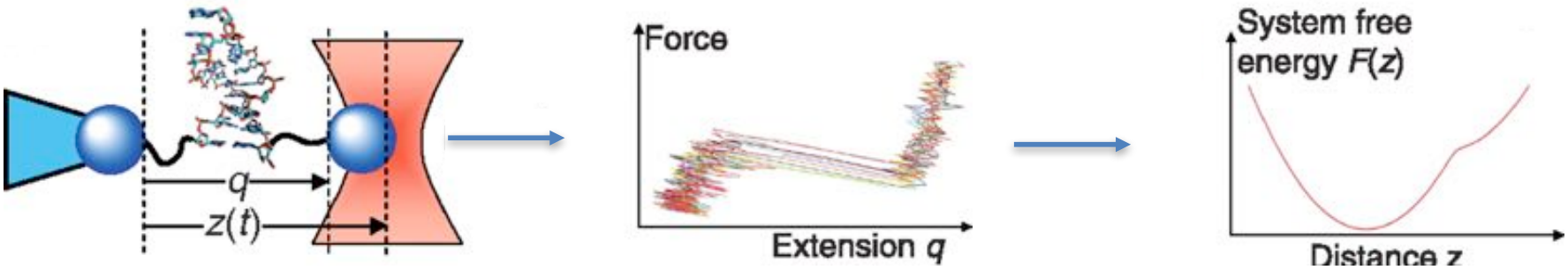


Example: Compute a free energy difference by stratifying the irreversible work, a path-dependent variable

Jarzynski's equality provides an exact relationship between the exponential of the accumulated work, $W^{(t)}$, in a driven process and the free energy difference.

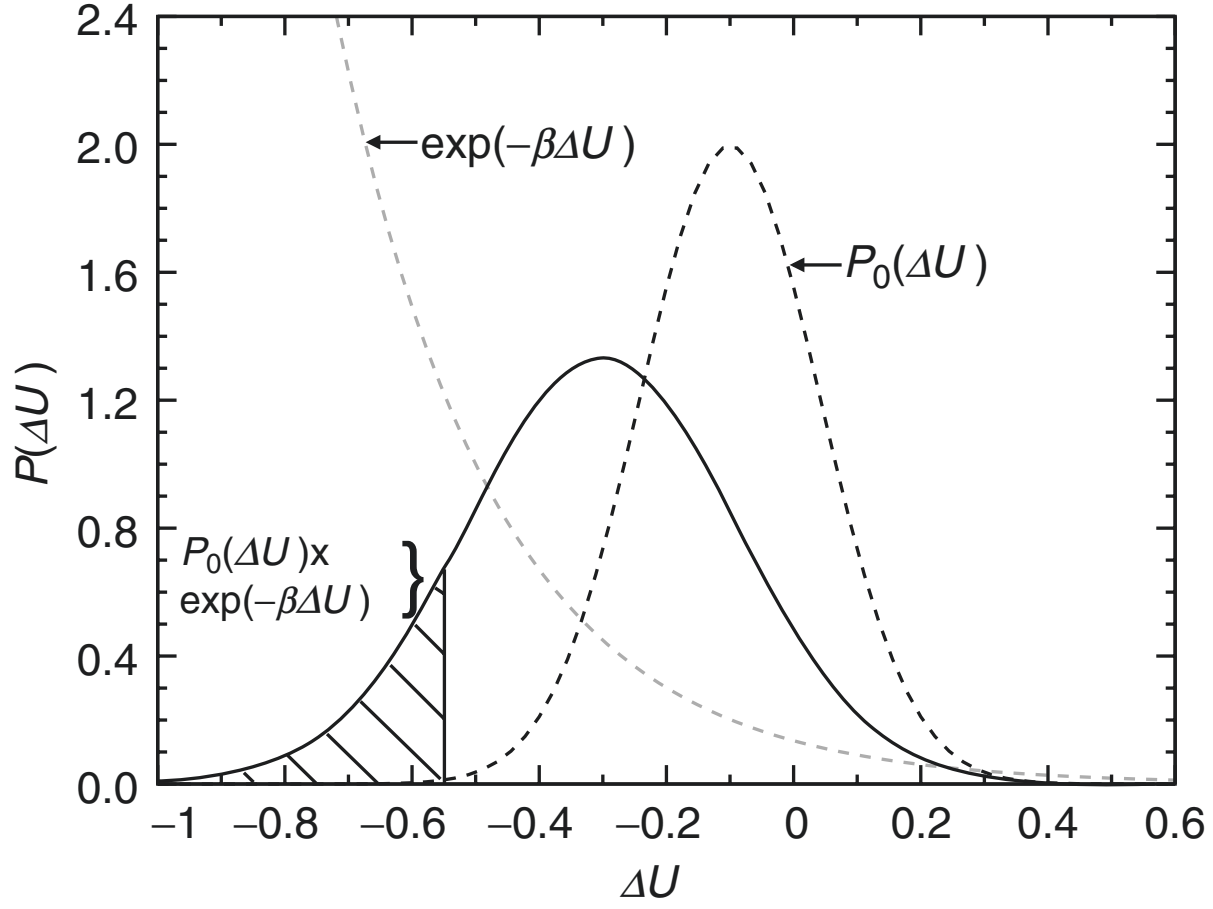
$$e^{-\beta \Delta F} = \left\langle e^{-\beta W^{(t)}} \right\rangle$$

This equality has been useful in measuring free energies in single-molecule force spectroscopy experiments as well as providing a computational strategy for computing free energies from steered molecular dynamics simulations.



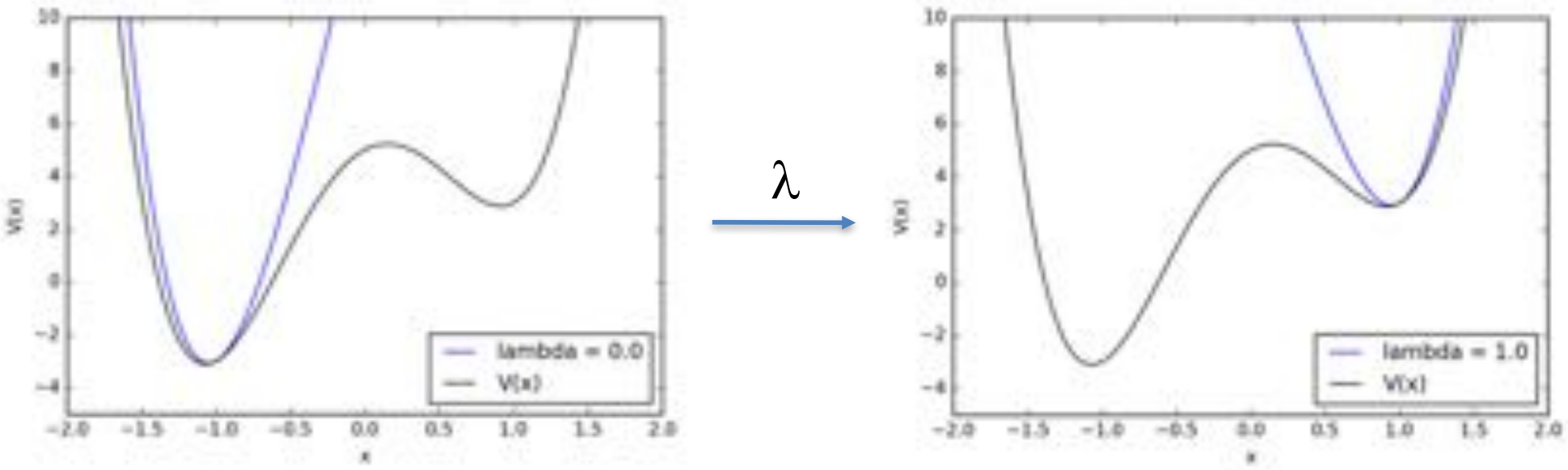
Example: Compute a free energy difference by stratifying the irreversible work, a path-dependent variable

In practice, such simulations suffer from severe statistical errors in the fast-switching regime because the low-work tail dominates the average.



Chipot and Pohorille, Free Energy Simulations, Springer, 2007.

Example: Compute a free energy difference by stratifying the irreversible work, a path-dependent variable

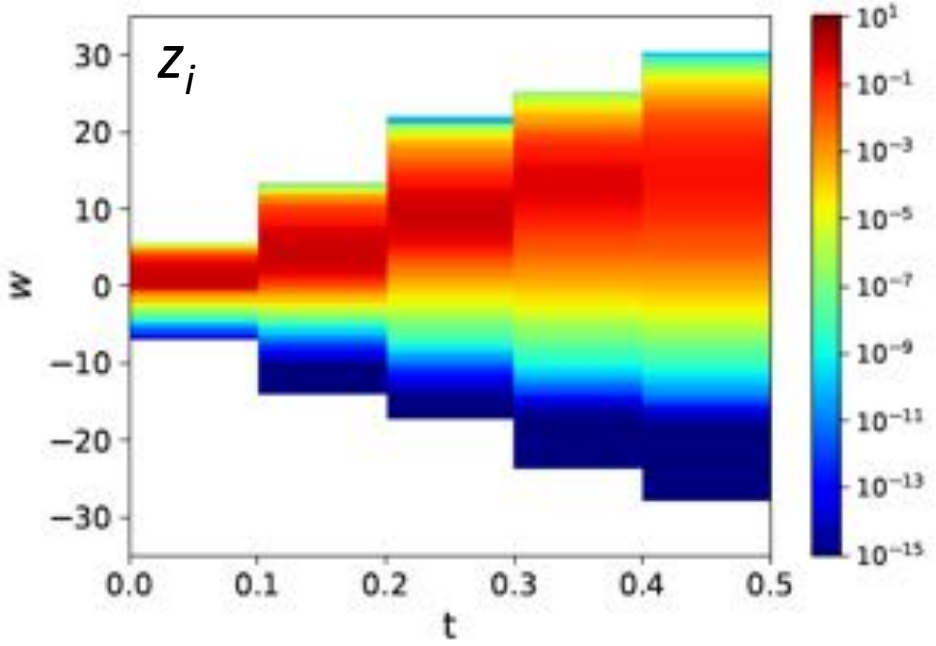


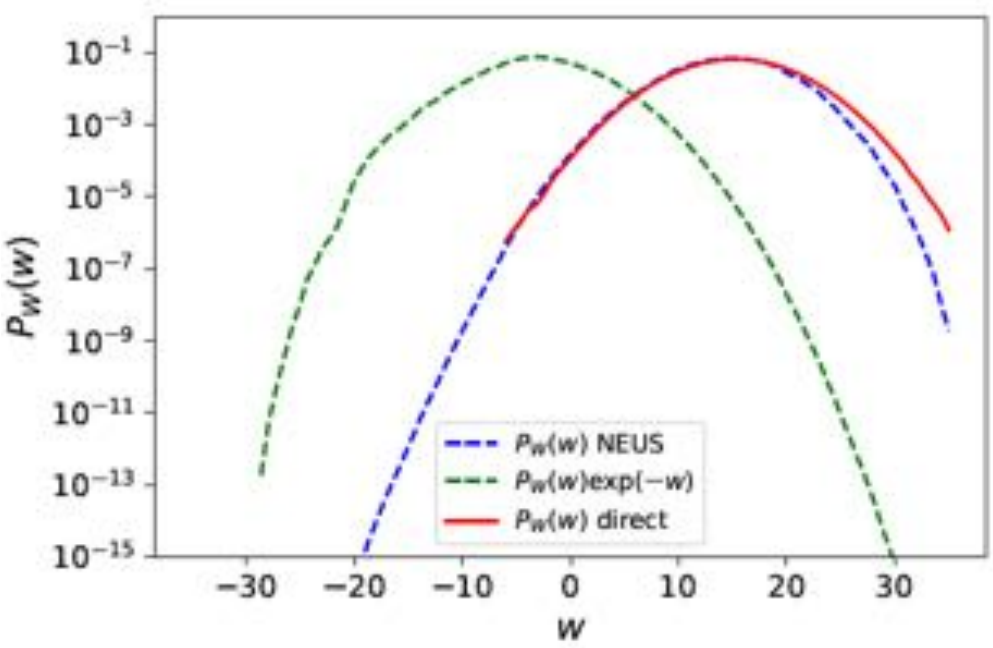
$$V(x,\lambda) = 5(x^2-1)^2 + 3x + 10(x-(2\lambda-1))^2$$

$$W^{(t)} = \sum_{\ell=0}^{t-1} V(\ell+1, X^{(\ell)}) - V(\ell, X^{(\ell)}), \qquad W^{(0)} = 0$$

$$\Delta F \approx -kT \ln \left[\frac{1}{N} \sum_{i=1}^{N} \exp(-\beta W_i^{(t)}) \right]$$

Example: Compute a free energy difference by stratifying the irreversible work, a path-dependent variable



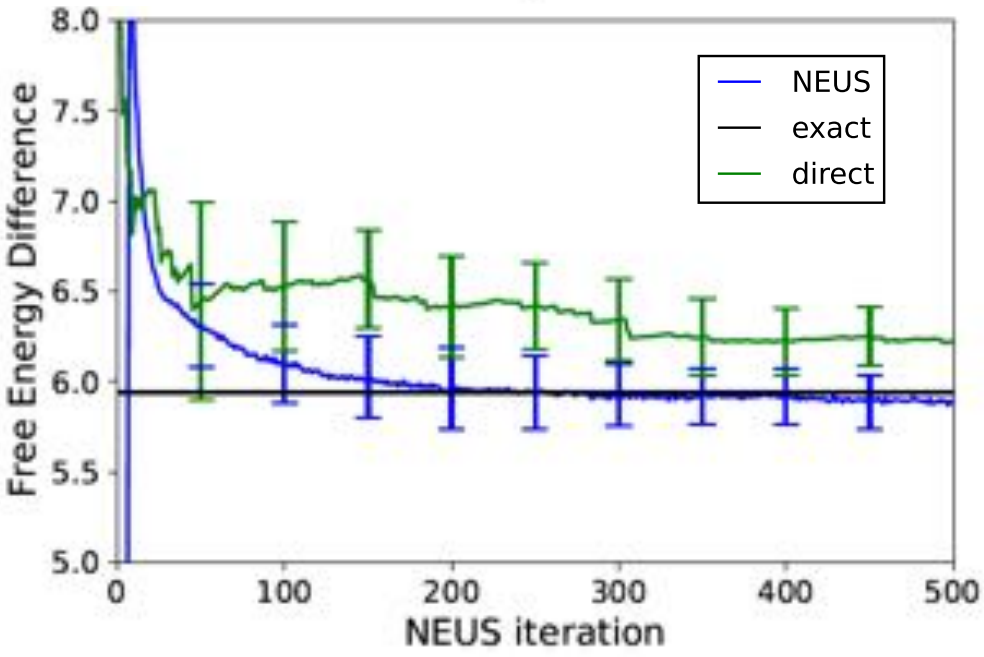


Direct

$$\Delta F \approx -kT \ln \left[\frac{1}{N} \sum_{i=1}^{N} \exp(-\beta W_i^{(t)}) \right]$$

NEUS

$$\langle \exp(-\beta W^{(t)}) \rangle = \sum_{i=1}^{n} z_i \langle \exp(-\beta W^{(t)}) \rangle_i$$



NEUS Summary

- We cast NEUS as an affine eigenproblem with a structure that parallels equilibrium US.

$$z^{\mathrm{T}} G + a^{\mathrm{T}} = z^{\mathrm{T}}$$

- Solving it involves fixed point iteration of window-to-window transition probabilities and entry-point distributions (reflecting that NEUS is both a splitting and a stratification method).

$$\big(\mathcal{G}(\bar{G},\gamma),\Gamma(\bar{G},\gamma)\big) = (\bar{G},\gamma)$$

- Present algorithm allows sampling new classes of dynamical averages.