

Stratification for Rare Event Problems

Jonathan Weare

Aaron Dinner, Andrey Kravstov, Charles Matthews, Jonathan Mattingly,
Jeremy Tempkin, Erik Thiede, and Brian Van Koten

Department of Statistics &
James Franck Institute
University of Chicago

April 27th, 2017

The rare event problem

Can sample $\pi \propto e^{-V/\varepsilon}$

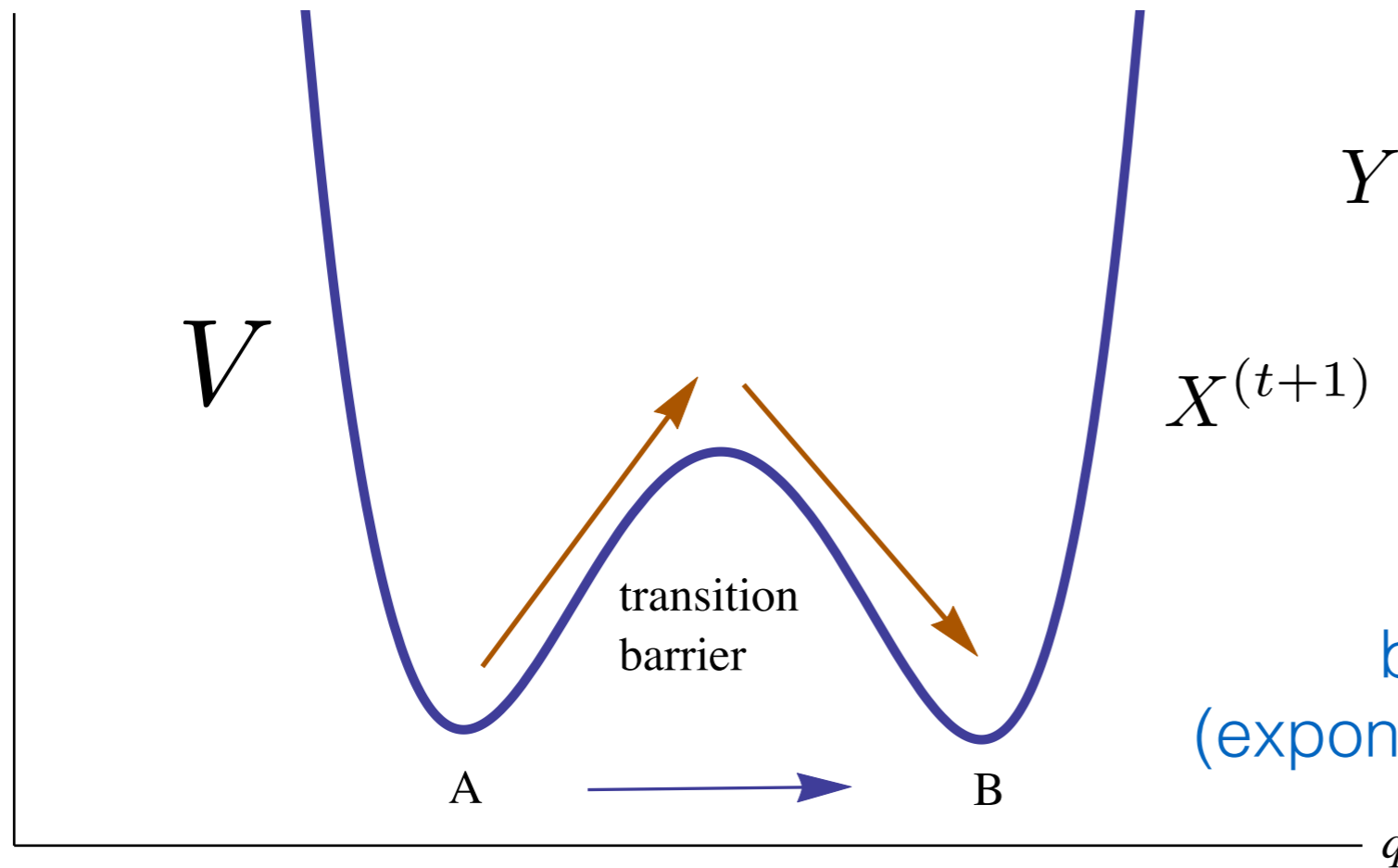
with e.g. Metropolis:

$$Y^{(t+1)} = X^{(t)} + \mathcal{N}(0, \sigma)$$

$$X^{(t+1)} = \begin{cases} Y^{(t+1)}, & \text{w.p. } 1 \wedge \frac{\pi(Y^{(t+1)})}{\pi(X^{(t)})} \\ X^{(t)}, & \text{otherwise} \end{cases}$$

but very slow convergence
(exponentially in ε^{-1}) for most averages

Energy

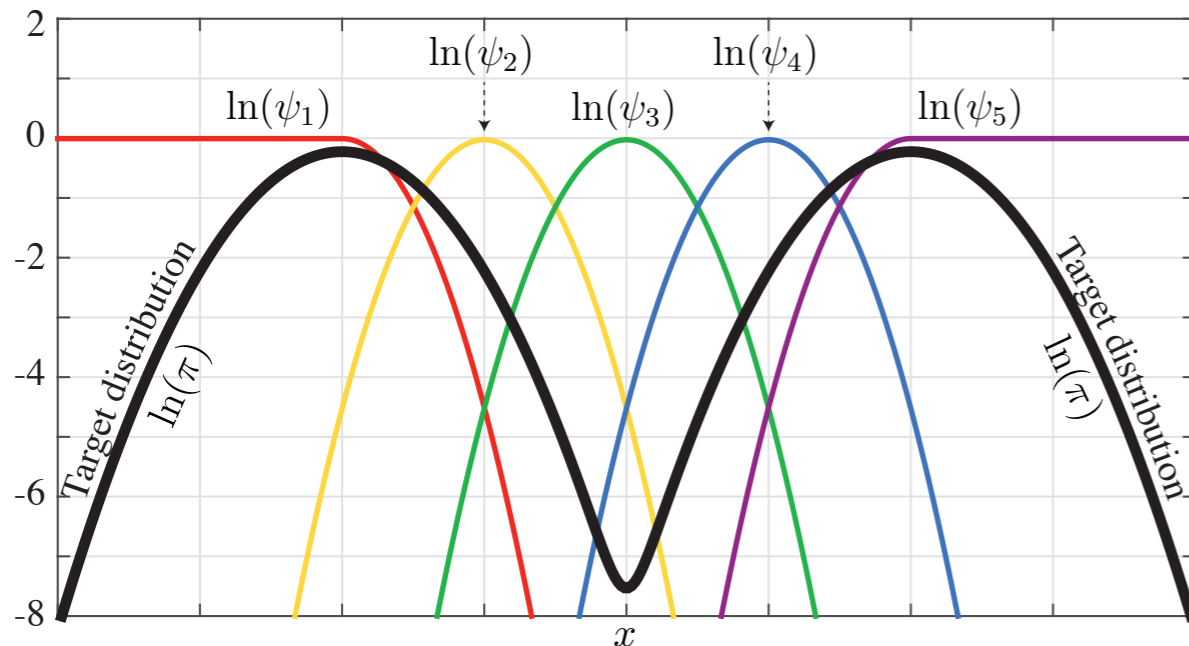


G. M. Torrie and J. P. Valleau, *J. Comp. Phys.* **23**, 187–199 (1977)

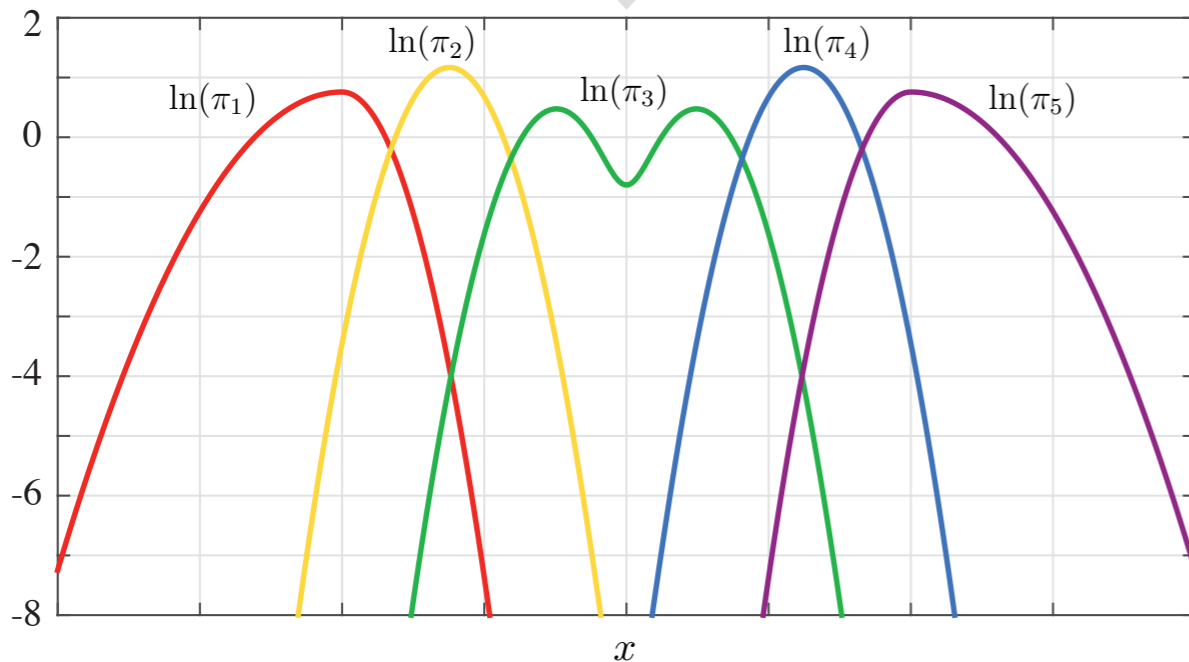
C. Pangali, M. Rao, and B. J. Berne, *J. Chem. Phys.* **71**, 2975-2981 (1979)

E. Thiede, B. Van Koten, A. Dinner, and **JW**, *J. Chem. Phys.* **145**(8), 084115 (2016)

Umbrella Sampling



↓
Biasing



1. Introduce “bias” functions

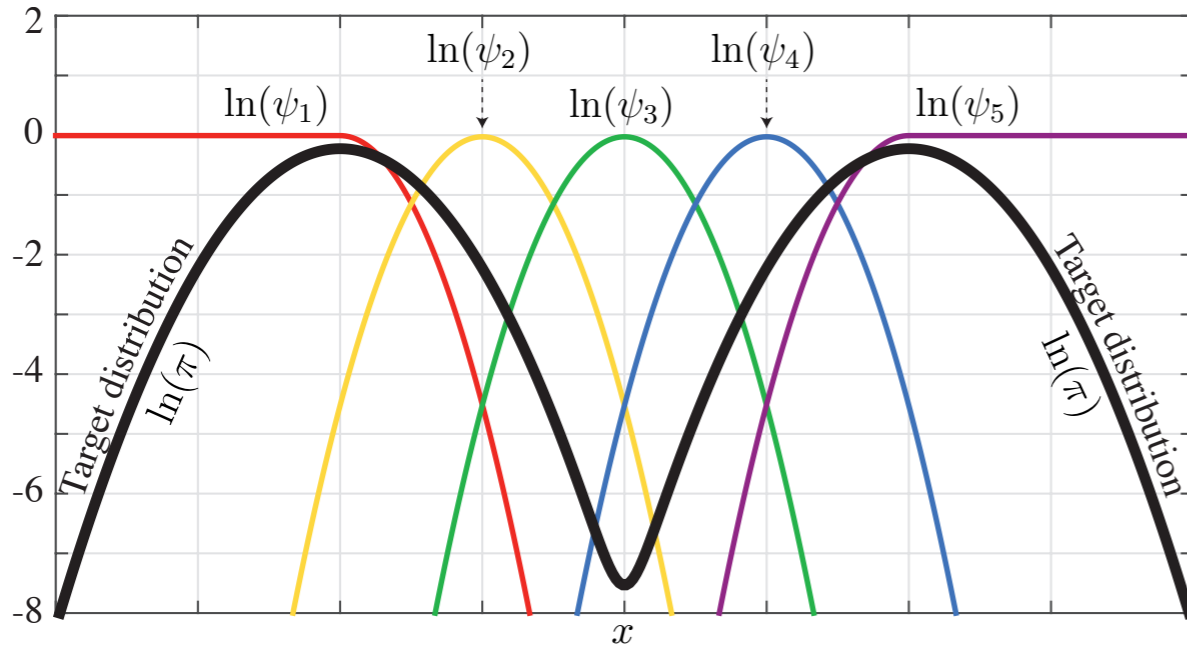
$$\psi_i \geq 0 \quad \sum_i \psi_i = 1$$

G. M. Torrie and J. P. Valleau, *J. Comp. Phys.* **23**, 187–199 (1977)

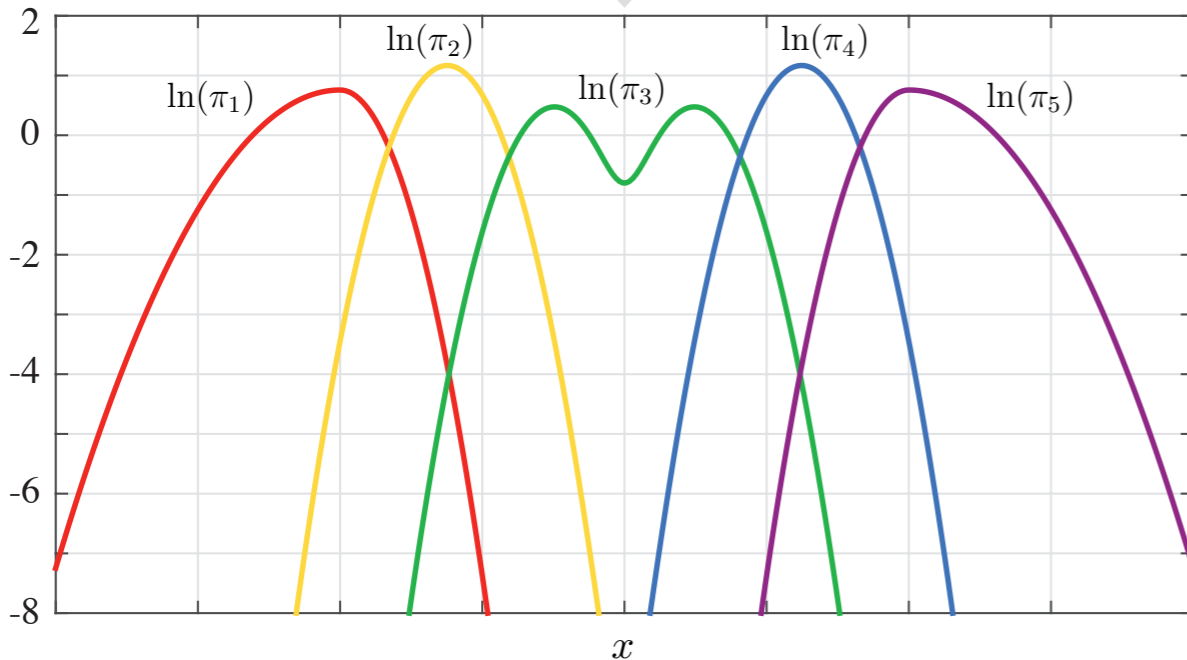
C. Pangali, M. Rao, and B. J. Berne, *J. Chem. Phys.* **71**, 2975-2981 (1979)

E. Thiede, B. Van Koten, A. Dinner, and **JW**, *J. Chem. Phys.* **145**(8), 084115 (2016)

Umbrella Sampling



Biassing



1. Introduce “bias” functions

$$\psi_i \geq 0 \quad \sum_i \psi_i = 1$$

2. Define “restrained” distributions

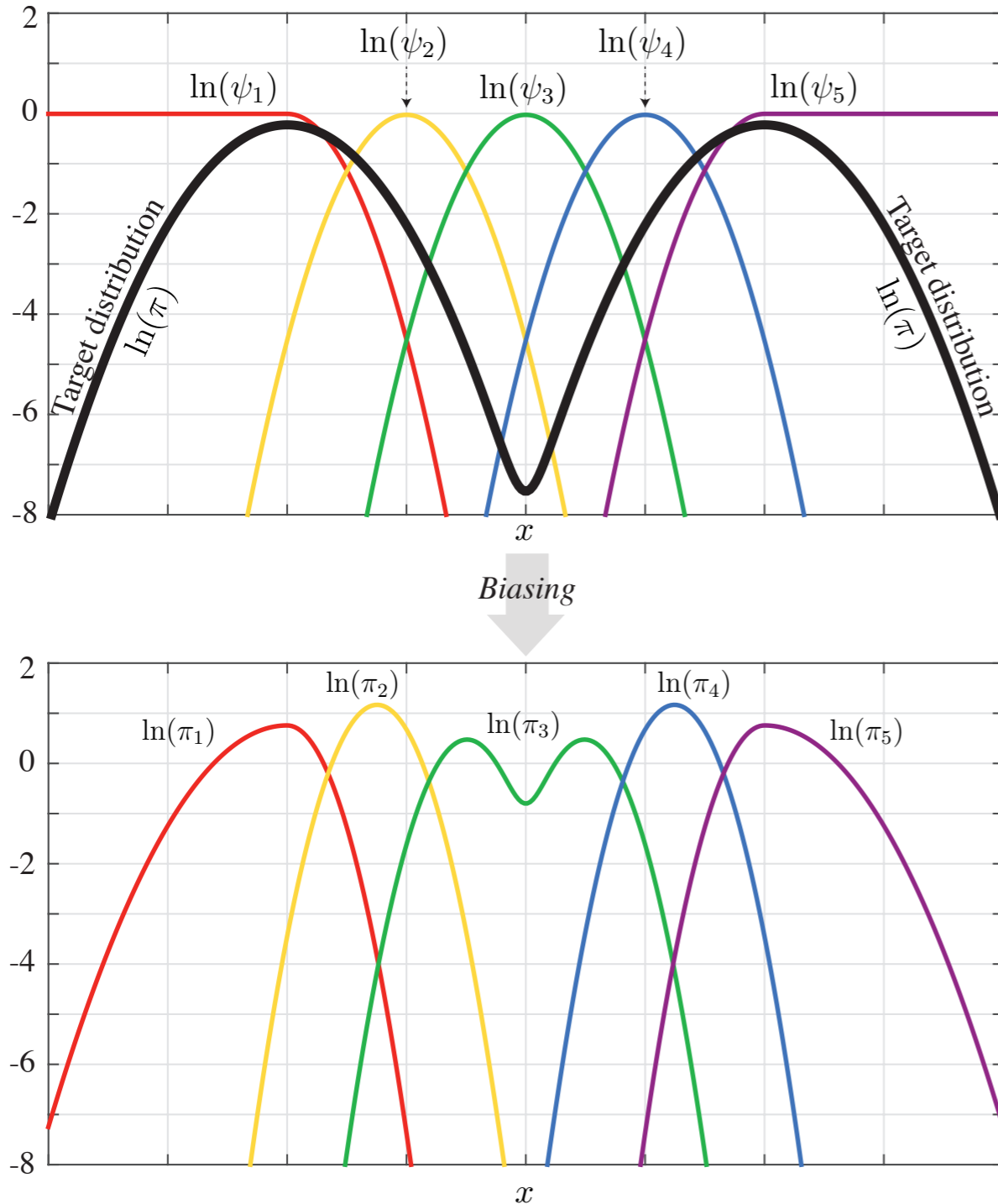
$$\pi_i \propto \psi_i \pi$$

G. M. Torrie and J. P. Valleau, *J. Comp. Phys.* **23**, 187–199 (1977)

C. Pangali, M. Rao, and B. J. Berne, *J. Chem. Phys.* **71**, 2975-2981 (1979)

E. Thiede, B. Van Koten, A. Dinner, and **JW**, *J. Chem. Phys.* **145**(8), 084115 (2016)

Umbrella Sampling



1. Introduce “bias” functions

$$\psi_i \geq 0 \quad \sum_i \psi_i = 1$$

2. Define “restrained” distributions

$$\pi_i \propto \psi_i \pi$$

3. Define normalization constants

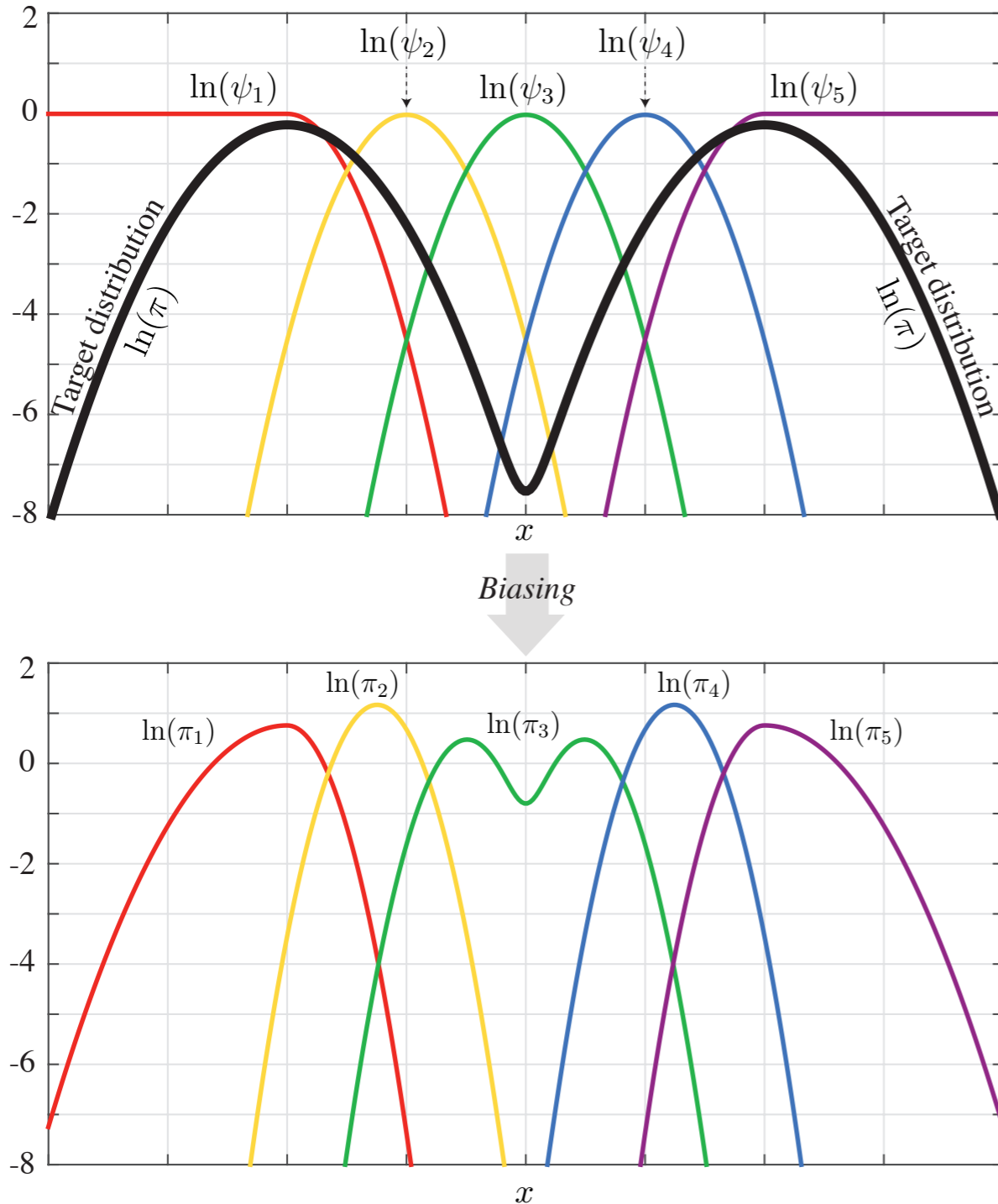
$$z_i = \int \psi_i(x) \pi(dx)$$

G. M. Torrie and J. P. Valleau, *J. Comp. Phys.* **23**, 187–199 (1977)

C. Pangali, M. Rao, and B. J. Berne, *J. Chem. Phys.* **71**, 2975-2981 (1979)

E. Thiede, B. Van Koten, A. Dinner, and **JW**, *J. Chem. Phys.* **145**(8), 084115 (2016)

Umbrella Sampling



1. Introduce “bias” functions

$$\psi_i \geq 0 \quad \sum_i \psi_i = 1$$

2. Define “restrained” distributions

$$\pi_i \propto \psi_i \pi$$

3. Define normalization constants

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

4. Assemble general averages

$$\int f(x) \pi(dx) = \sum_i z_i \int f(x) \pi_i(dx)$$

G. M. Torrie and J. P. Valleau, *J. Comp. Phys.* **23**, 187–199 (1977)

C. Pangali, M. Rao, and B. J. Berne, *J. Chem. Phys.* **71**, 2975-2981 (1979)

E. Thiede, B. Van Koten, A. Dinner, and **JW**, *J. Chem. Phys.* **145**(8), 084115 (2016)

Umbrella Sampling

$$\int f(x) \pi(dx) = \sum_i z_i \int f(x) \pi_i(dx)$$

so

$$z_j = \int \psi_j(x) \pi(dx) = \sum_i z_i \int \psi_j(x) \pi_i(dx)$$

i.e. the weights $\{z_i\}$ solve the **eigenproblem**

$$z^T = z^T F, \quad \sum_i z_i = 1$$

with

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

Umbrella Sampling

1. Use your favorite MC scheme to sample from π_i
2. Estimate the entries of F and each $\int f(x)\pi_i(dx)$
3. Solve the eigenproblem for Z
4. Assemble $\int f(x)\pi(dx) = \sum_i z_i \int f(x)\pi_i(dx)$

A detailed analysis

We establish a central limit theorem with a detailed expression for the asymptotic variance along with a general bound:

$$\sigma^2(f) \leq 2 \sum_i \frac{1}{\kappa_i} \left\{ \begin{array}{l} \text{contribution from} \\ \text{restricted average of } f \\ \tau_i z_i^2 \\ \text{contribution from} \\ \text{estimate of } F \\ \pi \|f\|^2 \text{tr}(R^i) \sum_{\substack{j \neq i \\ F_{ij} \neq 0}} \frac{\text{var}_{\pi_i}(\psi_j)}{\mathbf{P}_i[t_j < t_i]} \end{array} \right\}$$

$$R_{jk}^i = \frac{\sigma_{jk}^i}{\sqrt{\text{var}_{\pi_i}(\psi_j)} \sqrt{\text{var}_{\pi_i}(\psi_k)}}$$

σ_{jk}^i is the asymptotic covariance of the estimators of F_{ij} and F_{ik}

$\mathbf{P}_i[t_j < t_i]$ is the probability that the chain defined by F and starting from index i and hitting index j before returning to i

An illustrative example

Consider the probability vector:

$$\pi(i) := \frac{\exp(-LV(i/L))}{\sum_{k=1}^L \exp(-LV(k/L))}$$

$$\text{with } V(x) = \frac{1}{4\pi} \cos(4\pi x)$$

on the periodic 1D lattice $\mathbb{Z}/L\mathbb{Z}$.

Nearest neighbor Metropolis MCMC generates a chain with exponentially small spectral gap in L

An illustrative example

Consider umbrella sampling with $\psi_i(j) = \frac{1}{2}(\delta_i(j) + \delta_{i+1}(j))$

$$\text{Then } F_{ij} = \begin{cases} \frac{1}{2} \frac{\pi(i+1)}{\pi(i) + \pi(i+1)}, & j = i + 1 \\ \frac{1}{2} \frac{\pi(i)}{\pi(i) + \pi(i+1)}, & j = i - 1 \\ \frac{1}{2}, & j = i \\ 0, & \text{otherwise} \end{cases}$$

$$\text{and } z_i = \frac{1}{2}(\pi(i) + \pi(i + 1))$$

We can sample $\pi_i(j) = \frac{\pi(i)\delta_i(j) + \pi(i + 1)\delta_{i+1}(j)}{\pi(i) + \pi(i + 1)}$ e.g. by Metropolis

with simple random walk proposals

An illustrative example

The inequality:
$$\left| \log \frac{\pi(i)}{\pi(i+1)} \right| = L \left| V \left(\frac{i+1}{L} \right) - V \left(\frac{i}{L} \right) \right| \leq 1$$

implies that sampling of π_i is not hard, i.e.:

$$\text{tr}(R^i) \sim \frac{1}{\# \text{ of MCMC steps in window } i}$$

and also that

$$F_{i,i\pm 1} = \frac{1}{2} \frac{1}{1 + \exp \left[\pm L \left(V \left(\frac{i+1}{L} \right) - V \left(\frac{i}{L} \right) \right) \right]} \geq \frac{1}{2} \frac{1}{1 + e}$$

An illustrative example

And $\mathbf{P}_i[t_j < t_i] \geq F_{ij}$ implies that

$$\sum_{\substack{j \neq i \\ F_{ij} \neq 0}} \frac{\text{var}_{\pi_i}(\psi_j)}{\mathbf{P}_i[t_j < t_i]} \sim 1$$

So

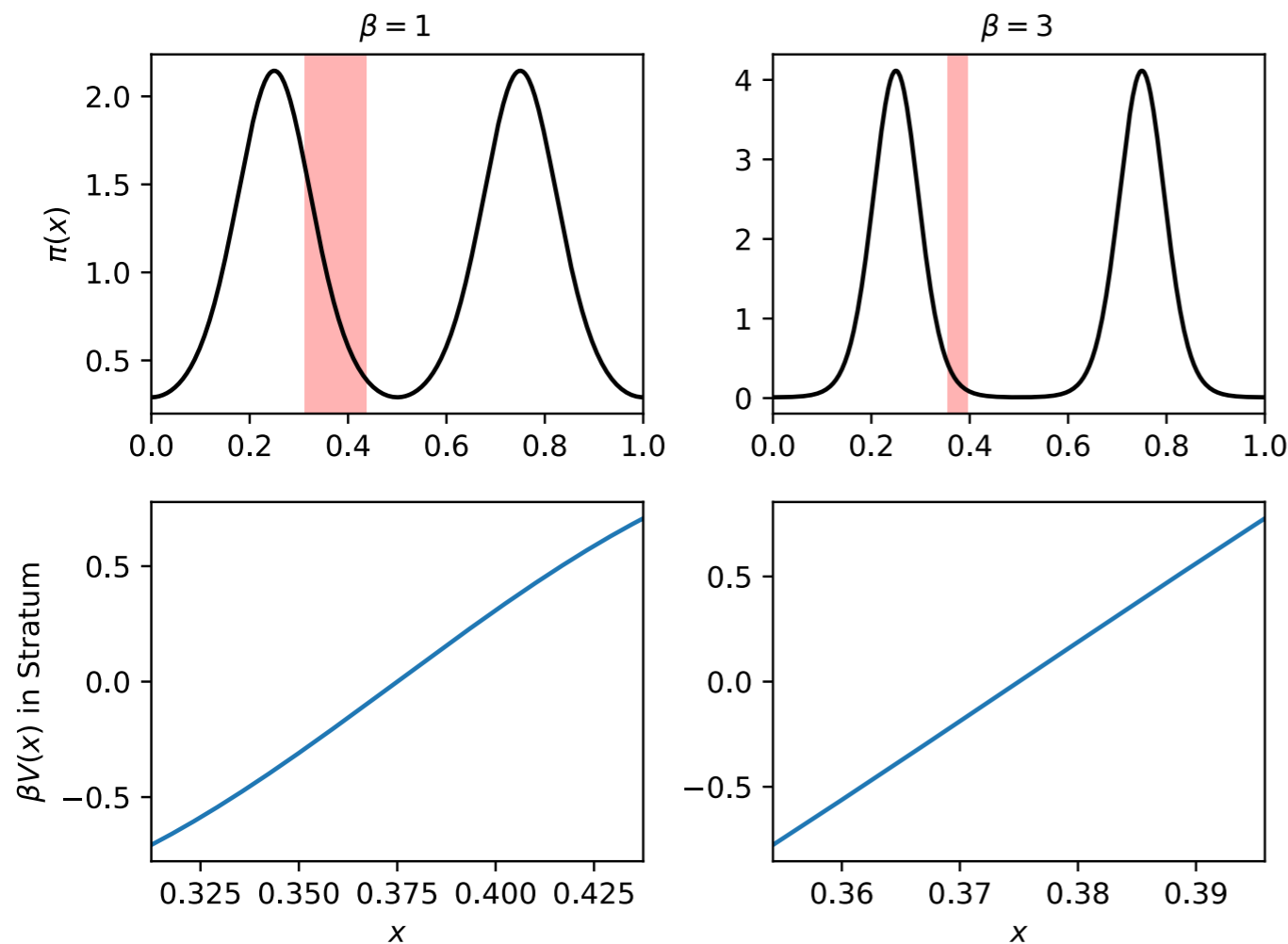
$$\sigma^2(f) \sim \frac{L}{\# \text{ of MCMC steps per window}}$$

and total cost to achieve fixed accuracy scales about like L^2

i.e. not exponentially!

A detailed analysis

US can reduce cost to sample $\pi \propto e^{-V/\varepsilon}$ from exponential in ε^{-1} to algebraic.



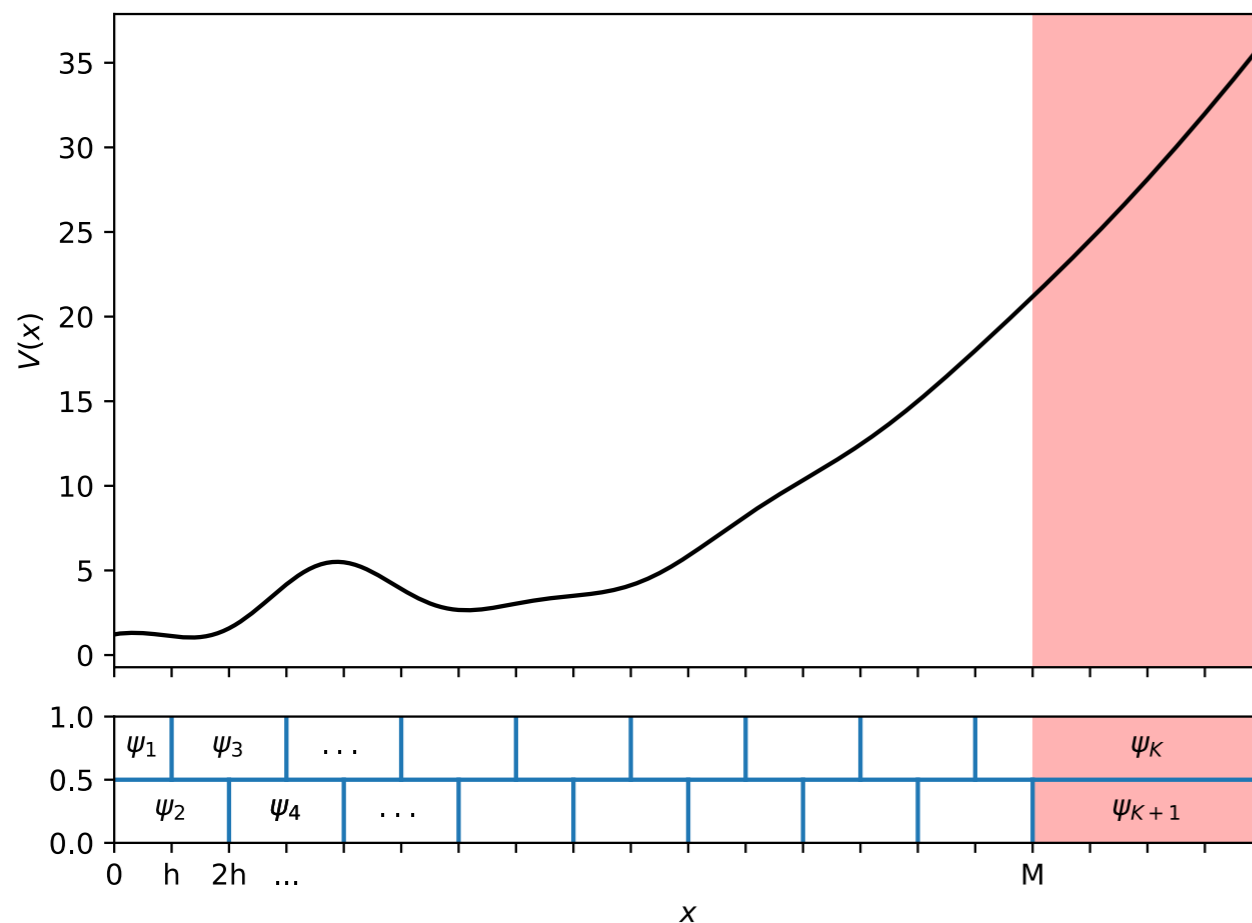
The restrained distributions can be much easier to sample. E.g. if ψ_i is constant on a strip Z_h of width h and if π_i is sampled using reflected overdamped Langevin, then

$$\sigma_h^2(f) \leq \frac{\Lambda h^2 \text{var}_h(f)}{\varepsilon} \exp\left(\frac{\max_{Z_h} V - \min_{Z_h} V}{\varepsilon}\right)$$

But: You have to think about how the eigenproblem amplifies errors in F

A detailed analysis

US can reduce the cost to compute tail probabilities $\mathbf{P}[X \geq M]$ from exponential in M to algebraic.



Instead of trying to estimate the small probability with relative accuracy, US estimates the relatively large entries in F .

But: Again, you have to think about how the eigenproblem amplifies errors in F .

A key ingredient

We need to understand the error in the z_i resulting from sampling error in a stochastic matrix F

$$z^T = z^T F, \quad \sum_i z_i = 1$$

Existing perturbation bounds for Markov matrices blow up as (one over) the spectral gap of F

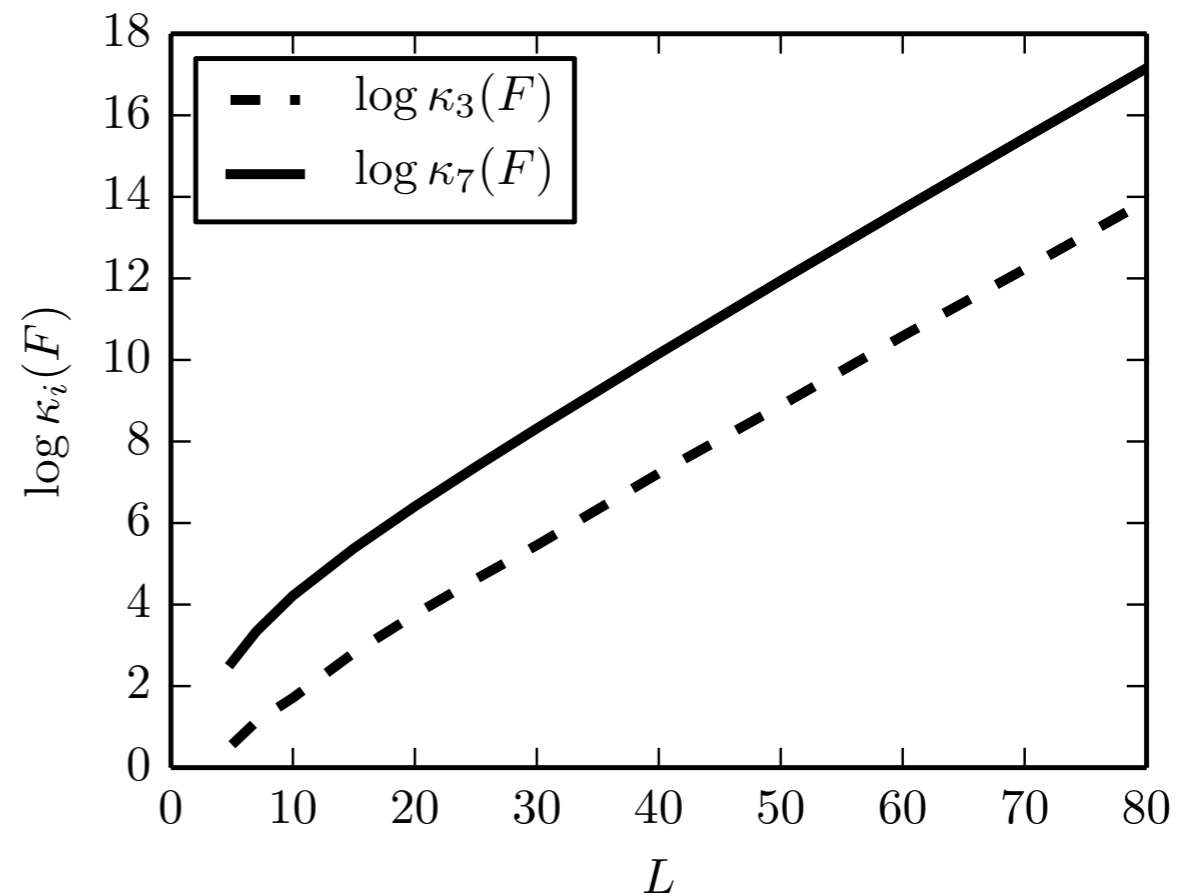
In US applications the spectral gap of F is typically extremely small.

A key ingredient

We are far from the first to consider this problem

Some bounds of the form $\|\tilde{z} - z\| \leq \kappa_i \|F - F\|$ (from [Chao and Meyer 2001](#)):

$\kappa_1 = \ Z\ _\infty$	Schweitzer [17]
$\kappa_2 = \ A^\#\ _\infty$	Meyer [12]
$\kappa_3 = \frac{\max_j (a_{jj}^\# - \min_i a_{ij}^\#)}{2}$	Haviv and van Heyden [5]
$\kappa_4 = \max_{i,j} a_{ij}^\# $	Kirkland et al. [9] Funderlic and Meyer [4]
$\kappa_5 = \frac{1}{1 - \tau_1(P)}$	Seneta [21]
$\kappa_6 = \tau_1(A^\#) = \tau_1(Z)$	Seneta [22]
$\kappa_7 = \frac{\min_j \ A_{(j)}^{-1}\ _\infty}{2}$	Ipsen and Meyer [6] Kirkland et al. [9]
$\kappa_8 = \frac{1}{2} \max_j \left[\frac{\max_{i \neq j} m_{ij}}{m_{jj}} \right]$	Cho and Meyer [1]



Existing bounds blow up as the spectral gap decreases

A key ingredient

We establish new more detailed bounds that show that error can be small even when spectral gap is small

If $z^T = z^T F$ and $\tilde{z}^T = \tilde{z}^T \tilde{F}$ with $F_{ij}, \tilde{F}_{ij} \geq S_{ij}$ for some sub-stochastic matrix S then

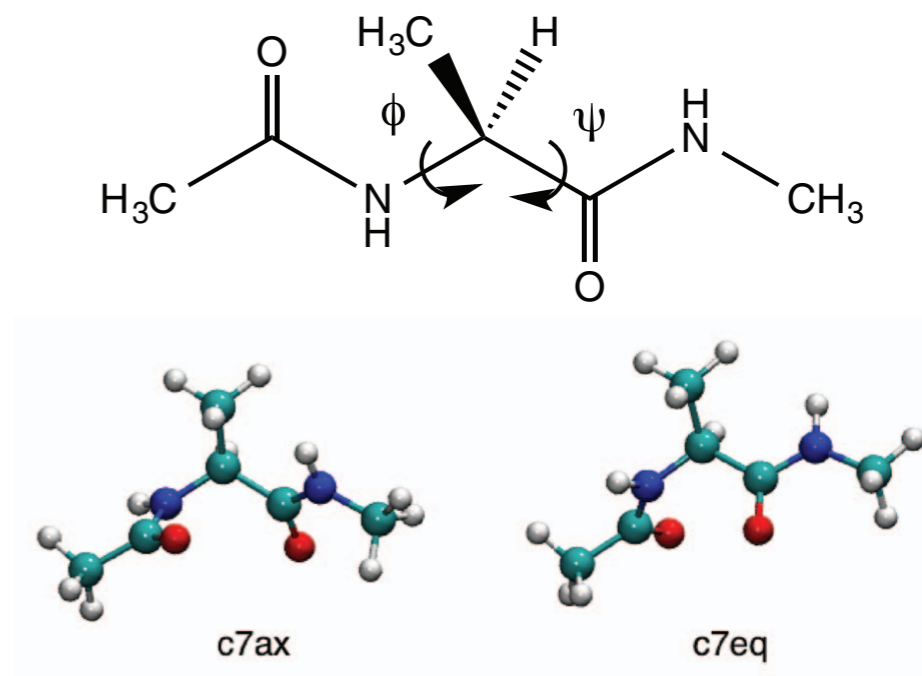
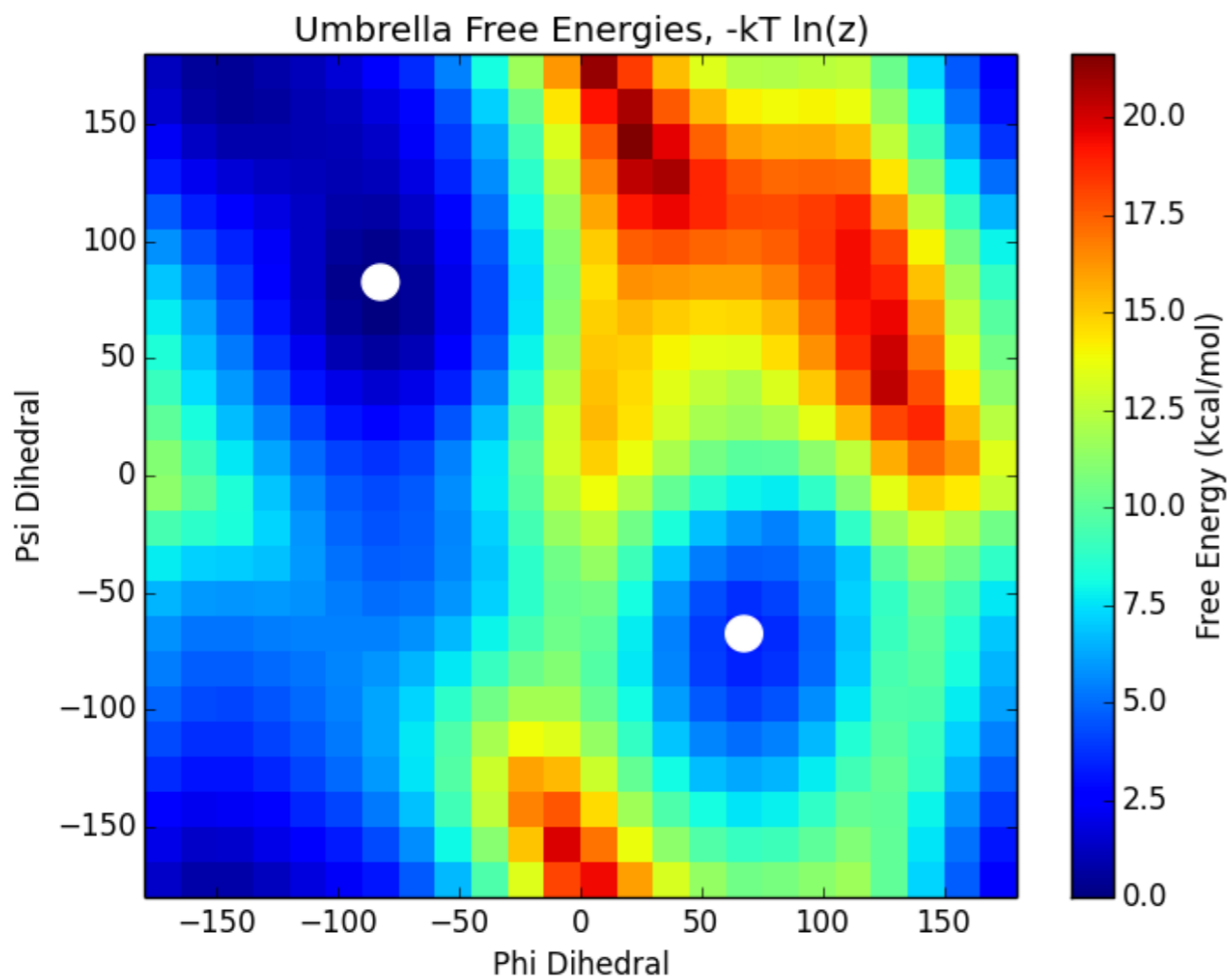
$$\max_k |\log z_k - \log \tilde{z}_k| \leq \sum_{i \neq j} |\log(F_{ij} + \gamma_{ij}) - \log(\tilde{F}_{ij} + \gamma_{ij})|$$

where

$$\gamma_{ij} = P_i(\tau_j < \min\{\tau_i, \tau\}) - S_{ij}$$

and τ_i and τ are respectively the first time to arrive at state i and absorption time for the absorbing Markov process governed by S

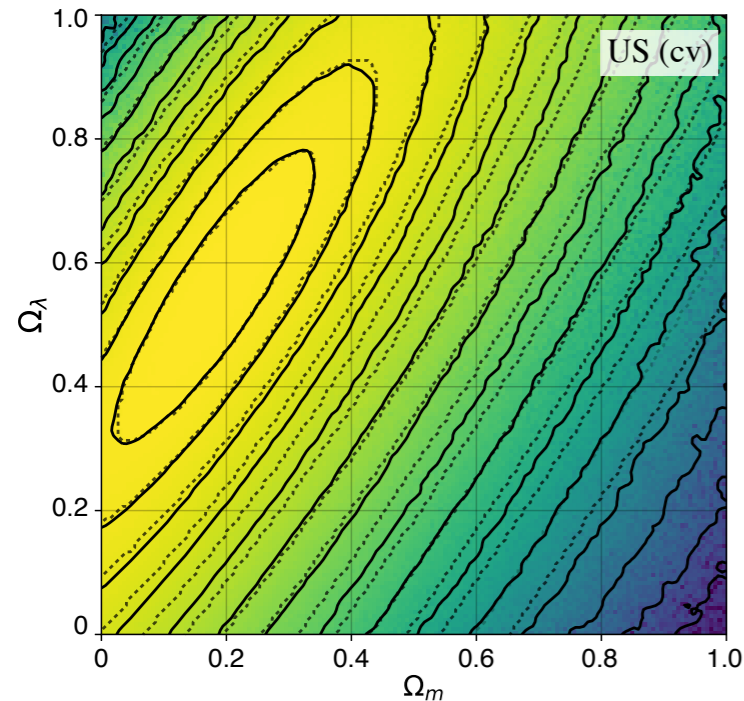
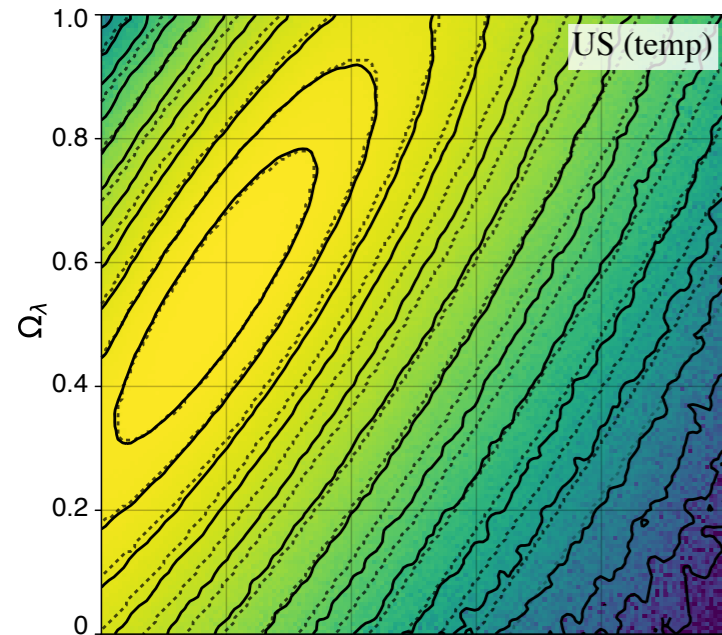
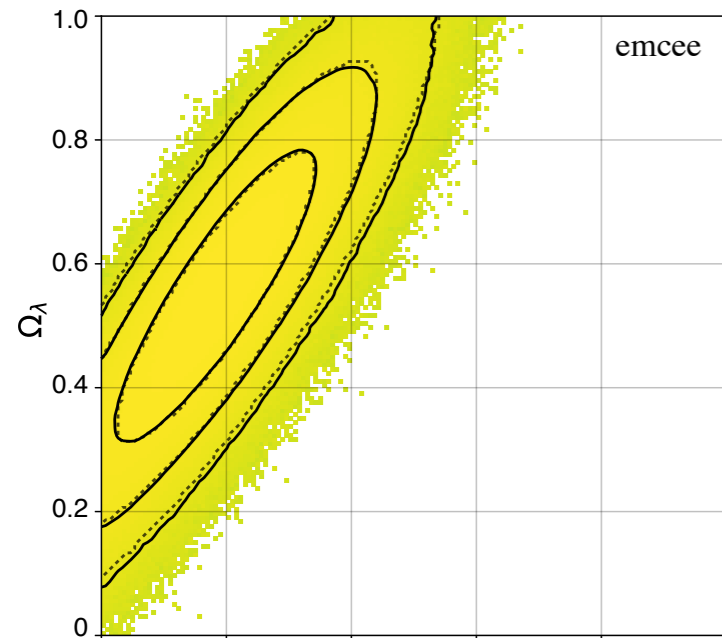
A typical US application: A free energy surface for the alanine dipeptide



Joint posterior of cosmological values

Ω_m : mean dimensionless matter

Ω_λ : vacuum energy density



US with two different choices of bias functions:

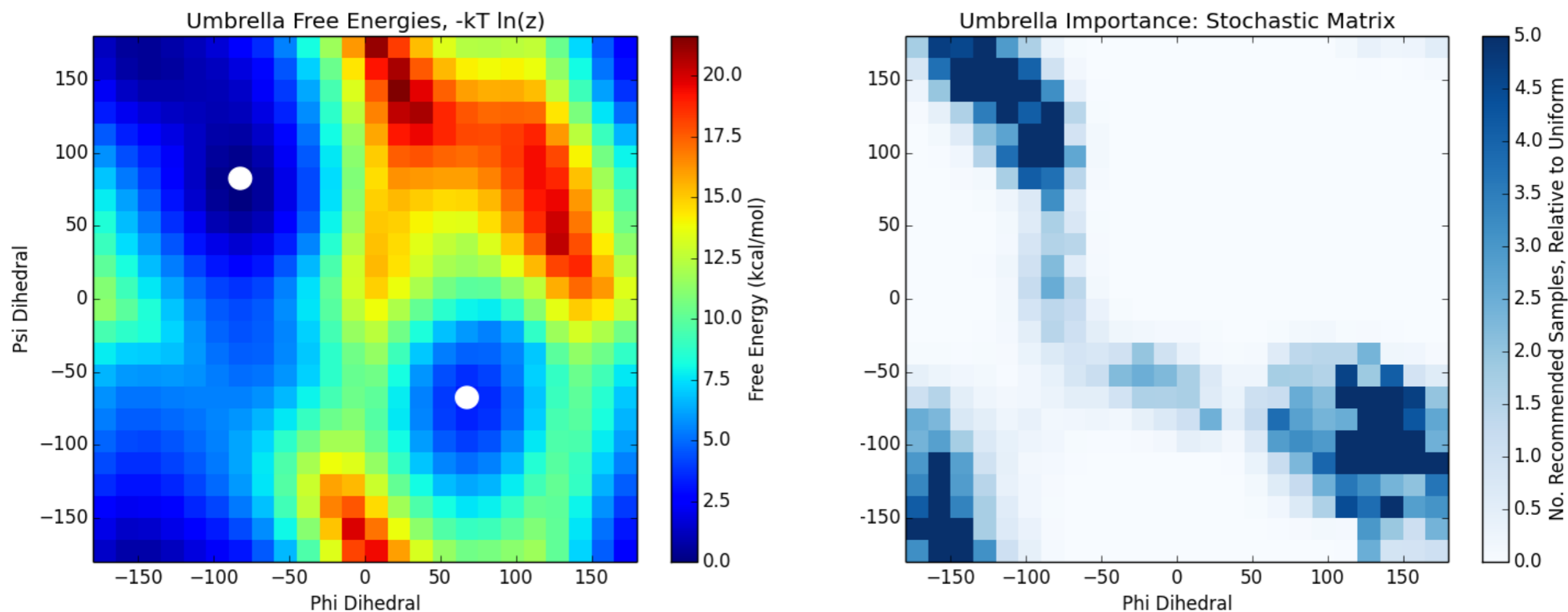
(tempering) $\pi_i \propto (\pi)^{\alpha_i}$ $\alpha_i \in (0, 1]$

(collective variable) bias functions restrict along strips parallel with long axis of level sets

In both cases US gives much more accurate representation of the tails of the posterior

In particular we obtain a more accurate estimate of the probability that the expansion of the universe is decelerating, contradicting previous (higher) estimates.

new applications + improvements for current applications



Trajectory Stratification Framework

We estimate expectations of the form:

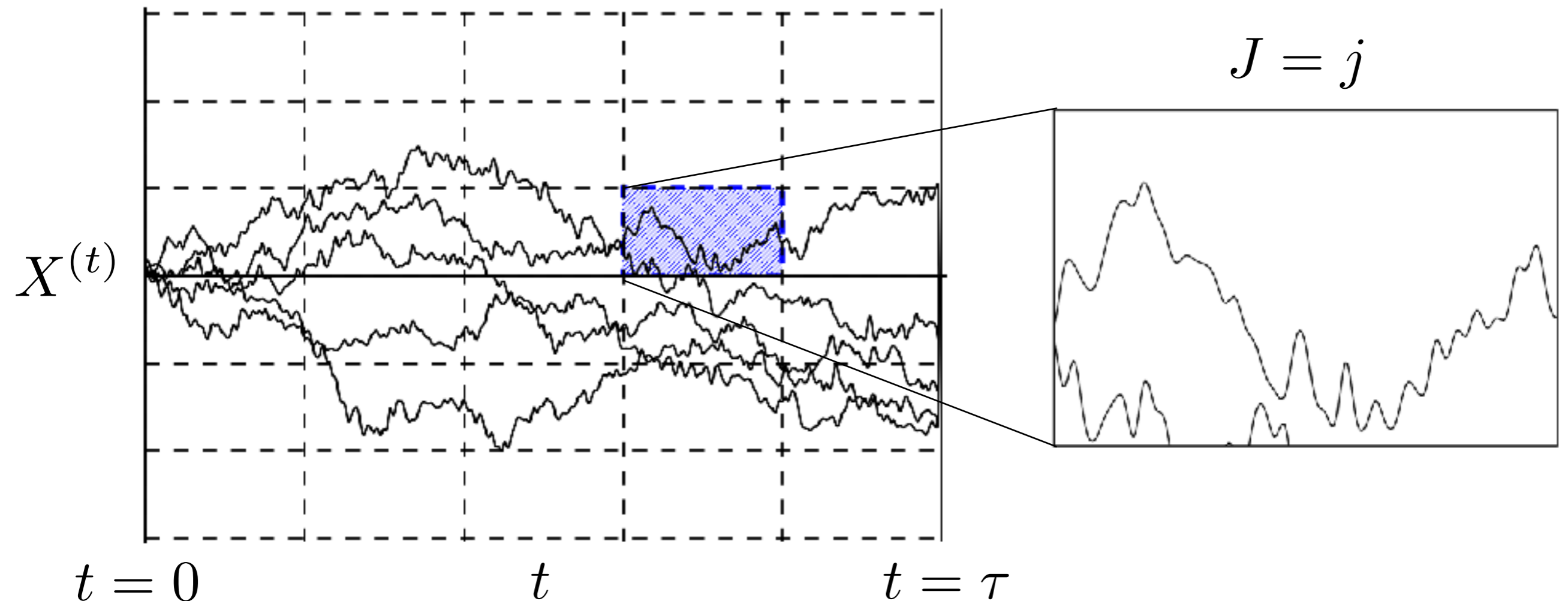
$$\mathbf{E} \left[\sum_{t=0}^{\tau-1} f(t, X^{(t)}) \right]$$

τ is a first hitting time for $(t, X^{(t)})$

Associate with $X^{(t)}$ an index process

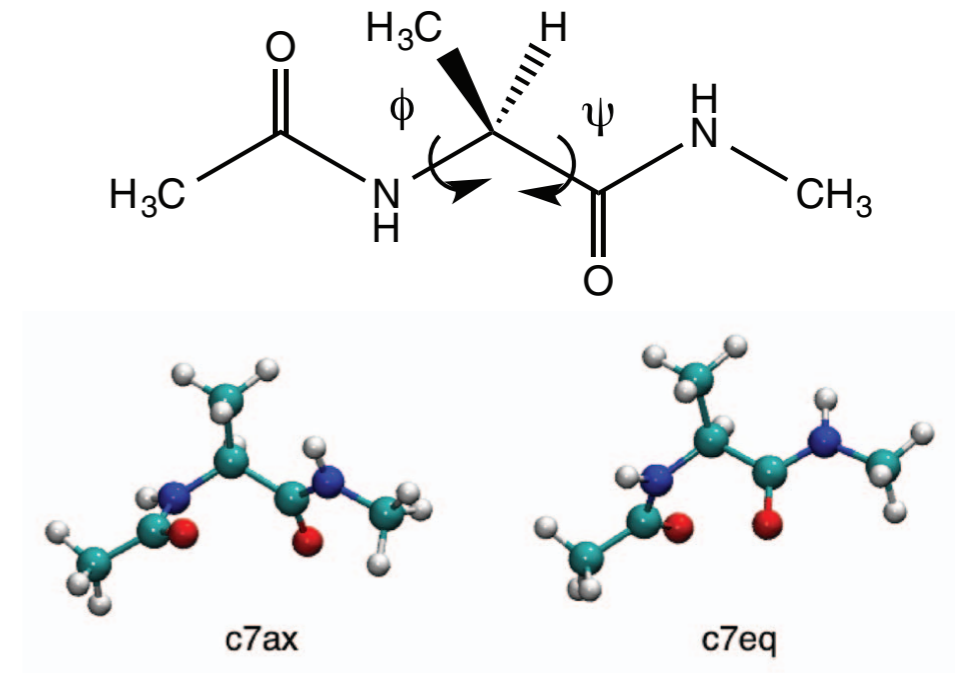
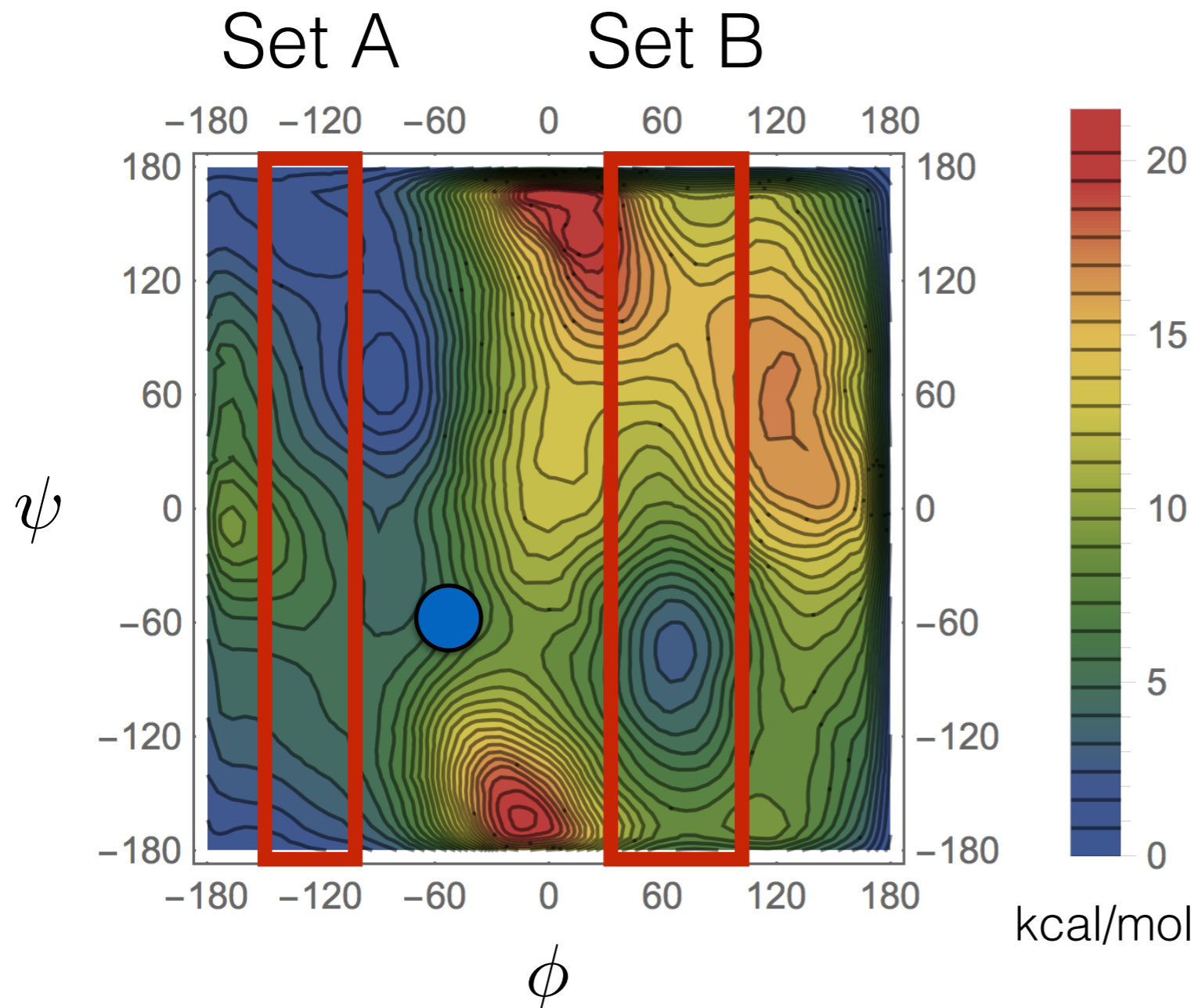
$$J^{(t)} \in \{0, 1, 2, \dots, n\}$$

that labels regions in both space and time:



Illustrative example: Computing set hitting probabilities in the isomerization of the alanine dipeptide

Compute the probability $P_{BA} = \mathbf{P}_x[\tau_B < \min\{\tau_A, \tau_{\max}\}]$ that the system hits the set B before hitting the set A and before time τ_{\max}

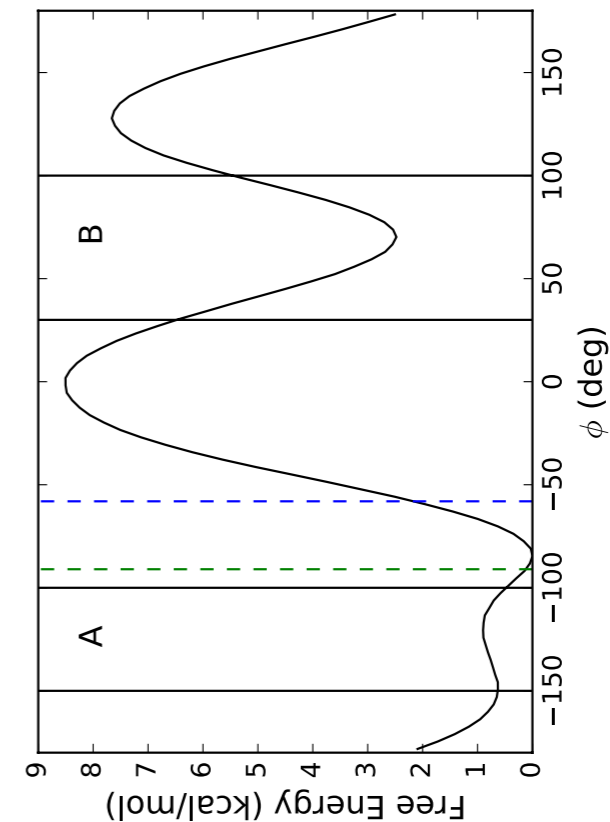
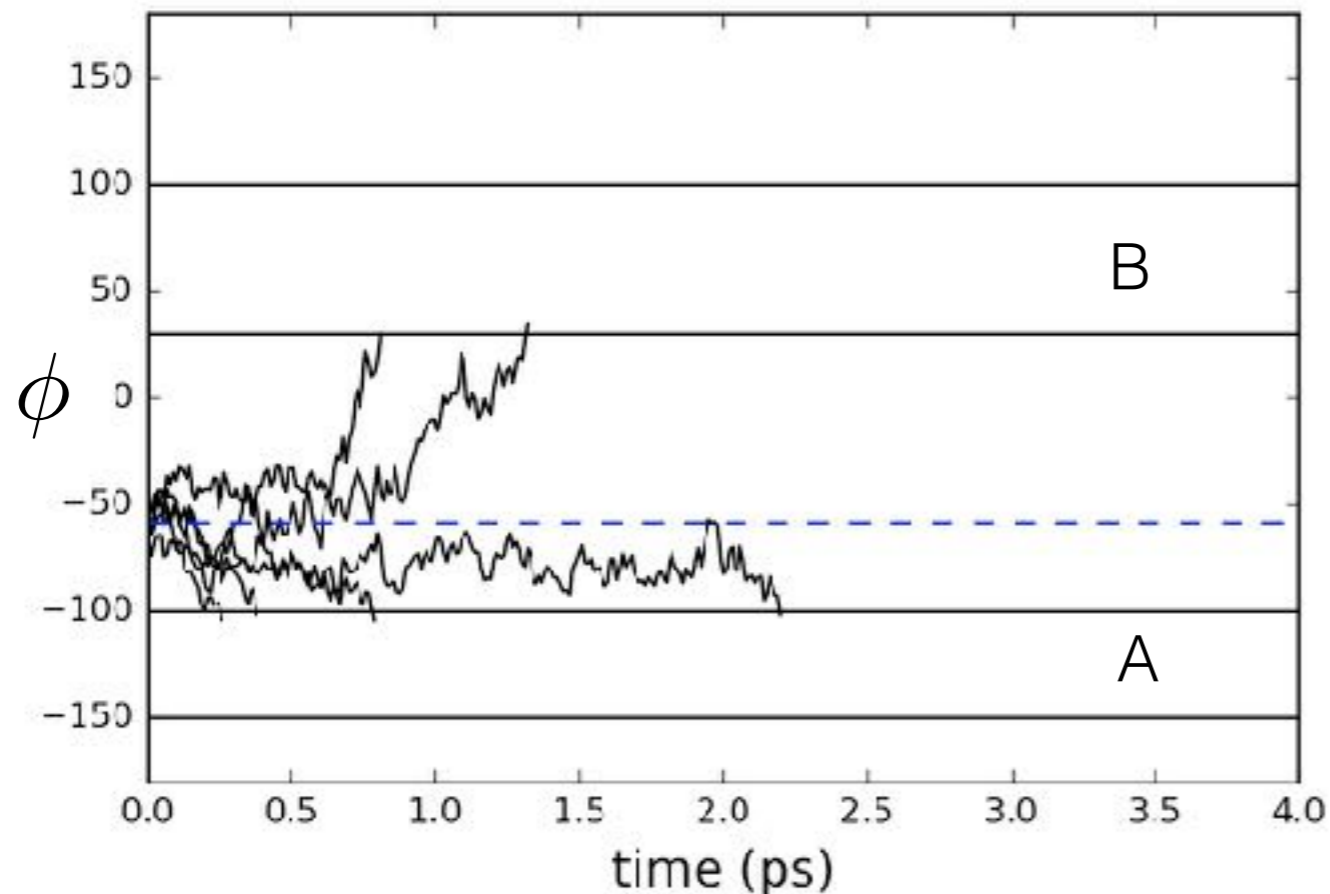


Illustrative example: Computing set hitting probabilities in the isomerization of the alanine dipeptide

$$\phi^{(0)} = -58.0^\circ \quad P_{BA} \approx \frac{N_B}{N}$$

$$f(t, X^{(t)}) = \mathbf{1}_B(X^{(t)})$$

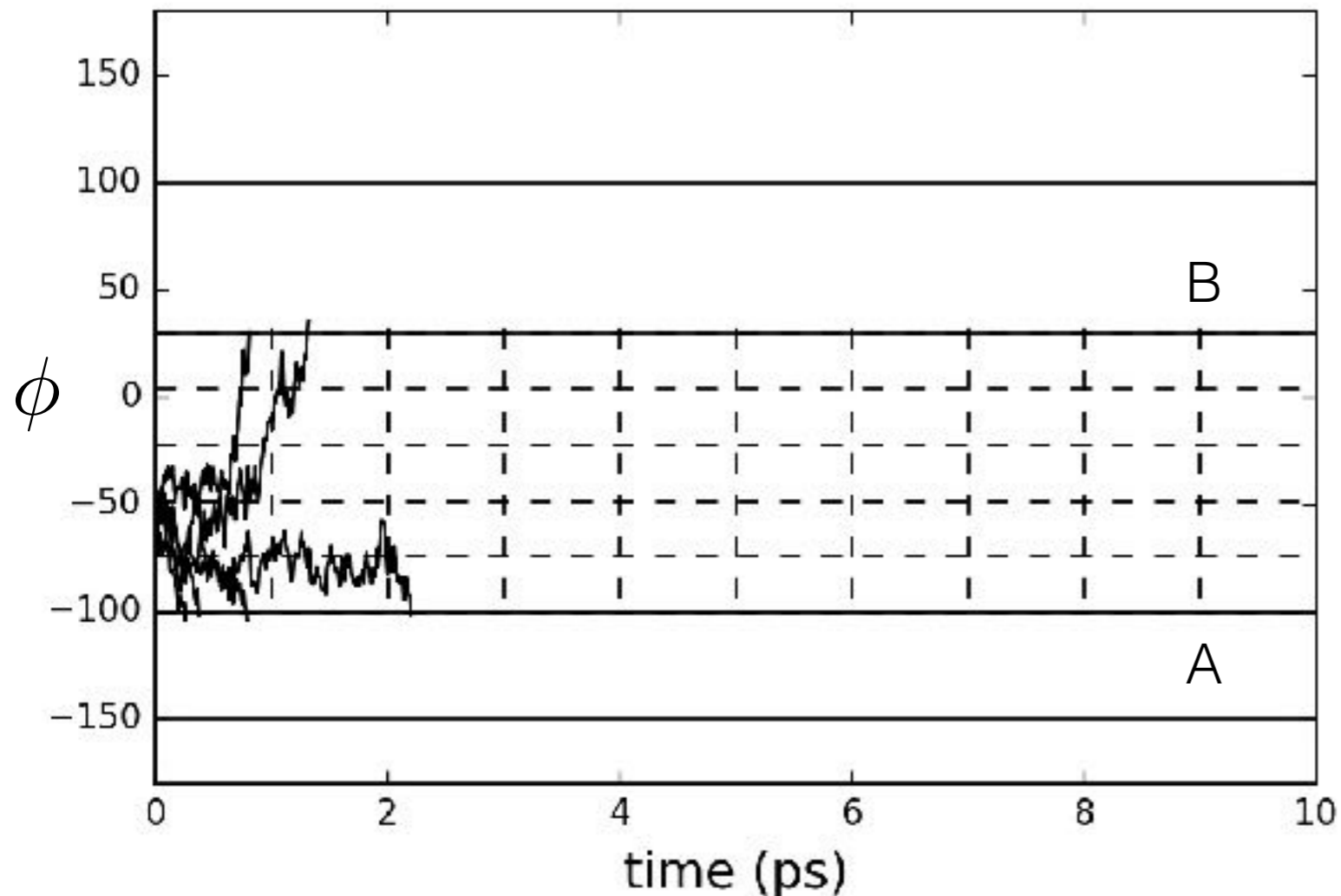
$$\tau - 1 = \min\{\tau_A, \tau_B, \tau_{\max} - 1\}$$



This probability can be very difficult to estimate when the probability of hitting B is very small.

Illustrative example: Computing set hitting probabilities in the isomerization of the alanine dipeptide

To compute P_{BA} using the non-stationary trajectory stratification, we divide the trajectory space along the phi dihedral and along time:

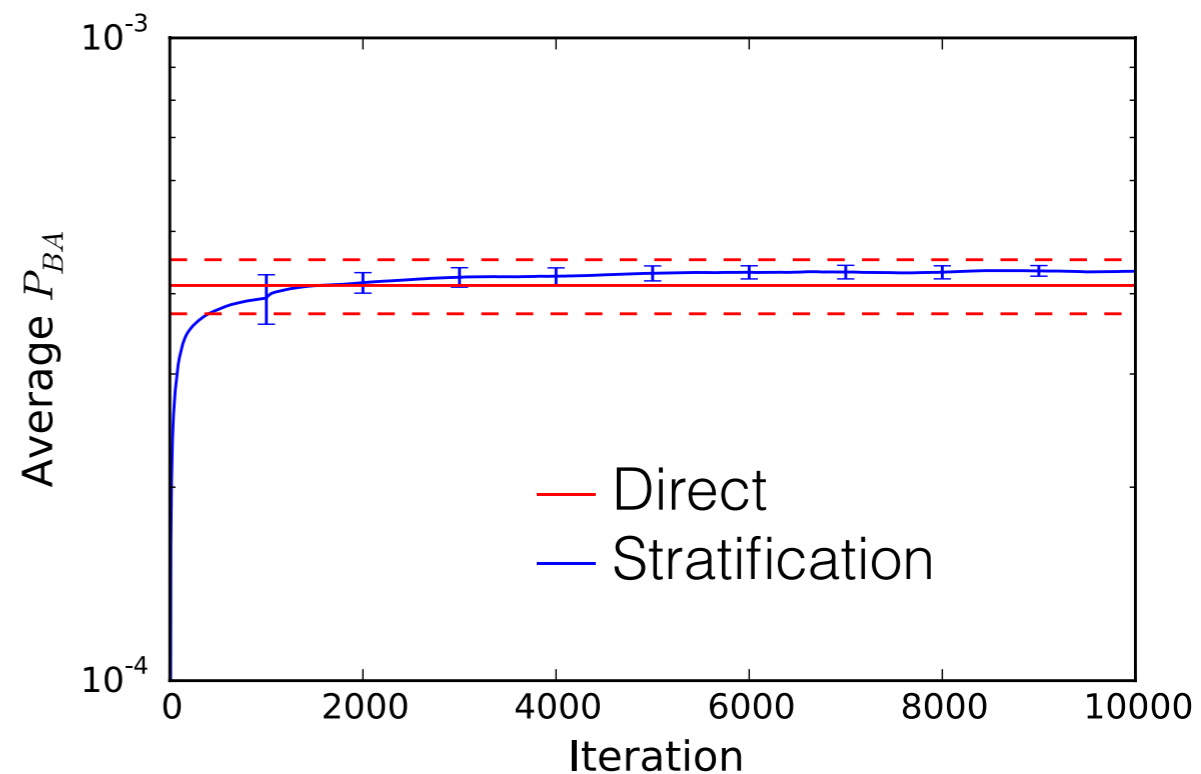
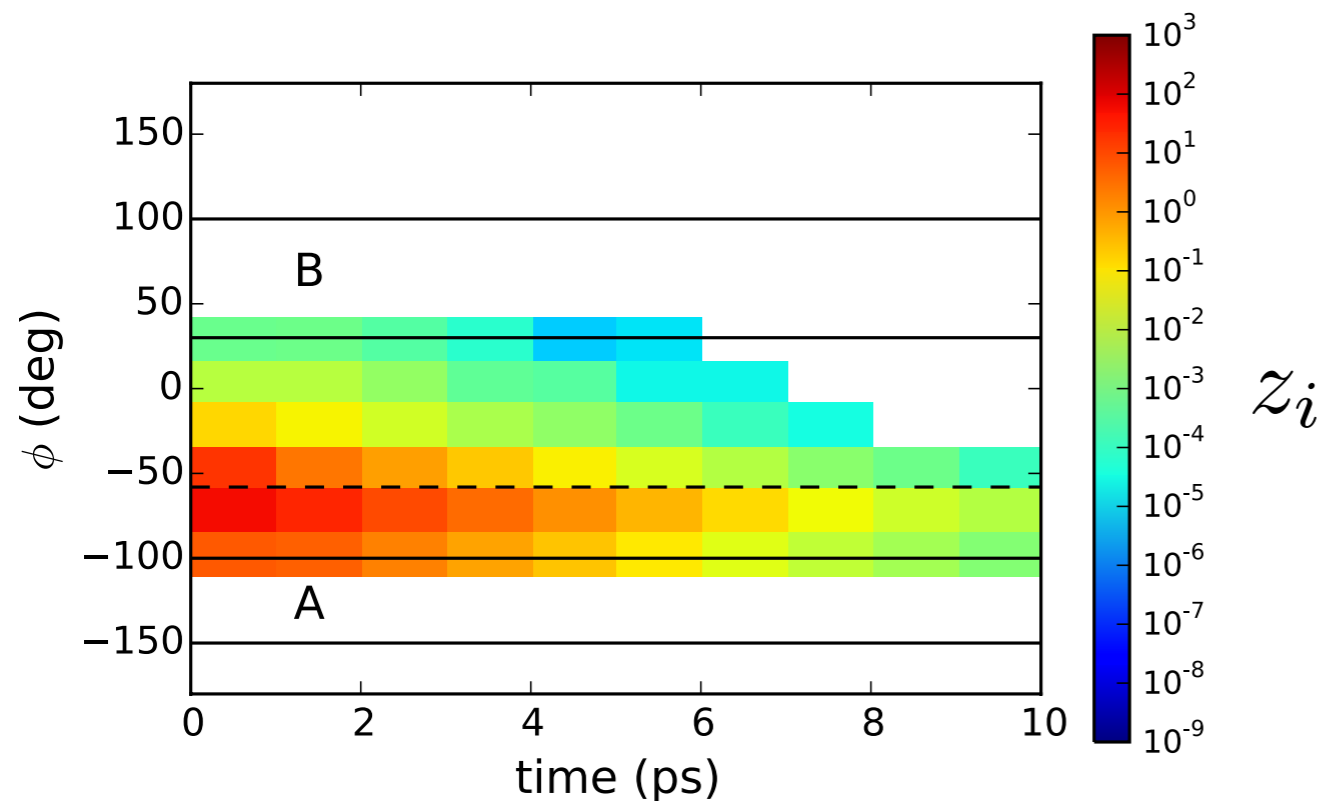


$$f(t, X^{(t)}) = \mathbf{1}_B(X^{(t)})$$

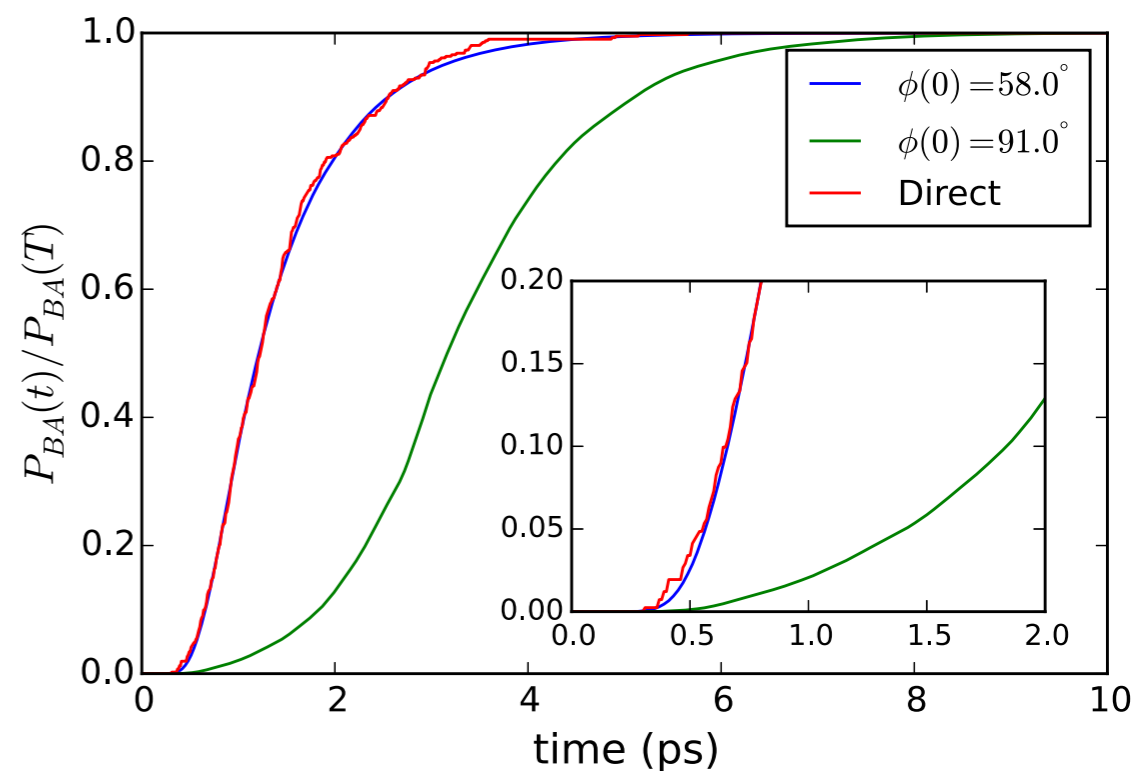
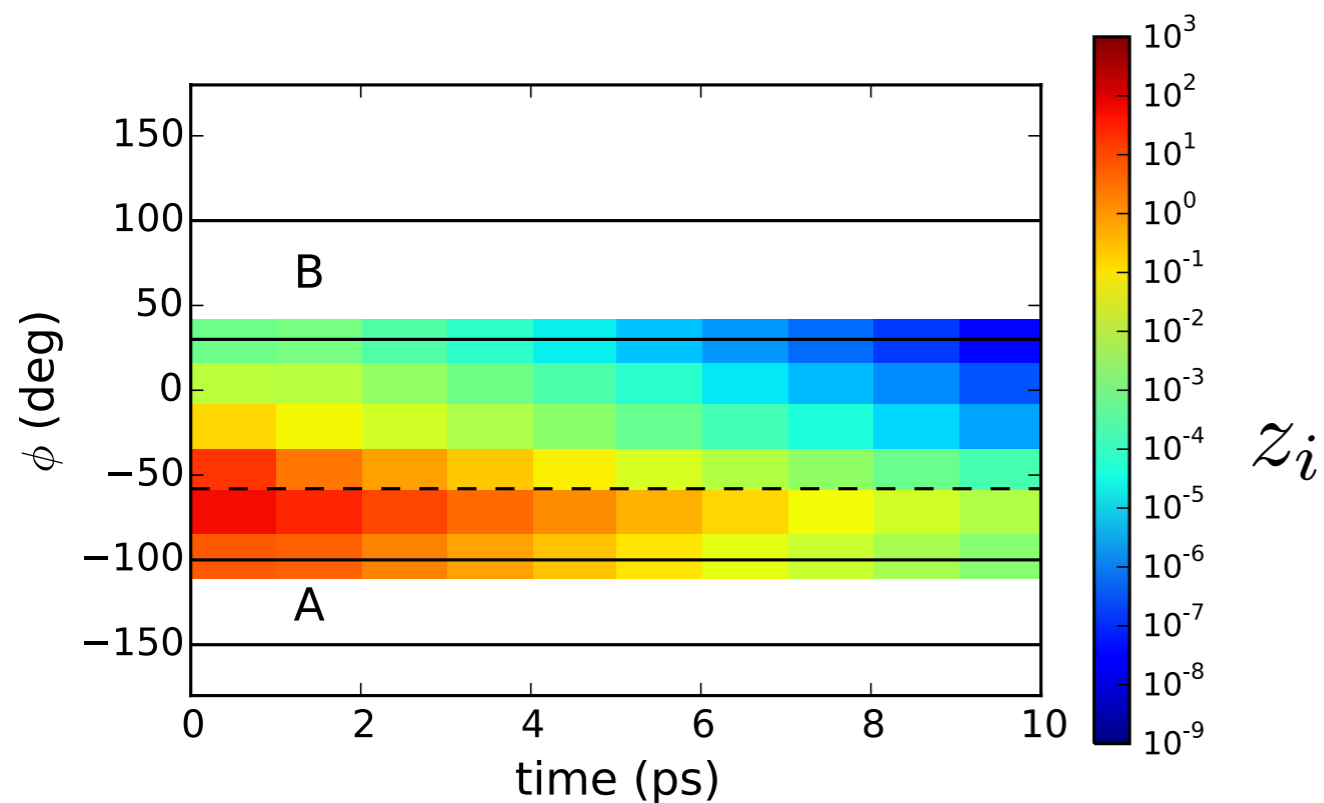
$$\tau - 1 = \min\{\tau_A, \tau_B, \tau_{\max} - 1\}$$

$$E \left[\sum_{t=0}^{\tau-1} f(t, X(t)) \right] = \sum_{i=1}^n z_i \langle f \rangle_i$$

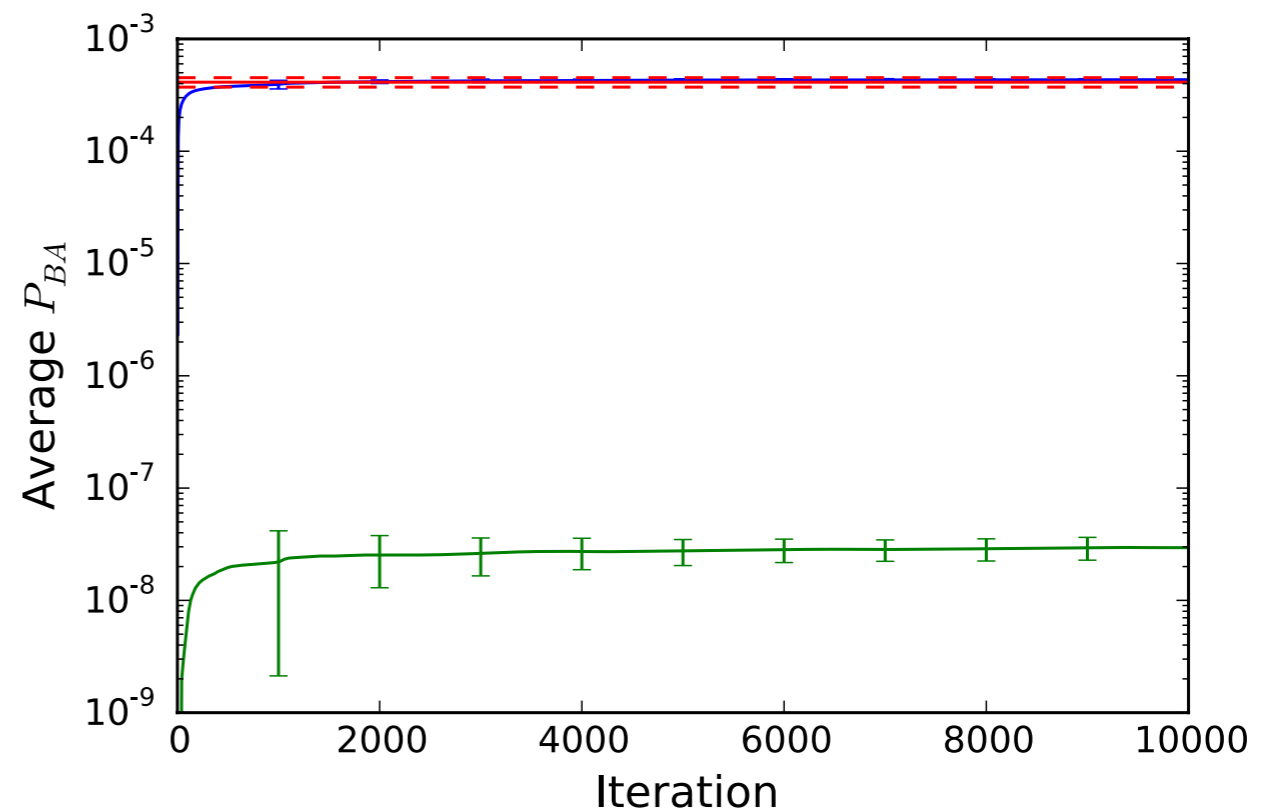
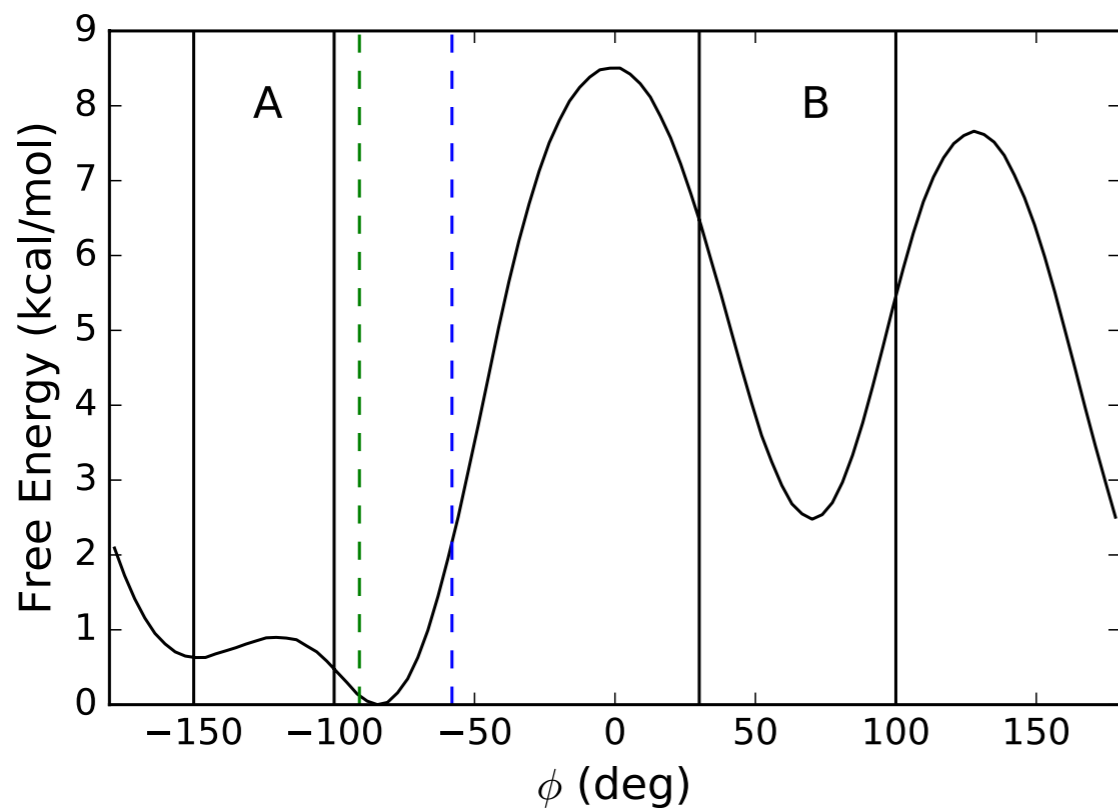
Repeated direct integration (10^6 independent simulations)



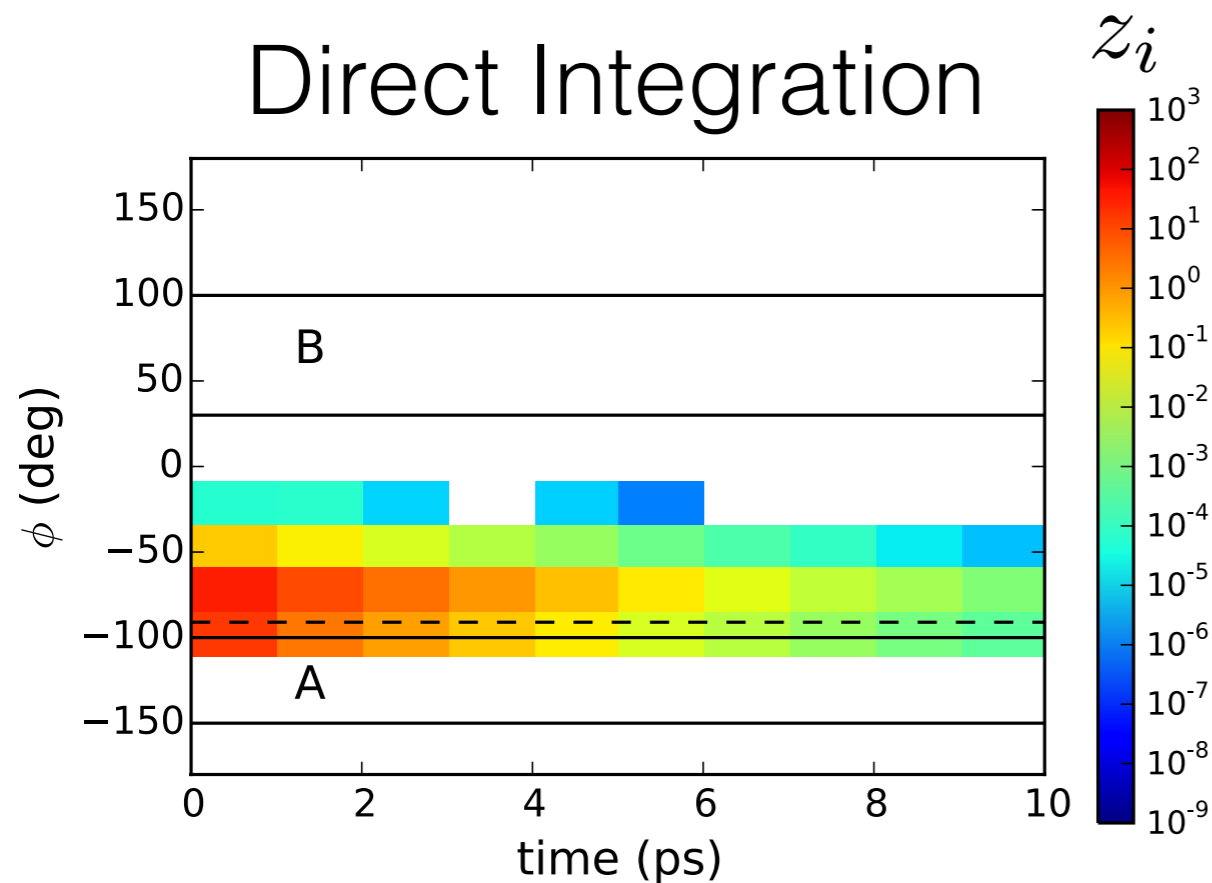
NEUS stratification



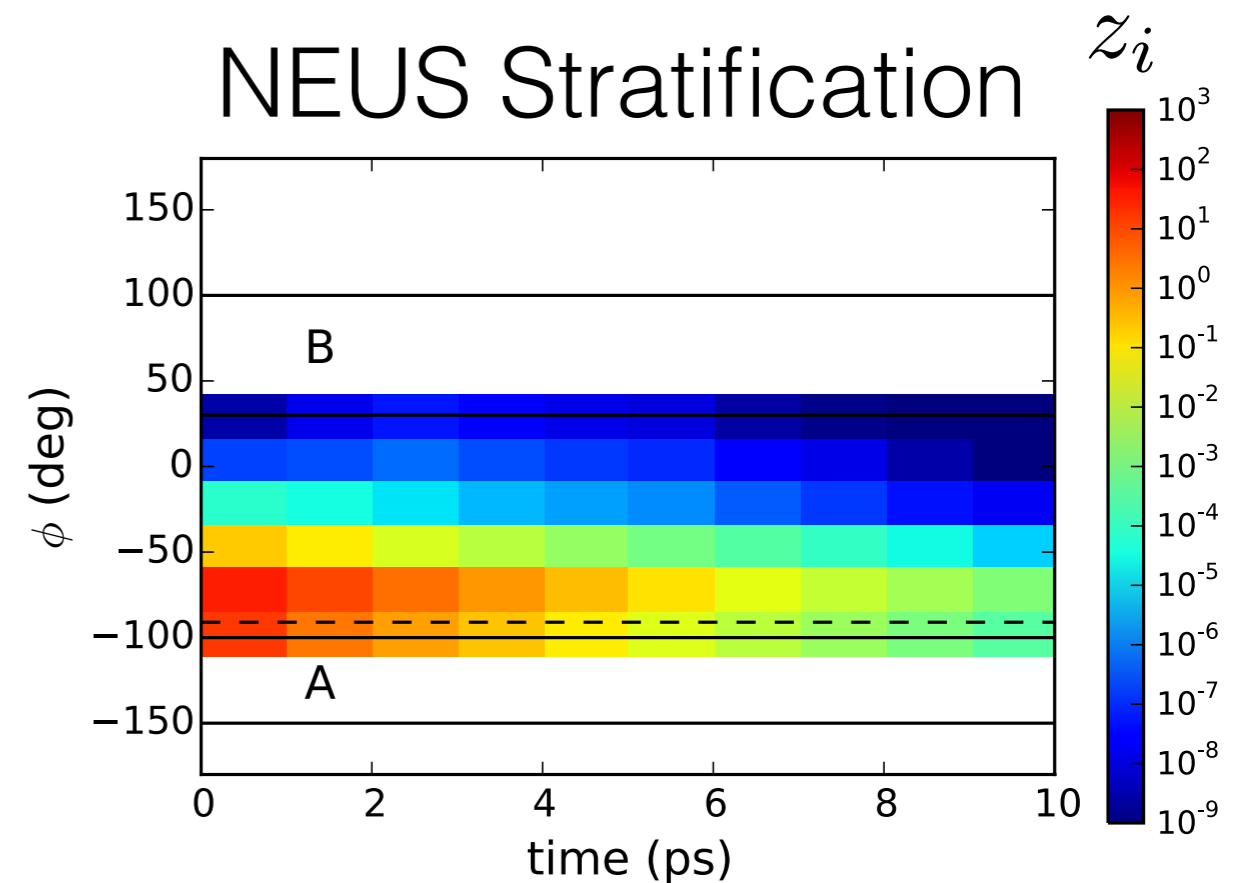
Choosing $\phi^{(0)} = -91.0$ close to A makes estimating P_{BA} a hard rare event problem



Direct Integration



NEUS Stratification



Software Tools: Enhanced Sampling Toolkit

I've developed a set of software tools designed to facilitate rapid prototyping of enhanced sampling algorithms.

Enhanced Sampling Toolkit



Enhanced Sampling Algorithms:

- Allows developers to write algorithms in 100% Python
- Enables rapid prototyping in developer friendly languages
- Enables clean and concise writing of algorithmic code

MD Engines:

- LAMMPS, OpenMM, CHARMM, GROMACS (potential future implementation)
- Dynamics are executed in widely available and popular MD codes
- Computationally expensive part of the code is still executed in fast MD codes