Coarse Graining in Conformational Space

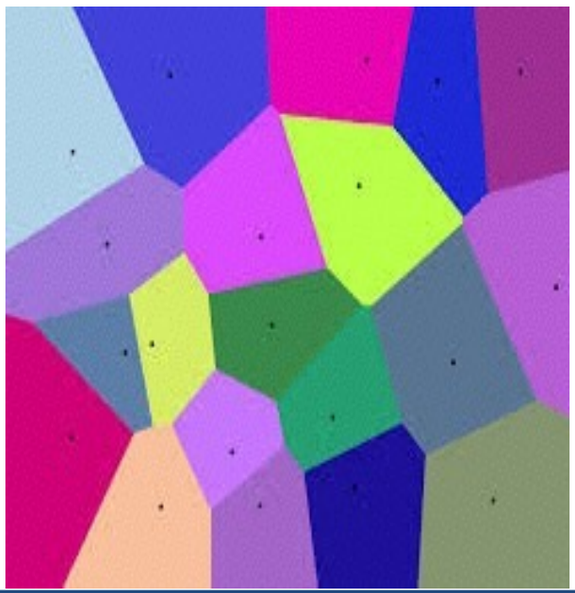
Feliks Nüske and Cecilia Clementi
Rice University
feliks.nueske@rice.edu
Sep 14, 2017

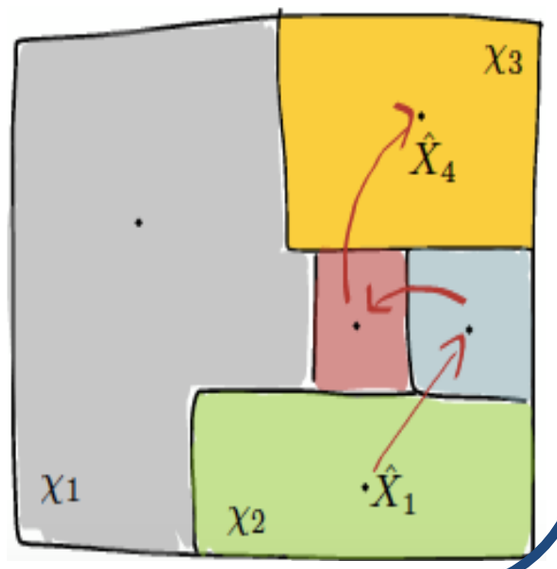
Coarse-Graining

RICE
Unconventional Wisdom

$\mathbb{R}^{3N}$

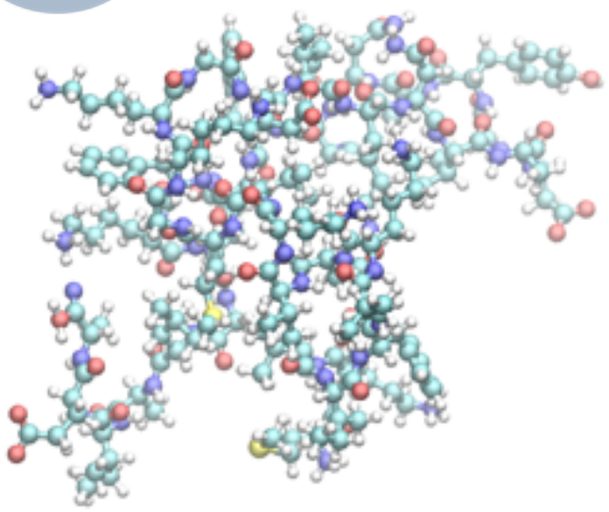
coarse-graining in conformation space

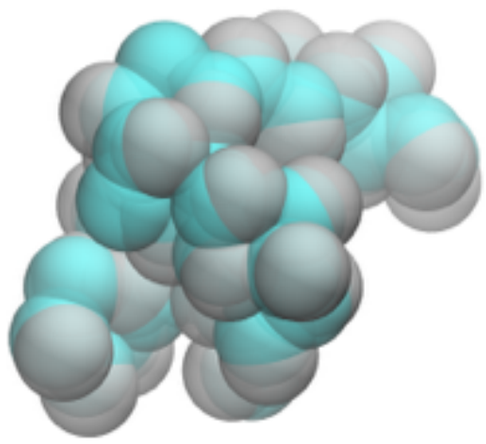


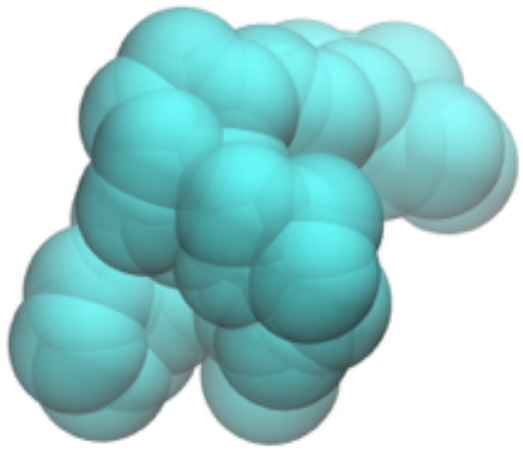


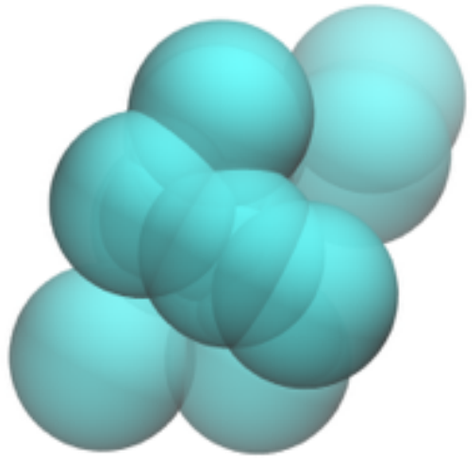
$\mathbb{R}^{3}$

coarse-graining in structural space



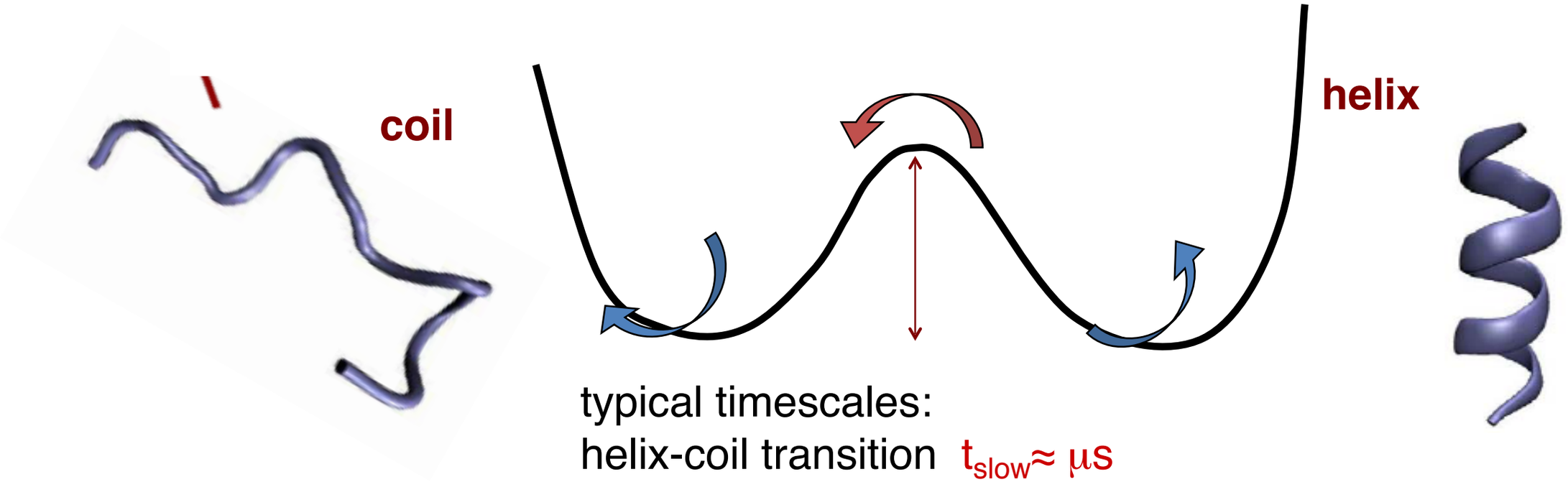






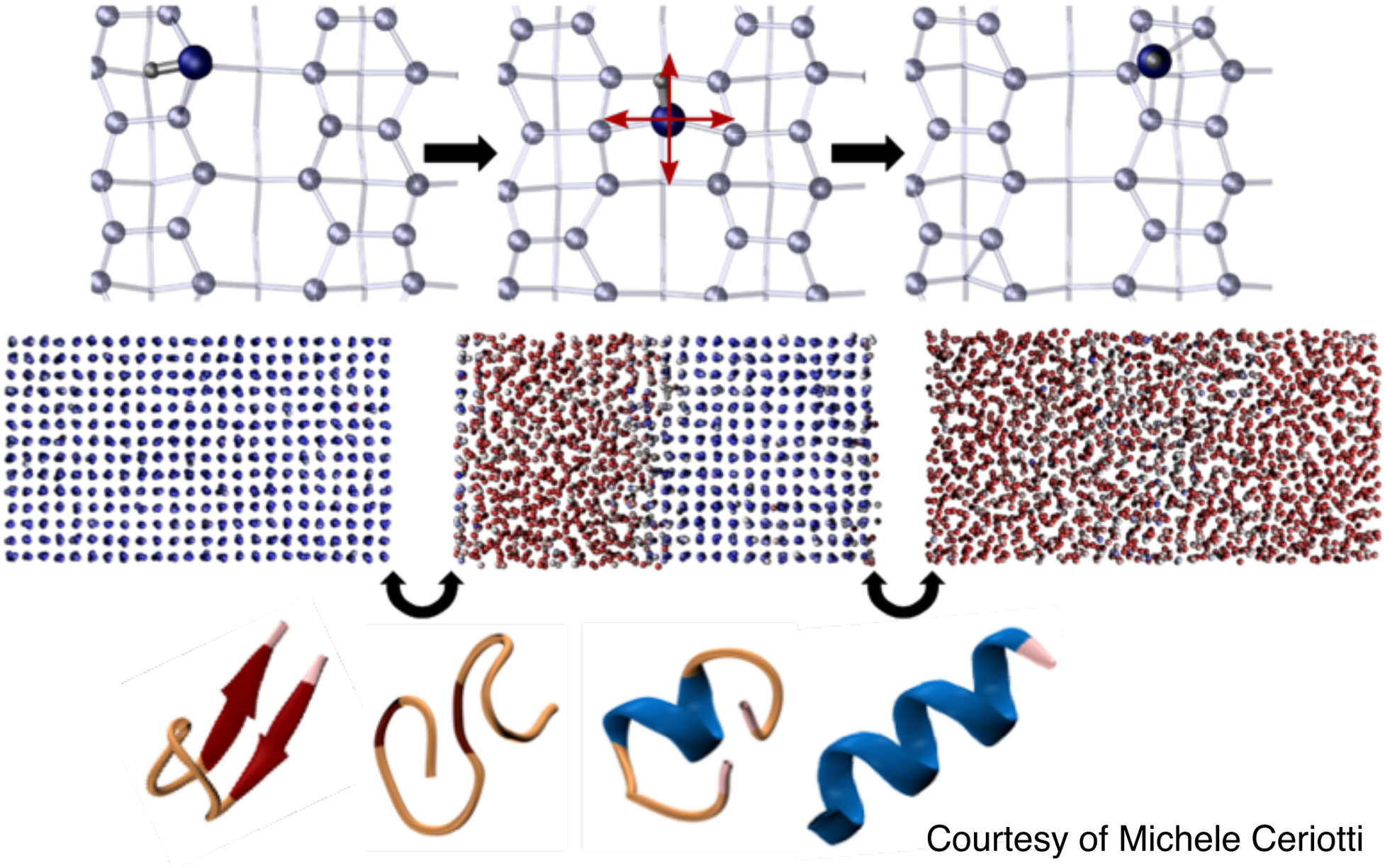
Biophysical systems are often characterized by different timescales, separated by large gaps

Example: helix-coil transition



typical timescales:
helix-coil transition $t_{slow} \approx \mu s$
diffusion within a state $t_{fast} \approx ns$

Reaction Coordinates should capture the (rare) barrier-crossing events.



Courtesy of Michele Ceriotti

I. Optimal Reaction Coordinates

Basic Assumptions

Dynamical Model is a Markov process

$$X_t, t \geq 0$$

The process is **ergodic** if spatial and temporal averages coincide:

$$\mathbb{E}_\mu(f(X_0, X_\tau, \ldots, X_{L\tau})) =$$

$$= \lim_{K\to\infty} \frac{1}{K-L} \sum_{k=0}^{K-L-1} f(X_{k\tau}), \ldots, X_{(k+L)\tau})$$

The process is **reversible** if there is no preferred direction:

$$\mathbb{P}_\mu(X_0 \in A, X_t \in B) = \mathbb{P}_\mu(X_0 \in B, X_t \in A).$$

Basic Assumptions

Dynamical Model is a Markov process

$$X_t, t \geq 0$$

A distribution is stationary if it is invariant under the dynamics:

$$\mathbb{P}_\mu(X_t \in A) = \mu(A)$$

For an ergodic process, there is a unique stationary distribution.

Transfer Operators

Transfer Operators describe the evolution of probability distributions:

$$\int_A \left[\mathcal{T}_t f\right](x)\mu(\mathrm{dx}) = \int_S p_t(x, A) f(x)\mu(\mathrm{dx}).$$

Here, A is a set, μ is a measure, f is a density w.r.t. μ, and $p_t(x, A)$ is the stochastic transition function.

Lasota, Mackey: *Chaos, Fractals and Noise*, Springer (1993)

Properties

If μ is stationary, transfer op. are defined on all spaces L^p_μ.

Semigroup property:

$$\mathcal{T}_{s+t} = \mathcal{T}_t \mathcal{T}_s,\ s,t \geq 0.$$

For reversible processes, the transfer operators are self-adjoint on L^2_μ.

Lasota, Mackey: *Chaos, Fractals and Noise*, Springer (1993)

Spectral Decomposition

Using the **infinitesimal Generator**:

$$\mathcal{L}f = \lim_{t \to 0} \frac{\mathcal{T}_t f - f}{t},$$

and its spectral components:

$$\mathcal{L}\psi_j = -\kappa_j \psi_j$$

we have the spectral mapping theorem:

$$e^{\tau\sigma(\mathcal{L})} \subset \sigma(\mathcal{T}_\tau) \subset e^{\tau\sigma(\mathcal{L})} \cup \{0\},$$

Pazy: *Semigroups of linear operators and applications to partial differential equations*, Springer (1983)

and the spectral decomposition:

$$\mathcal{T}_\tau f = \sum_{j=1}^{\infty} e^{-\kappa_j \tau} \langle f, \psi_j \rangle_\mu \psi_j$$

Slow Processes

In many biophysical applications, we have:

$$0 = \kappa_0 < \kappa_1 < \ldots < \kappa_M \ll \kappa_{M+1}$$

such that, for $\tau \gg \dfrac{1}{\kappa_{M+1}}$

$$\mathcal{T}_\tau f \approx \sum_{j=1}^{M} e^{-\kappa_j \tau} \langle f, \psi_j \rangle_\mu \psi_j$$

The presence of separated timescales is directly linked to the structure of the spectrum.

Slow Processes

Using the spectral decomposition, we find for the auto-correlation function:

$$\begin{aligned}\mathbb{E}_t(f(X_t)f(X_{t+\tau})) &= \mathbb{E}_\mu(f(X_0)f(X_\tau)) \\ &= \langle \mathcal{T}_\tau f, f\rangle_\mu \\ &= \sum_{j=1}^{\infty} e^{-\kappa_j \tau} \langle f, \psi_j\rangle_\mu^2\end{aligned}$$

Variational Theorem

Variational Theorem for self-adjoint operators:

$$\sum_{j=1}^{M} e^{-\kappa_j \tau} = \sup \sum_{j=1}^{M} \langle \mathcal{T}_\tau f_j, f_j \rangle_\mu,$$

$$= \sup \sum_{j=1}^{M} \mathbb{E}_t(f_j(X_t) f_j(X_{t+\tau})),$$

$$\langle f_j, f_{j'} \rangle_\mu = \delta_{j,j'}.$$

Dominant eigenfunctions are slowest to de-correlate.
They are optimal descriptors of slow dynamics.

Noé and Nüske, *SIAM Multiscale Model. Simul.* (2013)

Galerkin Projection

Proposition 3.2. *Let D be a space of N linearly independent ansatz functions f_i, $i = 1,\dots,N$. The set of $M \leq N$ mutually orthonormal functions $f_{\mathbf{a}_m}$, $m = 1,\dots,M$ which maximize the Rayleigh trace restricted to D, is given by the first M eigenvectors of the generalized eigenvalue problem*

$$\mathbf{C}^\tau \mathbf{a}_m \quad = \quad \hat{\lambda}_m \mathbf{C}^0 \mathbf{a}_m, \tag{3.1.7}$$

where the matrices $\mathbf{C}^\tau$, $\mathbf{C}^0$ are given by

$$\mathbf{C}^\tau(i,j) \quad = \quad \langle \mathcal{T}_\tau f_i, f_j \rangle_\mu \tag{3.1.8}$$

$$\mathbf{C}^0(i,j) \quad = \quad \langle f_i, f_j \rangle_\mu. \tag{3.1.9}$$

Noé and Nüske, *SIAM Multiscale Model. Simul.* (2013)

Galerkin Projection

By ergodicity of the process:

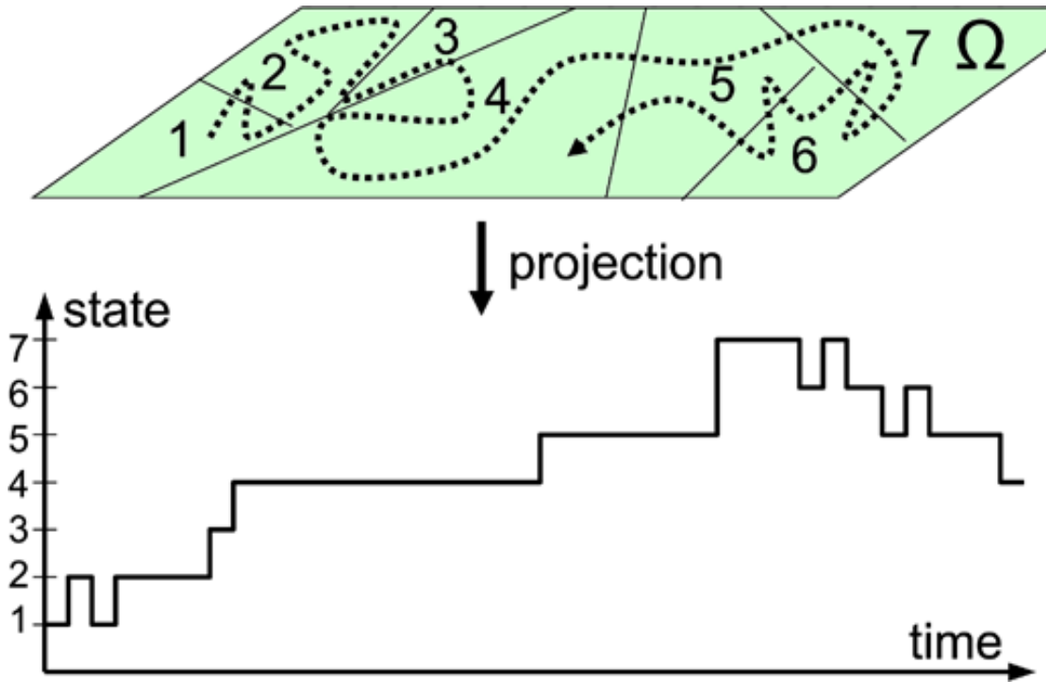
Theorem 3.4. *Let $\tau = L\Delta t$ be an integer multiple of the discrete time step. If the Markov process is initially distributed according to the unique invariant measure μ, then for $\mathbb{P}_{\Delta t}$-a.s. all trajectories $(X_{k\Delta t})_{k=0}^{\infty}$, we have:*

$$\mathbf{C}^{\tau}(i,j) = \lim_{K\to\infty} \frac{1}{K-L} \sum_{k=0}^{K-L-1} f_i(X_{k\Delta t}) f_j(X_{(k+L)\Delta t}), \tag{3.2.1}$$

$$\mathbf{C}^{0}(i,j) = \lim_{K\to\infty} \frac{1}{K-L} \sum_{k=0}^{K-L-1} f_i(X_{k\Delta t}) f_j(X_{k\Delta t}). \tag{3.2.2}$$

Noé and Nüske, *SIAM Multiscale Model. Simul.* (2013)

Markov State Models (MSM):



Apply previous formalism with a basis of step functions:

$$\chi_i(x) = \begin{cases} 1, \, x \in S_i \\ 0, \, \text{else} \end{cases}$$

$$c_{ij} = \frac{1}{T-\tau} \sum_{k=1}^{T-\tau} \chi_i(X_k)\chi_j(X_{k+\tau})$$

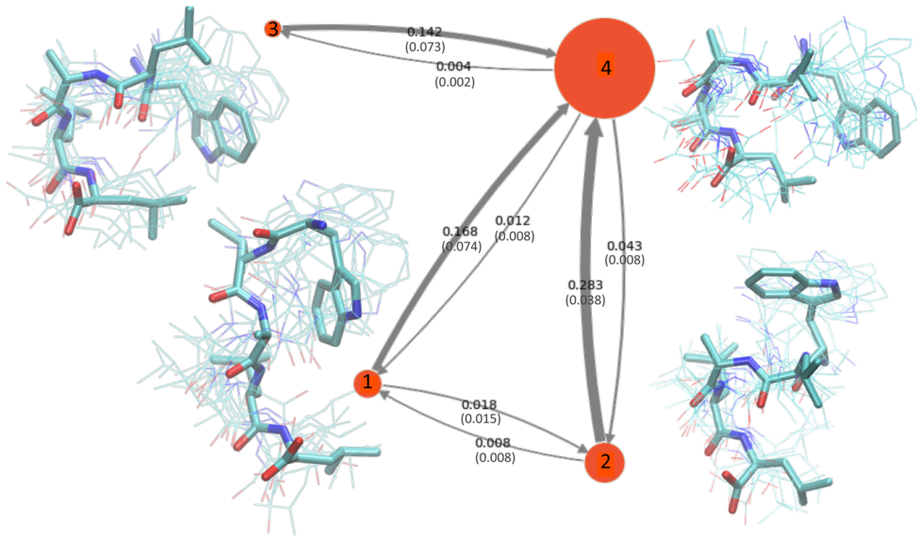
Normalize to stochastic matrix:

$$\mathbf{T}(\tau) = \left(\frac{c_{ij}}{c_i} \right)_{ij}$$

- Schütte et. al. *J. Comput. Phys.* (1999)
- Prinz et. al. *J. Chem. Phys.* (2011)
- Schütte and Sarich, *Metastability and Markov State Models in Molecular Dynamics: Modeling, Analysis, Algorithmic Approaches*, Courant Lecture Notes (2013)
- Bowman et. al. (Eds) *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*, Springer (2014)

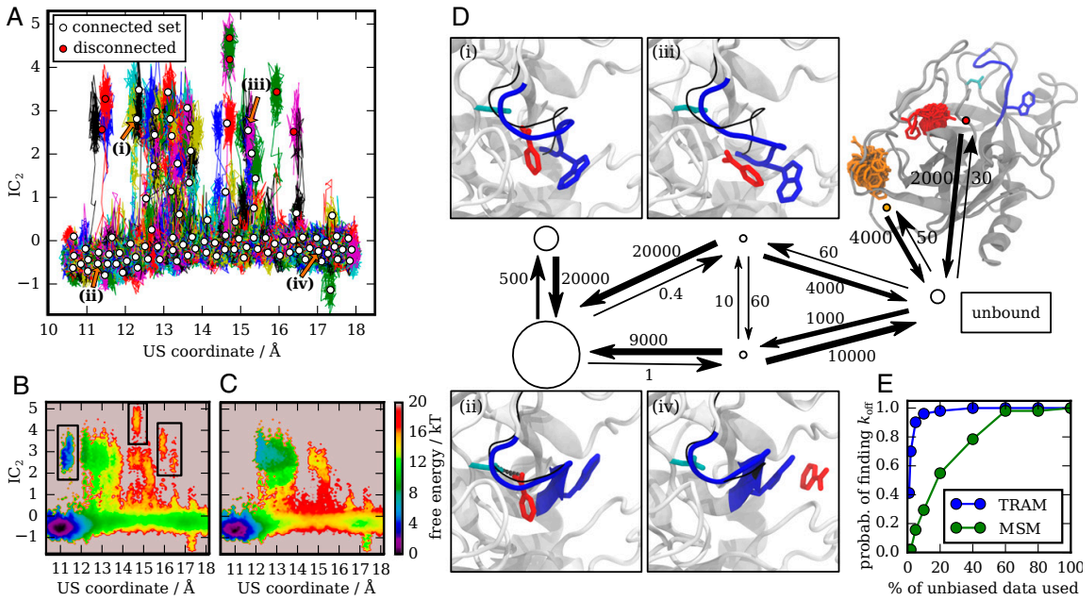
MSMs

Identification of metastable states:



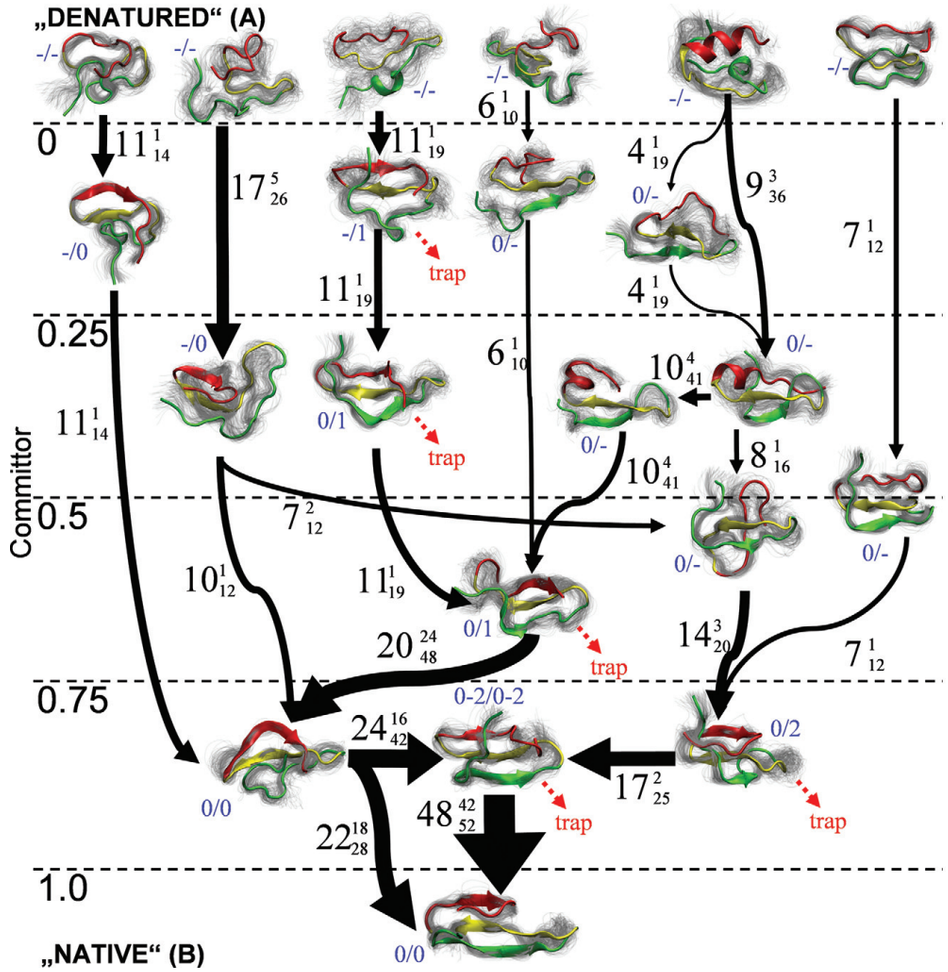
Deuflhard and Weber, *Linear Algebra and its Applications* (2005)

Multi-Ensemble MSMs:



Wu et al. *PNAS* (2016)

Transition Path Theory

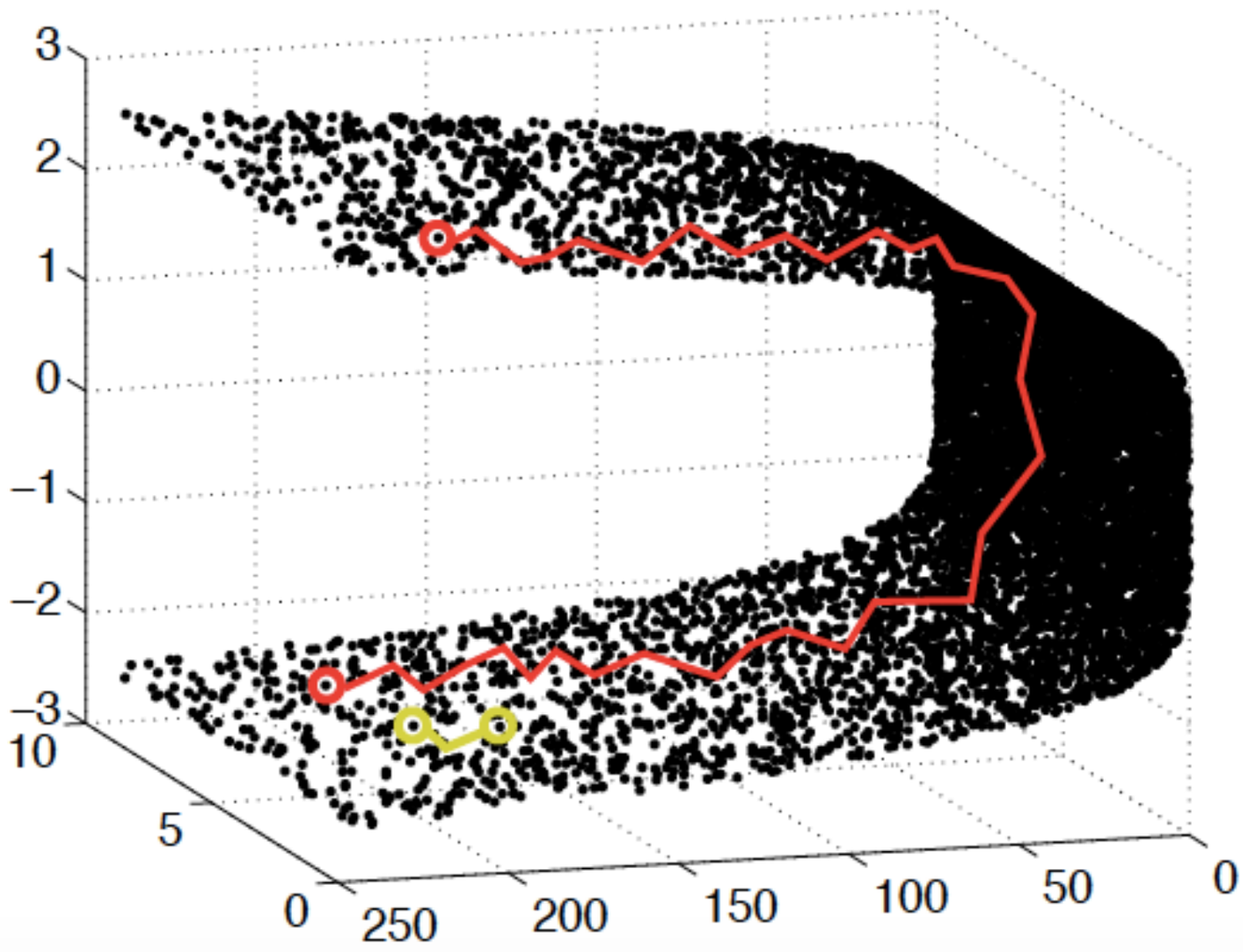


Noé et al. *PNAS* (2009)

...

Diffusion Maps

Diffusion maps idea: Build a Markov chain on the data points. Base jump probabilities on Euclidean distance, but only allow local jumps. Then compute distances based on how the Markov chain traverses the dataset.



Courtesy of Ralf Banisch

Diffusion Maps

Pick a symmetric, positive kernel k. Define a Markov chain by

$$p(x,y) = \frac{k(x,y)}{d(x)}. \quad d(x) = \int_X k(x,y)\,\mathrm{d}\mu(y)$$

Introduce the diffusion distance:

$$D_t(x,y)^2 \triangleq \left\| p_t(x,\cdot) - p_t(y,\cdot) \right\|^2_{L^2(X,\mathrm{d}\mu/\pi)} = \int_X \left(p_t(x,u) - p_t(y,u)\right)^2 \frac{\mathrm{d}\mu(u)}{\pi(u)}.$$

By spectral expansion:

$$D_t(x,y) = \left(\sum_{l \geqslant 1} \lambda_l^{2t} \left(\psi_l(x) - \psi_l(y)\right)^2 \right)^{\frac{1}{2}}.$$

In the following space, Euclidean distance equals diffusion dist.

$$\Psi_t(x) \triangleq \begin{pmatrix} \lambda_1^t \psi_1(x) \\ \lambda_2^t \psi_2(x) \\ \vdots \\ \lambda^t_{s(\delta,t)} \psi_{s(\delta,t)}(x) \end{pmatrix}.$$

Coffman et. al. Appl. *Comput. Harmon. Anal.* (2006)

Diffusion Maps

(1) Fix $\alpha \in \mathbb{R}$ and a rotation-invariant kernel $k_\varepsilon(x, y) = h\left(\frac{\|x-y\|^2}{\varepsilon}\right)$.
(2) Let

$$q_\varepsilon(x) = \int_X k_\varepsilon(x, y) q(y) \, \mathrm{d}y$$

and form the new kernel

$$k_\varepsilon^{(\alpha)}(x, y) = \frac{k_\varepsilon(x, y)}{q_\varepsilon^\alpha(x) q_\varepsilon^\alpha(y)}.$$

(3) Apply the weighted graph Laplacian normalization to this kernel by setting

$$d_\varepsilon^{(\alpha)}(x) = \int_X k_\varepsilon^{(\alpha)}(x, y) q(y) \, \mathrm{d}y$$

and by defining the anisotropic transition kernel

$$p_{\varepsilon,\alpha}(x, y) = \frac{k_\varepsilon^{(\alpha)}(x, y)}{d_\varepsilon^{(\alpha)}(x)}.$$

Coffman et. al. Appl. *Comput. Harmon. Anal.* (2006)

Diffusion Maps

Theorem 2. *Let*

$$L_{\varepsilon,\alpha} = \frac{I - P_{\varepsilon,\alpha}}{\varepsilon}$$

be the infinitesimal generator of the Markov chain. Then for a fixed $K > 0$, we have on E_K

$$\lim_{\varepsilon \to 0} L_{\varepsilon,\alpha} f = \frac{\Delta(f q^{1-\alpha})}{q^{1-\alpha}} - \frac{\Delta(q^{1-\alpha})}{q^{1-\alpha}} f.$$

For $\alpha = 0.5$, this operator becomes the generator of a diffusion process

Coffman et. al. Appl. *Comput. Harmon. Anal.* (2006)

II. Effective Dynamics on Reaction Coordinates

How can we define dynamics on the space of reaction coordinates?

Assume the process is given by an SDE

$$dx_t = b(x_t) + \sqrt{2\beta^{-1}}\boldsymbol{\sigma} dB_t.$$

Then the generator is a differential operator:

$$\mathcal{L}f(x) = \sum_{i=1}^{n} b_i(x) \frac{\partial f}{\partial x_i}(x) + \frac{1}{\beta} \sum_{i=1}^{n} \boldsymbol{\sigma}_{ii}^2 \frac{\partial^2 f}{\partial x_i^2}(x).$$

Reaction coordinates are represented by smooth mapping:

$$\xi : \mathbb{R}^n \mapsto \mathbb{R}^m$$

Projection Formalism

Projection operator:

$$\mathcal{P}f(x) = \frac{1}{\nu(z)} \int_{\boldsymbol{\Sigma}_z} f(x)\mu(x)J^{-1/2}(x)\,\mathrm{d}\sigma_{\mathrm{z}}(\mathrm{x}).$$

Projection is an orthogonal projection onto the space:

$$H_0 = \left\{ f \in L^2_\mu : f = f(z) = f(\xi(x)) \right\}$$

Zhang et. al., *Faraday Discuss.* (2016)

Projection Formalism

Projected generator is again the generator of a diffusion:

$$\mathcal{L}^\xi = \mathcal{P}\mathcal{L}\mathcal{P},$$
$$\mathcal{L}^\xi f(z) = \sum_{l=1}^{m} \mathcal{P}(\mathcal{L}\xi_l)(z)\frac{\partial f}{\partial z_l} + \frac{1}{\beta}\sum_{l,r=1}^{m} \mathcal{P}\left(\sum_{i=1}^{n} \boldsymbol{\sigma}_{ii}^2 \frac{\partial \xi_l}{\partial x_i}\frac{\partial \xi_r}{\partial x_i}\right)\frac{\partial^2 f}{\partial z_l \partial z_r}.$$

Effective drift and diffusion can be estimated from data:

$$\widetilde{b}_l(z) = \lim_{s\to 0+} \mathbf{E}\Big(\frac{\xi_l(x(s)) - z_l}{s} \;\Big|\; x(0)\sim \mu_z\Big), \quad 1\le l\le m,$$
$$\widetilde{a}_{lk}(z) = \frac{\beta}{2}\lim_{s\to 0+} \mathbf{E}\Big(\frac{(\xi_l(x(s)) - z_l)(\xi_k(x(s)) - z_k)}{s} \;\Big|\; x(0)\sim \mu_z\Big)$$

Legoll and Lelièvre, *Nonlinearity* (2010)
Zhang et. al., *Faraday Discuss.* (2016)

Approximation Quality

Do these effective dynamics preserve the dominant timescales of the original process?

If we use the optimal reaction coordinates, yes! More precisely, if the eigenfunctions only depend on z:

$$\psi_j = \psi_j(z)$$

the dominant timescales are exactly preserved:

$$\mathcal{P}\mathcal{L}\mathcal{P}\psi_j(z) = -\kappa_j \psi_j(z)$$

Zhang et. al., *Faraday Discuss.* (2016)

Approximation Quality

RICE
Unconventional Wisdom

What if we only have an approximation?

Proposition 1 *If, for each dominant eigenfunction ψ_i, $i = 2, \ldots, M$, there is a function $g_i \in H_0$ that approximates ψ_i in H^1-norm, i.e.*

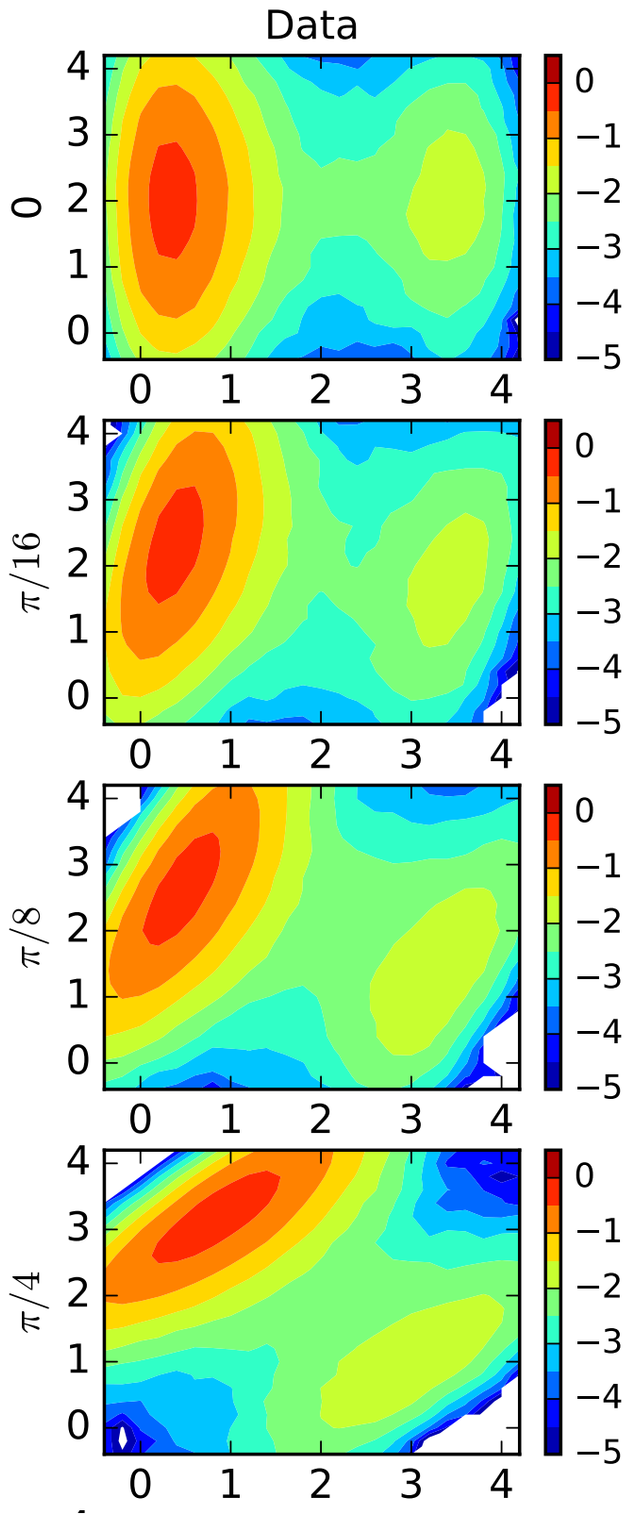
$$\|g_i - \psi_i\|_{H^1}^2 = \|g_i - \psi_i\|_{L^2}^2 + \sum_{j=1}^{N} \|\frac{\partial g_i}{\partial x_j} - \frac{\partial \psi_i}{\partial x_j}\|_{L^2}^2 \leq \epsilon^2, \tag{31}$$

then the maximal relative error between the dominant eigenvalues of $\mathcal{L}^\xi$ and $\mathcal{L}$ is bounded by

$$\max_{i=2,\ldots,M} \frac{\omega_i - \kappa_i}{\omega_i} \leq (M-1)\kappa_2^{-1/2} \|\boldsymbol{\sigma}\| \sqrt{\beta^{-1}} \epsilon. \tag{32}$$

Nüske and Clementi (in preparation)

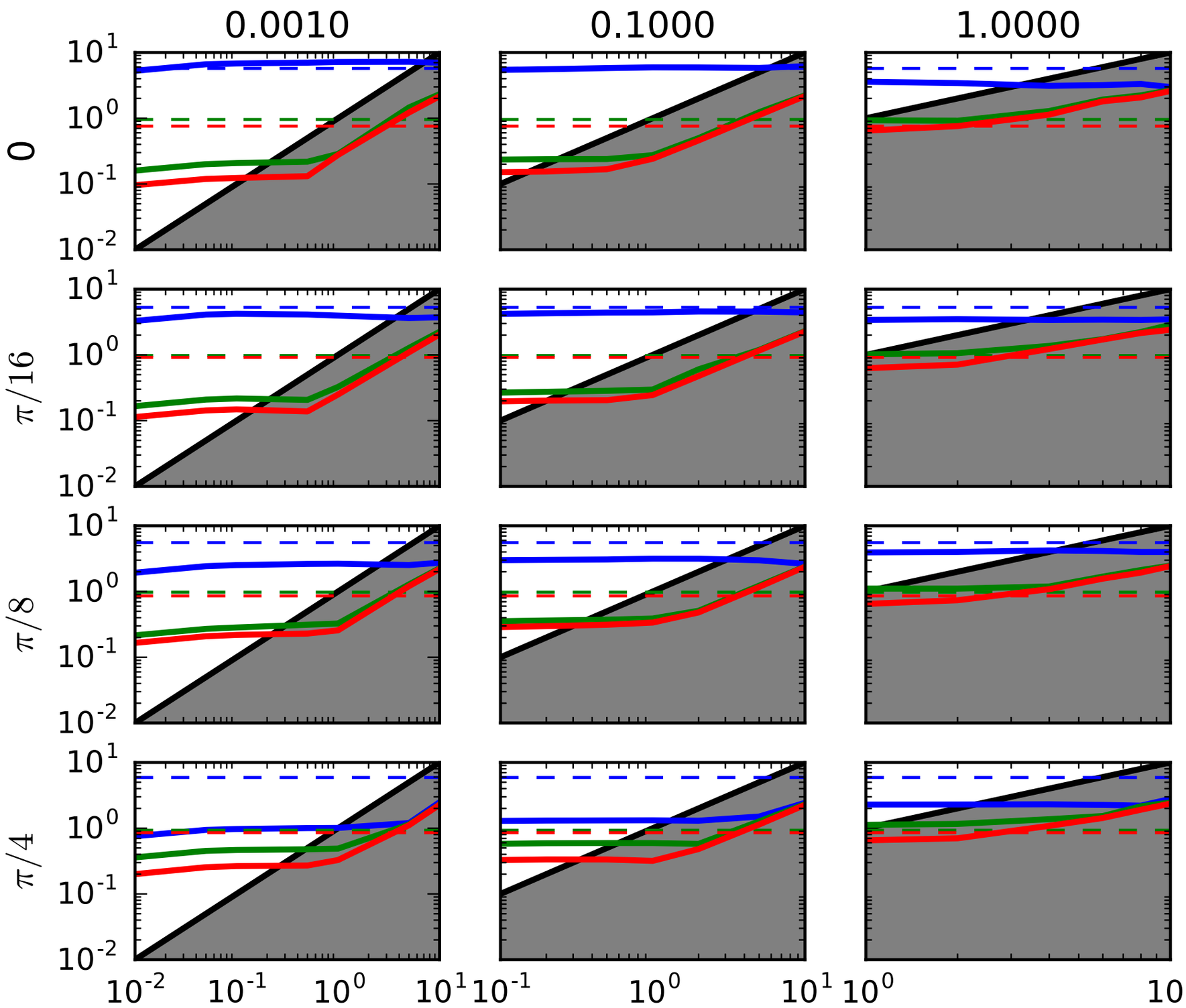
Numerical Example



- Two-d landscape
- Dynamics in x-direction is slow.
- Study several rotations of the system.
- Always use x as reaction coordinate.

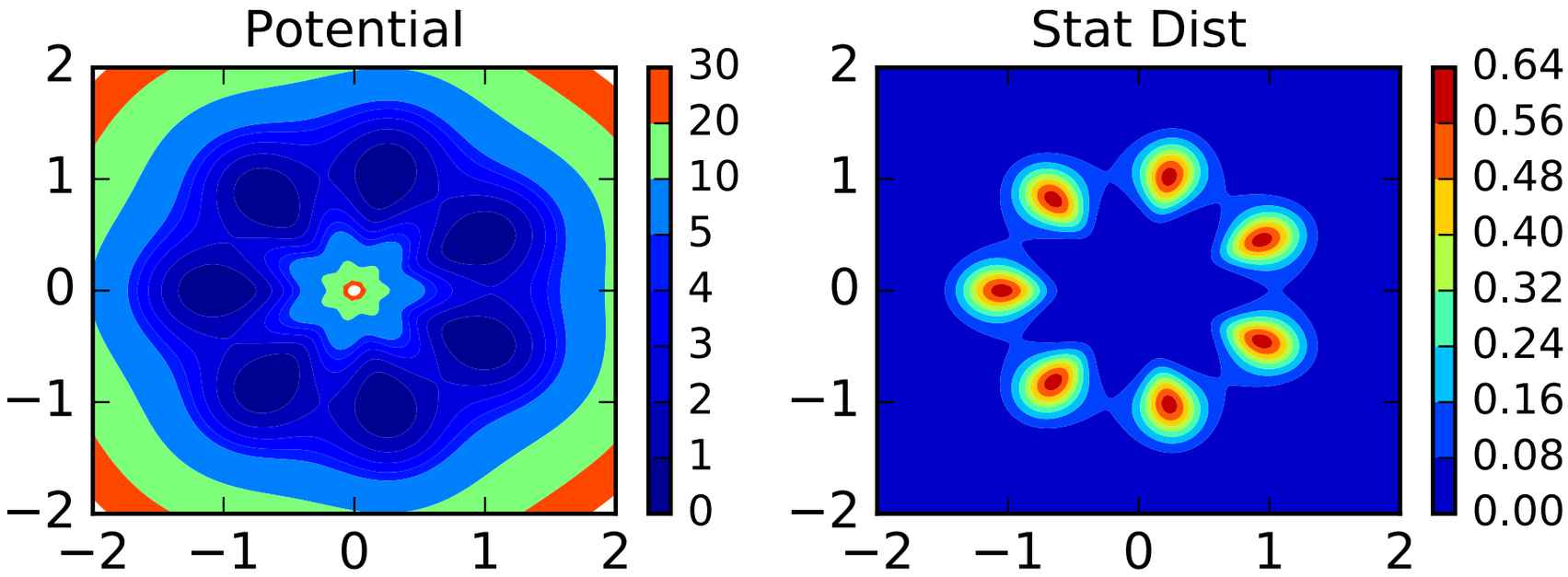
Nüske and Clementi (in preparation)

Numerical Example



Nüske and Clementi (in preparation)

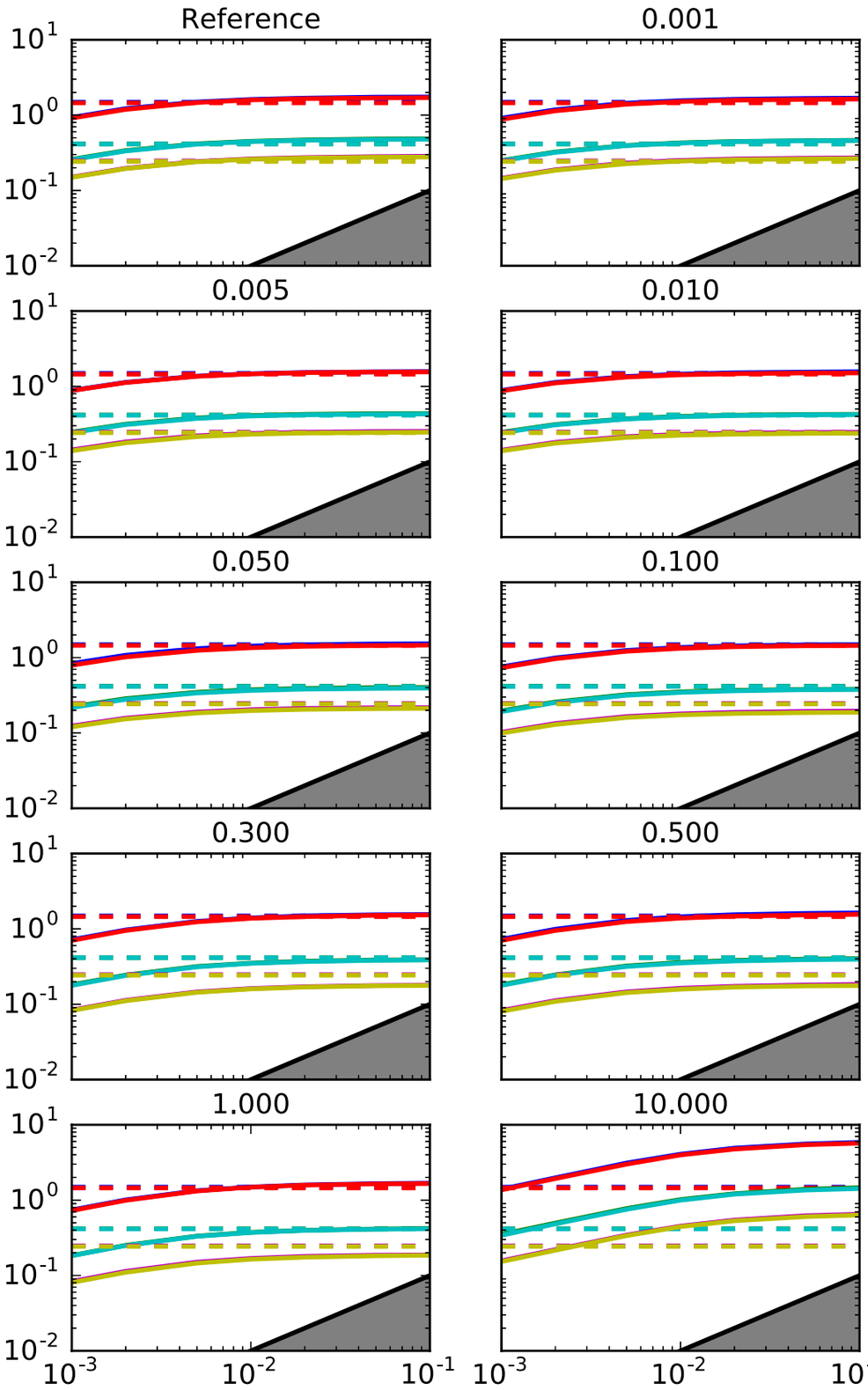
Numerical Example II



Use the polar angle as reaction coordinate.

Nüske and Clementi (in preparation)

Numerical Example II



Nüske and Clementi (in preparation)

ACKNOWLEDGMENTS

Prof. Cecilia Clementi

Dr. Giovanni Pinamonti
Lorenzo Boninsegna
Alexander Kluber
Justin Chen
Eugen Hruska

Our collaborators:

Dr. Ralf Banisch (Freie Universität Berlin)
Dr. Péter Koltai (Freie Universität Berlin)
Prof. Frank Noé (Freie Universität Berlin)