Geometric Methods for the approximation of high-dimensional dynamical systems

Mauro Maggioni
Department of Mathematics,
Electrical and Computer Engineering and
Computer Science,
Duke University

July 25th., 2014
I.P.A.M.

Geometry of Data I: Low-dimensional “Models”

Principal Component Analysis

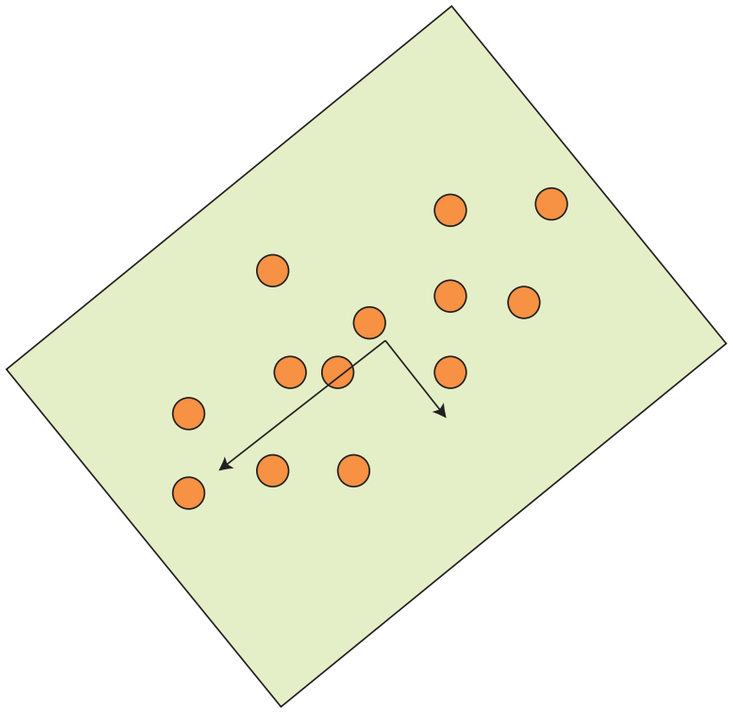
Curse of dimensionality: sample size for important statistical tasks scales exponentially in the ambient dimension. Hope: data concentrates near low-dimensional sets.

Represent data as a matrix $X \in \mathbb{R}^{D \times n}$, one sample per column.
If data lies on a k-dimensional hyperplane, compute the Singular Value Decomposition $X = U\Sigma V^T$: then only k singular values are nonzero, and the first k columns of U span the desired hyperplane.

U: orthogonal $D \times D$ ← system of coordinates for points

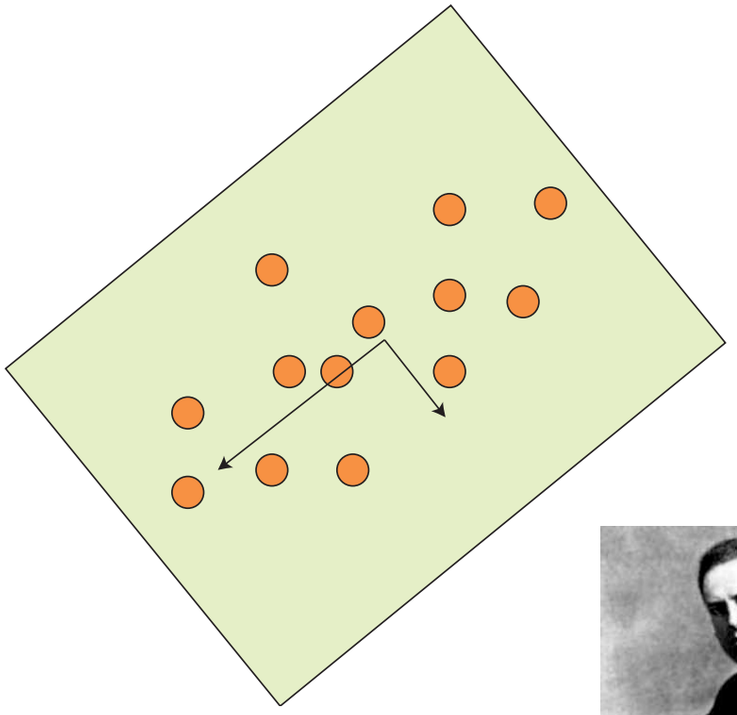
Σ: diagonal $D \times n$
Diagonal entries $\sigma_1 \geq \sigma_2 \geq \cdots \geq 0$ are called singular values.

V: orthogonal $n \times n$ ← system of coordinates for features

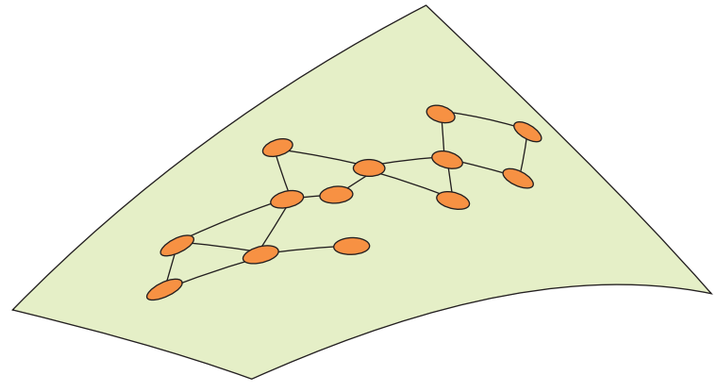


1901, K. Pearson

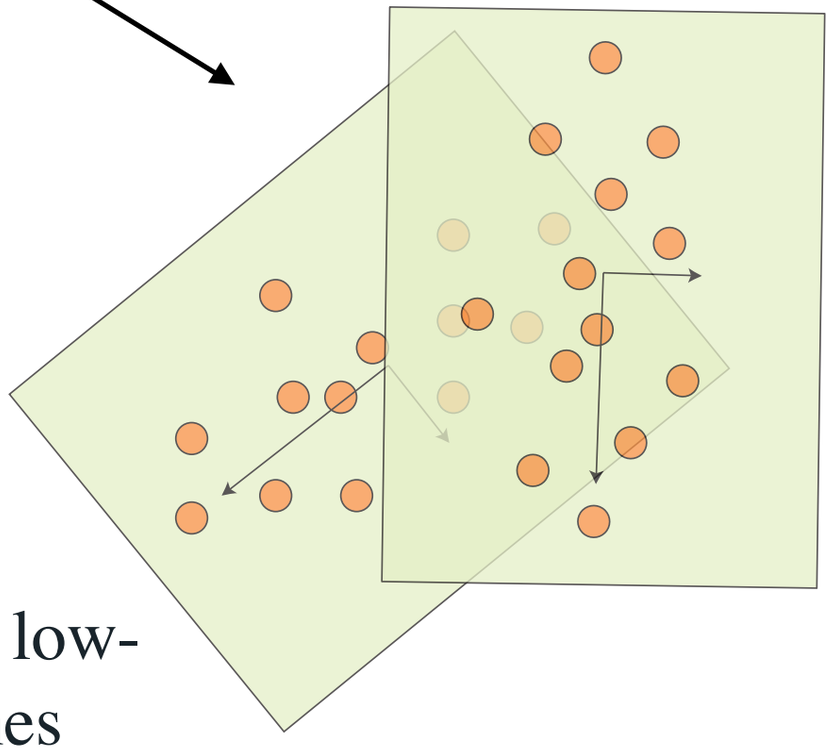
Beyond PCA



Data on low-dimensional plane (Pearson, 1901)



Data on low-dimensional curved structure



Data on multiple low-dimensional planes

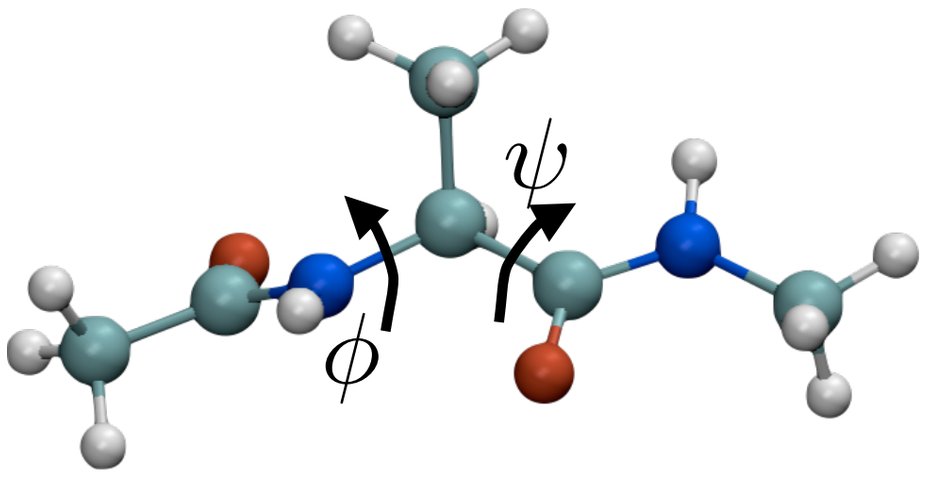
Geometry of Data II:
Simple models for M.D.

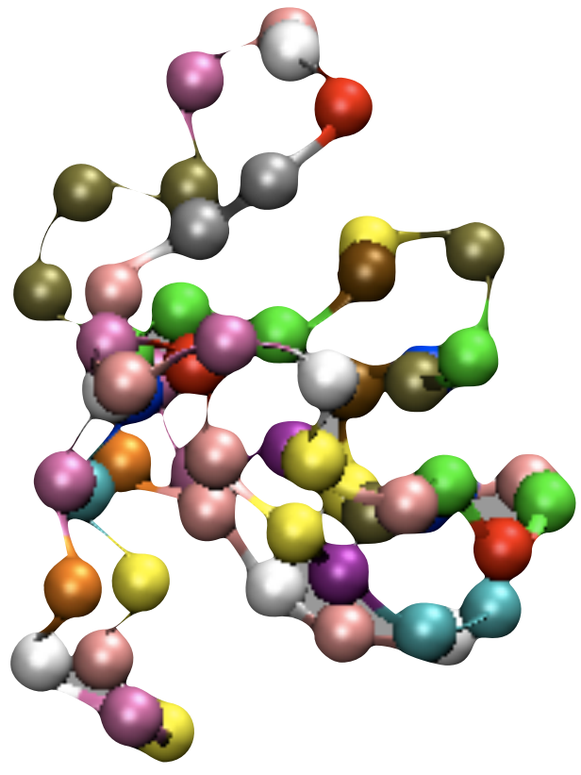
Molecular Dynamics Data

The dynamics of a small peptide (12 atoms with H-atoms removed) in a bath of water molecules, is approximated by a Langevin system of stochastic equations

$$\dot{x} = -\nabla U(x) + \dot{w}$$

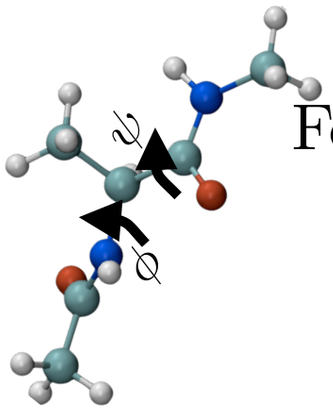
The set of configurations is a set of points in $\mathbb{R}^{3N}$; in the state space we take the RMSD metric, which quotients out the Galileo group.





Molecular Dynamics & F-P. equation

R.R.Coifman, I.G.Kevrekidis, S.Lafon,
MM, B.Nadler, *Multiscale Model. Simul.*



Fokker-Planck equation & eigenfunctions

$$\frac{\partial p}{\partial t} = -\sum_{i}^{3N} \frac{\partial}{\partial x_i}\left(\frac{1}{\beta}\frac{\partial}{\partial x_i} + \frac{\partial E}{\partial x_i}\right) p = -\mathbf{H}_{\mathrm{FP}} p$$

$\beta = 1/(k_B T)$, k_B is Boltzmann's constant

Under suitable conditions, $\mathbf{H}_{\mathrm{FP}}$ has discrete spectrum $0 = \lambda_0 < \lambda_1 \leq \ldots \lambda_k \ll \lambda_{k+1} \leq \ldots$, and fundamental solution with eigen-expansion

$$p_t(x,y) = \phi_0(x) + \sum_{j=1}^{+\infty} \psi_j(y)\phi_j(x)e^{-\lambda_j t}\,.$$

The dual system of eigenfunctions, which we pick as reaction coordinates, is

$$\psi_j(x) = \phi_j(x)/\phi_0(x)\,.$$

Diffusion Distance at time t

$$d^{(t)}(x,y) = ||p_t(x,\cdot) - p_t(y,\cdot)||_{L^2} = \sqrt{\sum_j e^{-\lambda_j t}|\psi_j(x) - \psi_j(y)|^2}$$

Diffusions maps for MD

R.R. Coifman, I.G. Kevrekidis, S. Lafon, MM, B. Nadler

$$\dot{x} = -\nabla U(x) + \sqrt{\frac{2}{\beta}}\dot{w}\,,\; x \in \mathbb{R}^{3N}$$

$$\frac{\partial p}{\partial t} = \mathcal{L}p = \tfrac{1}{\beta}\Delta p + \nabla\cdot(p\nabla U) = -\mathbf{H}_{\mathrm{FP}}p \qquad \mathbf{H}_{\mathrm{FP}}\varphi_j = \lambda_j\varphi_j \qquad \mathbf{H}^*_{\mathrm{FP}}\psi_j = \lambda_j\psi_j$$

Theorem: *The optimal k-dimensional approximation of $p(x,t|y)$ that minimizes the mean squared norm of the approximation error*

$$\min_{m_k} \mathbb{E}_y\left\{\|p(x,t|y) - m_k(x,t|y)\|^2_{L_2(\Omega,w)}\right\}$$

where averaging is over all initial points y sampled according to the equilibrium density $\varphi_0(y)$, is given by by the truncated sum

$$p_k(x,t|y) = \sum_{j=0}^{k-1} e^{-\lambda_j t}\psi_j(y)\varphi_j(x).$$

How do we compute/learn the above?

This is simply an SVD approximation w.r.t. the diffusion kernel.

Empirical approximations, given samples distributed according to φ_0, can be obtained by suitably modifying diffusion maps (with guarantees).

Random walks on graphs and data

R. Coifman, S. Lafon, MM, B. Nadler

We may construct weighted graphs from data: given

. Data $X = \{x_i\}_{i=1}^n \subset \mathbb{R}^D$.

. Local similarities via a kernel function $W(x_i, x_j) \geq 0$.

Simplest example: $W_\sigma(x_i, x_j) = e^{-||x_i - x_j||^2/\sigma}$.

Given a weighted graph (G, E, W): vertices represent data points, edges connect x_i, x_j with weight $W_{ij} := W(x_i, x_j)$, when positive (or above a threshold). Let $D_{ii} = \sum_j W_{ij}$ and

$$P = D^{-1}W, \qquad P_{ij} = \text{prob. of jumping } x_i \to x_j$$

Let $1 = \lambda_0 \geq \lambda_1 \geq \dots$ and φ_i the eigenval.'s and eigenvec.'s of P, i.e. $P\varphi_i = \lambda_i \varphi_i$. We consider the map

$$G \ni x \mapsto (\varphi_1(x), \dots, \varphi_m(x)) \in \mathbb{R}^m$$

Random walks on graphs and data

R. Coifman, S. Lafon, MM, B. Nadler, Y. Kevrekidis

Given a weighted graph (G, E, W): vertices represent data points, edges connect x_i, x_j with weight $W_{ij} := W(x_i, x_j)$, when positive (or above a threshold). Let $D_{ii} = \sum_j W_{ij}$ and

$$P = D^{-1}W, \qquad P_{ij} = \text{prob. of jumping } x_i \to x_j$$

Let $1 = \lambda_0 \geq \lambda_1 \geq \dots$ and φ_i the eigenval.'s and eigenvec.'s of P, i.e. $P\varphi_i = \lambda_i \varphi_i$. We consider the map

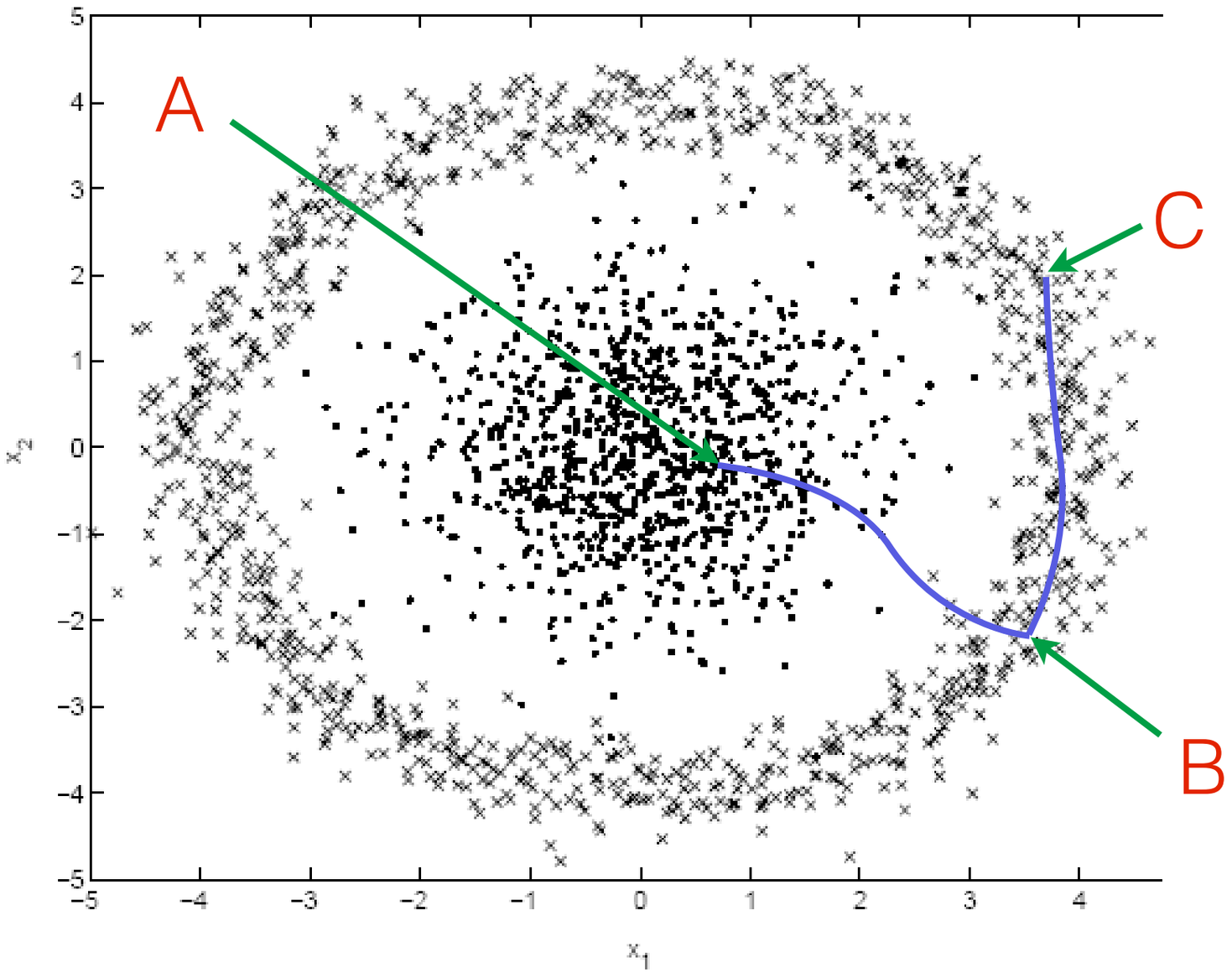
$$G \ni x \mapsto (\varphi_1(x), \dots, \varphi_m(x)) \in \mathbb{R}^m$$

With suitable modifications, one can show that as $n \to +\infty$, and if the samples $\{x_i\}$ are drawn according to the stationary distribution of the process, P converges to the propagator of the Fokker-Planck equation.

Clustering, diffusion distances

Joint with R. Coifman and S. Lafon

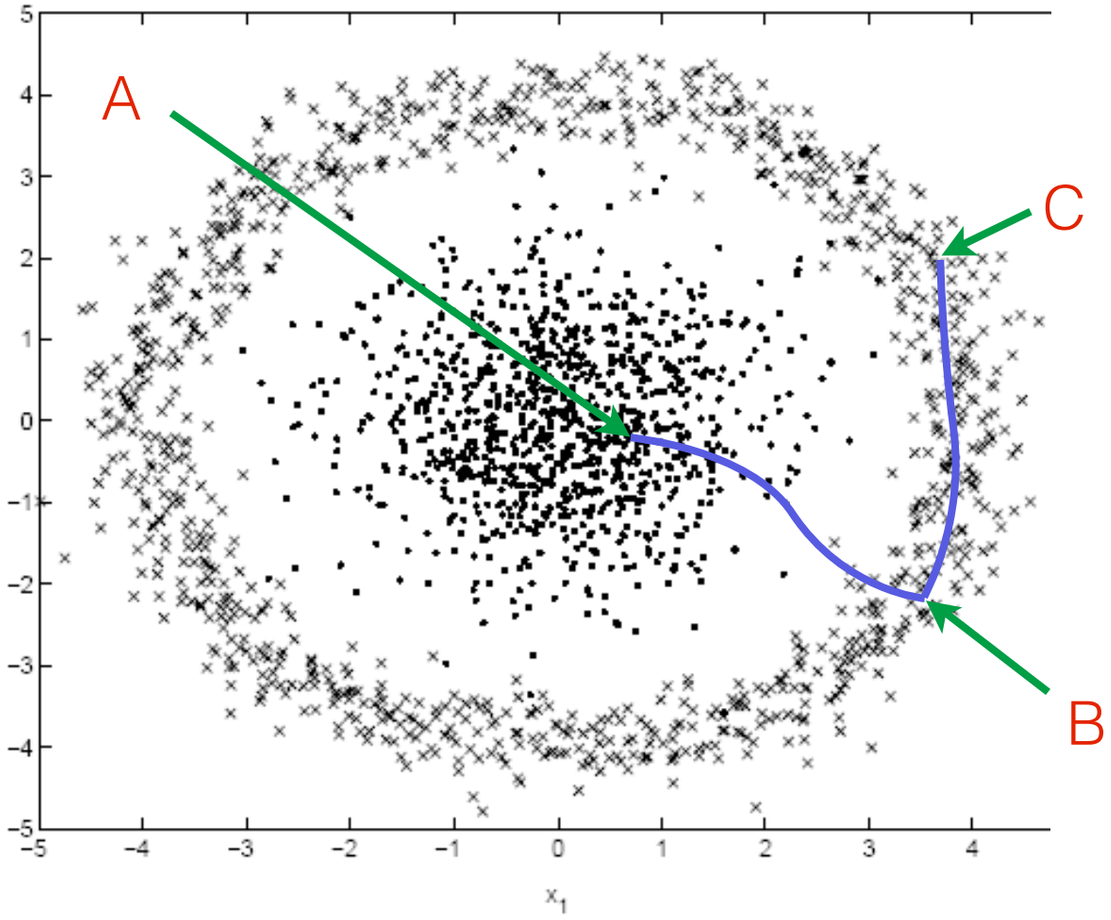
In some cases the geodesic distance d_G may not capture geometric information of interest. For example here $d_G(A, B) \sim d_G(B, C)$. The diffusion distance is sensitive to connectivity between clusters.



Shortest paths have comparable length

Spectral Clustering in one slide

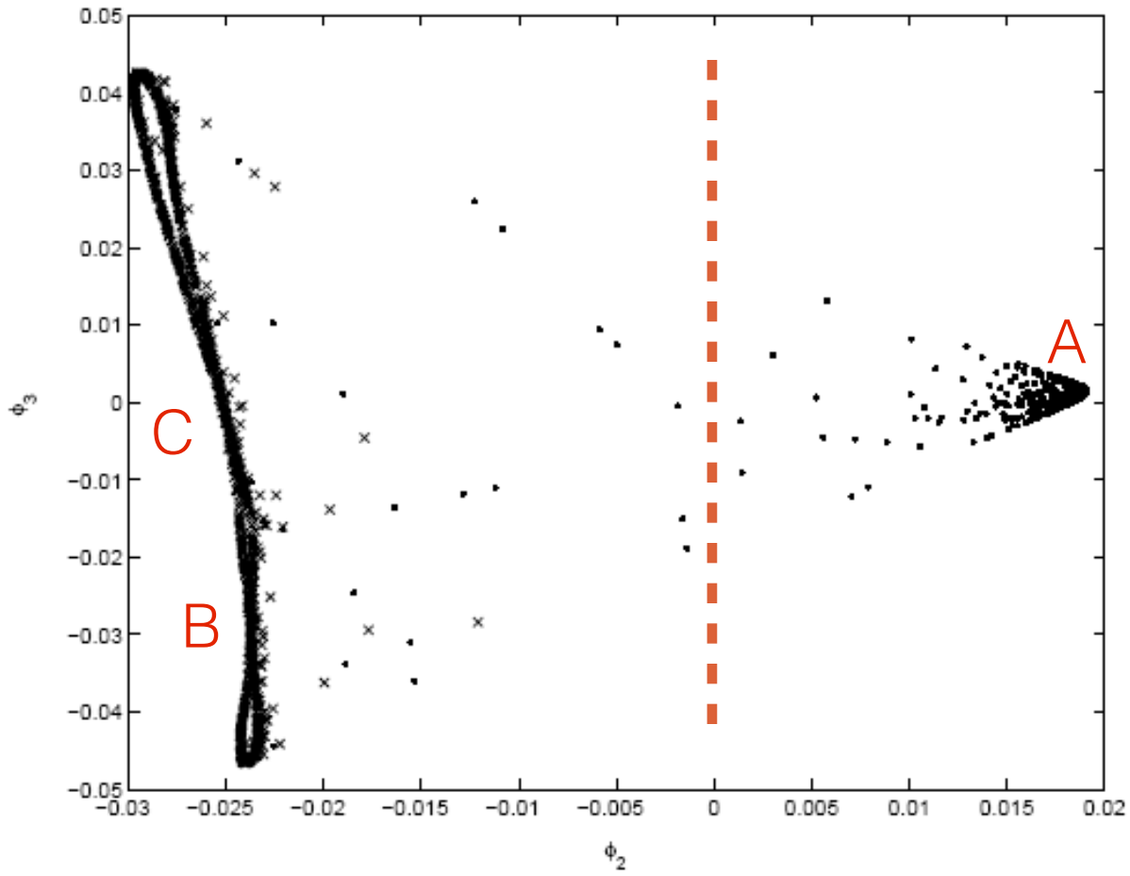
$$x \mapsto (\phi_2(x), \phi_3(x))$$



Original space

Every point is connected to its 5 nearest neighbors, obtaining a graph.

Flexibility + robustness



Diffusion space

$\phi_2 = 0$ is a good cut!

$$\langle Lf, f \rangle = \sum_x \sum_{y \sim x} W(x,y) \left(\frac{f(x)}{\sqrt{d_x}} - \frac{f(y)}{\sqrt{d_y}} \right)^2$$

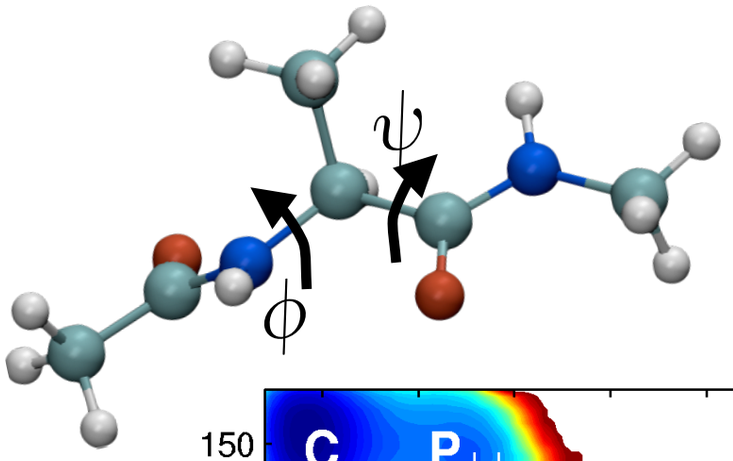
Example: Molecular Dynamics Data

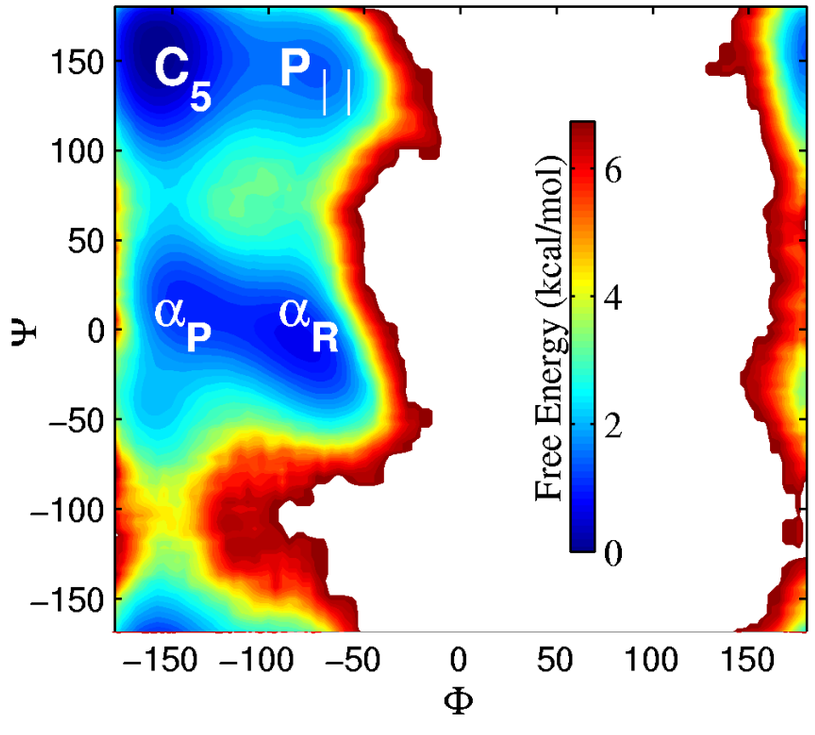
Joint with C. Clementi, M. Rohrdanz, W. Zheng

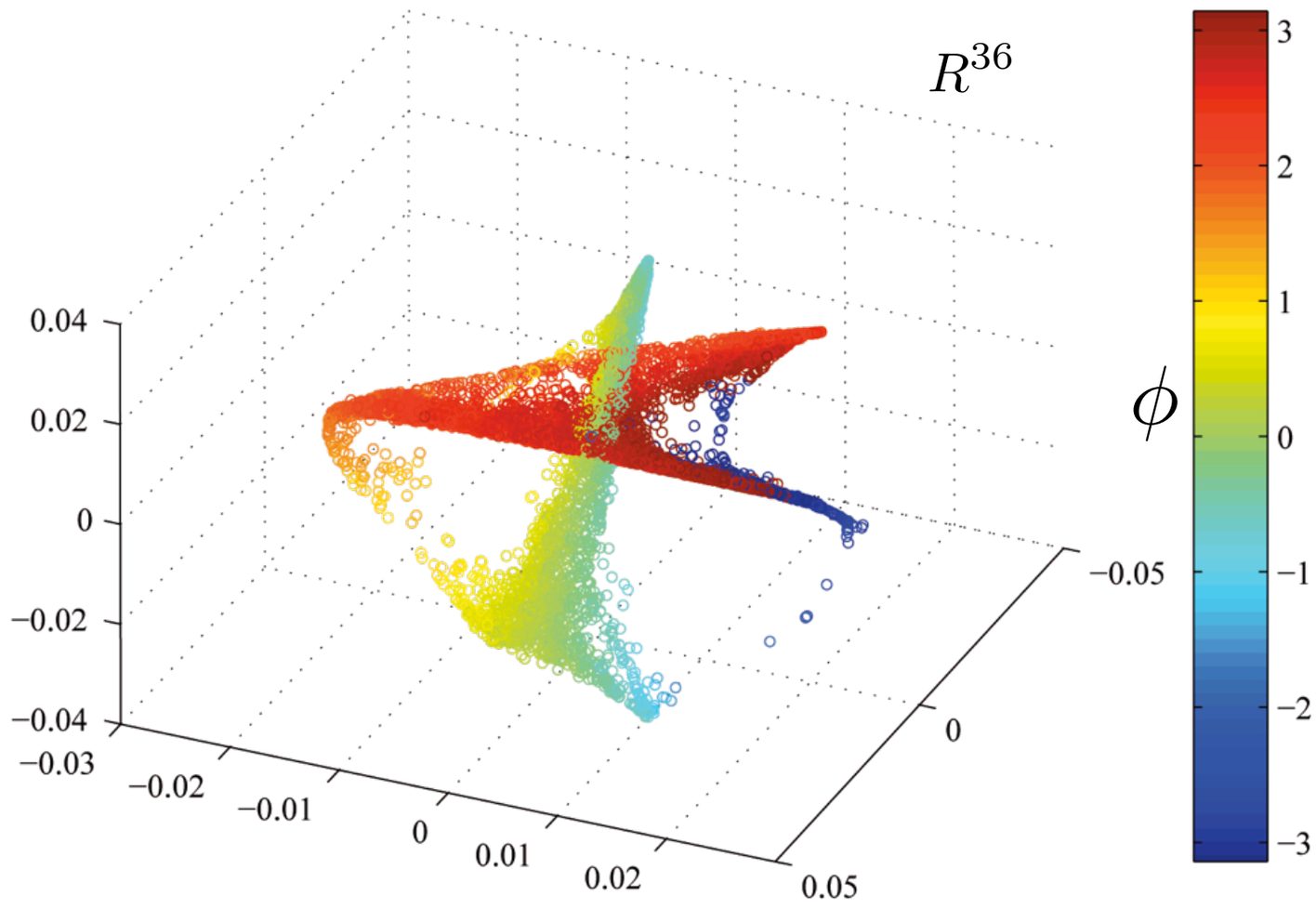
The dynamics of a small peptide (12 atoms with H-atoms removed) in a bath of water molecules, is approximated by a Langevin system of stochastic equations

$$\dot{x} = -\nabla U(x) + \dot{w}$$

The set of configurations is a point cloud in $\mathbb{R}^{12\times 3}$.







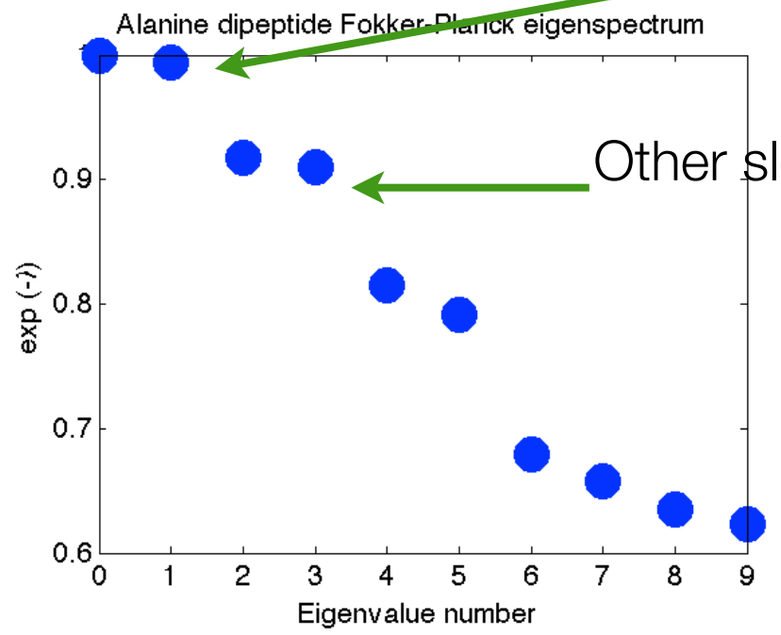
Molecular Dynamics data for alanine

M. Rohrdanz, W. Zheng, MM, C. Clementi, JCP

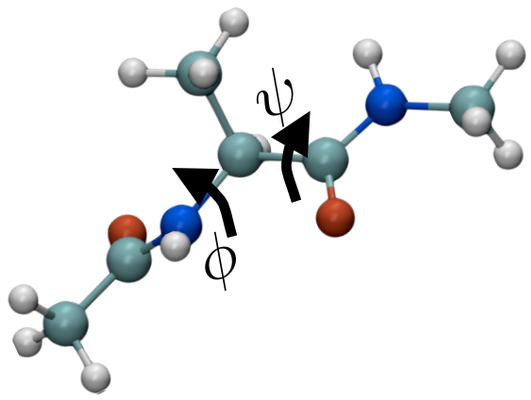
Given only trajectory data, we construct an empirical approximation to the generator of the Fokker-Planck, and compute its eigenvalues/vectors to obtain a low-dimensional embedding and reaction coordinates.

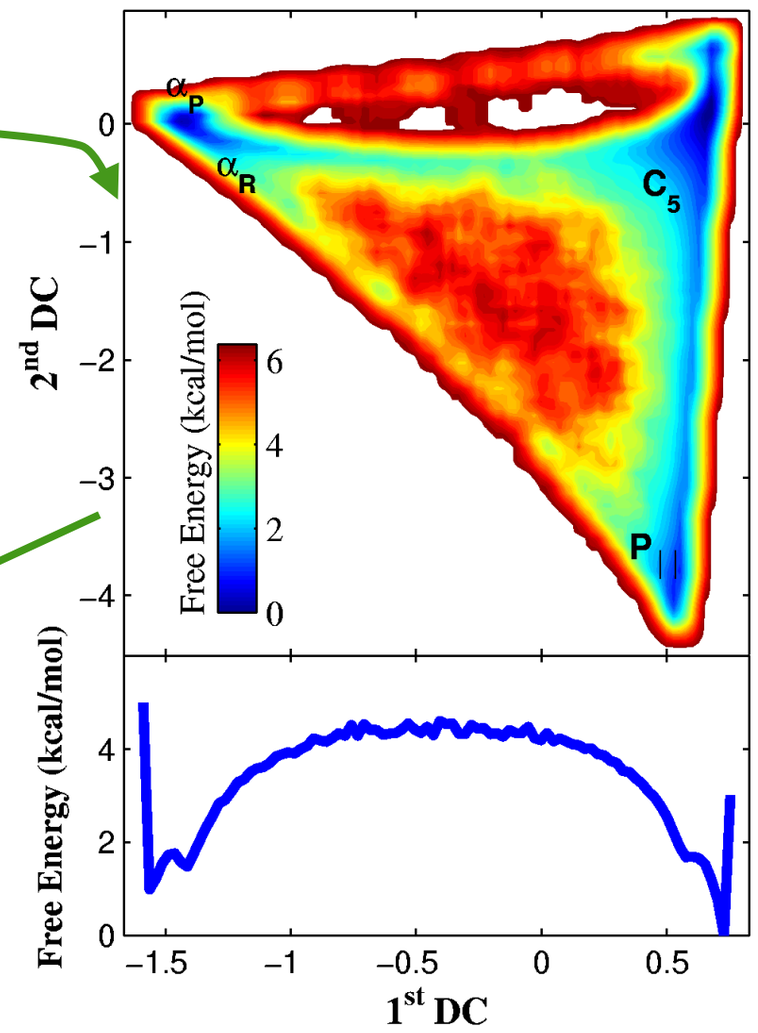
$$\frac{\partial p}{\partial t} = -\sum_{i}^{3N} \frac{\partial}{\partial x_i}\left(\frac{1}{\beta}\frac{\partial}{\partial x_i} + \frac{\partial E}{\partial x_i}\right) p = -\mathbf{H}_{\mathrm{FP}} p$$

Slowest mode



Other slow modes





Coordinate	C_5, $P_{\|} \rightarrow \alpha_R\ \alpha_P$	$\alpha_R\ \alpha_P \rightarrow C_5$, $P_{\|}$
From simulation[b]	0.023	0.047
1st DC	0.023 ± 0.001	0.048 ± 0.003
Ψ	0.020 ± 0.001	0.040 ± 0.003

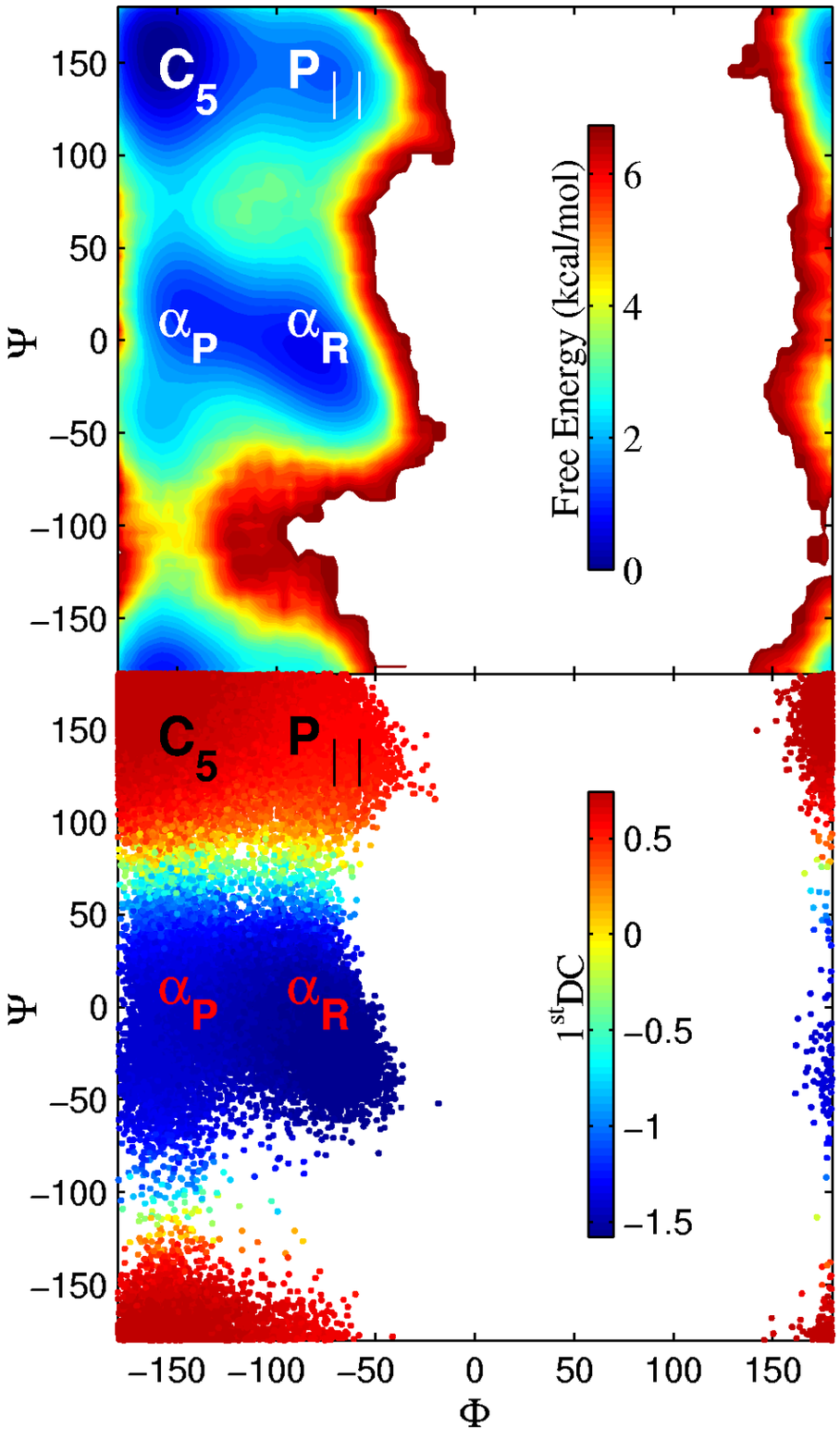
Reduced model with reasonably good rates

$$\text{rate} = \left(\int_{\text{barrier}} \frac{e^{\beta F(x)}}{D(x)} dx \int_{\text{well}} e^{-\beta F(x')} dx'\right)^{-1}$$

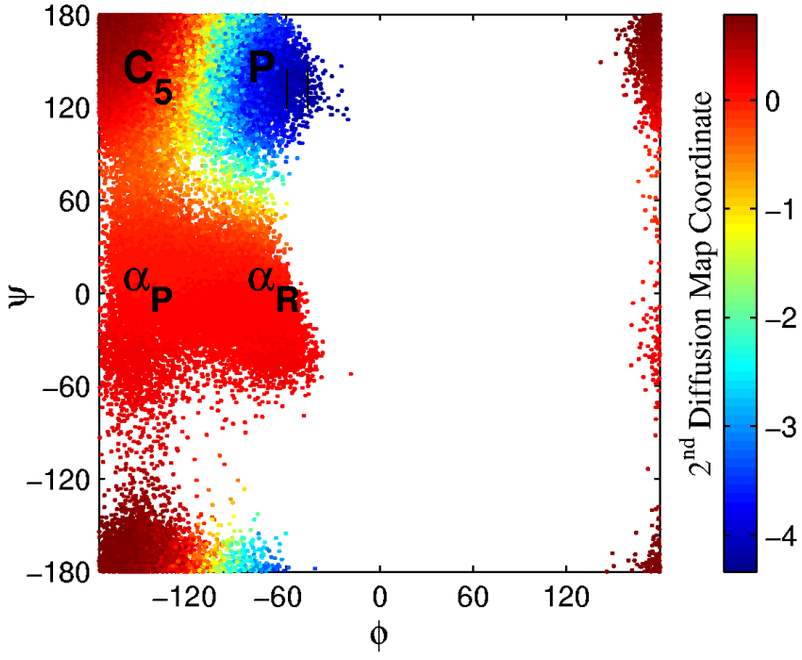
G. Hummer, 2005

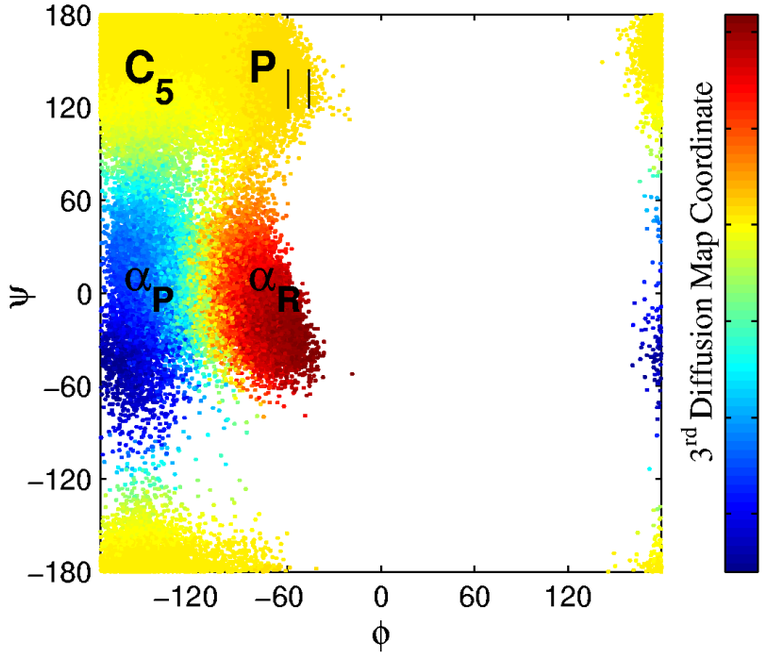
Diffusion coord.'s - empirical coord.'s

M. Rohrdanz, W. Zheng, MM, C. Clementi, JCP



We may plot the diffusion coordinates as functions of the physical observables given by the angles and notice they are essentially in one-to-one correspondence, with the diffusion coordinates emphasizing energy barriers separating minima.





Random walks on graphs and data

R. Coifman, S. Lafon, MM, B. Nadler, Y. Kevrekidis

We may construct weighted graphs from data: given

. Data $X = \{x_i\}_{i=1}^n \subset \mathbb{R}^D$.

. Local similarities via a kernel function $W(x_i, x_j) \geq 0$.

Simplest example: $W_\sigma(x_i, x_j) = e^{-||x_i - x_j||^2/\sigma}$.

Given a weighted graph (G, E, W): vertices represent data points, edges connect x_i, x_j with weight $W_{ij} := W(x_i, x_j)$, when positive (or above a threshold). Let $D_{ii} = \sum_j W_{ij}$ and

$$P = D^{-1}W, \qquad P_{ij} = \text{prob. of jumping } x_i \to x_j$$

Let $1 = \lambda_0 \geq \lambda_1 \geq \dots$ and φ_i the eigenval.'s and eigenvec.'s of P, i.e. $P\varphi_i = \lambda_i \varphi_i$. We consider the map

$$G \ni x \mapsto (\varphi_1(x), \dots, \varphi_m(x)) \in \mathbb{R}^m$$

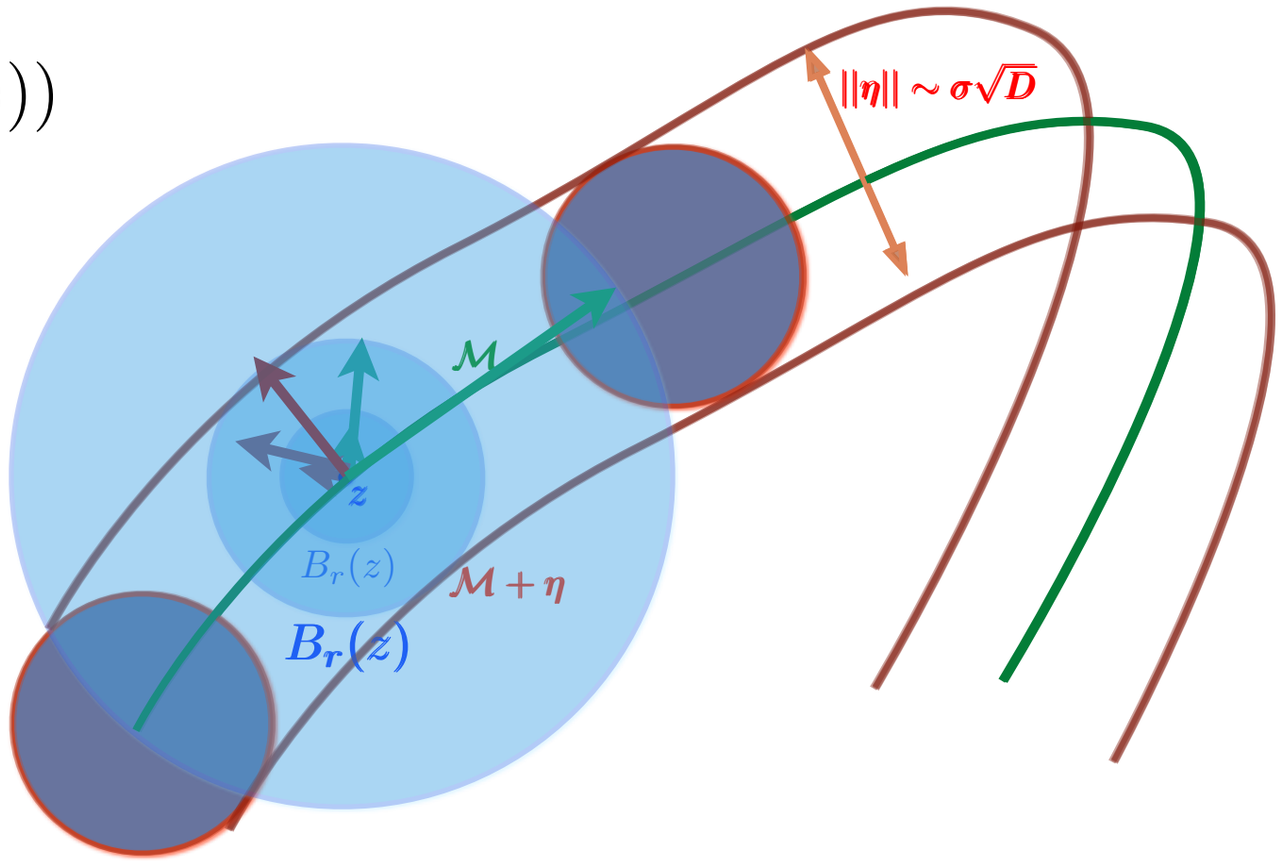
Geometry of Data III: Intrinsic Dimension

Multiscale SVD: $\mathrm{cov}(\mu|_{B_z(r)})$

Model: data $\{x_i\}_{i=1}^n$ is sampled from a manifold $\mathcal{M}$ of dimension k, embedded in $\mathbb{R}^D$, with $k \ll D$. We receive $\tilde{X}_n := \{x_i + \eta_i\}_{i=1}^n$, where $\eta_i \sim_{\text{i.i.d}} N$ is D-dimensional noise (e.g. Gaussian). *Objective: estimate k.*

Consider $\sigma_i^{\tilde{z},r} = \lambda_i(\mathrm{cov}(\tilde{X}_n \cap B_{\tilde{z}}(r)))$ for all $r > 0$, and all $i = 1, \ldots, D$.

Green: where data is
Red: where noisy data is
Blue: volume in ball



Only need number of samples *linear* in the intrinsic dimension, robust to noise
Turns out to be consequence of a blessing of dimensionality

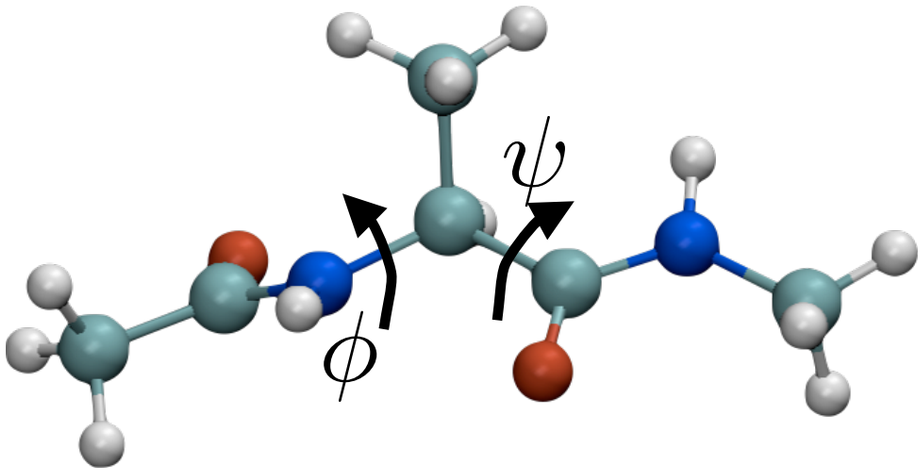
Molecular Dynamics Data

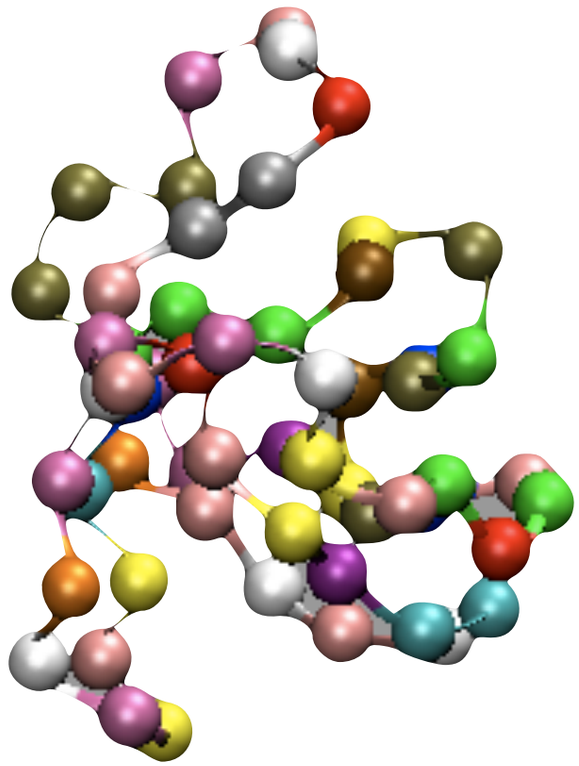
Joint with C. Clementi, M. Rohrdanz, W. Zheng

The dynamics of a small peptide (12 atoms with H-atoms removed) in a bath of water molecules, is approximated by a Langevin system of stochastic equations

$$\dot{x} = -\nabla U(x) + \dot{w}$$

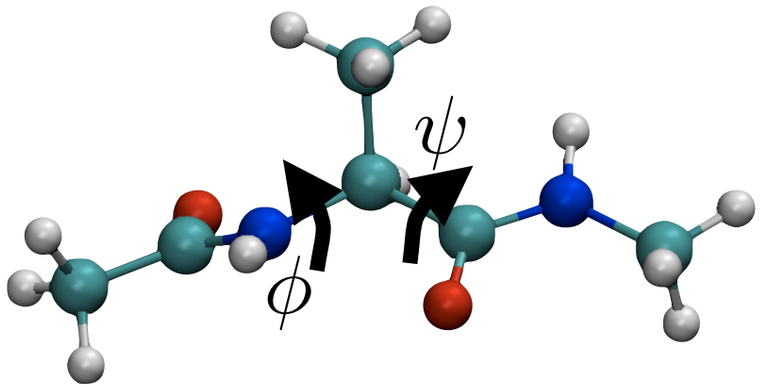
The set of configurations is a point cloud in $\mathbb{R}^{12\times 3}$.



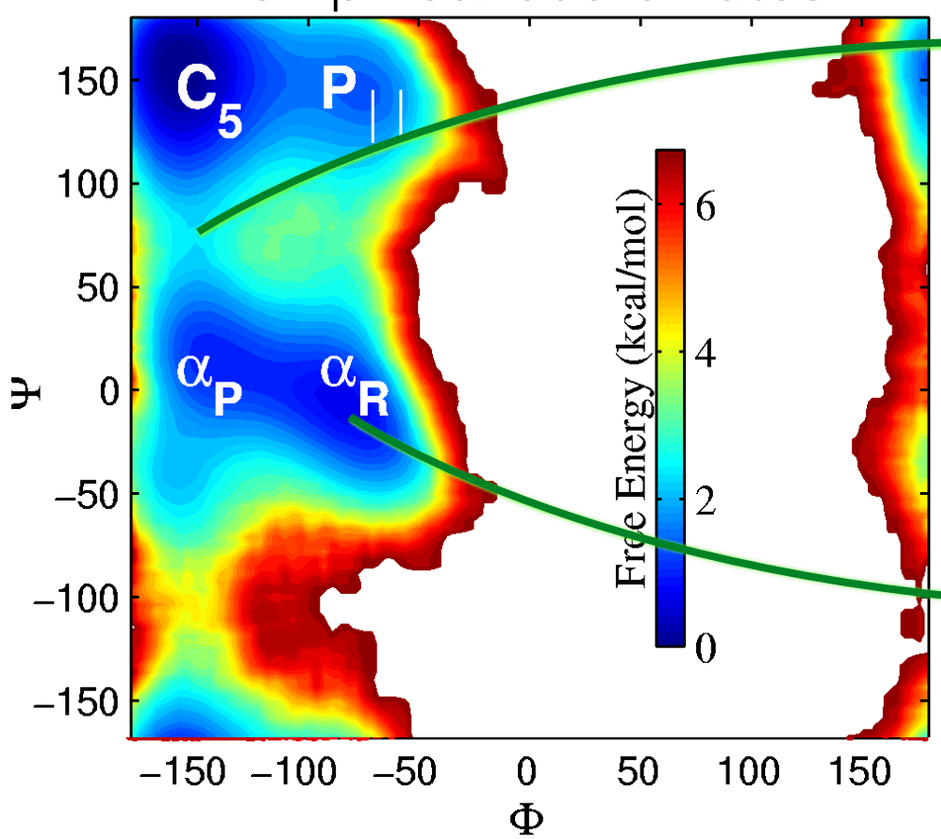


Example: Alanine dipeptide

M. Rohrdanz, W. Zheng, MM, C. Clementi, J. Chem. Phys. 2011

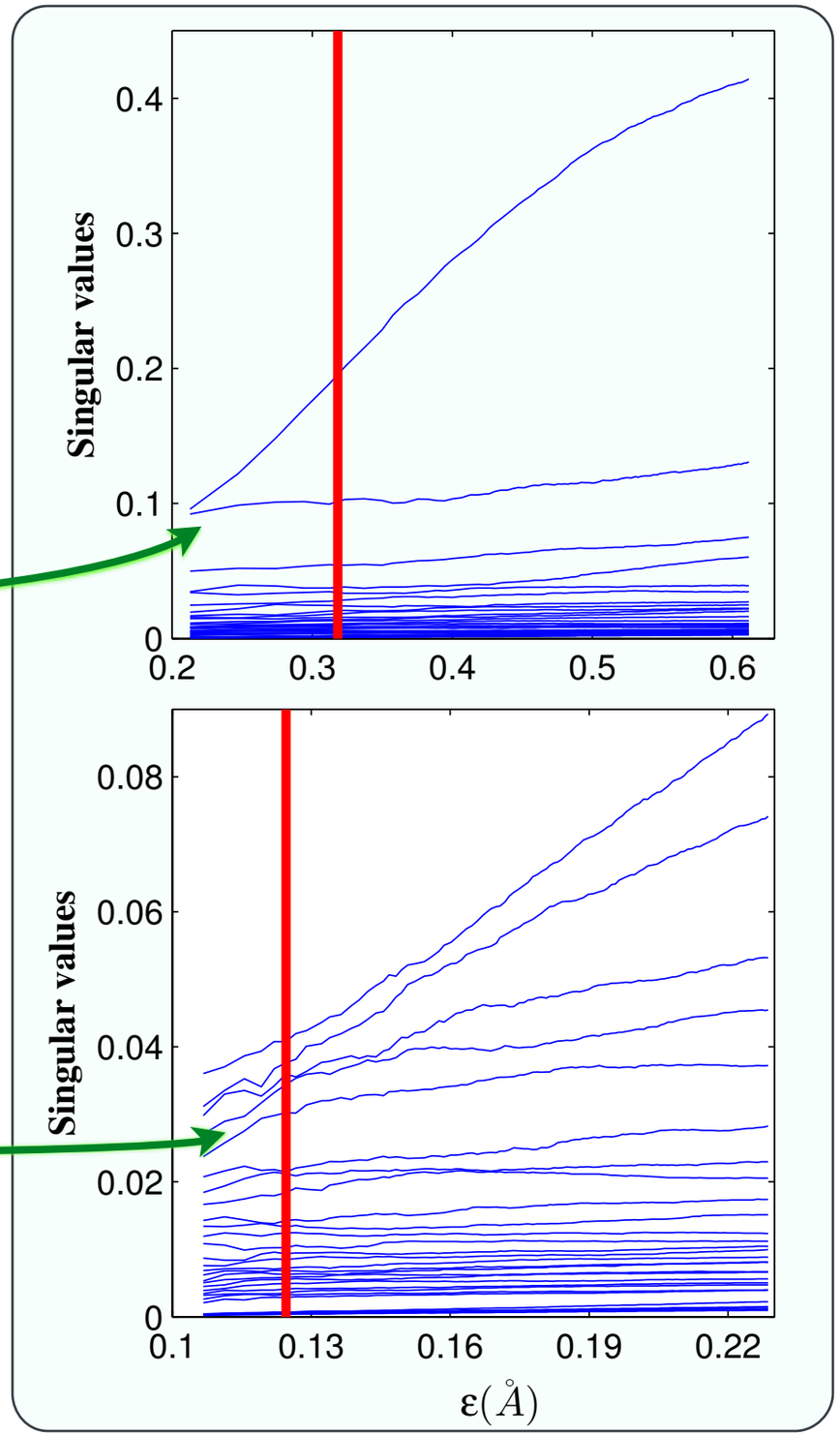


Free energy in terms of empirical coordinates



MSVD near transition state

MSVD near free energy minimum



Random walks on graphs and data

R. Coifman, S. Lafon, MM, B. Nadler, Y. Kevrekidis

We may construct weighted graphs from data: given

. Data $X = \{x_i\}_{i=1}^n \subset \mathbb{R}^D$.

. Local similarities via a kernel function $W(x_i, x_j) \geq 0$.

Simplest example: $W_\sigma(x_i, x_j) = e^{-||x_i - x_j||^2/\sigma}$.

Given a weighted graph (G, E, W): vertices represent data points, edges connect x_i, x_j with weight $W_{ij} := W(x_i, x_j)$, when positive (or above a threshold). Let $D_{ii} = \sum_j W_{ij}$ and

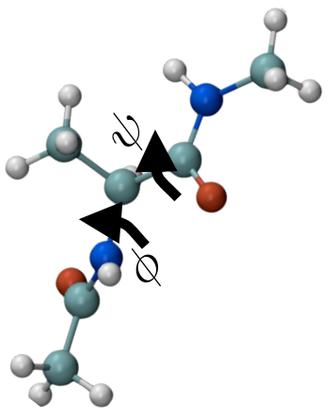
$$P = D^{-1}W, \qquad P_{ij} = \text{prob. of jumping } x_i \to x_j$$

Let $1 = \lambda_0 \geq \lambda_1 \geq \dots$ and φ_i the eigenval.'s and eigenvec.'s of P, i.e. $P\varphi_i = \lambda_i \varphi_i$. We consider the map

$$G \ni x \mapsto (\varphi_1(x), \dots, \varphi_m(x)) \in \mathbb{R}^m$$

Locally Scaled Diffusion Map

M. Rohrdanz, W. Zheng, MM, C. Clementi



. Construct the $N \times N$ matrix of transition probability kernels K, as

$$K_{ij} = e^{-\frac{d_{\mathrm{RMSD}}(x_i, x_j)^2}{2\epsilon_i \epsilon_j}},$$

for x_i and x_j molecular configurations, ϵ_i and ϵ_j their local scales.

. For each x_i, compute

$$P_i = \sum_{j=1}^{N} K_{ij},$$

which is proportional to a density estimation around x_i.

. Normalize the kernel as

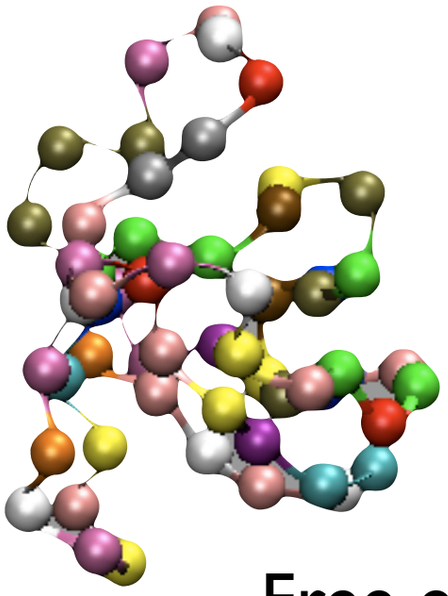
$$\tilde{K}_{ij} = P_i^{-\frac{1}{2}} K_{ij} P_j^{-\frac{1}{2}}.$$

. Define the diagonal matrix D as $D_i = \sum_{j=1}^{N} \tilde{K}_{ij}$, and construct a Markov matrix $M = D^{-1}\tilde{K}$,

$$M_{ij} = D_i^{-1} \tilde{K}_{ij}.$$

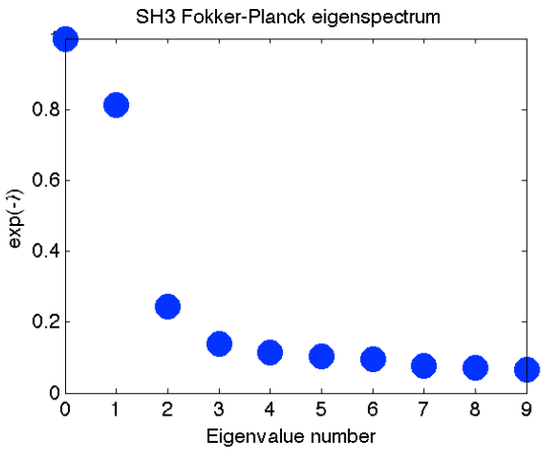
. Compute largest eigenvalues and corresponding right eigenvectors of M.

Example: SH-3

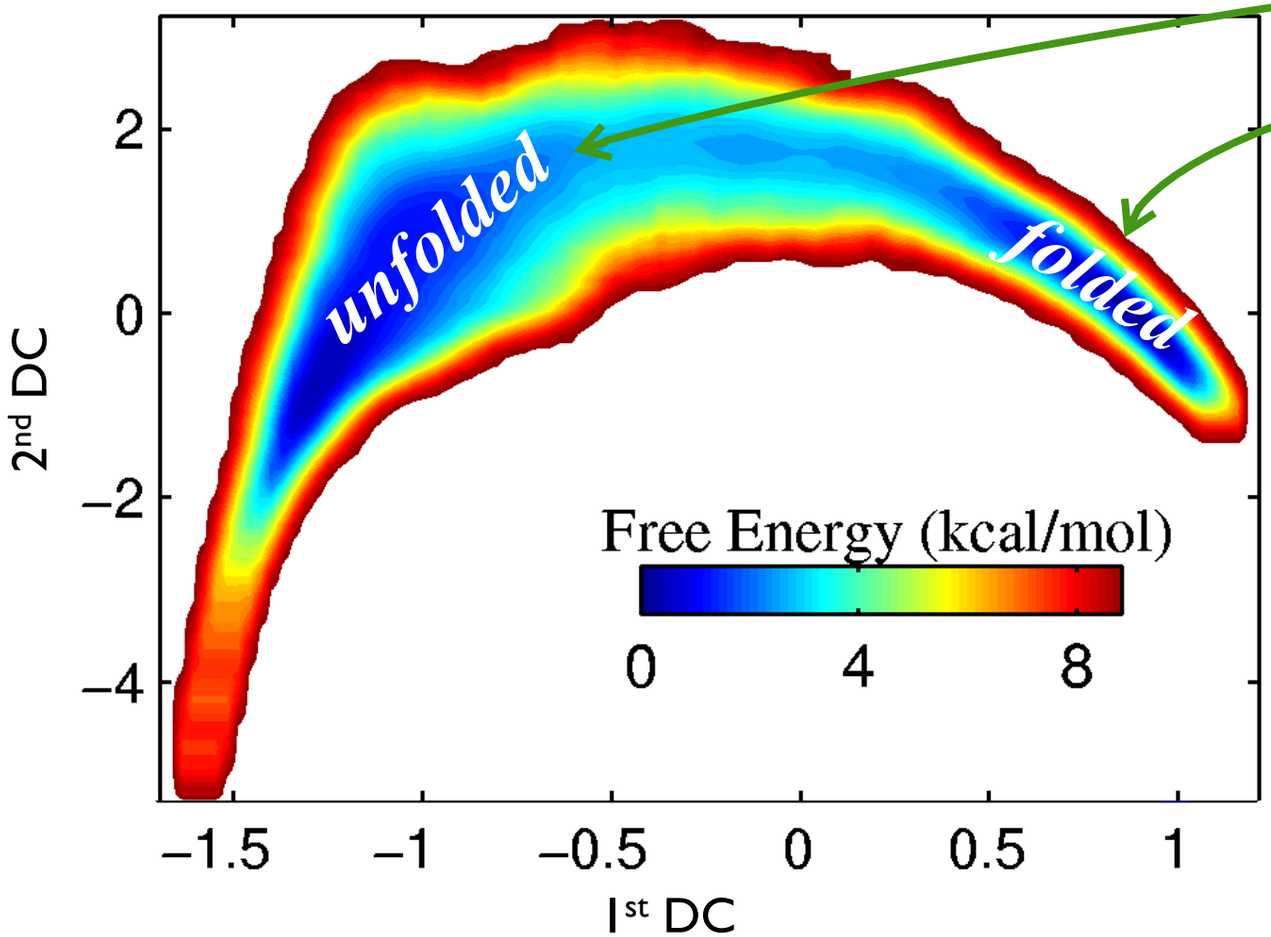


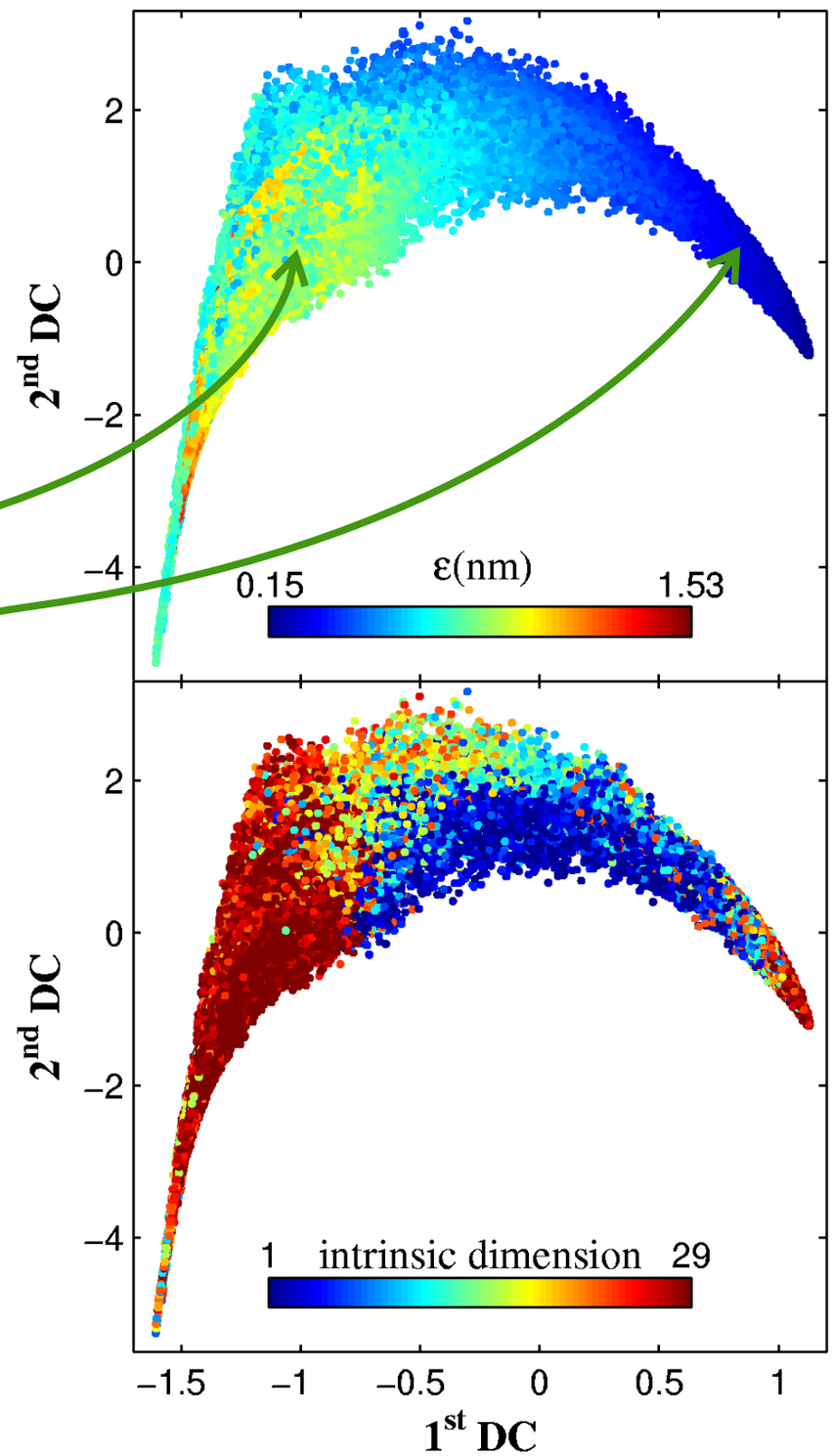
M. Rohrdanz, W. Zheng, MM, C. Clementi

$e^{-\lambda_i}$



Free energy in terms of diffusion coordinates



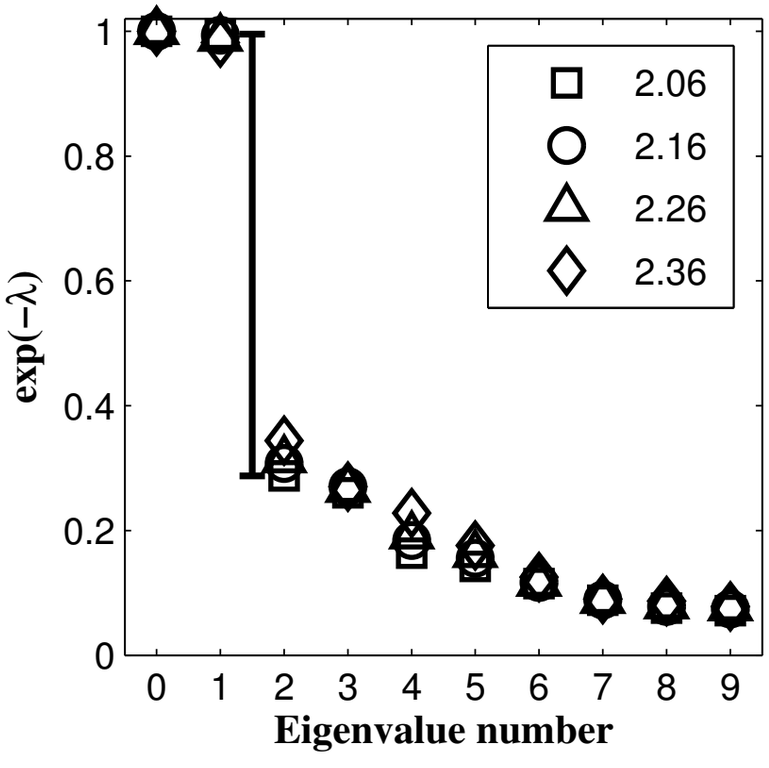


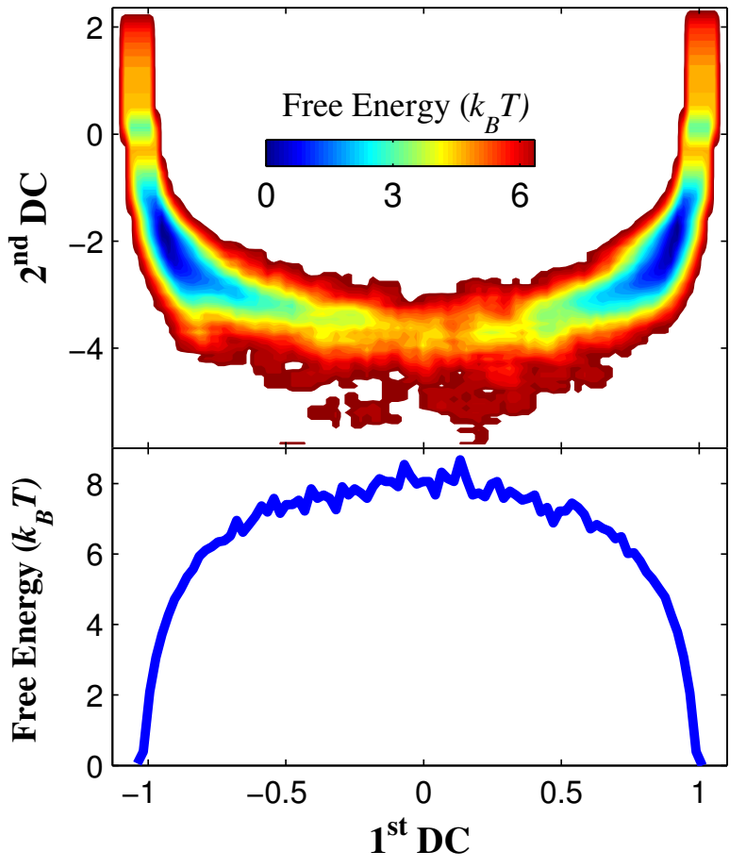
Small polymer reversal rate

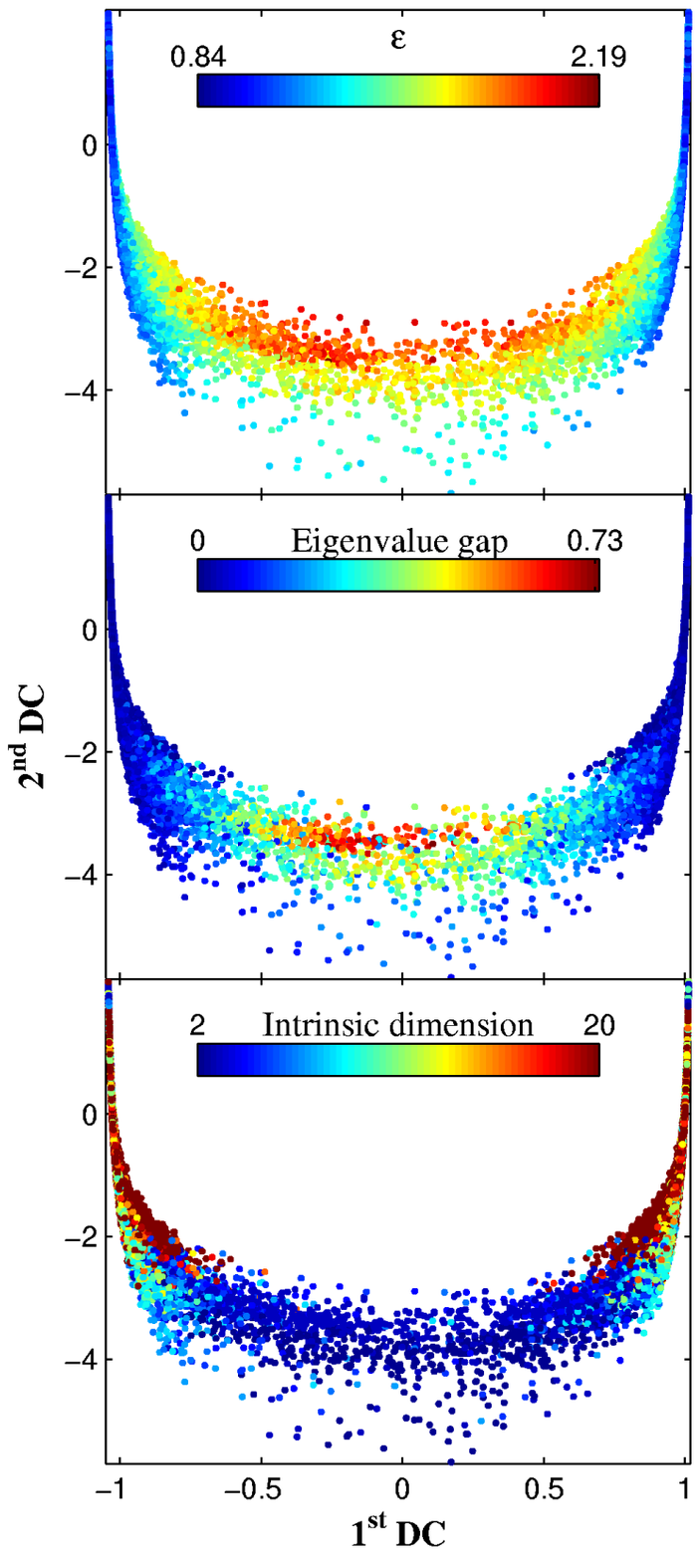
M. Rohrdanz, W. Zheng, MM, C. Clementi

20 bead polymer confined in nanopore; the axis *z* of nanopore is the reaction coordinate. Previous work by Huang and Makarov used distance between bead ends, projected onto *z*, as empirical reaction coordinates, and noticed poor approximation of the transition rates.

Locally scaled diffusion maps yield a reaction coordinate along which a more accurate transition rate is estimated; this coordinate corresponds quite well with *z*, while the second diffusion coordinate is related to radius of gyration of the polymer.







Summary, part I

"*Models*": high-dimensional data may often be approximated by low-dimensional geometric objects

Intrinsic dimension: use Multiscale SVD to measure intrinsic dimension

Coordinates: global low-dimensional parametrizations via diffusion maps. This can be adapted to Langevin dynamics for low-dimensional descriptions of large-time dynamics.

The above requires long trajectories - typically expensive

Geometry of Data IV: Learning Accelerated Dynamics

ATLAS: a geometric approach to learning stochastic systems in high dimensions

Miles Crosskey

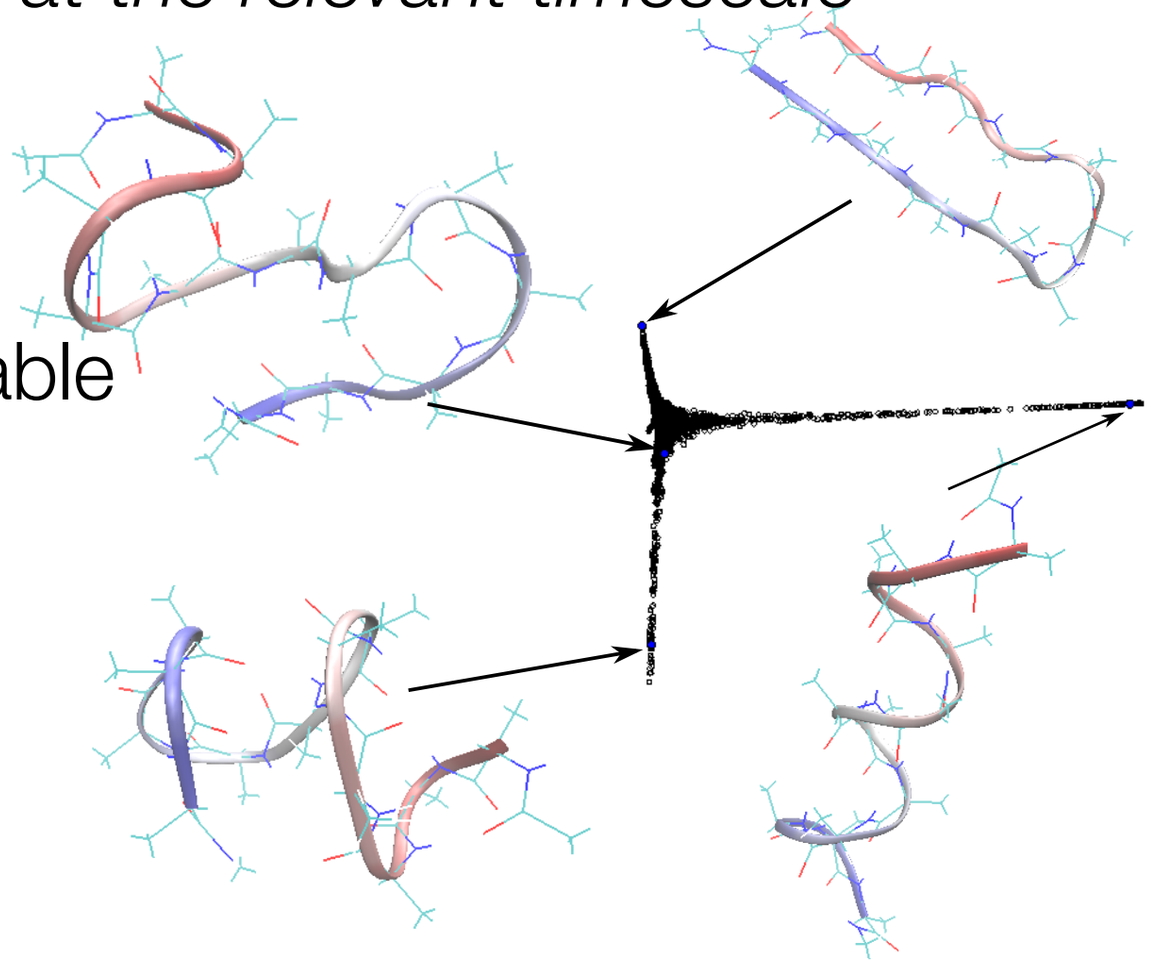
Interested in high-dimensional stochastic systems s.t.:

- the ***timescale*** of interest is medium-large

- are ***intrinsically low-dimensional***, *at the relevant timescale*

- are very expensive to simulate

- need an approximation with reasonable but not too high accuracy



Objectives

Given:

(i) stochastic simulator for the system $\mathcal{S}$ concentrated around a manifold $\mathcal{M}$ of dimension d

(ii) smallest "spatial" scale of interest δ

(iii) distance function d

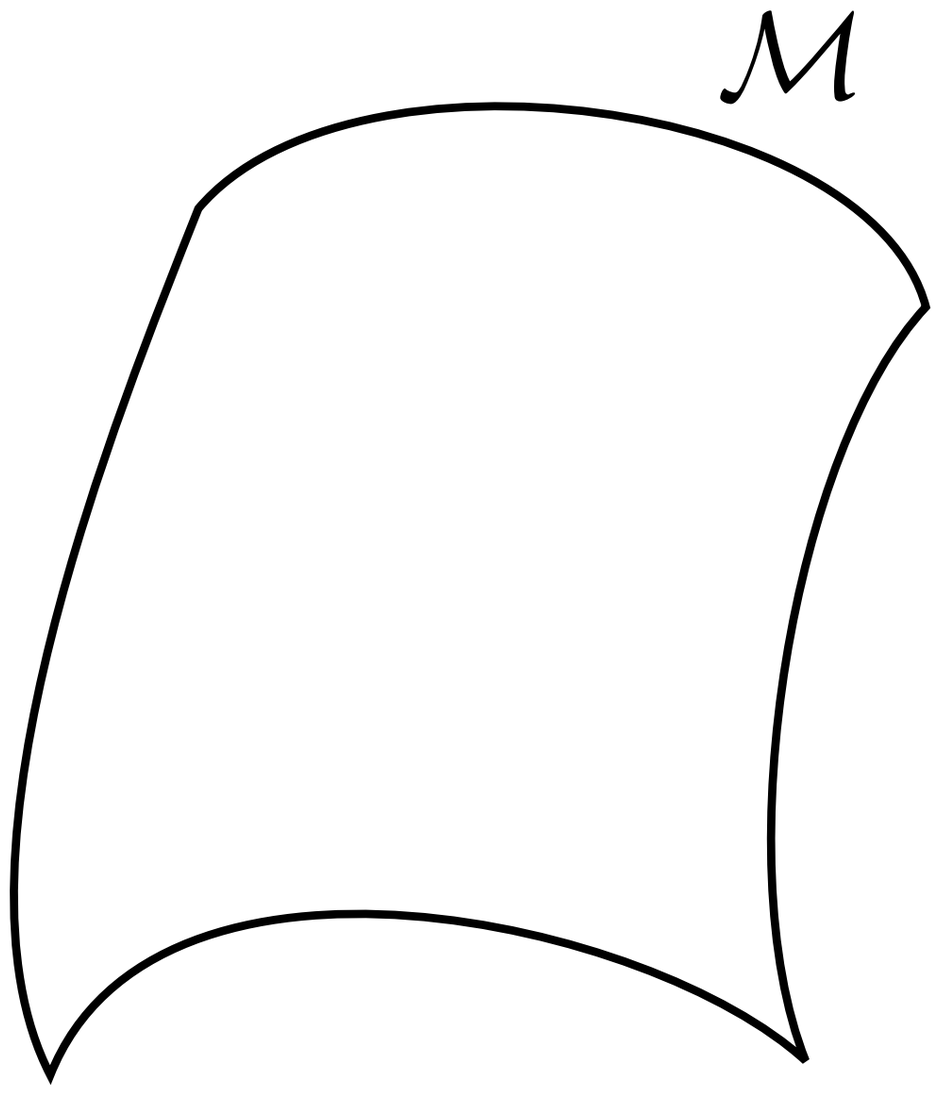
Return:

(i) a fast, continuous-time and continuous-space simulator (ATLAS) at all scales larger than δ, having accuracy $O(\delta)$

(ii) ATLAS is constructed using $O(d\delta^{-4})$ paths of length $O(\delta)$ from $\mathcal{S}$, collected in parallel

(iii) efficient storage of all paths

Main Ideas

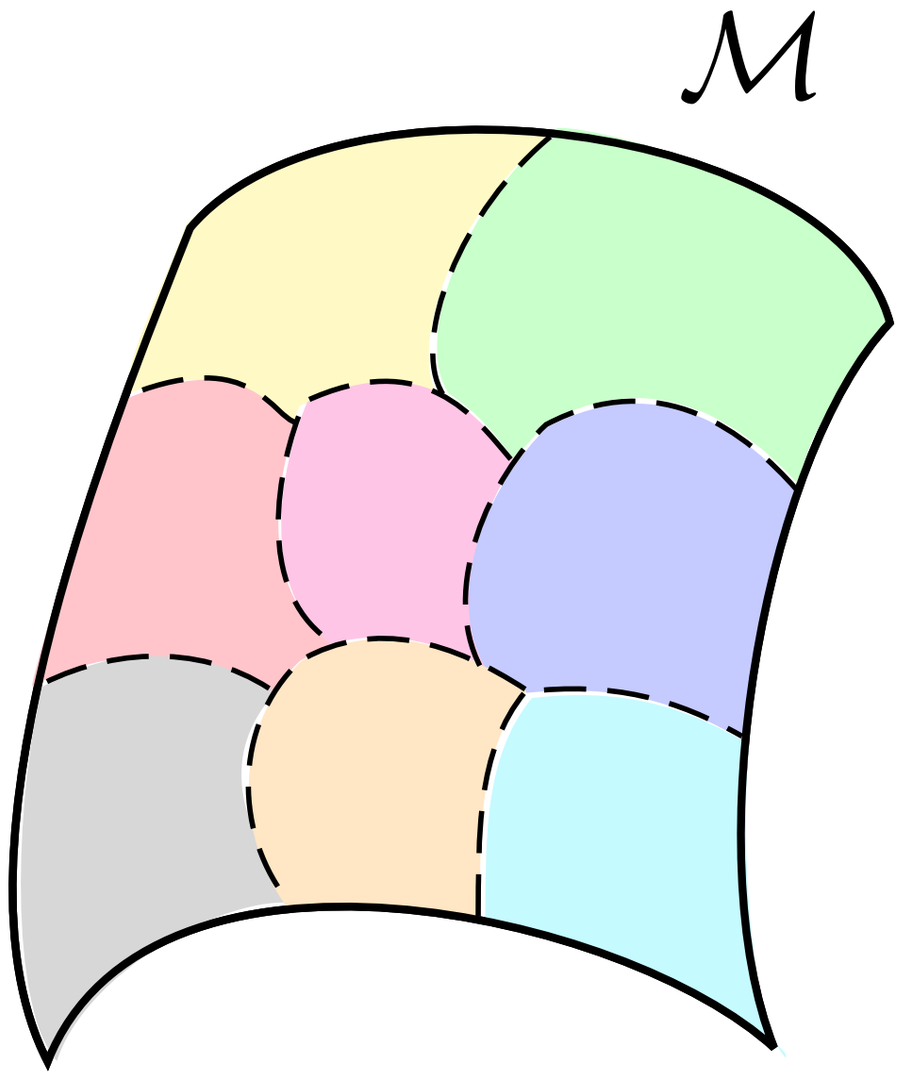


Main Ideas

. Divide configuration space using a δ-net

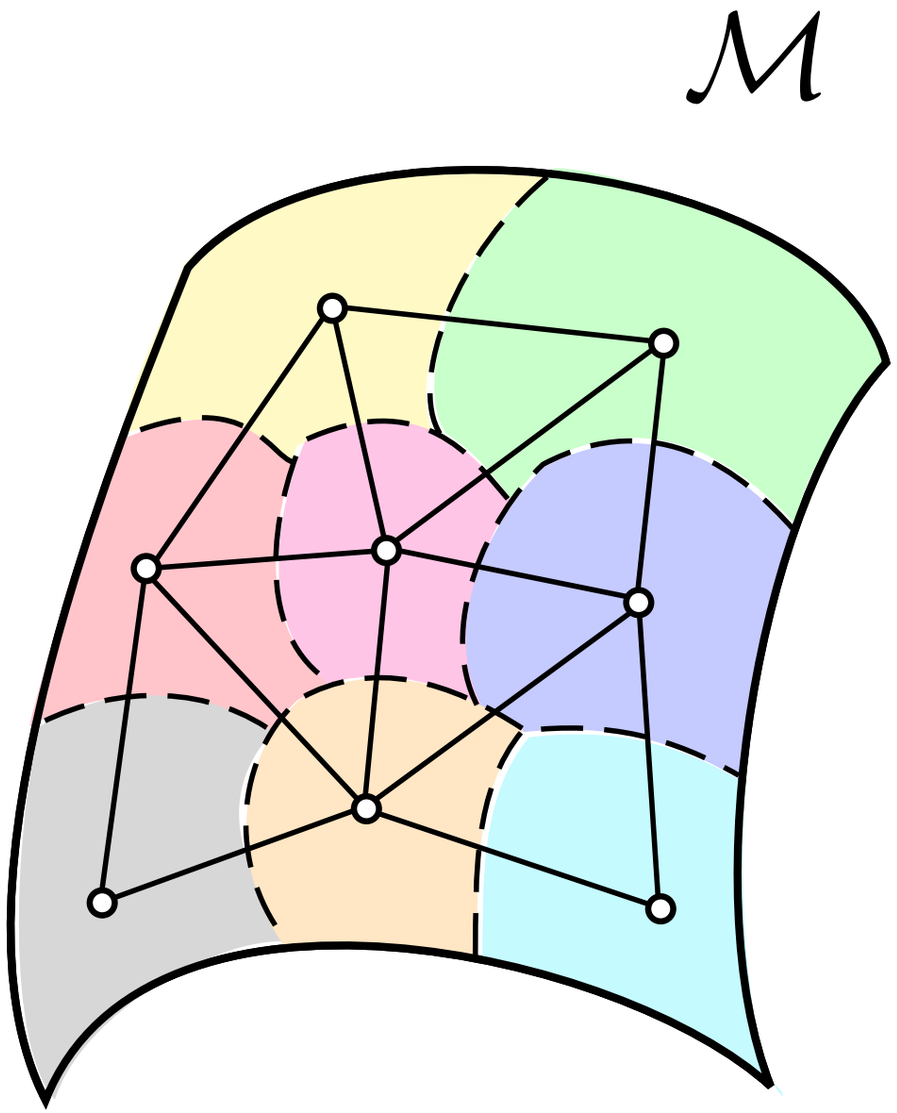
Γ a δ-net if $x \neq y \in \Gamma \implies d(x,y) > \frac{\delta}{2}$ and for every $x \in \mathcal{M}$ there is $y \in \Gamma$ with $d(x,y) < \delta$

Use cover trees to construct in online fashion in $O(C^d Dn \log n)$



Main Ideas

. Divide configuration space using a δ-net

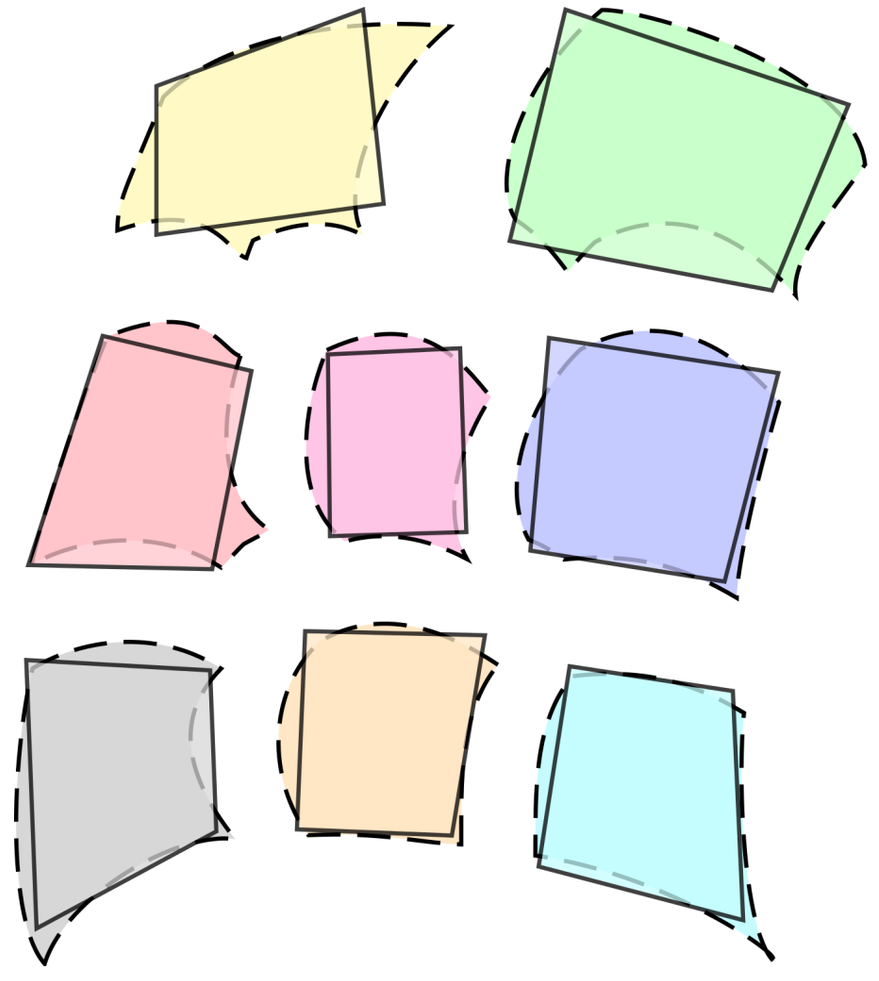


Connect each $y_k \in \Gamma$ to its neighbors.
This connectivity will be used to transition between local reduced simulators.
Also $O(C^d n \log n)$, and in parallel.

Main Ideas

. Divide configuration space using a δ-net

. Construct local Euclidean charts in each piece of partition



Use Multi-Dimensional Scaling
to obtain maps Φ_k from
neighborhood of $y_k \in \Gamma$
to $C_k \subset \mathbb{R}^d$
Completely local calculation,
Can use landmarks to speed up.

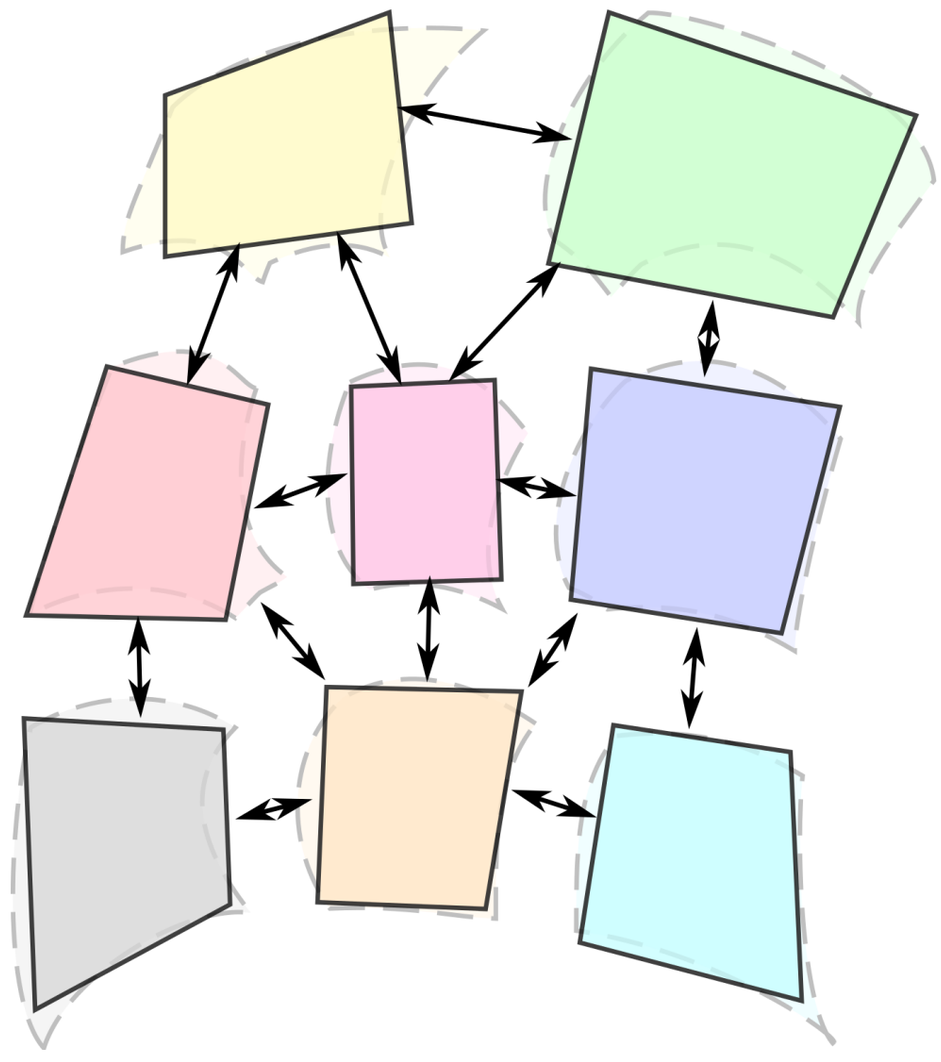
Main Ideas

. Divide configuration space using a δ-net

. Construct local Euclidean charts in each piece of partition

. Construct connections between charts

A transition map between C_k and $C_{k'}$ is learned whenever $k \sim k'$.
We use linear maps $S_{k,k'} : \mathbb{R}^d \to \mathbb{R}^d$.



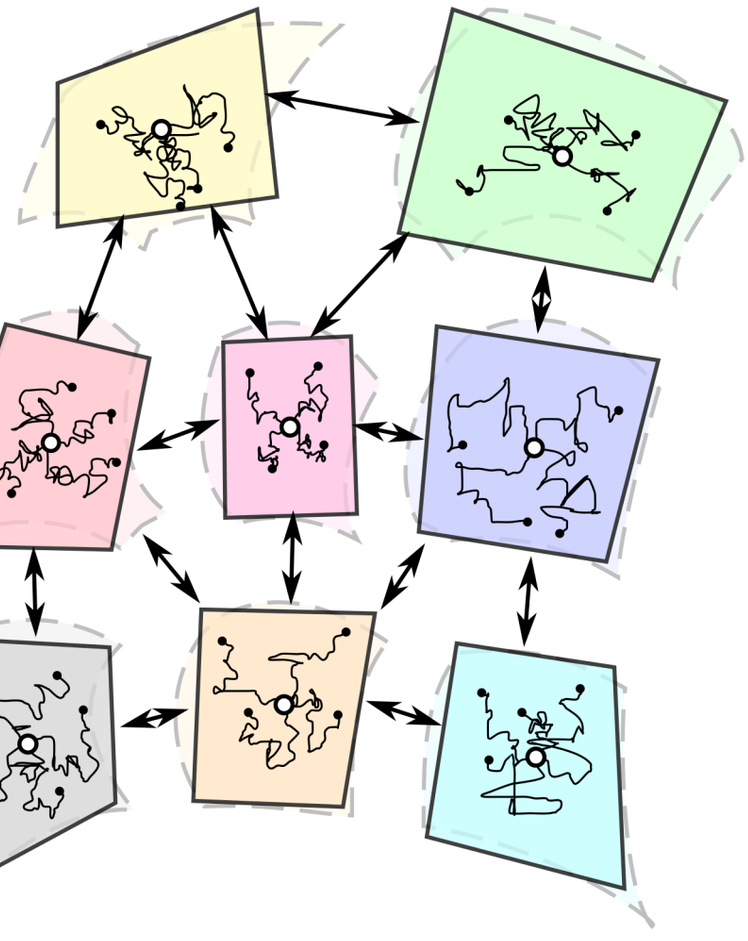
Main Ideas

. Divide configuration space using a δ-net

. Construct local Euclidean charts in each piece of partition

. Construct connections between charts

. Learn simulators on charts



In each chart we fit a constant coefficient Itô diffusion:
$d\overline{X}_t = \bar{b}dt + \overline{\sigma}dB_t$
We estimate $\bar{b}$ and $\overline{\sigma}$ by running
p paths of length $O(\delta)$.
Turns out we need $p = O(d\delta^{-4})$ in order to obtain accuracy δ.

Main Ideas

. Divide configuration space using a δ-net

. Construct local Euclidean charts in each piece of partition

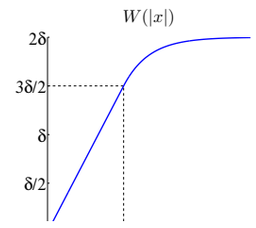
. Construct connections between charts

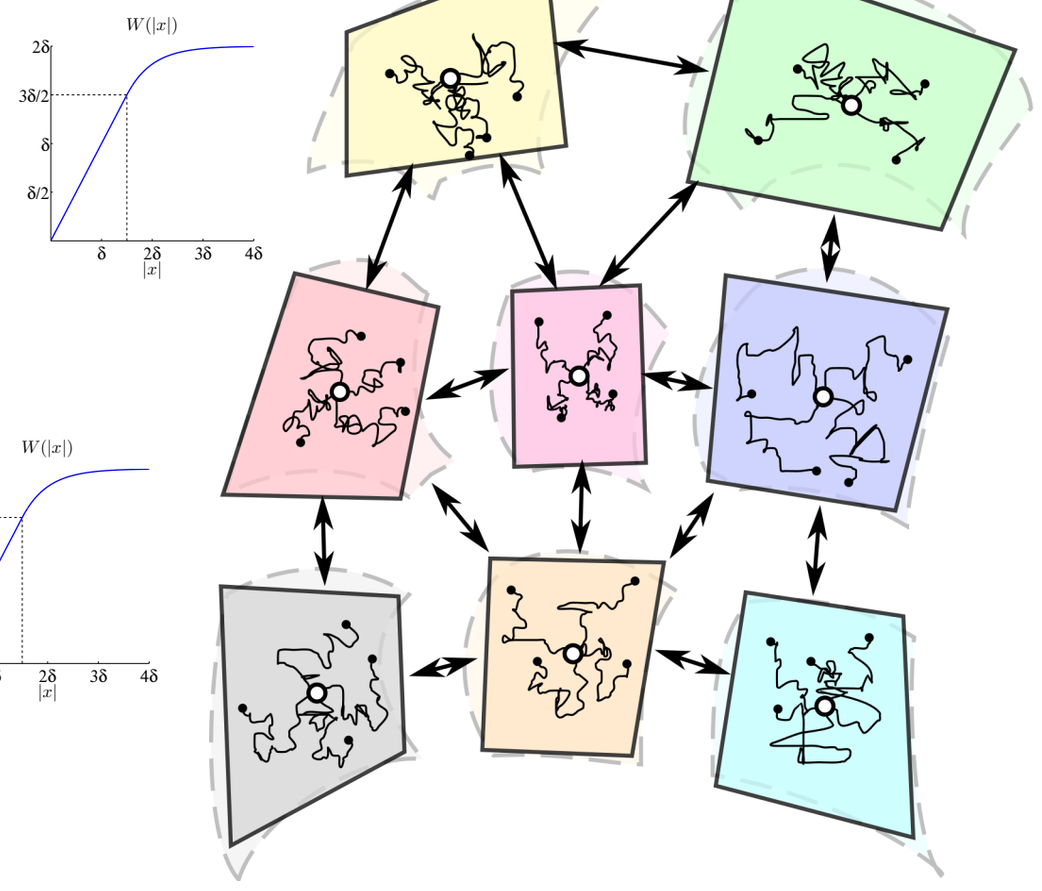
. Learn simulators on charts

. Glue simulators

ATLAS time step:

$$x \leftarrow x + \overline{b_k}\Delta t + \overline{\sigma}_k \Delta B$$
$$x \leftarrow W(x)$$
$$k' = \text{argmin}_{l \sim k} ||x - \Phi_k(y_l)||$$
$$x \to T_{k,k'}(x)$$





Large time guarantees

Theorem [Crosskey-MM] Let $\mathcal{M}$ be a nice manifold and suppose $X_t \in \mathcal{M}$ solves

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t$$

with b, σ Lipschitz, and σ uniformly elliptic on $\mathcal{M}$. Let q be stationary for X_t and $\widehat{q}$ be stationary for $\widehat{X}_t$. Then if δ is small enough and the number of sample paths $p > (\tau^2 + d)/\delta^4$,

$$||q - \Phi^{-1}(\widehat{q})||_{\mathbb{L}^1(\mathcal{M})} \leq C\delta \log(1/\delta)$$

with probability at least $1 - 2e^{-\tau^2}$.

X_t is the process that we are interested in approximating at time scale $> t_0$ - it may be itself an approximation of an "original system" given at a finer time-scale.

Here Φ is the map from $\mathcal{M}$ to the collection of approximate tangent spaces. It is invertible for δ small enough.

Large time guarantees

Theorem [Crosskey-MM] Let $\mathcal{M}$ be a nice manifold and suppose $X_t \in \mathcal{M}$ solves

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t$$

with b, σ Lipschitz, and σ uniformly elliptic on $\mathcal{M}$. Let q be stationary for X_t and $\widehat{q}$ be stationary for $\widehat{X}_t$. Then if δ is small enough and the number of sample paths $p > (\tau^2 + d)/\delta^4$,

$$||q - \Phi^{-1}(\widehat{q})||_{\mathbb{L}^1(\mathcal{M})} \leq C\delta \log(1/\delta)$$

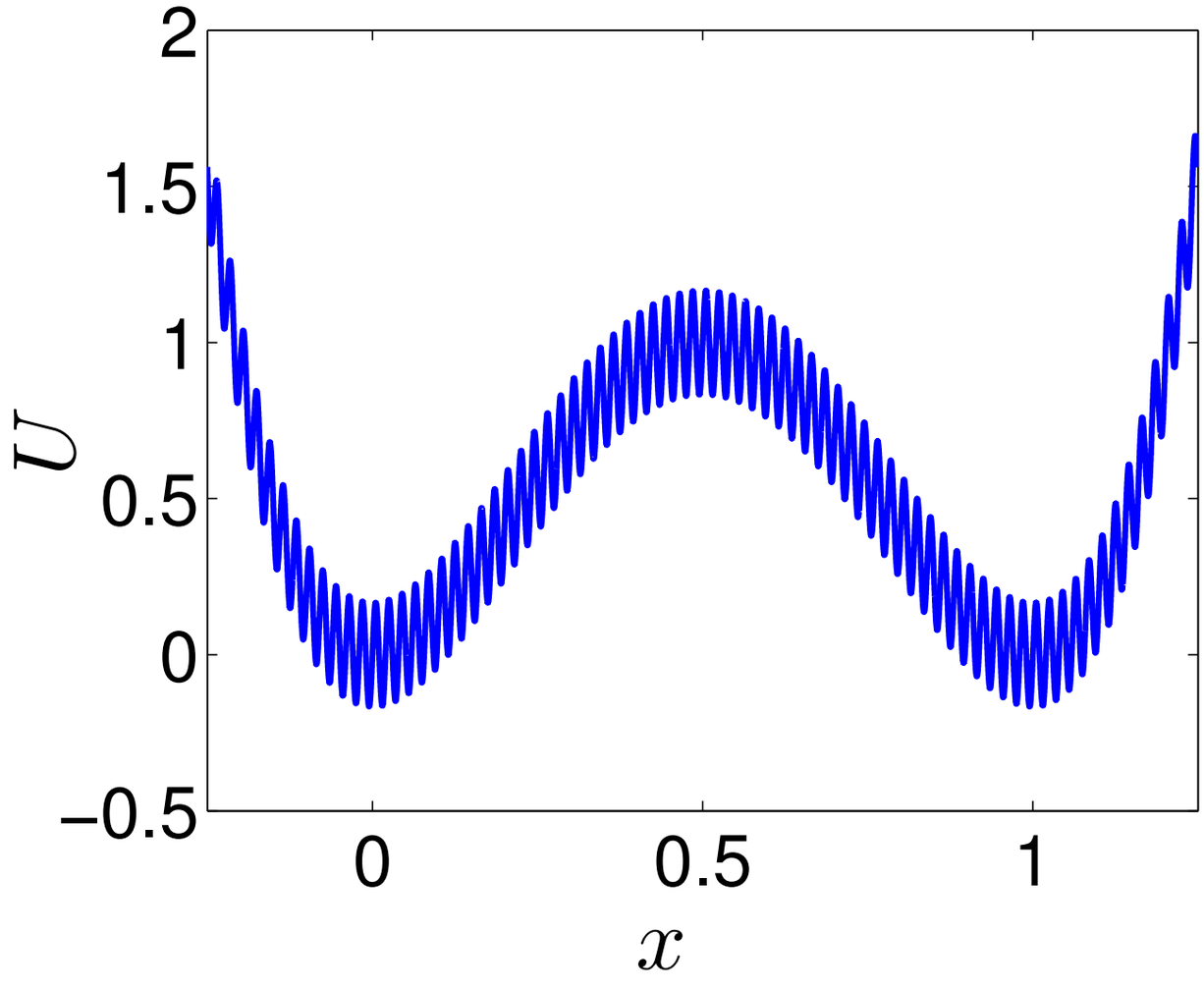
with probability at least $1 - 2e^{-\tau^2}$.

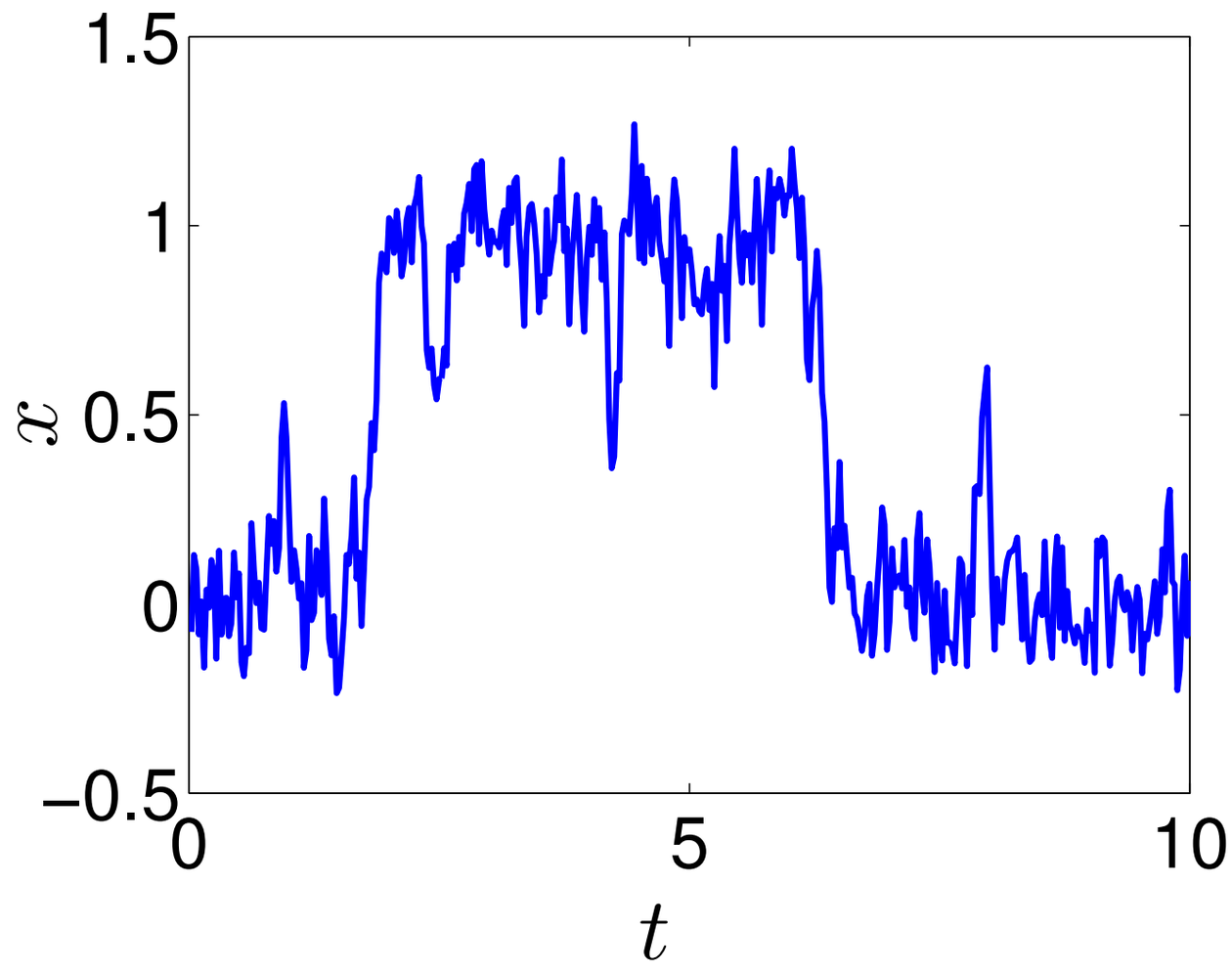
This result fits within the ideas that "short time accuracy implies long time accuracy" when averaging occurs and there is an underlying large-scale smoothness [J. Mattingly, A. Stuart, M. Tretyakov, E. Vanden-Eijnden, ...]

Examples: 1-D

Brownian motion in a potential well

$$dX_t = -\nabla U(X_t) + dB_t$$



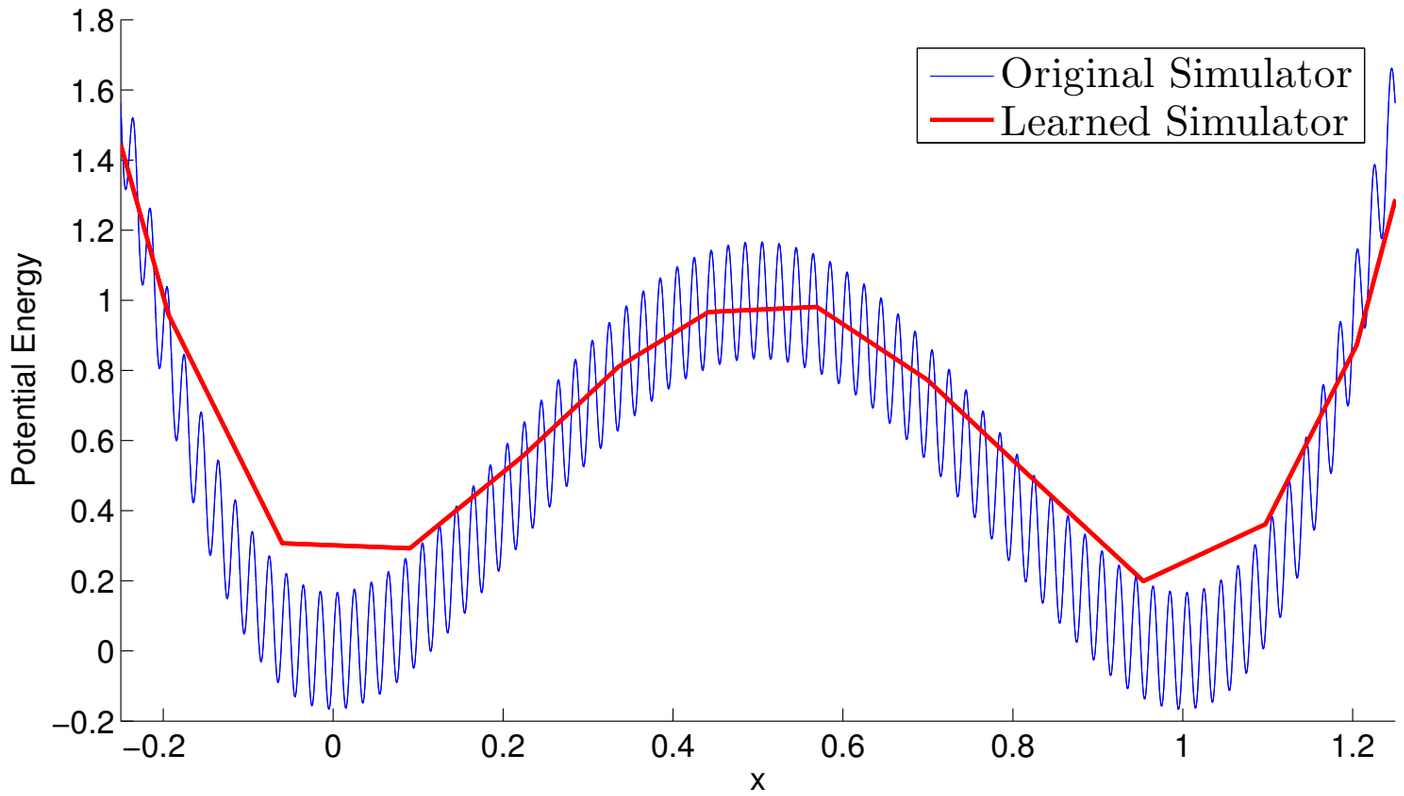


Examples: 1-D

Brownian motion in a potential well

$$dX_t = -\nabla U(X_t) + dB_t$$

- $\delta = 0.1$
- $t_0 = 10^{-2} = 500$ steps
- 14 charts
- $p = 10^4$ samples per chart



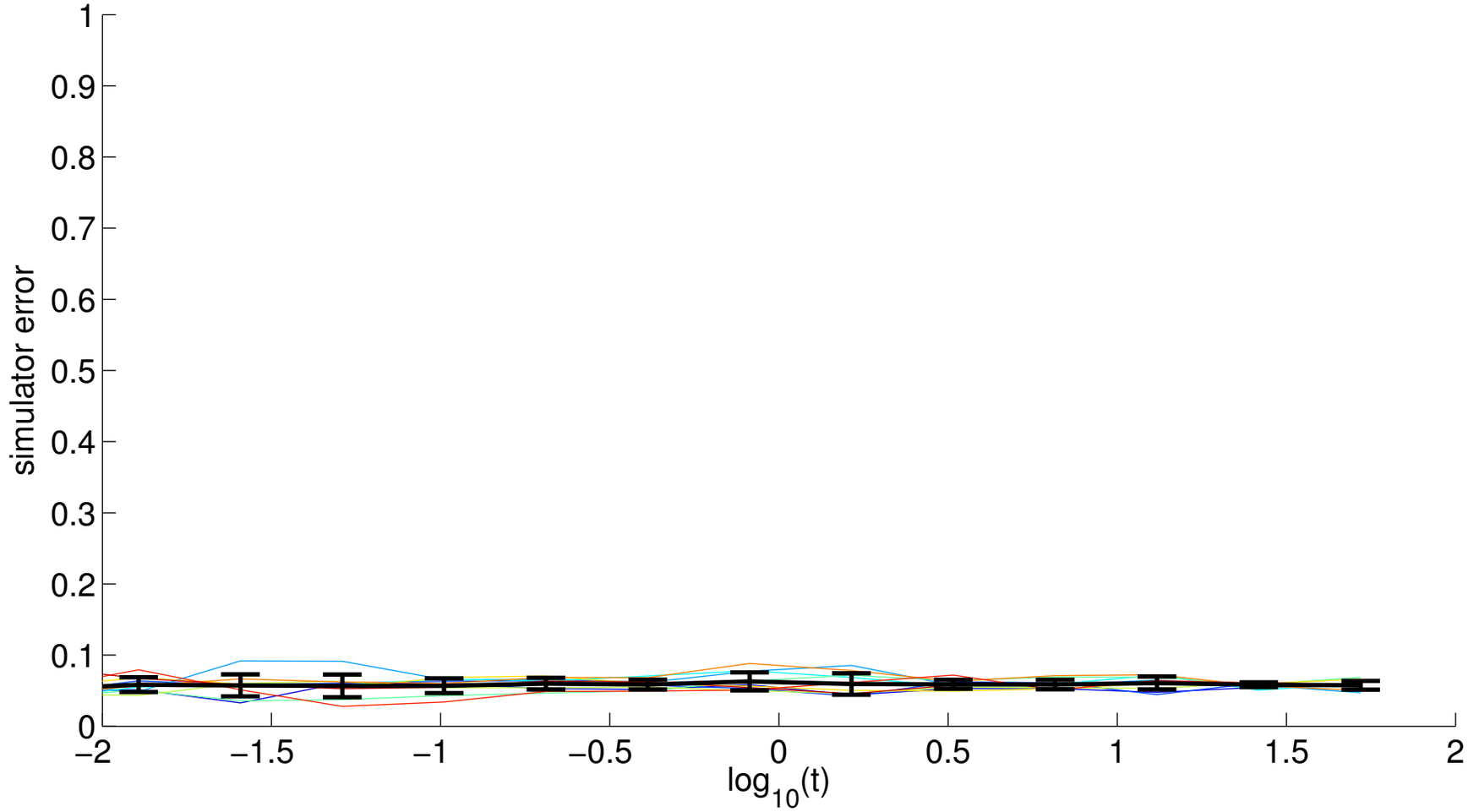
Recall: we should not (and do) obtain an approximation to the original potential U, but to a smooth effective potential which is accurate at spatial scales larger than δ.

Examples: 1-D

Measure of error:

$$||p_t(x,\cdot) - \hat{p}_t(x,\cdot)||_{L^1(\mathcal{M})} ,$$

where $p_t(x,\cdot)$ is the probability of being at $\cdot$ starting from x according to $\mathcal{S}$ and similarly for $\hat{p}$.
Look at the above, binned according to a partition associated with the net, averaged of x, as a function of t, aver all timescales.

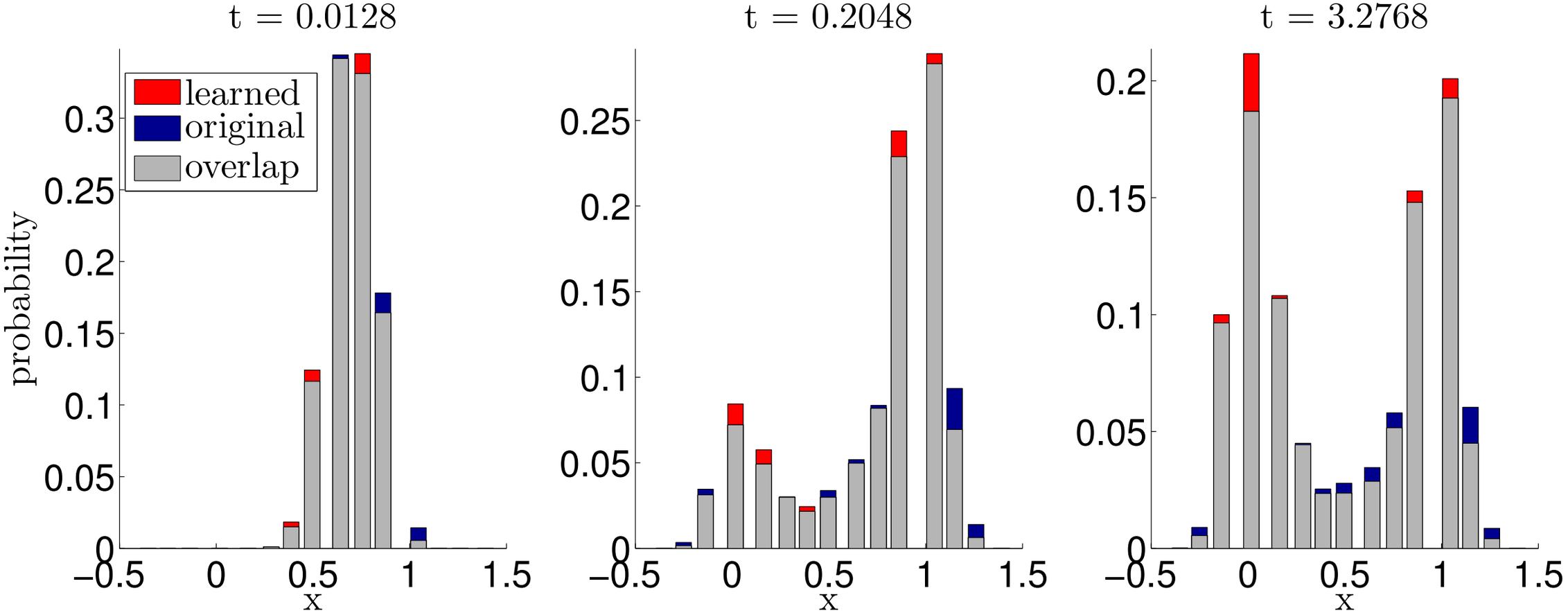


Examples: 1-D

Measure of error:

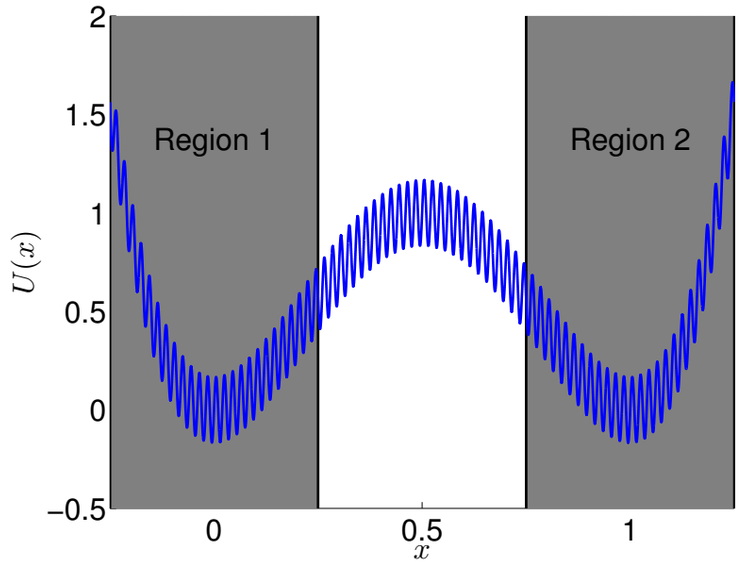
$$||p_t(x,\cdot) - \hat{p}_t(x,\cdot)||_{L^1(\mathcal{M})}\,,$$

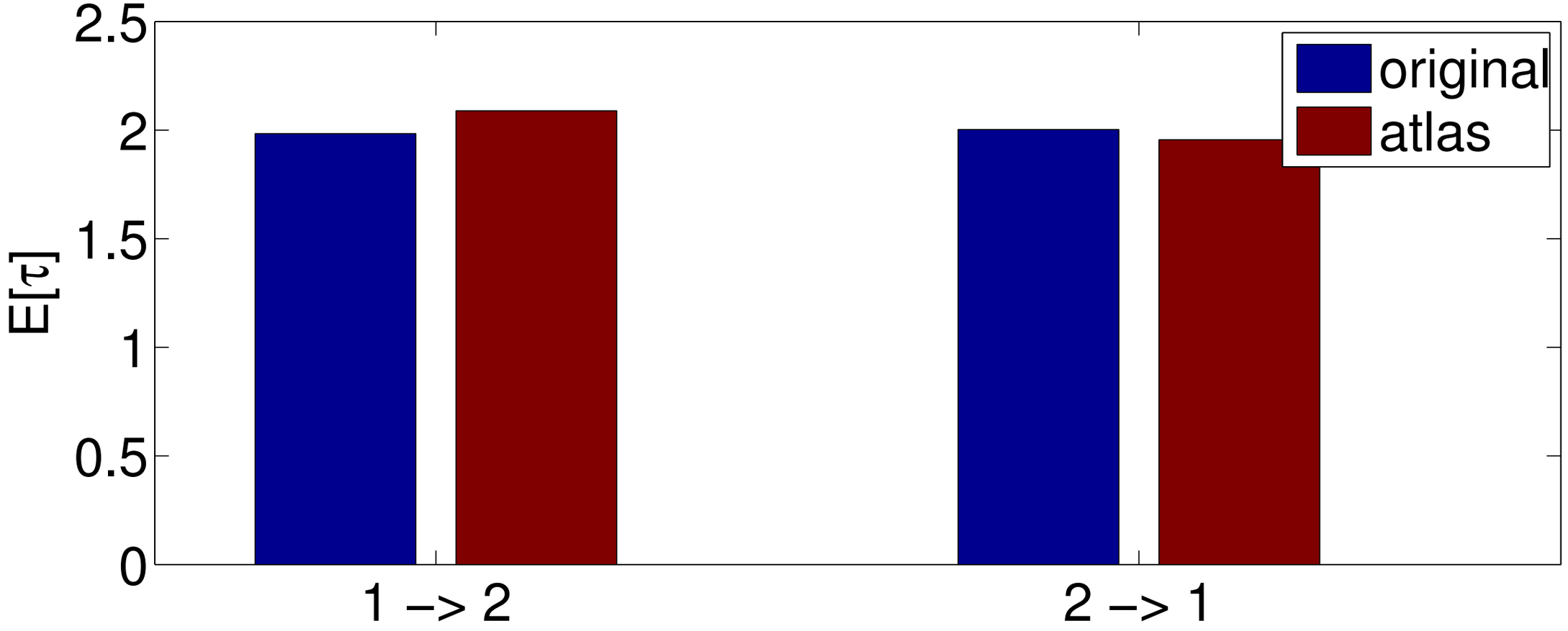
where $p_t(x,\cdot)$ is the probability of being at $\cdot$ starting from x according to $\mathcal{S}$ and similarly for $\hat{p}$.
We overlay $p_t(x,\cdot)$ and $\hat{p}_t(x,\cdot)$ for some x, binned; each plot correspond to a different choice of t.



Examples: 1-D

Transition rates:
$\tau_{1\to 2}$ is first hitting time
of 2 after last exit from 1.
Similarly for $\tau_{2\to 1}$.

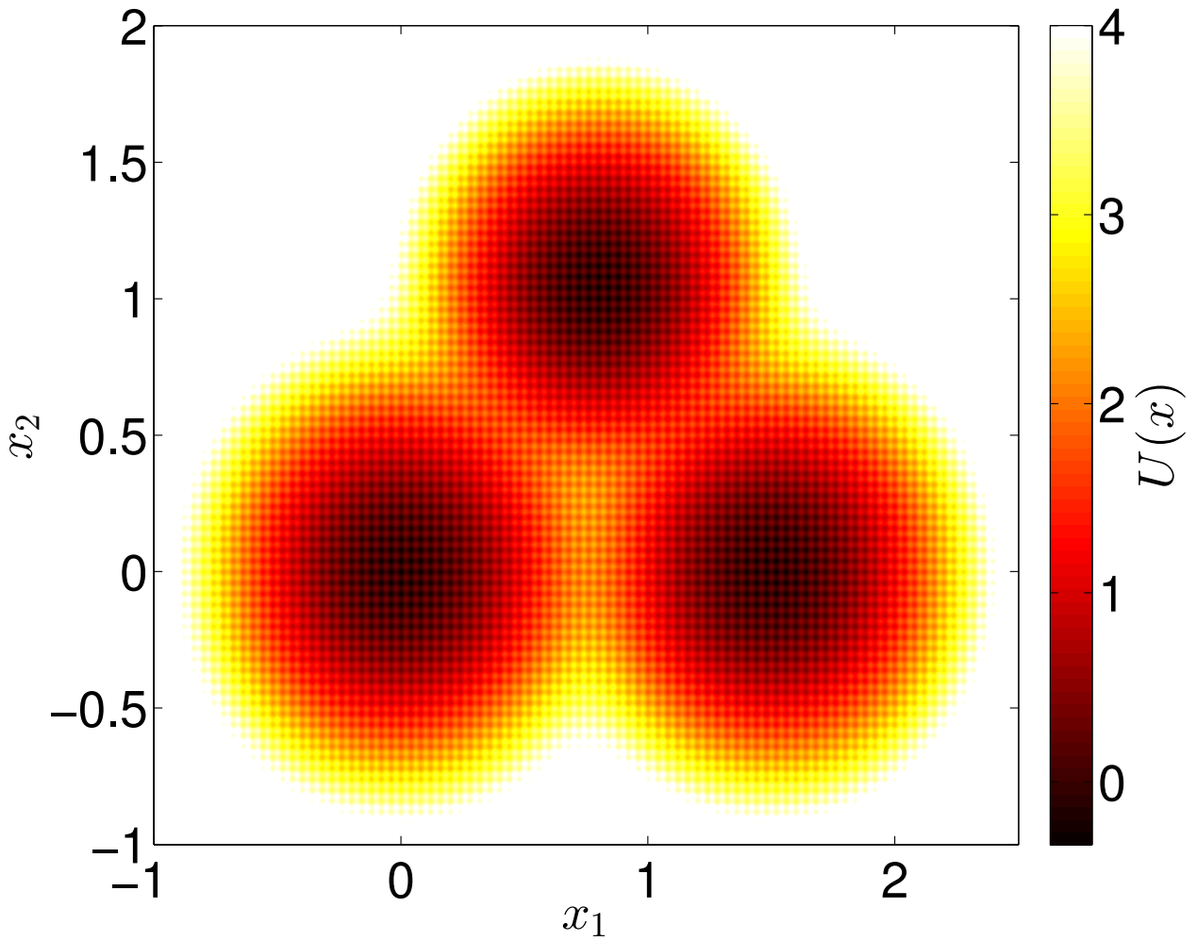


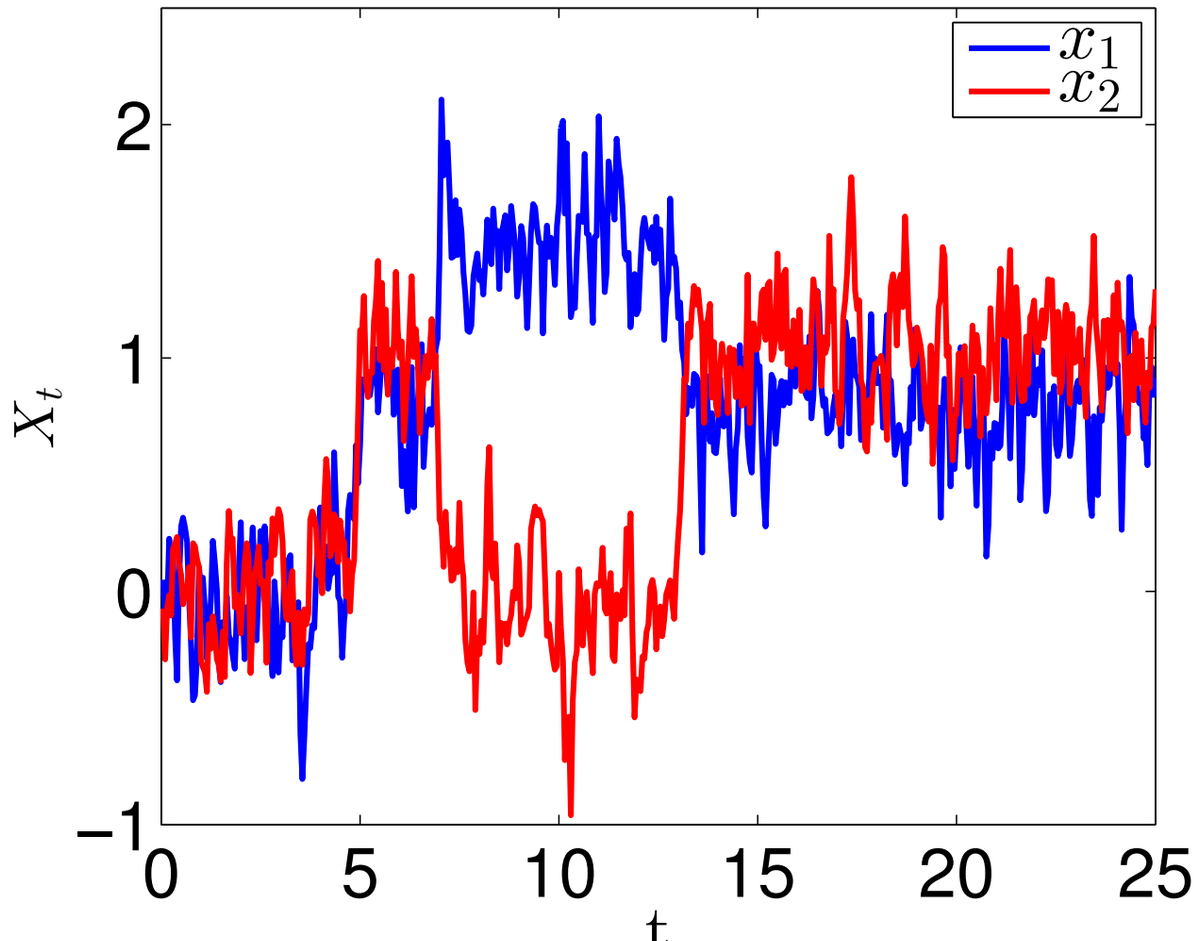


Examples: 2-D

Brownian motion in a potential well

$$dX_t = -\nabla U(X_t) + dB_t$$



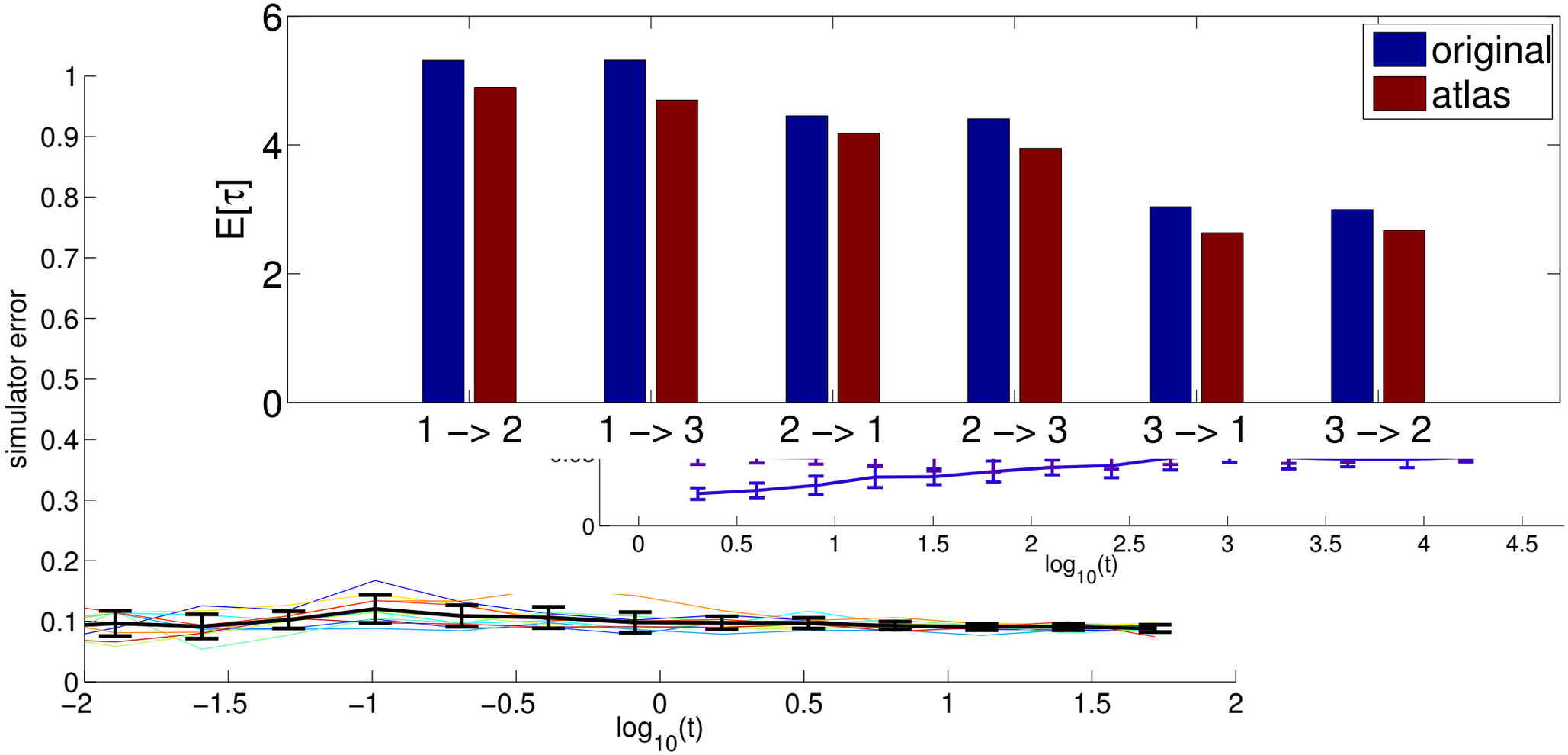


Examples: 2-D

Measure of error:

$$||p_t(x,\cdot) - \hat{p}_t(x,\cdot)||_{L^1(\mathcal{M})} ,$$

where $p_t(x,\cdot)$ is the probability of being at $\cdot$ starting from x according to $\mathcal{S}$ and similarly for $\hat{p}$.
Look at the above, binned according to a partition associated with the net, averaged of x, as a function of t, aver all timescales.



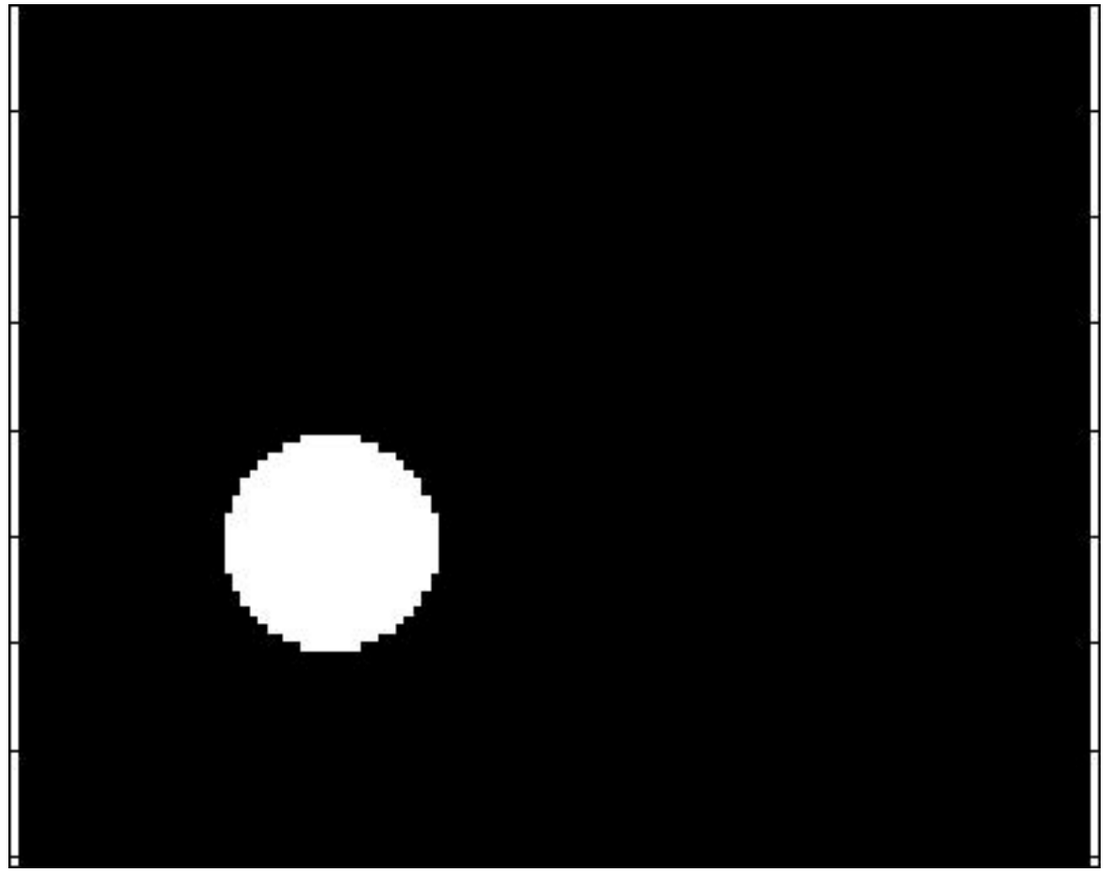
Examples: 12,500-D

Brownian motion in a potential well

$$dX_t = -\nabla U(X_t) + dB_t$$

obtained by mapping the 2-d rough potential to a (2-d) manifold in $\mathbb{R}^{12,500}$, endowed with L^1 distance, where each point is an image of a circle with center at the location corresponding to the 2-d example.

- $\delta = 0.2$
- $t_0 = 4 \cdot 10^{-2} = 800$ steps
- 230 charts
- $p = 2 \cdot 10^3$ samples per chart

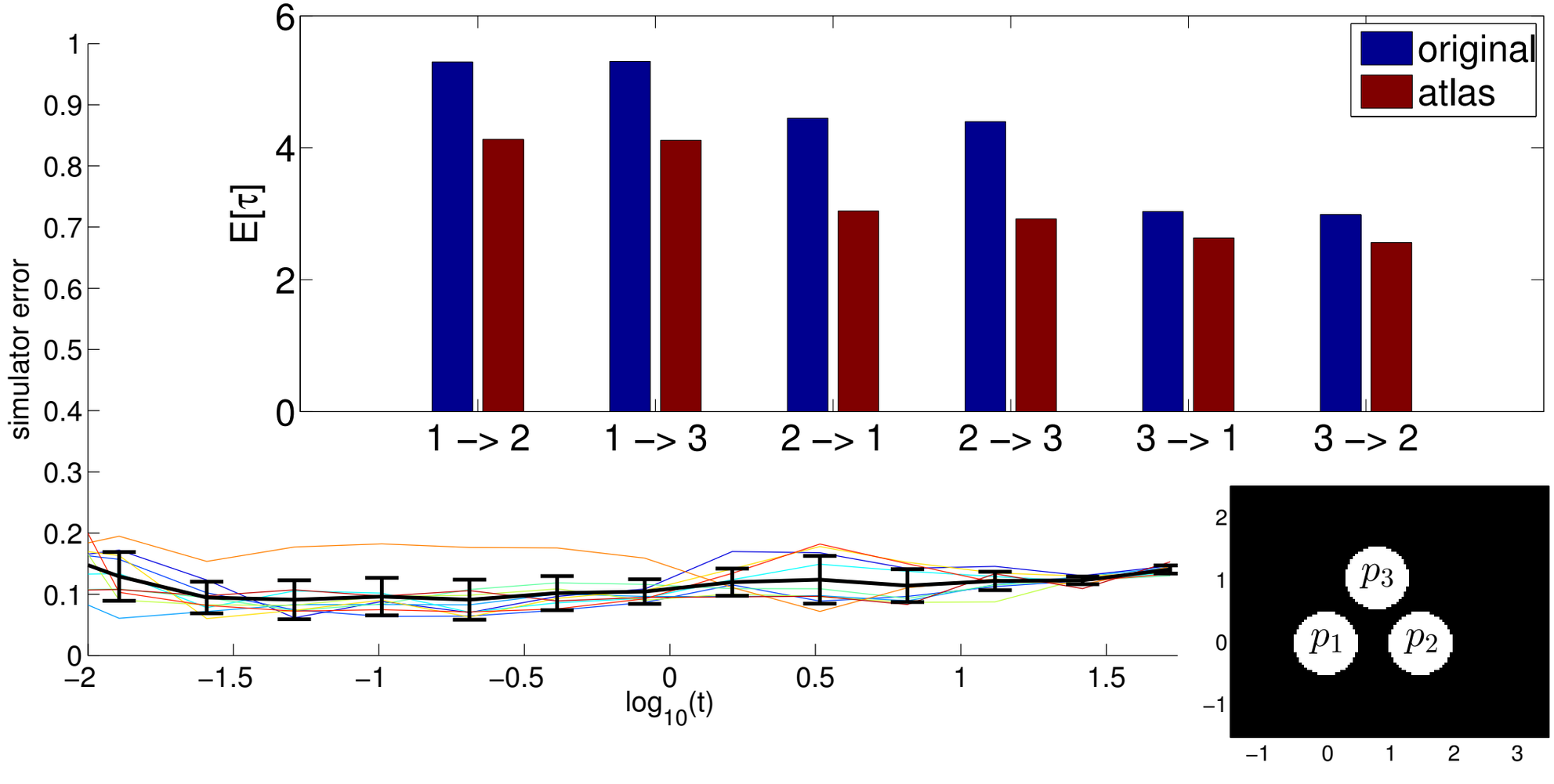


Examples: 12,500-D

Measure of error:

$$||p_t(x,\cdot) - \hat{p}_t(x,\cdot)||_{L^1(\mathcal{M})} ,$$

where $p_t(x,\cdot)$ is the probability of being at $\cdot$ starting from x according to $\mathcal{S}$ and similarly for $\hat{p}$.
Look at the above, binned according to a partition associated with the net, averaged of x, as a function of t, aver all timescales.



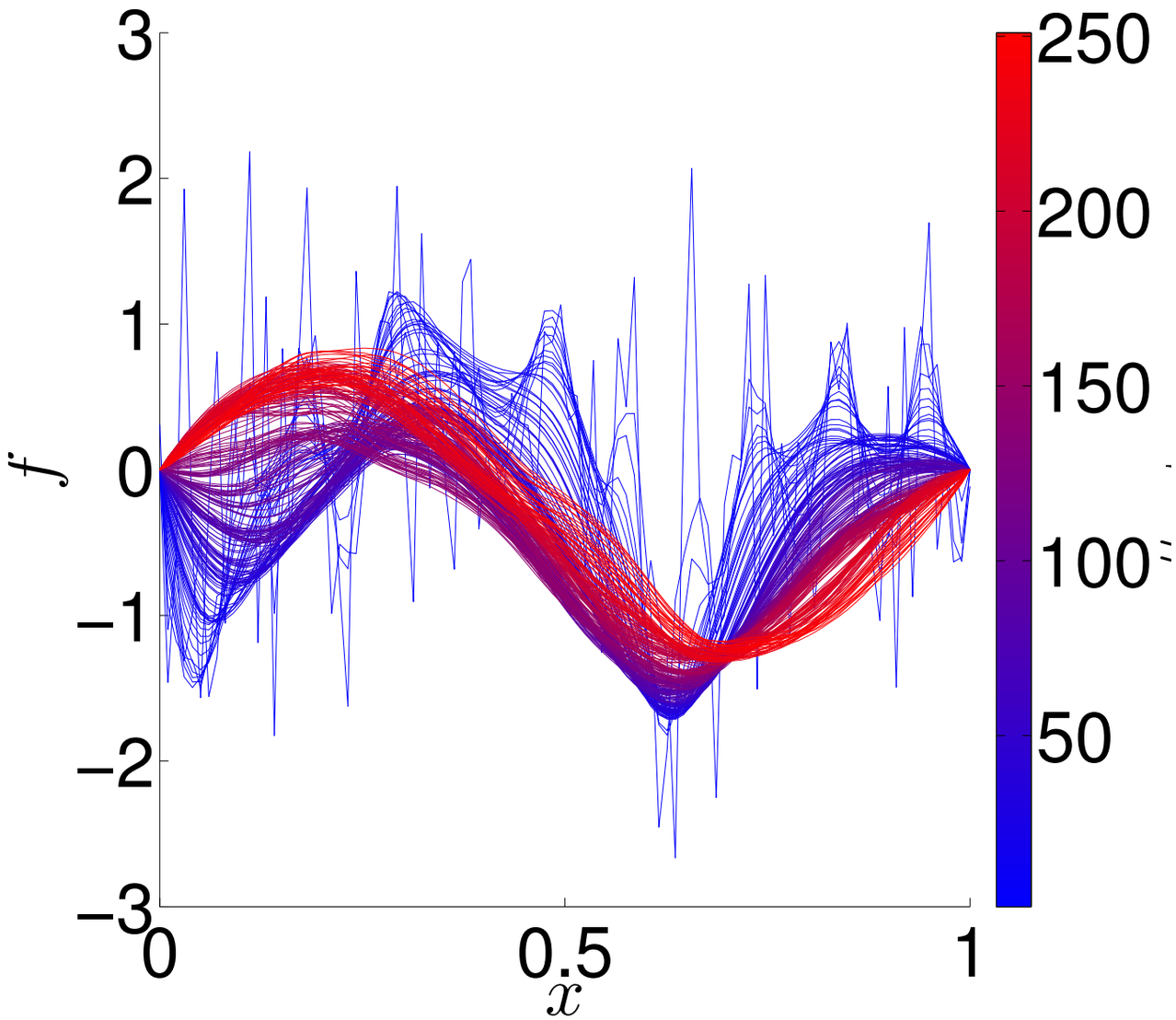
Examples: 100-D

$f = \mathcal{S}(f)$

$\%$ *simulate Brownian bridge*
$W = \text{cumsum}(\text{randn}(1{,}100))$
$W = W - W(1)$
$W = W - x * W(100)$

$\%$ *Add bridge to f, smooth and renormalize*
$f = f + (1/100) * W$
$f = \text{smooth}(f)$
$f = f * (f_{\text{norm}}/\text{norm}(f))$

- $\delta = 0.3$
- $t_0 = 2.5 \cdot 10^{-2} = 250$ steps
- ~ 70 charts
- $p = 2 \cdot 10^3$ samples per chart

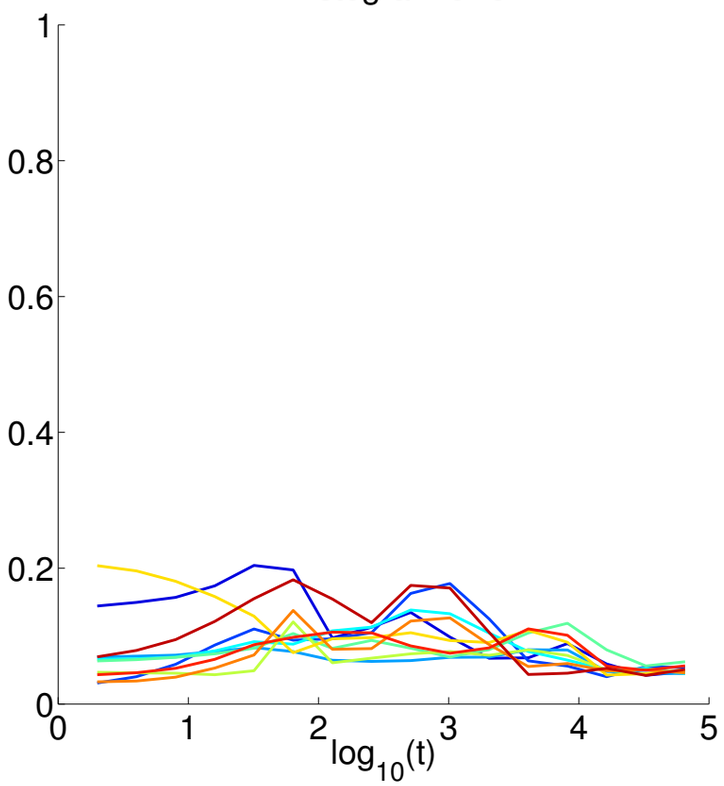


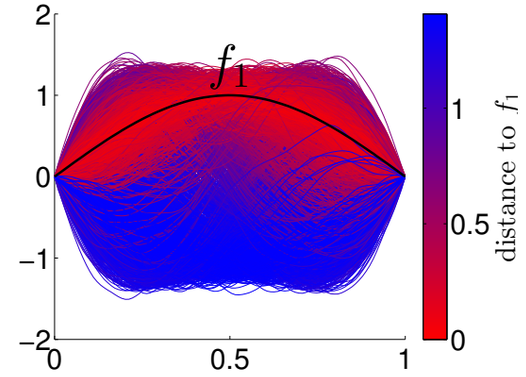
Examples: 12,500-D

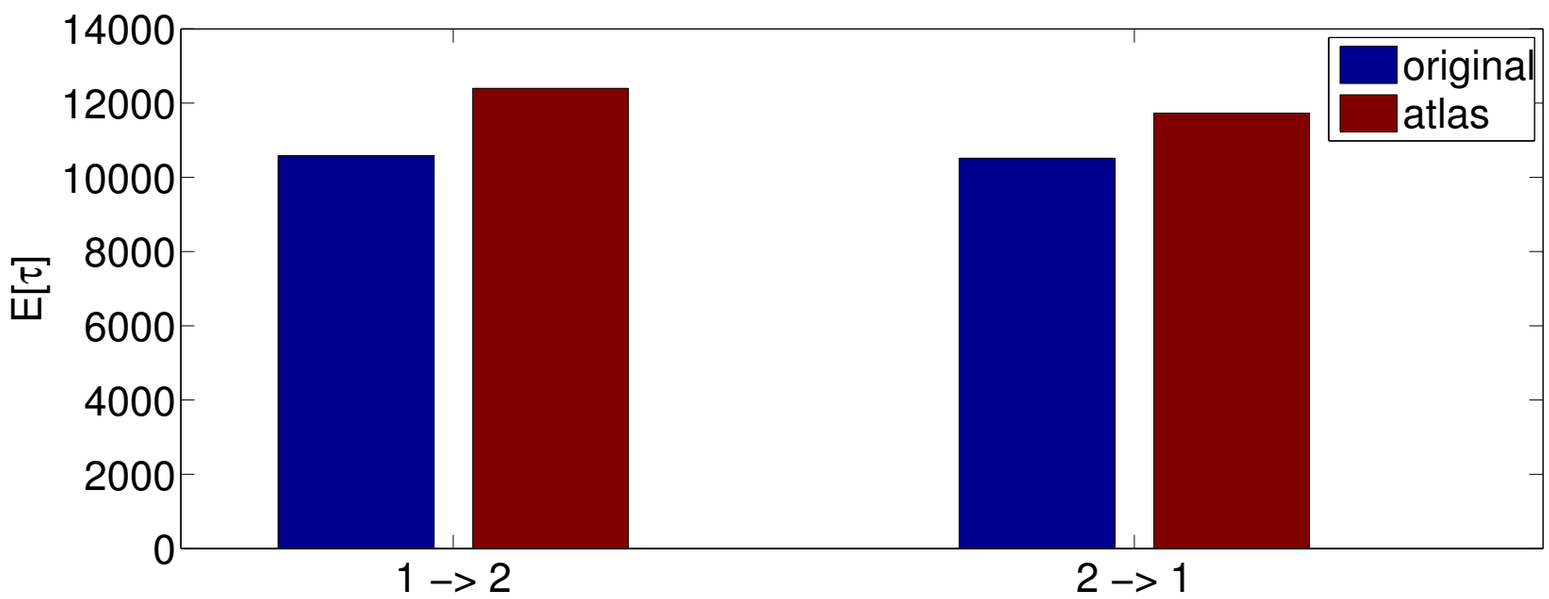
Measure of error:

$$||p_t(x,\cdot) - \hat{p}_t(x,\cdot)||_{L^1(\mathcal{M})}\,,$$

where $p_t(x,\cdot)$ is the probability of being at $\cdot$ starting from x according to $\mathcal{S}$ and similarly for $\hat{p}$.
Look at the above, binned according to a partition associated with the net, averaged of x, as a function of t, aver all timescales.







Examples: 82-D, chaotic

Noise may arise from ensembles of deterministic chaotic processes. Multiscale ODE with a scale ϵ:

$$\begin{cases} \dot{X}_t^\epsilon = \epsilon f(X_t^\epsilon) + g(Y_t), & X_0^\epsilon = x \\ \dot{Y}_t = h(Y_t) & Y_0 = y \end{cases}$$

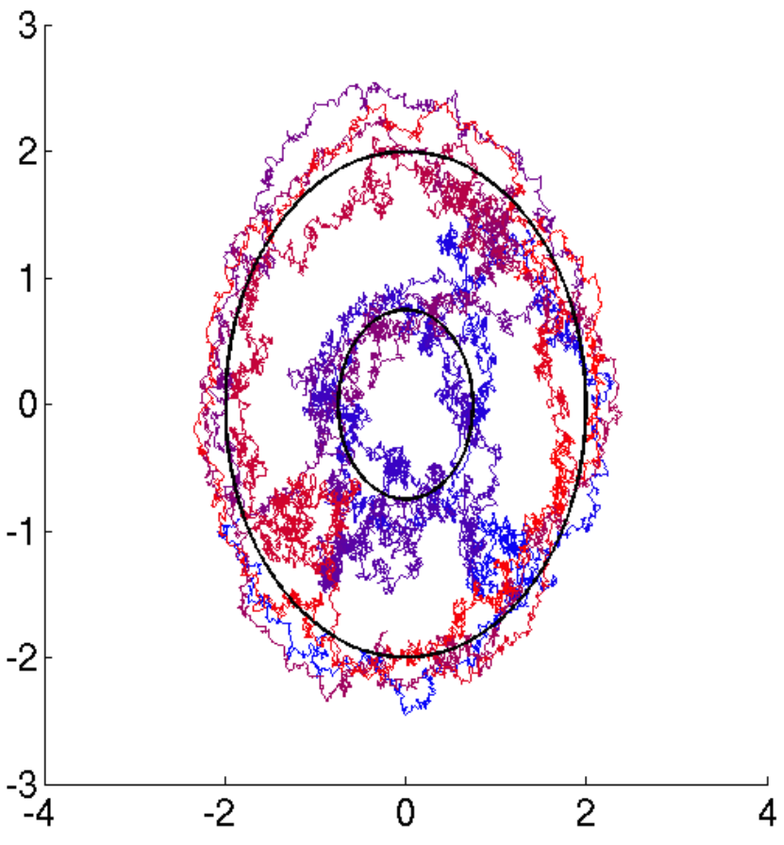
If the dynamics for Y_t alone admits an invariant measure μ, and $\mathbb{E}_\mu[f] = O(\epsilon)$, then the above behaves like the SDE $dX_s = b(X_s)ds + \sigma(X_s)dB_s$, on the timescale $s = \epsilon t$ in the limit $\epsilon \to 0$. For fixed ϵ, difficult to simulate directly due to the timescale separation.

We choose Y_t=Lorentz '96 with 80 dimensions, and $X_t \in \mathbb{R}^2$ so that there are two limit cycles consisting of two concentric circles.

. $\epsilon = 0.1$

. $\delta = 0.18 = 240$ steps

. $t_0 = \frac{1}{4}\epsilon^{-1}$



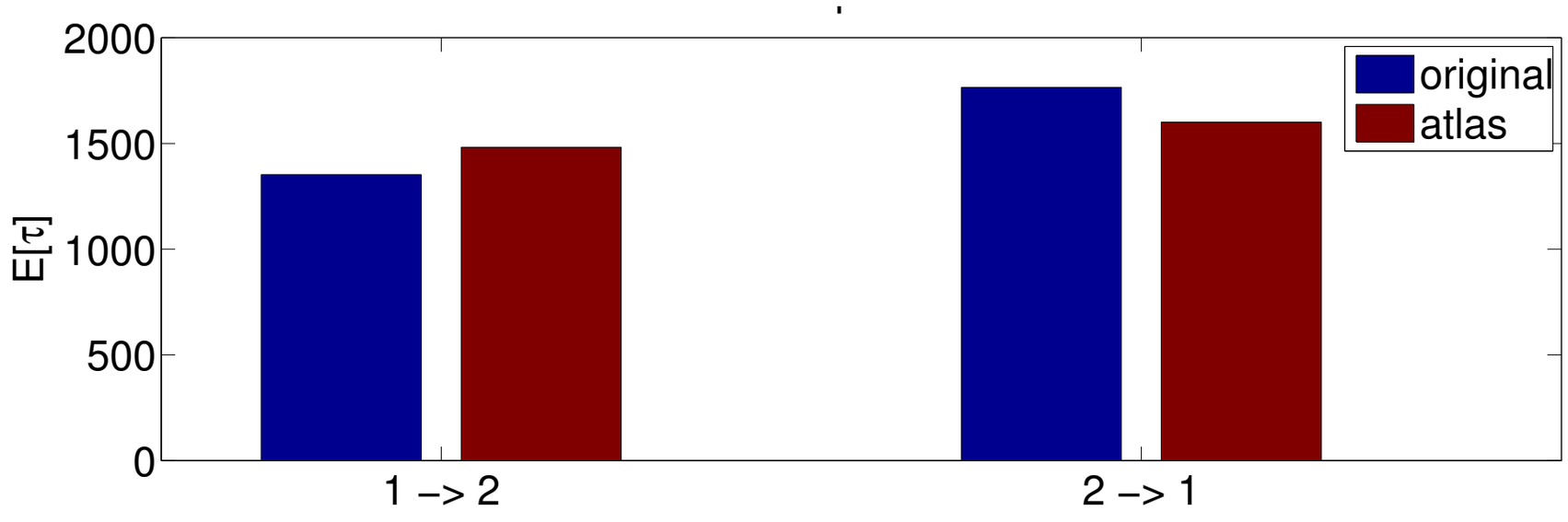
Examples: 82-D, chaotic

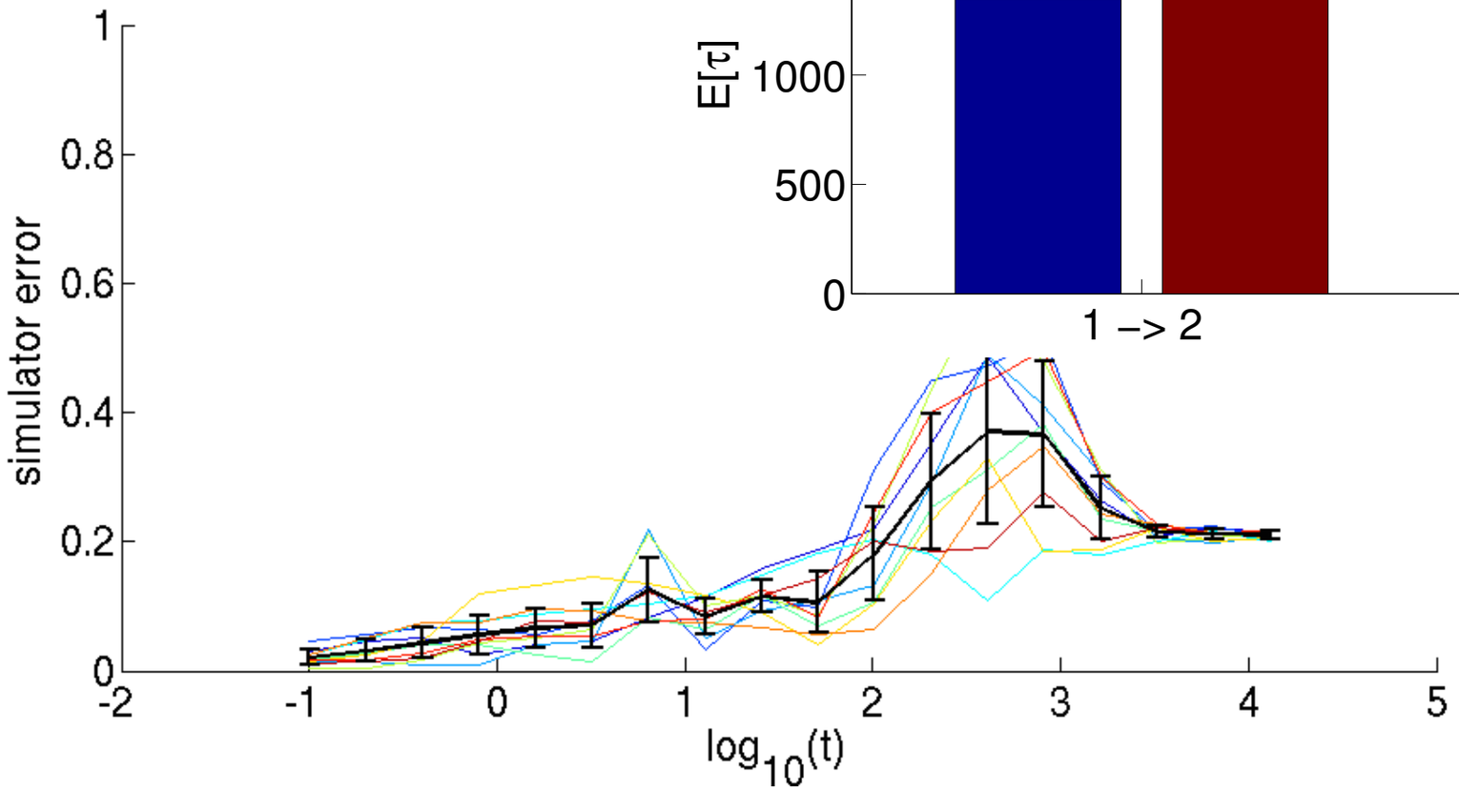
Measure of error:

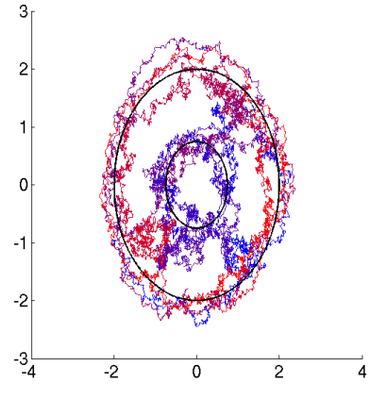
$$||p_t(x,\cdot) - \hat{p}_t(x,\cdot)||_{L^1(\mathcal{M})}\,,$$

where $p_t(x,\cdot)$ is the probability of being at $\cdot$ starting from x according to $\mathcal{S}$ and similarly for $\hat{p}$.

Look at the above, binned according to a partition associated with the net, averaged of x, as a function of t, aver all timescales.







Summary

"*Models*": high-dimensional data may often be approximated by low-dimensional geometric objects

Intrinsic dimension: use Multiscale SVD to measure intrinsic dimension

Coordinates: global low-dimensional parametrizations via diffusion maps. This can be adapted to Langevin dynamics for low-dimensional descriptions of large-time dynamics.

Learning dynamics: use parallel local learning of models for the dynamics, and careful piecing together local models can be done so that large time accuracy is achieved.

Current Work

Fully online mode with *exploration* [built into construction! Needs more code and theory]

Multiscale: choose scale and dimension adaptively [encouraging simulations with adaptive local scale, and changes in local dimension]

Molecular dynamics

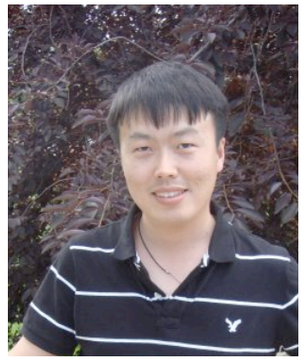
Generalize theory to *other large-time functionals* [transition rates and beyond]

Hypoellipticity (e.g. second order Langevin); *non-time-homogeneous systems*.

The Team

NOW HIRING!

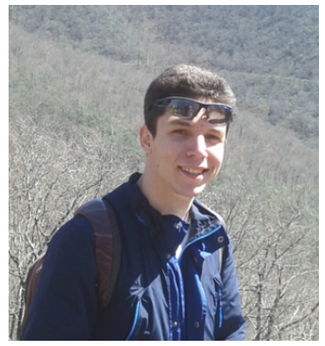
M. Crosskey



G. Chen*

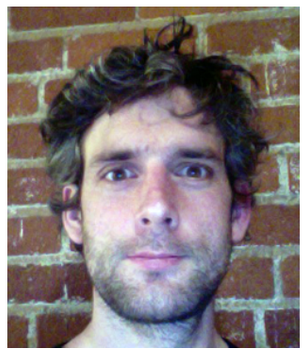
N. Strawn

W.K. Allard



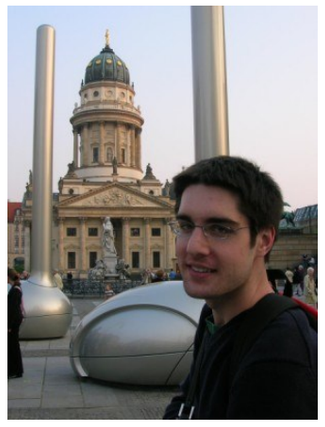
S. Minsker

E. Monson
Visualization



S. Gerber

M. Iwen*



D. Lawlor

P. E. Barbano

W. Liao

Collaborators: D. Brady, R.Brady (ECE, Duke), C. Clementi (Chem, Rice), R. Coifman, I. Daubechies (Math, Duke), P.W. Jones (Math, Yale), J. Harer, J. Mattingly, E. Monson (CS, Duke), S. Mukherjee (Stats, Duke), R. Rajae (ECE, U. Oregon), R. Schul (Math, Stony Brook), W. Willinger (AT&T).

Duke UNIVERSITY

THANK YOU!

www.math.duke.edu/~mauro

?