Markov state models

Theory, properties, estimation and validation

Simon Olsson

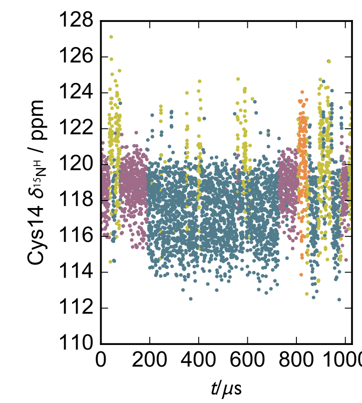
2020 PyEMMA Workshop
FU Berlin
Monday, Feb 17th

Motivation

$$
\begin{tikzcd}
\text{System of interest} \arrow[r, "\textit{experiment(s) / simulation(s)}"] \arrow[d, "\textit{questions}"'] & \text{Realization} \arrow[d, "\textit{estimation}"] \\
\text{Properties of interest} & \text{Model} \arrow[l, "\textit{predict}"]
\end{tikzcd}
$$

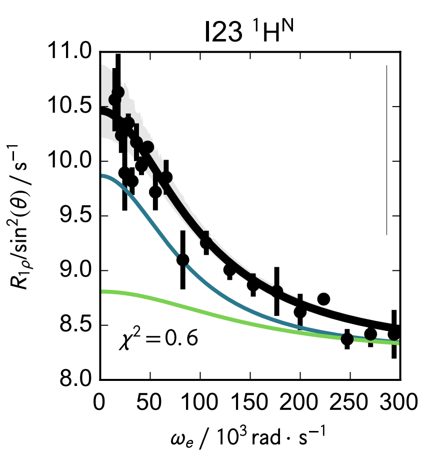
Motivation

experiment(s) / simulation(s)

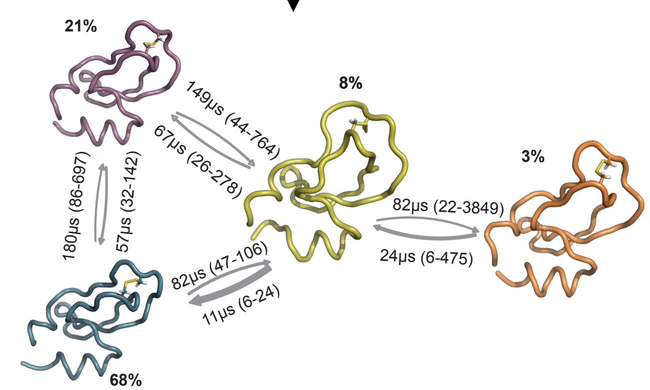


questions

estimation



predict

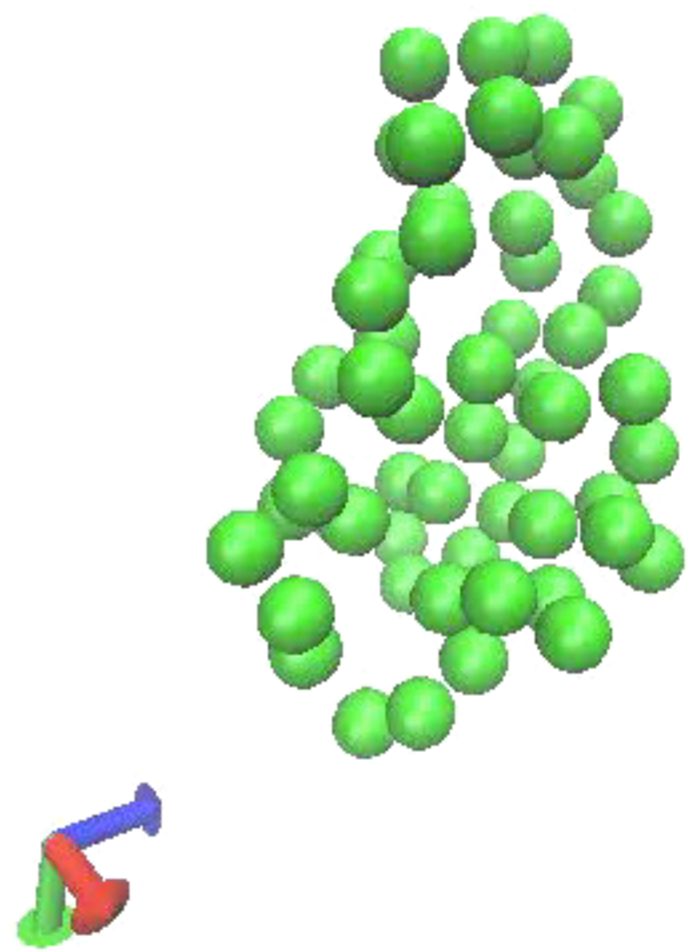


„Find properties of a system of interest

using a simple model parametrized from observations“

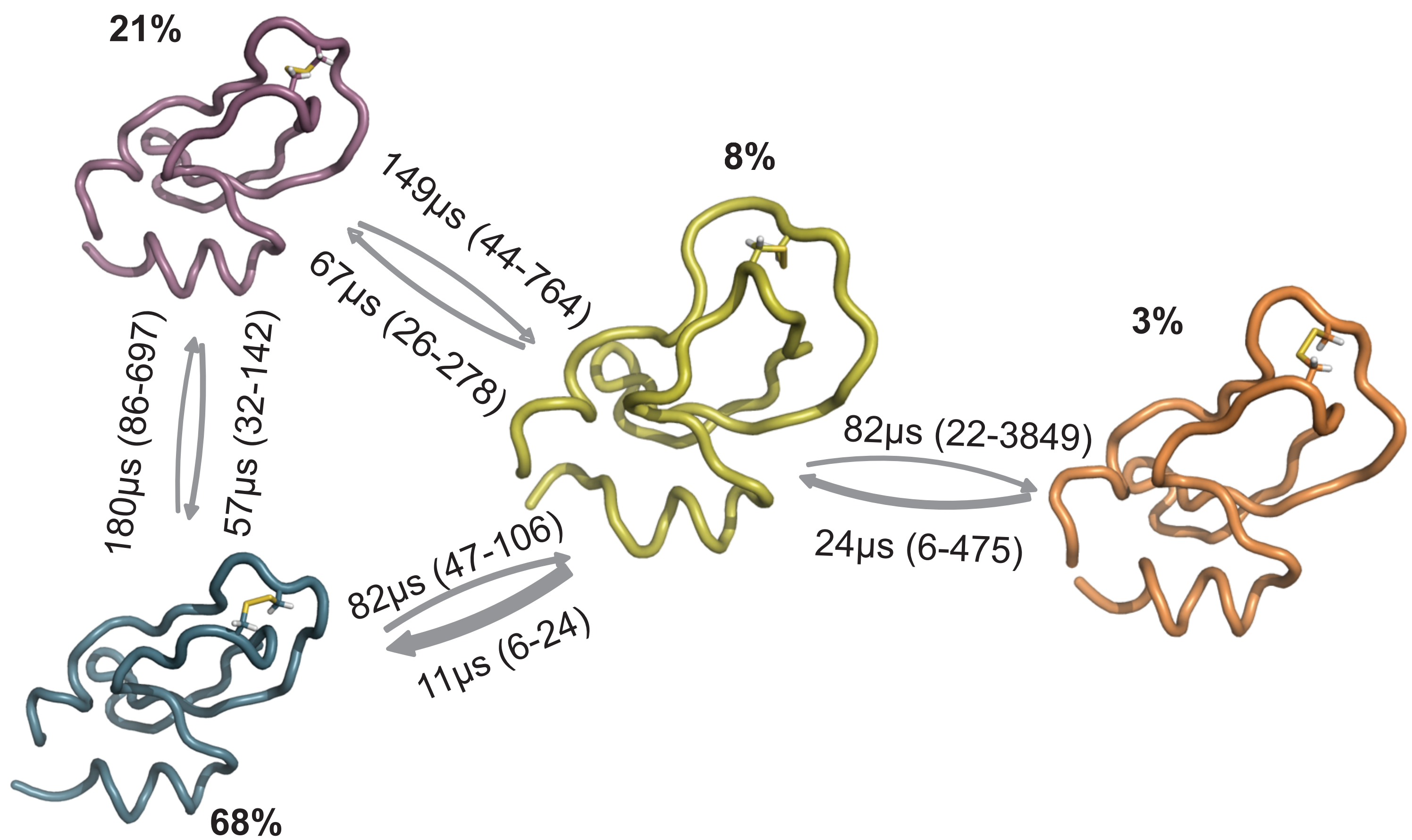
Example: BPTI

Simulation of BPTI



Shaw *et al.* **Science** 330, 6002, pp. 341-346 (2010)

Markov state models

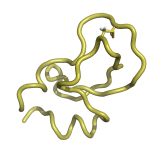
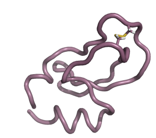
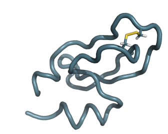
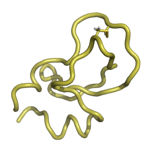
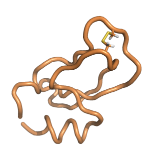
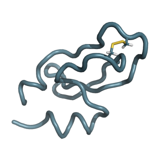


<u>Metastability of states</u> allow us to significantly simplify the dynamics of our system of interest

Markov state models

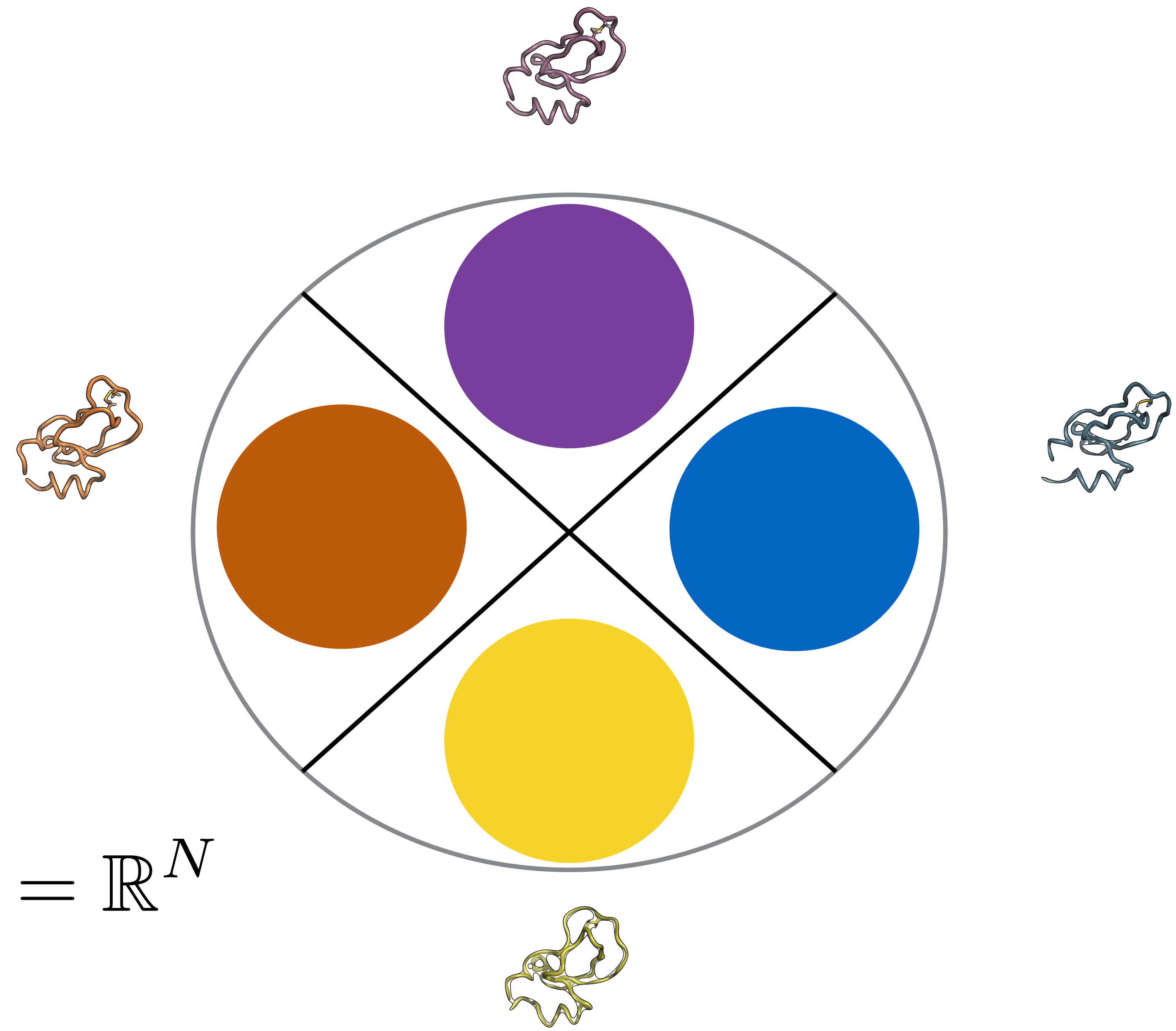
Final state

Initial state

				
	96%	1%	2%	1%
	5%	95%	0%	0%
	1%	0%	97%	2%
	1%	0%	2%	97%

A Markov state model describes the dynamics of a system as <u>conditional transition probabilities</u>

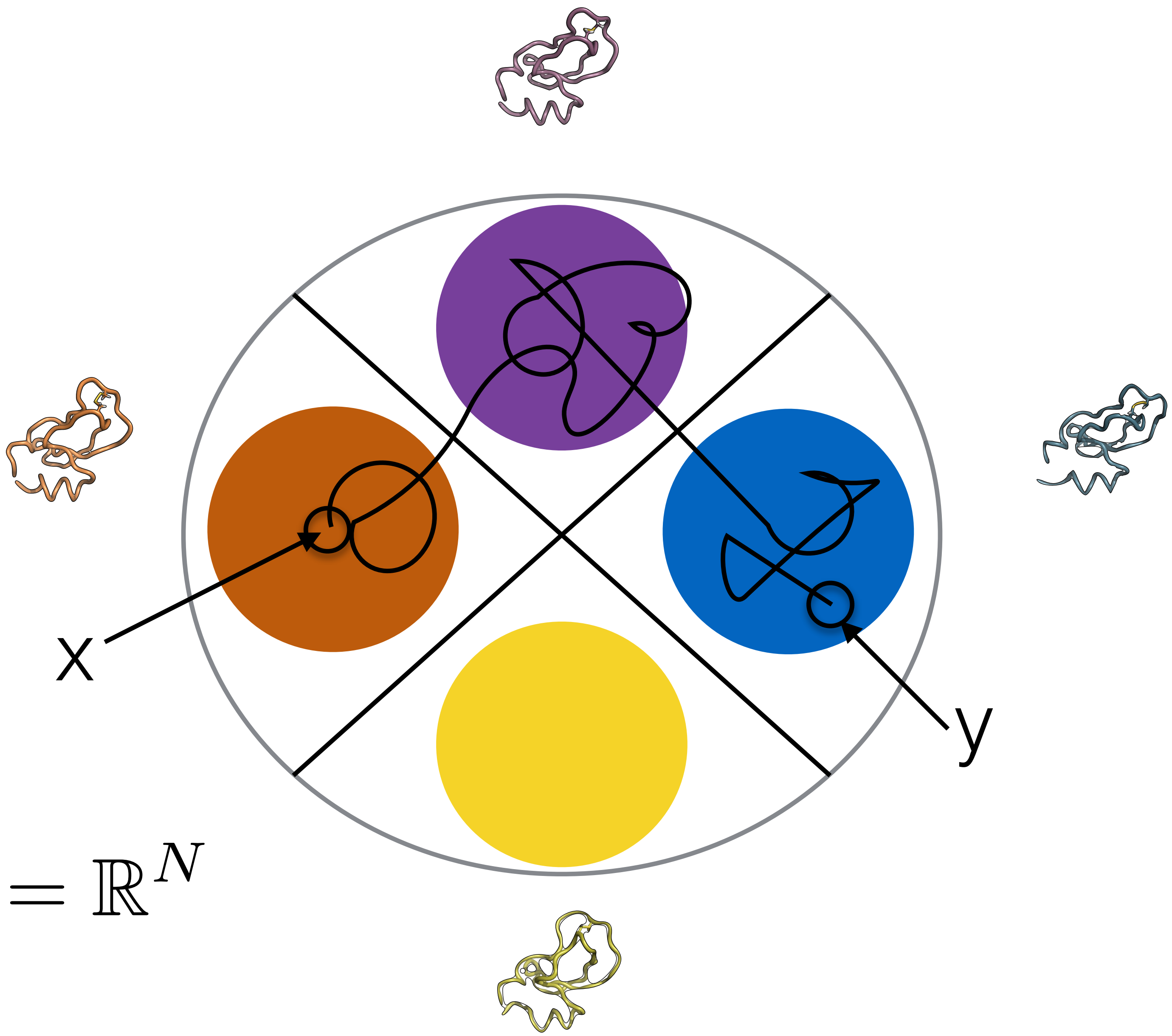
What is meta-stability?



$$x \in \Omega = \mathbb{R}^N$$

sets of configurations which are long-lived.
Markov state models assume these states, and exchange between them is important.

What is meta-stability?



$$x \in \Omega = \mathbb{R}^N$$

sets of configurations which are long-lived.
Markov state models assume these states, and exchange between them is important.

Molecular simulations

- Molecular simulations are realizations of stochastic process on Ω and are Markovian w.r.t. this space.

$$p(\mathbf{x}, \mathbf{y}; \tau)\, d\mathbf{y} = \mathbb{P}[\mathbf{x}(t+\tau) \in \mathbf{y} + d\mathbf{y} \mid \mathbf{x}(t) = \mathbf{x}]$$

$$\mathbf{x}, \mathbf{y} \in \Omega,\ \ \tau \in \mathbb{R}_{0+},$$

Transition probabilities are well defined

Prinz *et al.* (2011) JCP 134, 174105

Molecular simulations

- Molecular simulations are realizations of stochastic process on Ω and are Markovian w.r.t. this space.

$$p(\mathbf{x}, \mathbf{y}; \tau)\, d\mathbf{y} = \mathbb{P}[\mathbf{x}(t+\tau) \in \mathbf{y} + d\mathbf{y} \mid \mathbf{x}(t) = \mathbf{x}]$$

$$\mathbf{x}, \mathbf{y} \in \Omega, \;\; \tau \in \mathbb{R}_{0+},$$

Transition probabilities are well defined

$$\begin{aligned} p(\mathbf{x}, A; \tau) &= \mathbb{P}[\mathbf{x}(t+\tau) \in A | \mathbf{x}(t) = \mathbf{x}] \\ &= \int_{\mathbf{y} \in A} d\mathbf{y}\; p(\mathbf{x}, \mathbf{y}; \tau). \end{aligned}$$

Also applies for regions

Prinz *et al.* (2011) JCP 134, 174105

Molecular simulations (2)

Ergodicity

No two or more segments of the space Ω *are dynamically disconnected from each other.*

and

For an infinitely long simulation we will have visited every state $\mathbf{x} \in \Omega$ *infinitely many times.*

Prinz *et al*. (2011) JCP 134, 174105

Molecular simulations (3)

Reversibility

Simulations fulfill the detailed-balance condition:

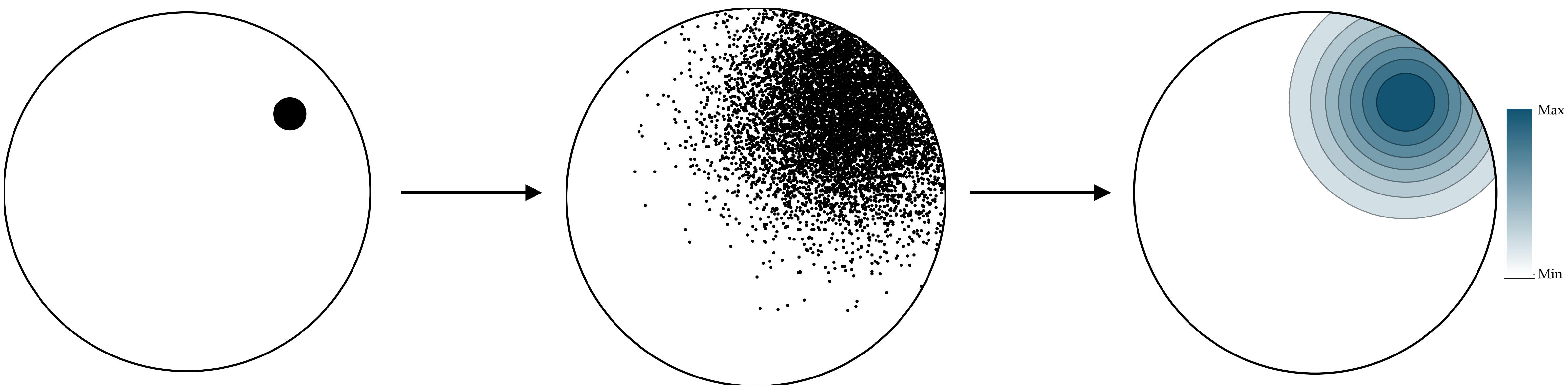
$$\mu(\mathbf{x})\ p(\mathbf{x}, \mathbf{y}; \tau) = \mu(\mathbf{y})\ p(\mathbf{y}, \mathbf{x}; \tau)$$

$$\mu(\mathbf{x}) = Z(\beta)^{-1} \exp\left(-\beta H(\mathbf{x})\right)$$

At equilibrium the probability of jumping from any x to any y is the same as jumping from y to x.

Prinz *et al*. (2011) JCP 134, 174105

An illustration of the transition density



Single — Ensemble — Density

Figure courtesy of JH Prinz

Assumptions about the full dynamics

Markovian

$$\mathbb{P}(x_{t+\tau} \in A \mid x_{t_1}, \dots, x_t = x) = \mathbb{P}(x_{t+\tau} \in A \mid x_t = x)$$

Factorization of the dynamics into conditional probabilities

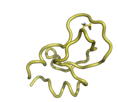
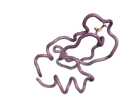
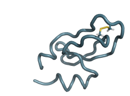
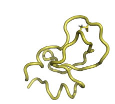
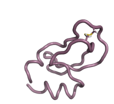
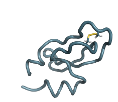
Chapman-Kolmogorov property

$$p_{\tau_1+\tau_2}(x, A) = \int_\Omega p_{\tau_1}(x, y) p_{\tau_2}(y, A) \, \mathrm{d}y$$

Direct combination of conditional probabilities with different lag-times

Final state

Initial state

				
	96%	1%	2%	1%
	5%	95%	0%	0%
	1%	0%	97%	2%
	1%	0%	2%	97%

Assumptions about the full dynamics

Irreducibility

All states of the state space can be reached from any other state in a finite time.
Ensures unique stationary distribution.

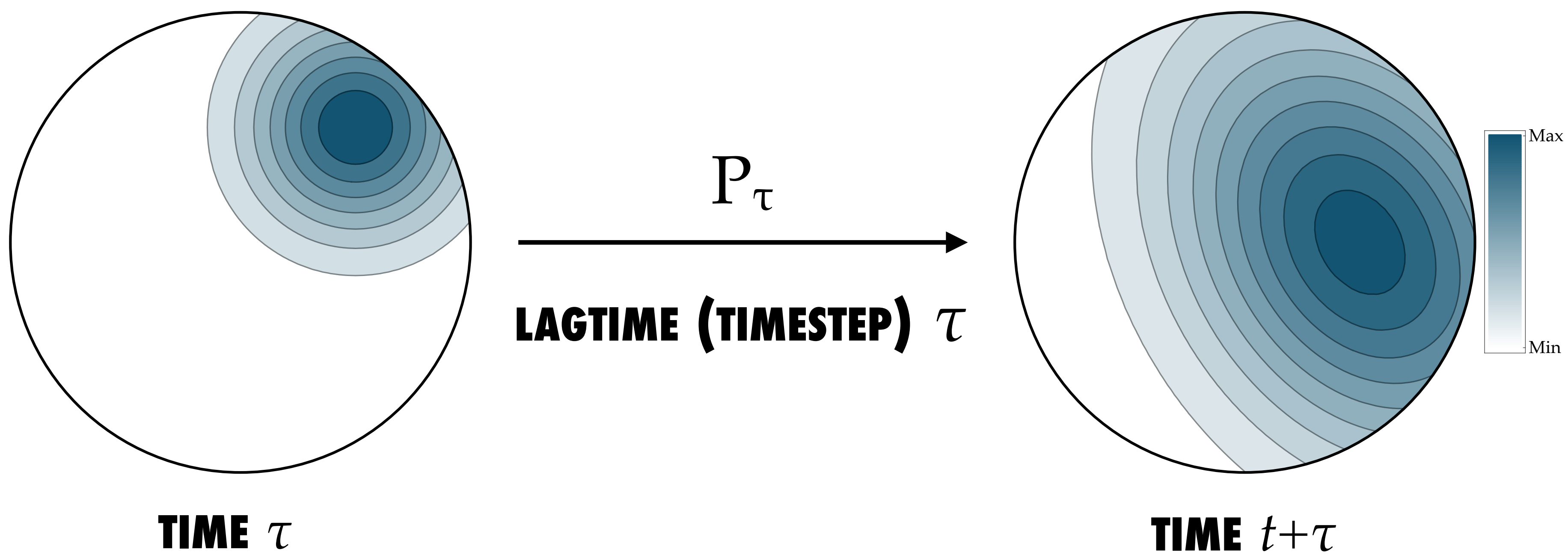
Ergodicity

No states are disconnected
No cyclic dynamics.
Ensures time and ensemble average properties are equal.

Reversibility

No net-probability flux at equilibrium. => no energy production/absorption => mass conservation.
Not strictly necessary for Markov models

Ensemble view of dynamics

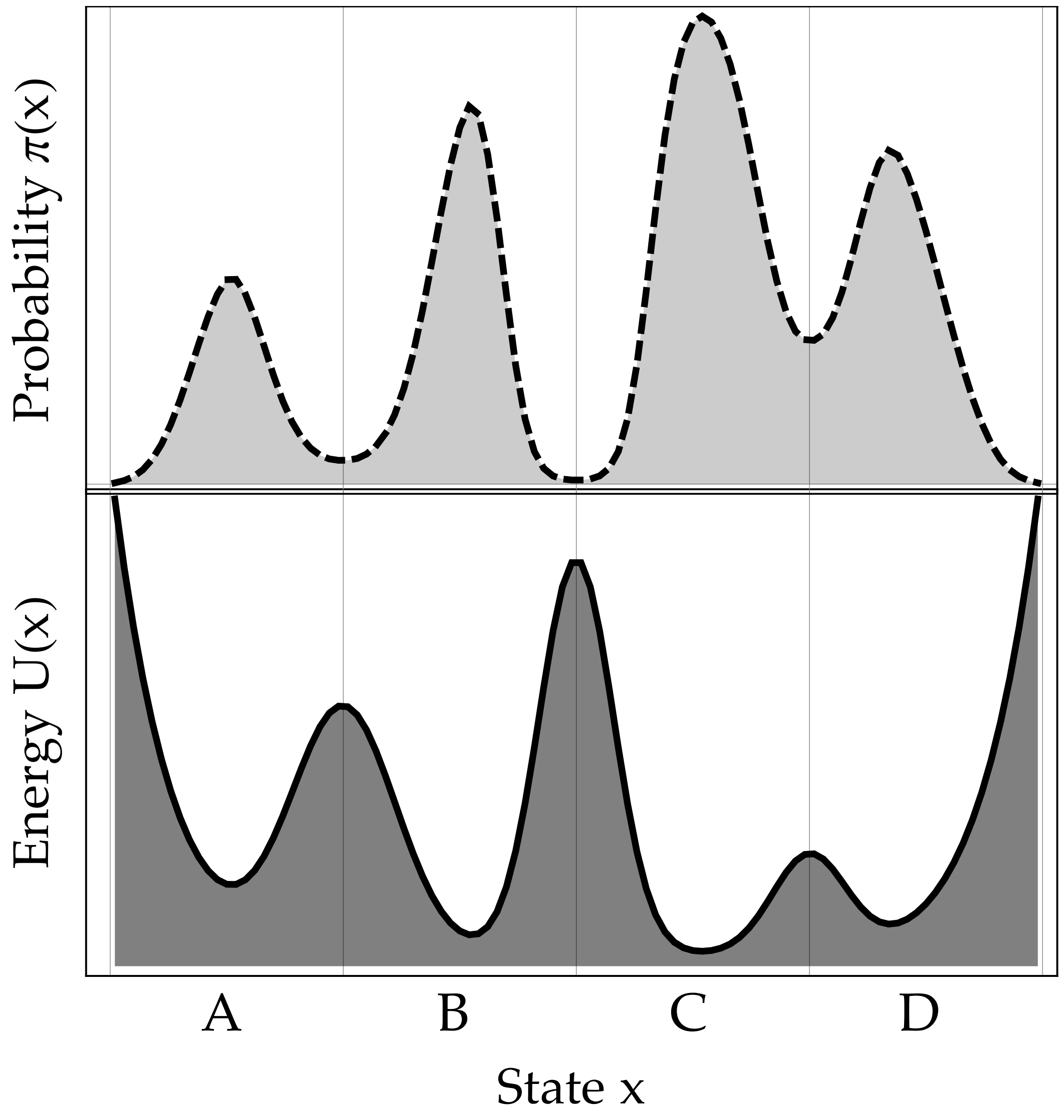


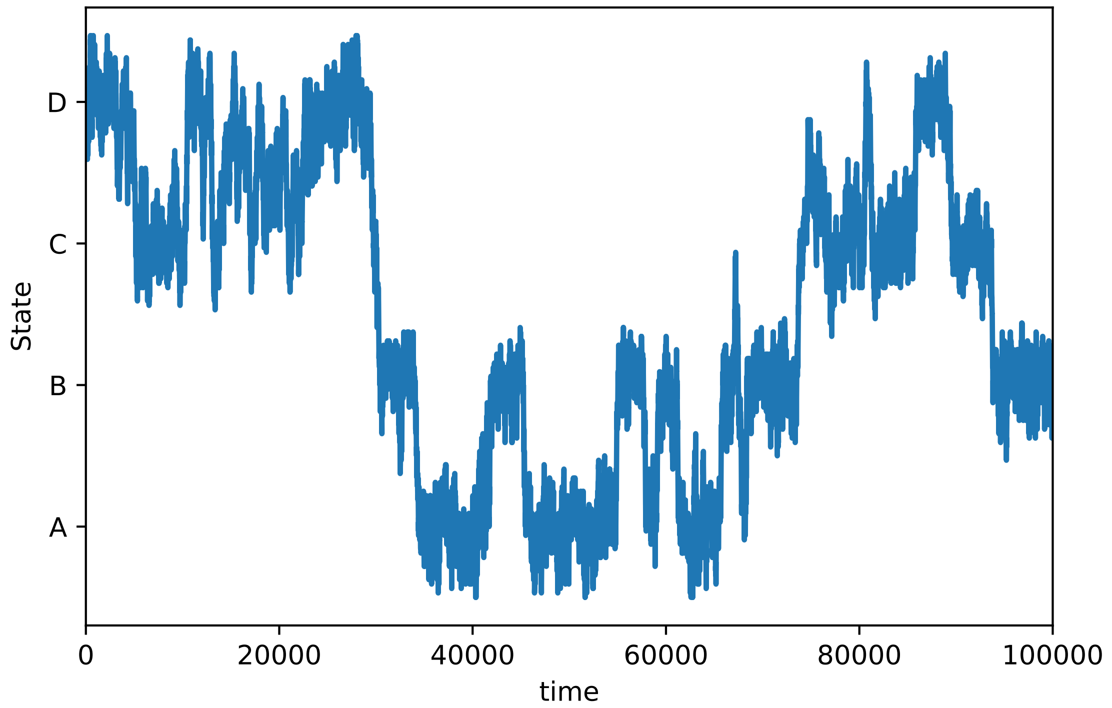
A propagator is an operator which transports probability densities in time

$$\mathrm{p}_{t+\tau}(x) = [\mathrm{P}_\tau \, \mathrm{p}_t](x) = \int_\Omega \mathrm{d}y \, \mathrm{p}_\tau(y, x) \mathrm{p}_t(y)$$

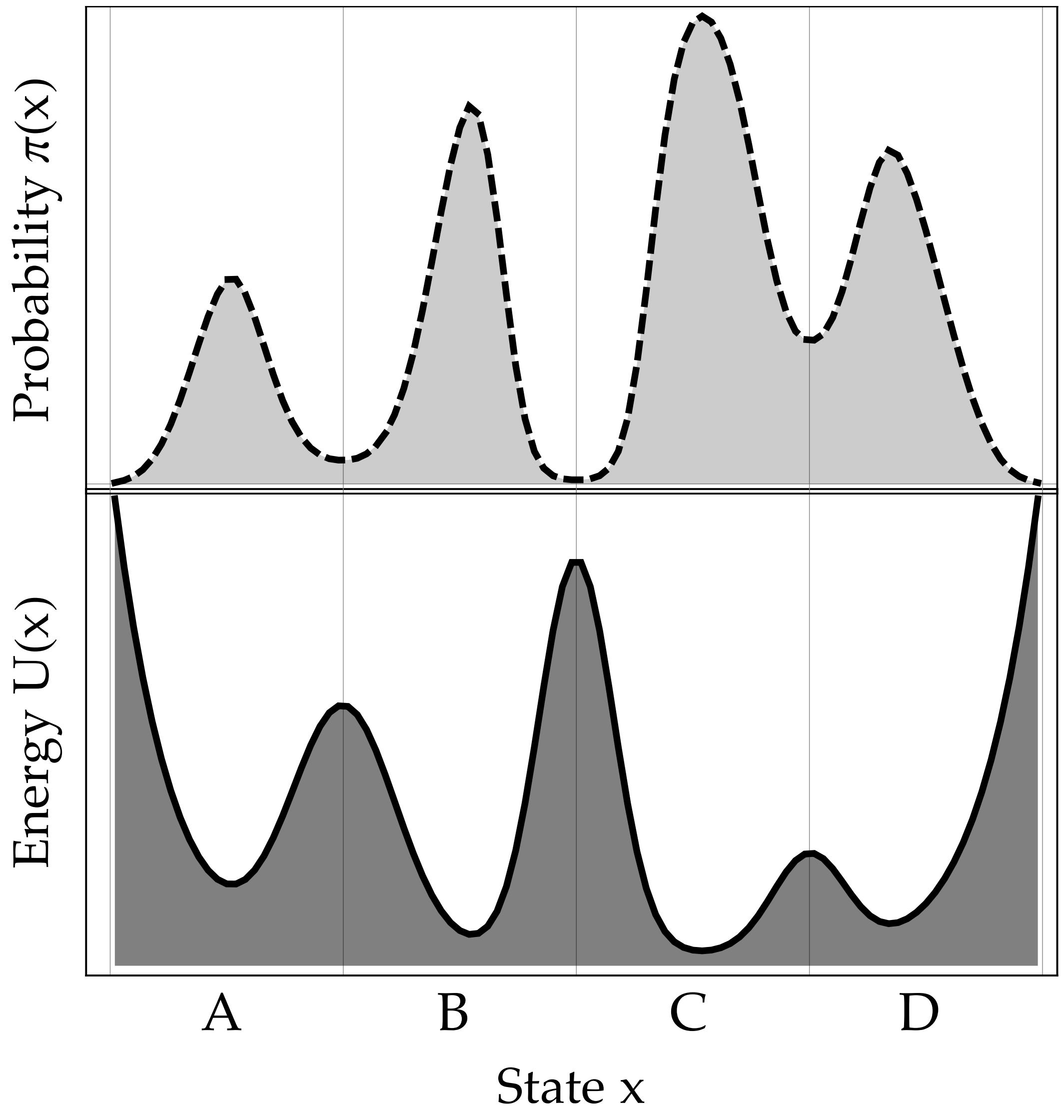
Figure courtesy of JH Prinz

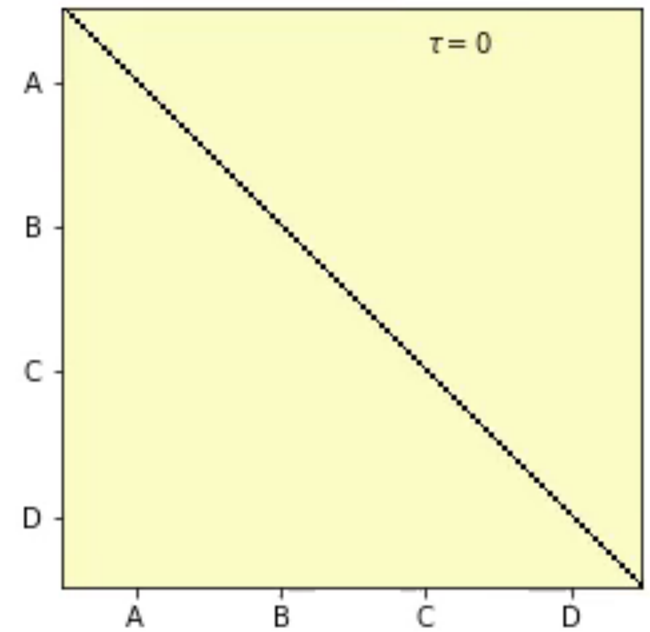
Example dynamics



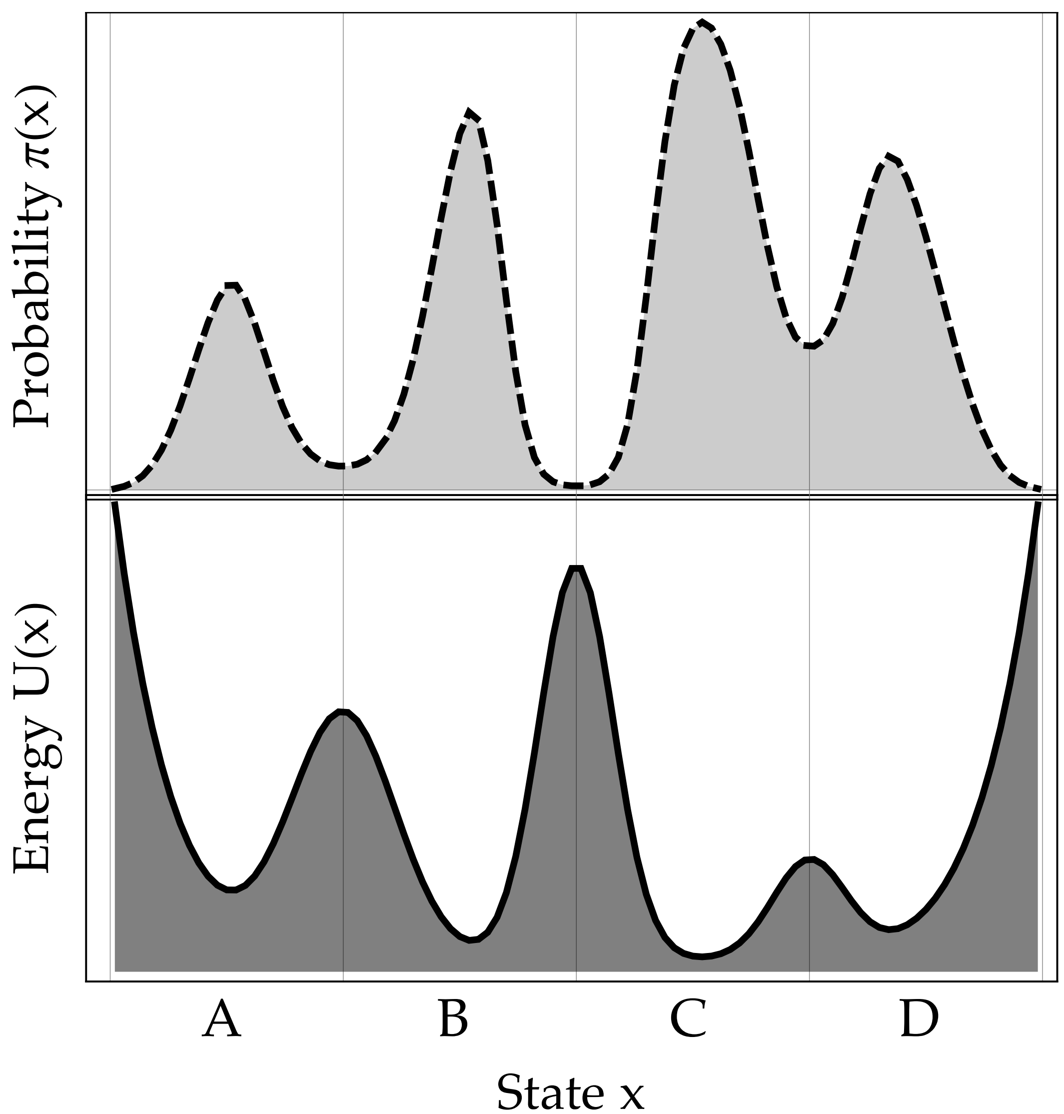


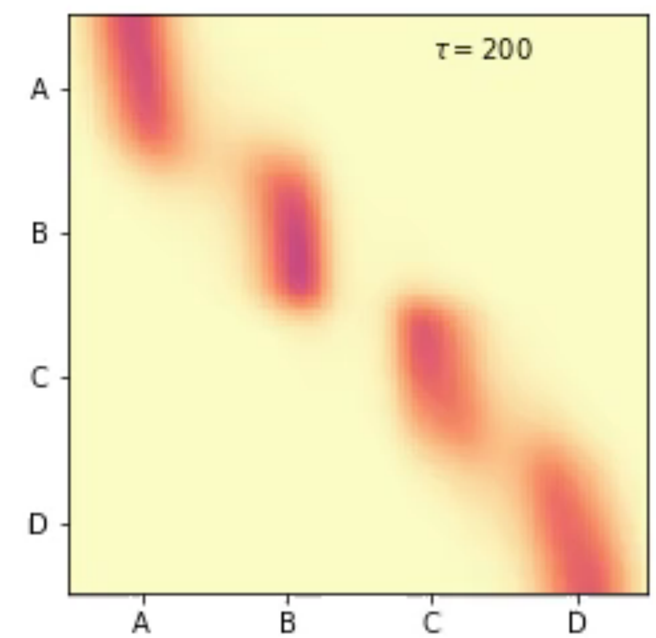
Propagator depends on lag time



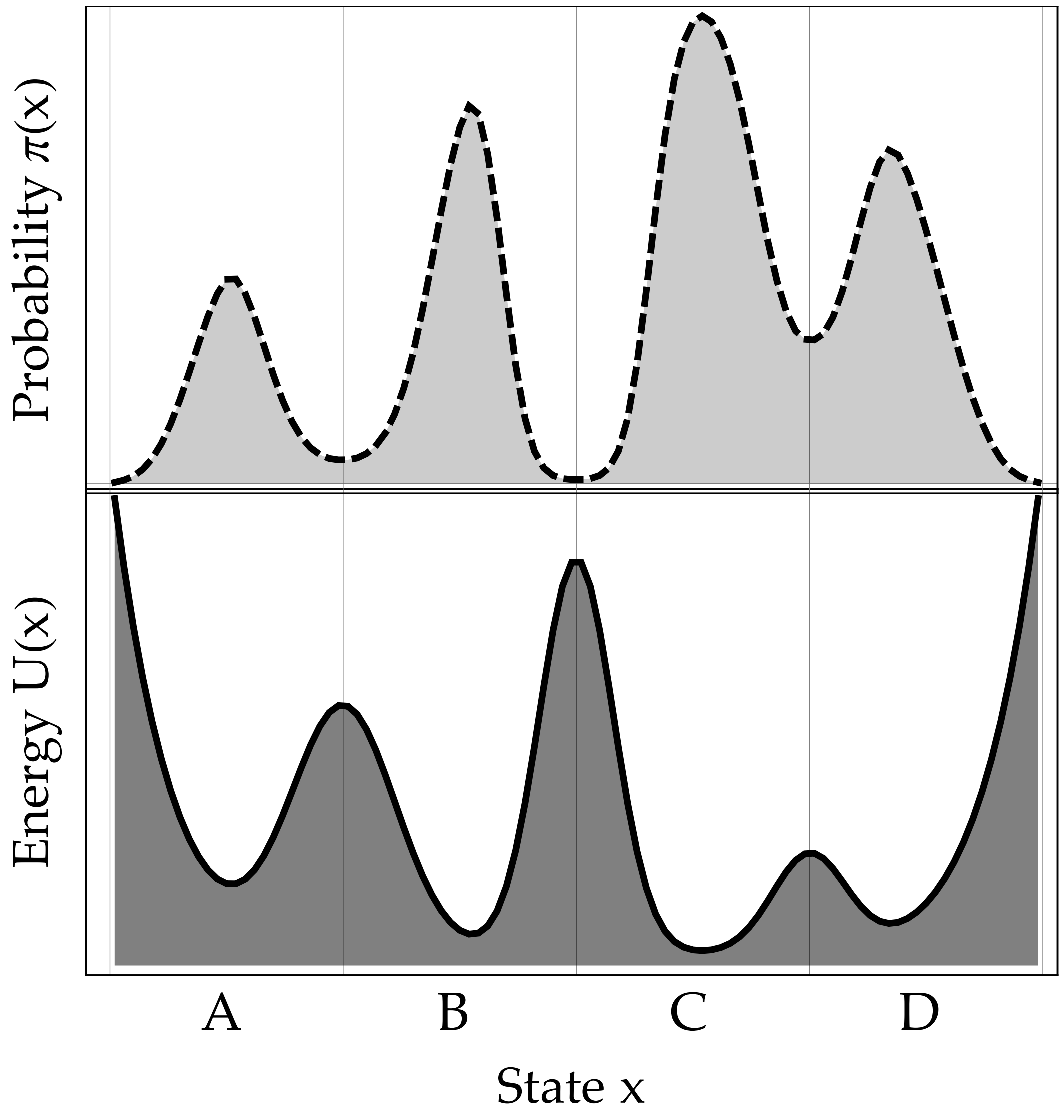


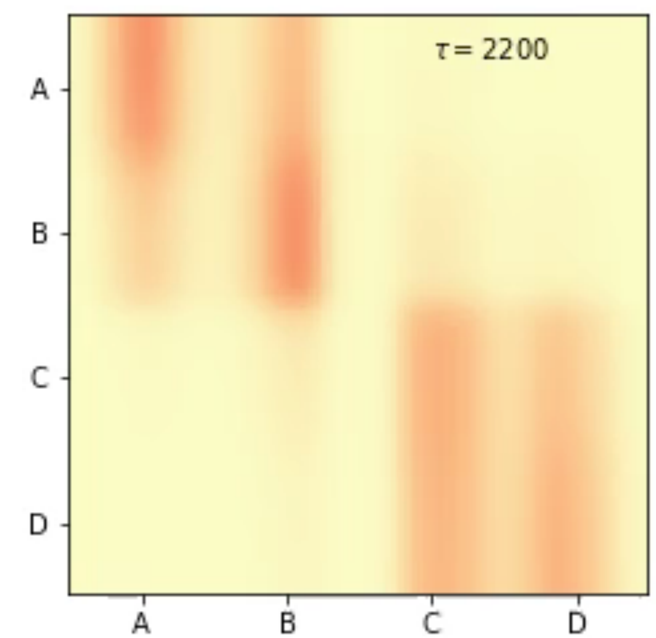
Propagator depends on lag time





Propagator depends on lag time





So why is this?

Implied time-scales

Eigenvalues of the propagator

$$P_\tau \phi_i = \lambda_i \phi_i$$

Chapman-Kolmogorov Implies exponential lag-time dependence

$$\lambda_i(k \cdot \tau) = \lambda_i^k(\tau)$$

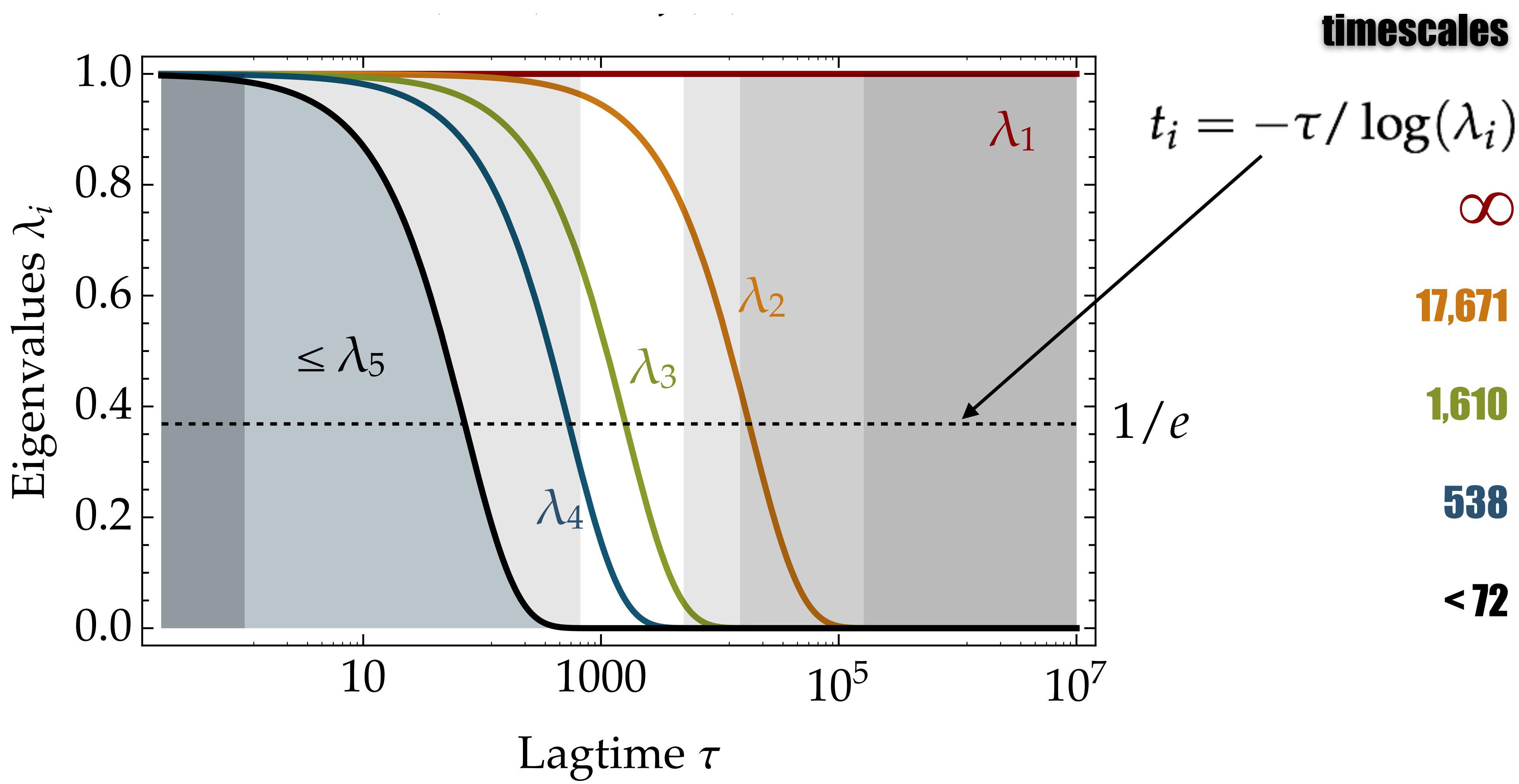


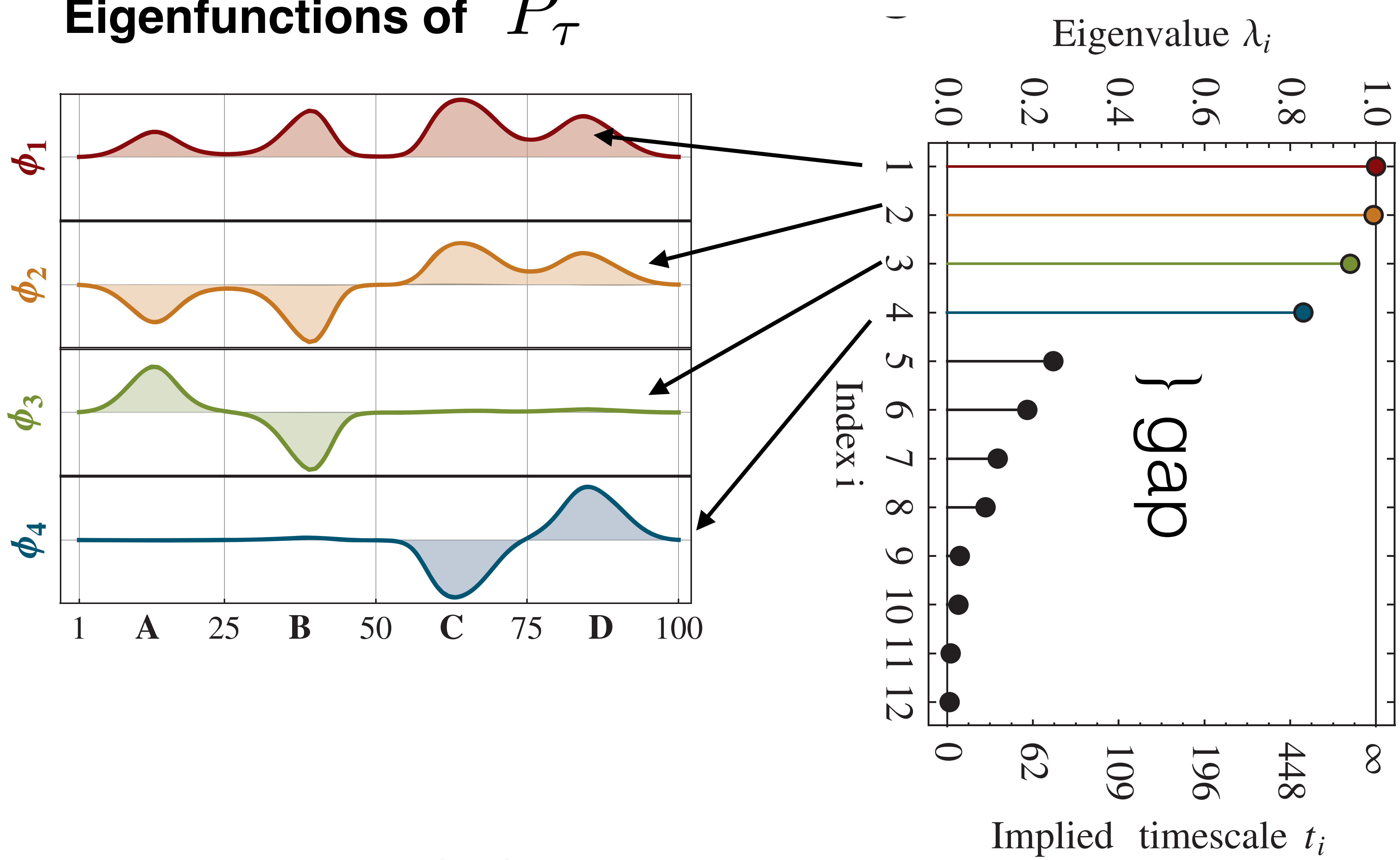
Figure courtesy of JH Prinz

Meta-stability

- We can approximate the propagator by a finite number of processes with non-zero Eigenvalues
- If we have a gap in the Eigenvalue spectrum, we can choose the lag-time in a manner such that we fulfill this assumption
- When we do this, processes faster than the lag-time ‘have decayed’ or ‘are not resolved’.

What do you mean by processes?

Eigenfunctions of P_τ



$$t_i = -\tau / \log(\lambda_i)$$

Prinz *et al.* (2011) JCP 134, 174105

Estimation

Discretization of Ω

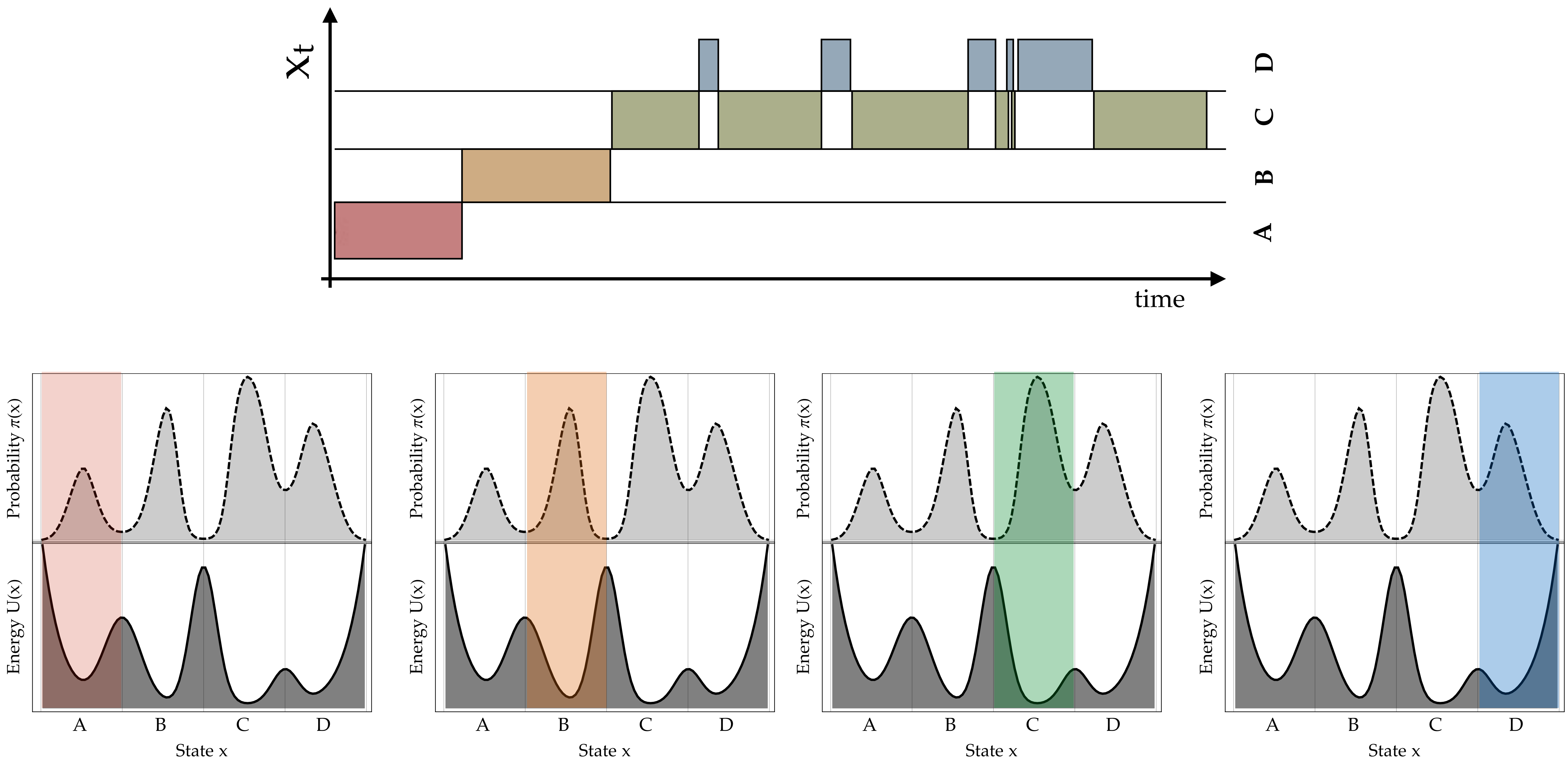


Figure courtesy of JH Prinz

Count matrix

$C_{ij}(1)$	A	B	C	D
A	9963	37	0	0
B	22	9974	4	0
C	0	2	9919	79
D	0	0	115	9885

$$C_{ij}(\tau) = \sum_{n=\tau}^{T} \delta(x_{n-\tau} = i, x_n = j)$$

Figure courtesy of JH Prinz

Maximum likelihood estimator

We can express the probability of the observed data - discrete trajectory - given a transition probability matrix of an MSM

$$
\begin{aligned}
\mathbb{P}(x_1, \ldots, x_t \mid P) &= \prod_{k=1}^{L} p_{x_{k-1}, x_k} \\
&= p_{x_0, x_1} \cdot \ldots \cdot p_{x_L - 1, L} \\
&= \prod_{ij} p_{ij}^{c_{ij}} \\
&= p_{11}^{c_{11}} \cdot \ldots
\end{aligned}
$$

The aim is then to find the *P* which maximizes this expression - That is, the *Maximum likelihood estimator.*

Analytical solution for Non-reversible case

- We enforce the constraint that the transition probability matrix is row-stochastic:

$$\sum_j p_{ij} = 1, \quad \forall i$$

- One can show the estimator is simply:

$$\hat{p}_{ij} = \frac{\hat{C}_{ij}}{\sum_j \hat{C}_{ij}}$$

Prinz *et al.* (2011) JCP 134, 174105

Reversible estimator

- Enforces the detailed balance condition.
- No exact analytical solution:
  - Fixed-point iteration algorithm available.
  - Approximate solutions.
- Implemented in PyEMMA

Prinz *et al.* (2011) JCP 134, 174105
Bowman et al. (2009) JCP 131, 124101

Bayesian inference of MSMs

- The less simulation data we have, the more ambiguous the solution of the likelihood problem will be.
- Consequently, if we limit ourselves to the MLE, we are *ignorant* as to how **robust** our inferred MSM is.
- One way to quantify the uncertainty of MSMs is through **Bayesian inference**

Bayesian inference of MSMs

Likelihood from before

$$\mathbb{P}(x_i, \ldots, x_t \mid P) = p(C \mid P) \propto \prod_{i,j=1}^{n} p_{ij}^{c_{ij}}$$

Bayesian inference of MSMs

Likelihood from before

$$\mathbb{P}(x_i, \ldots, x_t \mid P) = p(C \mid P) \propto \prod_{i,j=1}^{n} p_{ij}^{c_{ij}}$$

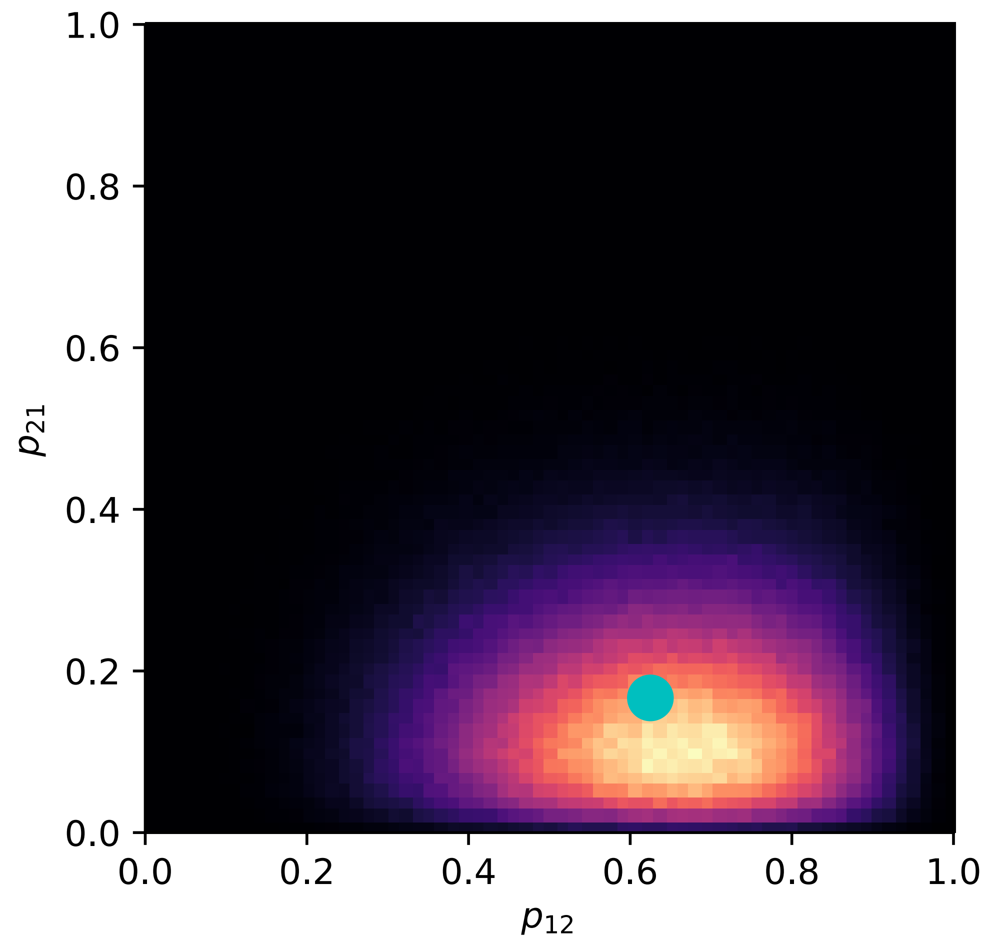
Introduction of prior information

$$p(P \mid C) \propto p(C \mid P) p(P)$$

The prior can encode useful constraints: row-stochasticity, reversibility, fixed stationary distribution, sparsity etc

Bayesian inference of MSMs

Inference is done by MCMC sampling

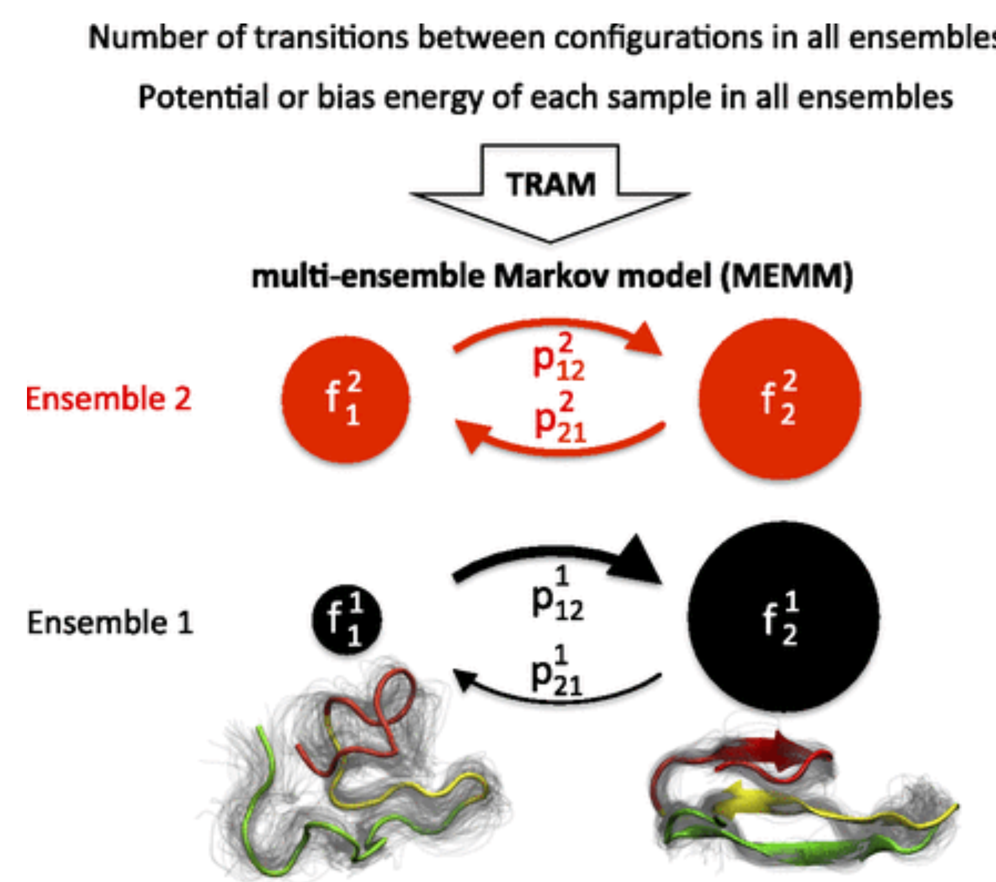


Noé (2008) JCP 128, 244103
Trendelkamp-Schroer & Noé (2013) JCP 138, 164113

Alternative estimators

Transition(-based) Reweighting Analysis Method

- Allows taking into account simulation data from multiple thermodynamic ensembles.
- That means, we can **use data from enhanced sampling simulations together with unbiased simulation data to generate models more efficiently.**
- More about this wednesday.

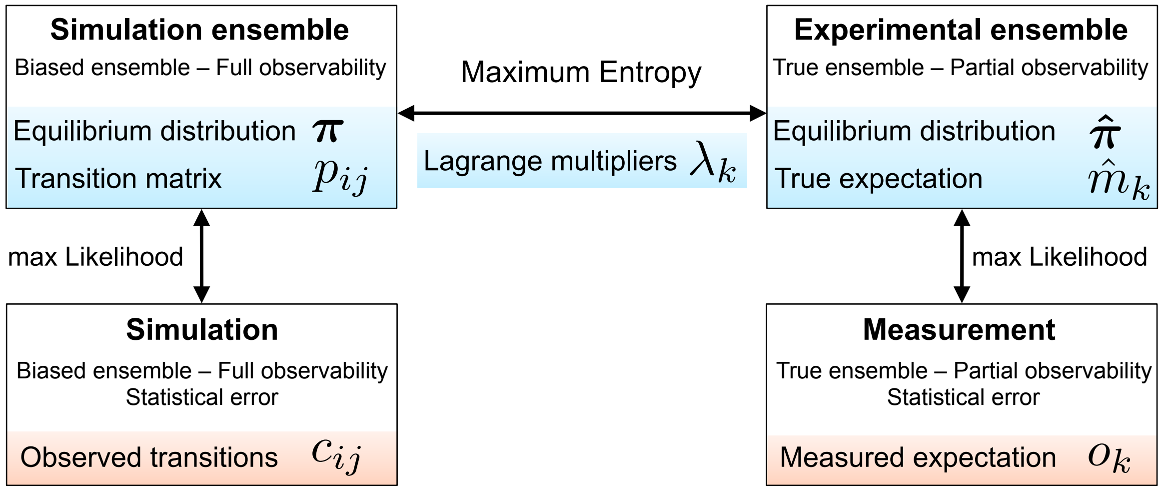


Wu et al. *PNAS* 2016, 113(23), E3221–E3230

Implemented in PyEMMA

Augmented Markov models

- Enables integration of external information into the estimation of Markov state models.
- Fx use of experimental constraints from biophysical experiments such as NMR.
- A notebook tutorial distributed with PyEMMA 2.5 and up.



Olsson et al. *PNAS* 2017, 114(31), pp. 8265-8270. doi: 10.1073/pnas.1704803114

Implemented in PyEMMA

Analysis of our estimate

$P_{ij}(1)$	A	B	C	D
A	0,9963	0,0037		
B	0,0022	0,9974	0,0004	
C		0,0002	0,9919	0,0079
D			0,0115	0,9885

projected timescales	original timescales
∞	∞
2,746	17,671
165	1,610
51	538

Time-scales are always under-estimated

Figure courtesy of JH Prinz

Increasing the lag-time

COUNT MATRIX

$C_{ij}(100)$	A	B	C	D
A	9533	477	40	0
B	1644	8014	262	80
C	0	40	9025	935
D	0	0	1366	8634

projected timescales	original timescales
∞	∞
15,397	17,671
1211	1,610
379	538

May improve estimates of predicted time-scales

Figure courtesy of JH Prinz

Projection/discretization error

$$t_i = -\tau / \log(\lambda_i)$$

metastable region

GOOD PROJECTION

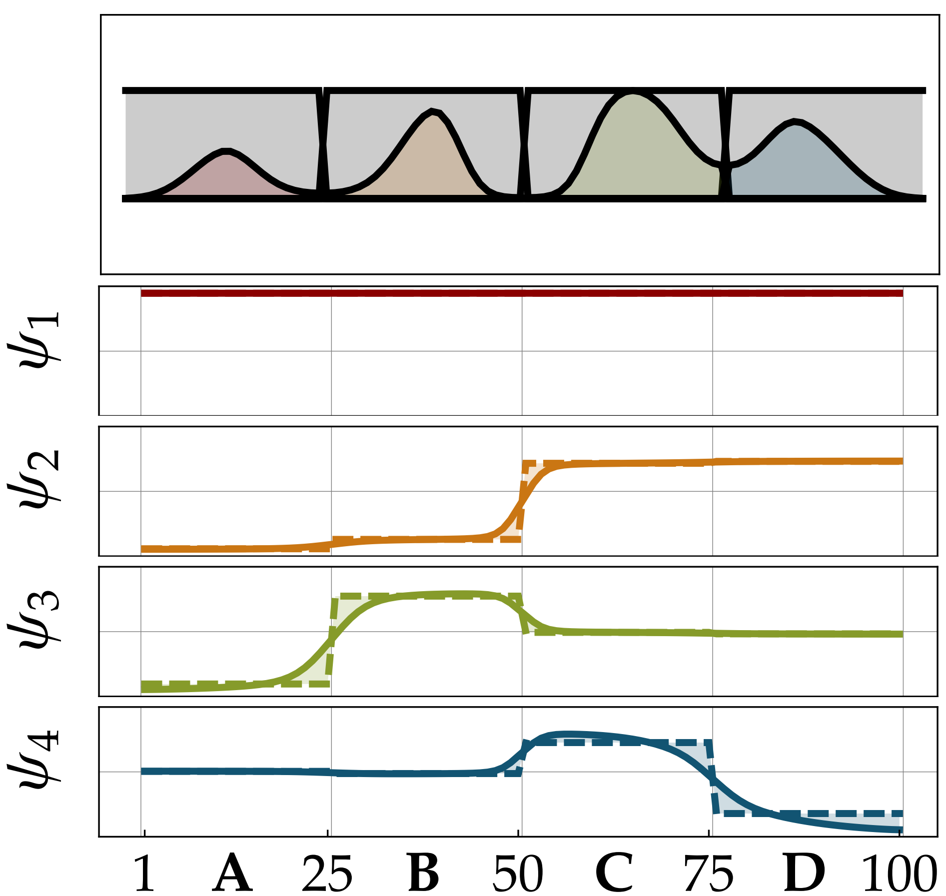


Figure courtesy of JH Prinz

Projection/discretization error

metastable region

$$t_i = -\tau / \log(\lambda_i)$$

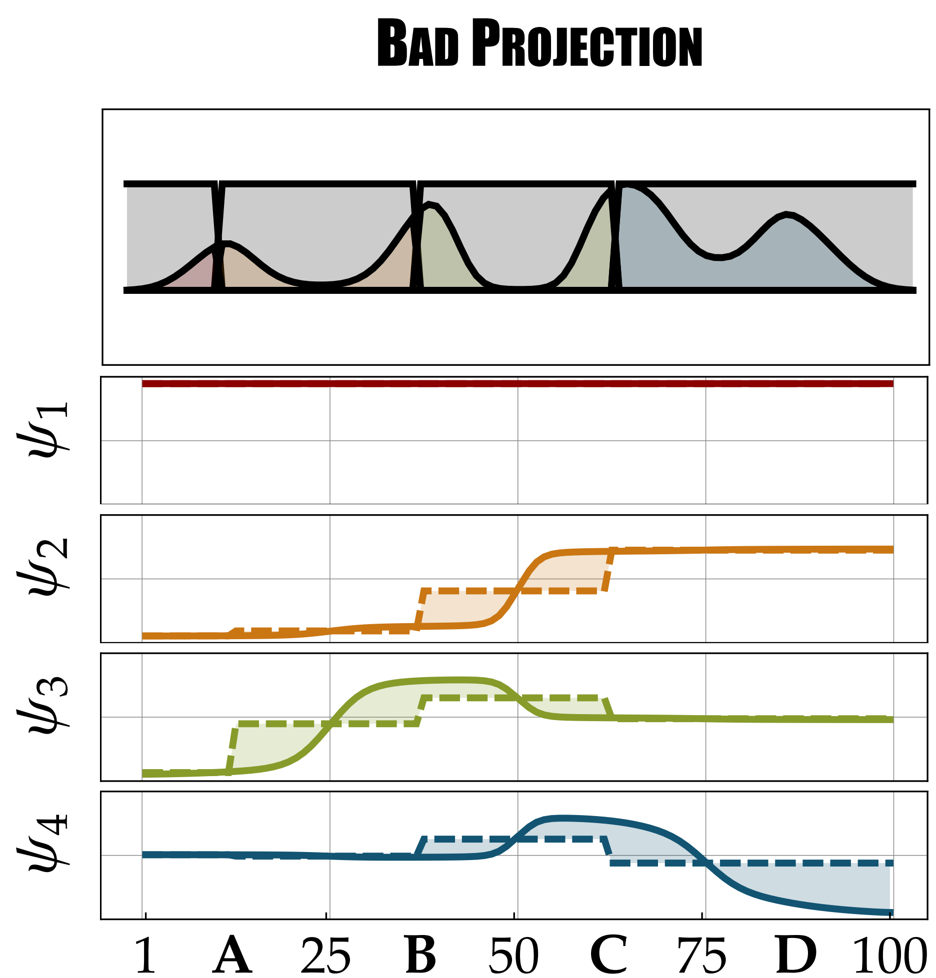


Figure courtesy of JH Prinz

Known problems

- Observations (projections, discretizations) are in many cases <u>not Markovian</u>
- However, we are often interested in understanding the full system not just the observation.
- Since we often have a lot of freedom to choose the projections and discretization, it is important to chose one which is as Markovian as possible.

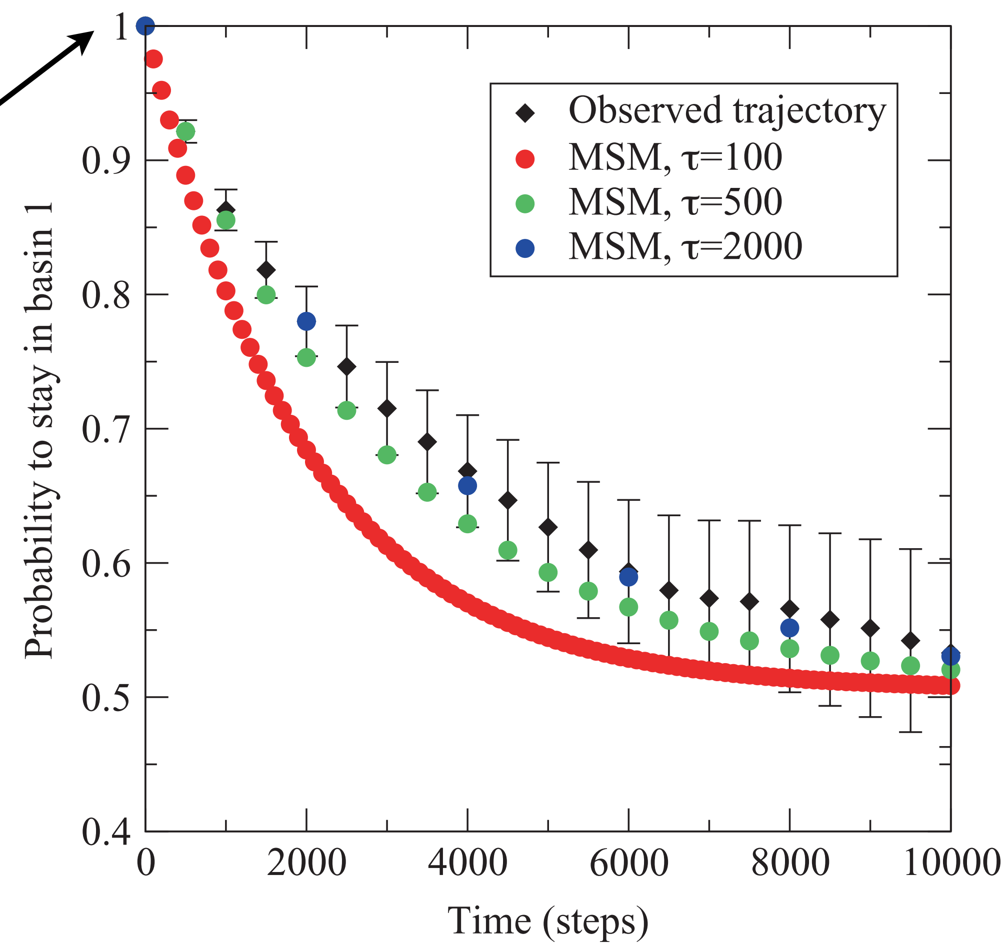
Validation

Chapman-Kolmogorov test

Compare the evolution of the data with the model

$$\underbrace{T^k(\tau)}_{\text{Markov model prediction}} \approx \underbrace{T(k\tau)}_{\text{estimation from data}}$$

assume
any initial distribution



General scheme for Markov state model generation

- Discretize a suitable projection of your data.
- Construct a transition matrix.
- Estimate the number of meta-stable states (time-scale gap)
- Perform Chapman-Kolmogorov test.

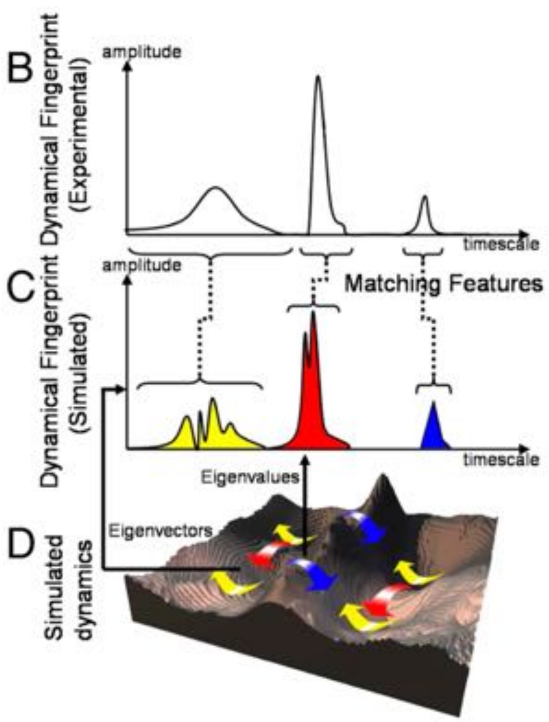
Analysis

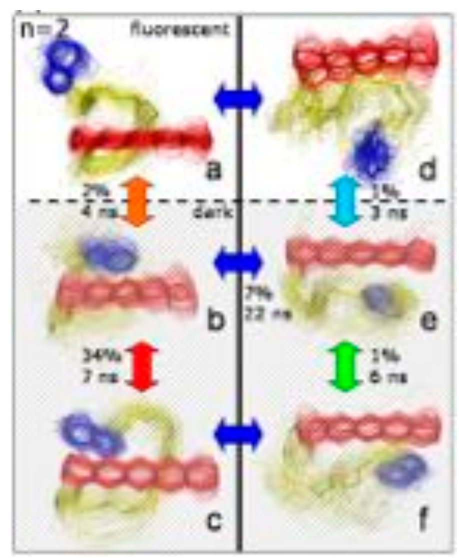
Useful predictions from a MSM

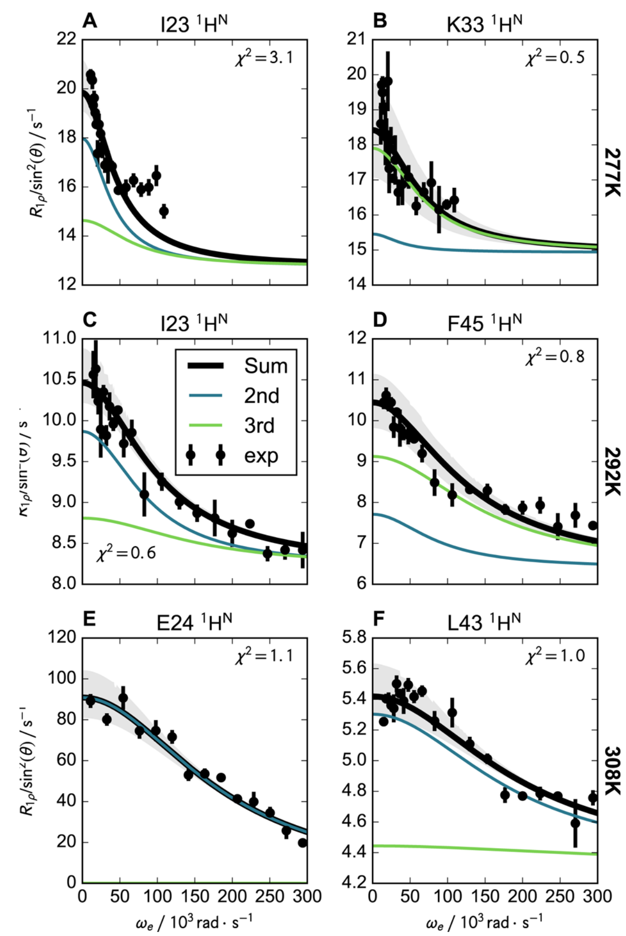
Common properties

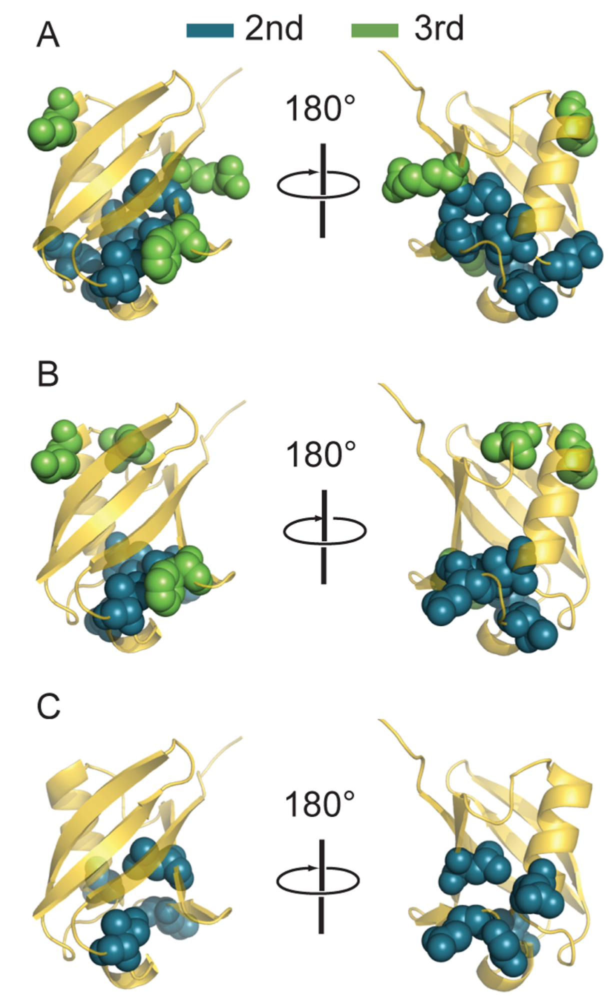
- Relaxation time-scales
- Dominant processes
- Stationary distribution (thermodynamics)
- Meta-stable sets (more about this later)
- Correlation functions (spectroscopic observables)
- Mean first passage times
- Path probabilities

Spectroscopic observables









Noé et al. *Dynamical fingerprints for probing individual relaxation processes in biomolecular dynamics with simulations and kinetic experiments*. Proc. Nat. Acad. Sci. USA 108, 4822–4827 (2011).

Olsson & Noé *Mechanistic Models of Chemical Exchange Induced Relaxation in Protein NMR. 139, 200–210 JACS (2017)*

Summary

- Markov state models are derived coarse-grained models of the full original (Markovian) dynamics .
- MSMs may be parameterized (estimated/learned) from simulation data to compute properties of interest.
- MSMs are particularly useful if the projection/ discretization error can be minimized: then the predicted quantities match the original.

Questions?