Introduction to MD Workflows and Tools

Prof. Vijay S. Pande, PhD

Departments of Chemistry, Computer Science, and Structural Biology

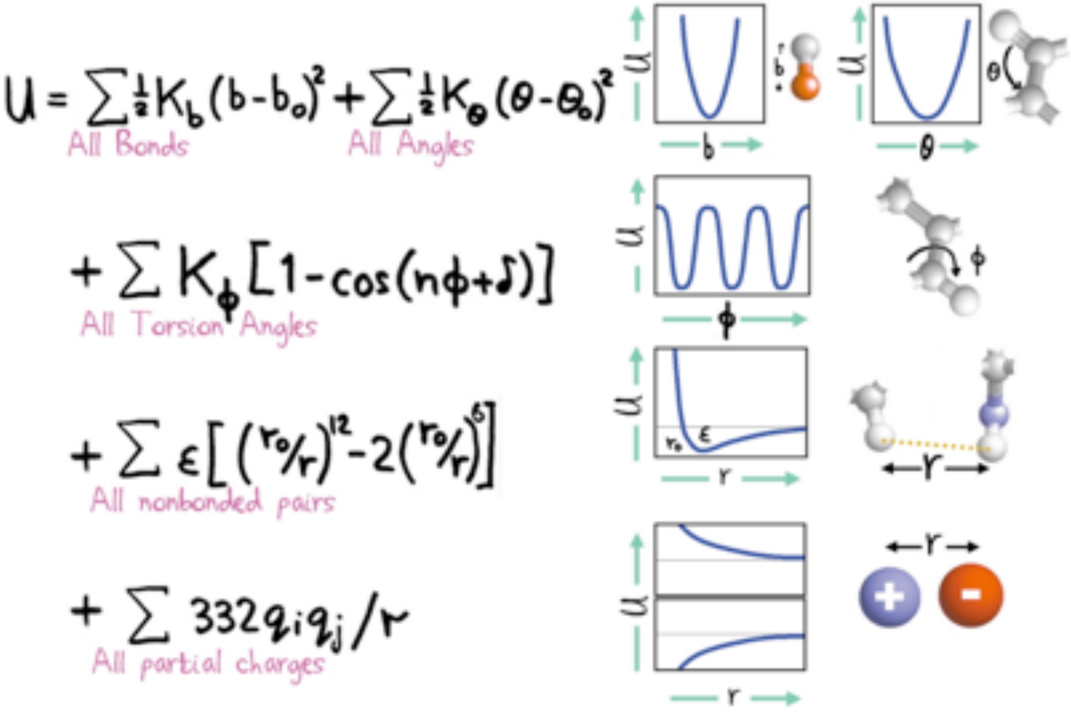
Director, Biophysics Program

Director, Folding@home Distributed Computing Project

Stanford University

The dream: simulating molecular dynamics

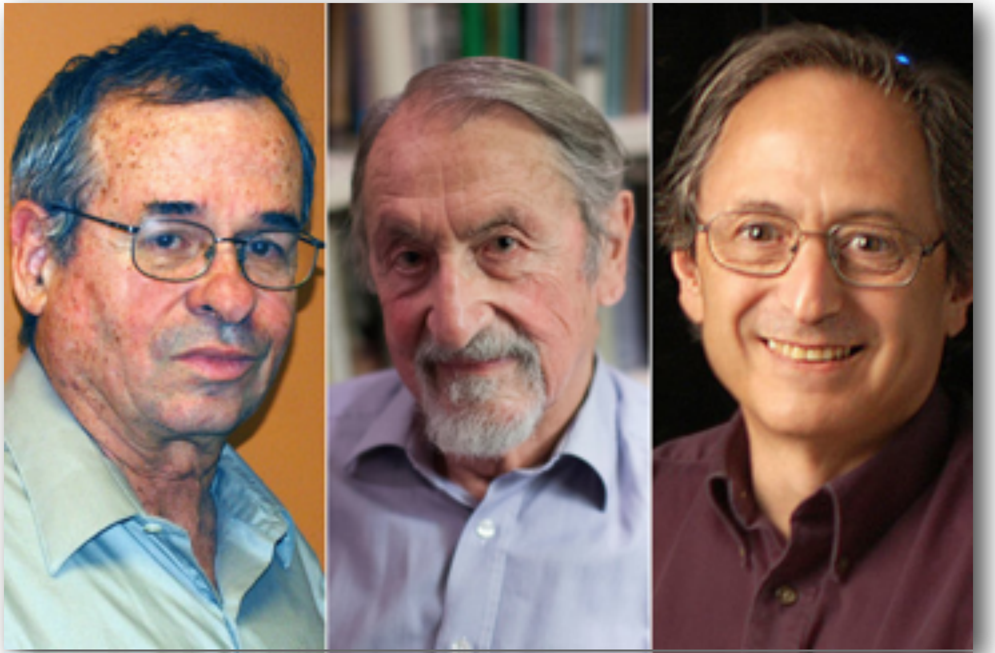
Basic idea: calculate forces between atoms, then numerically integrate Newton's Equations


$U = \sum \frac{1}{2} K_b (b-b_0)^2 + \sum \frac{1}{2} K_\theta (\theta-\theta_0)^2$
All Bonds
All Angles
$+ \sum K_\phi [1-\cos(n\phi+\delta)]$
All Torsion Angles
$+ \sum \epsilon [(r_0/r)^{12} - 2(r_0/r)^6]$
All nonbonded pairs
$+ \sum 332 q_i q_j / r$
All partial charges
U
b
θ
ϕ
r_0
ε
r
+
-

M. Levitt, *Nature Structural Biology* **8** 392 (2001)

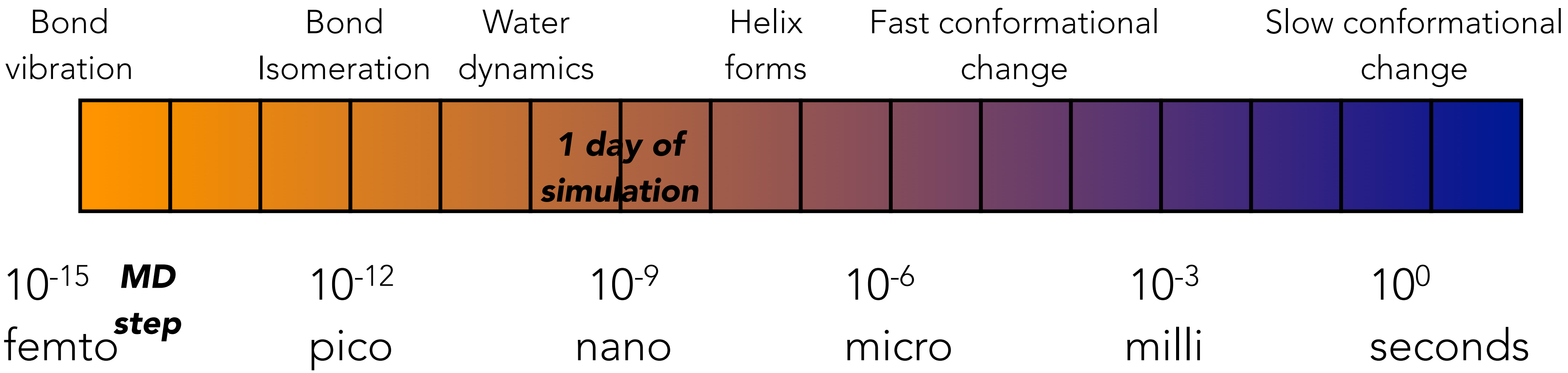
The dream: simulating molecular dynamics

Basic idea: calculate forces between atoms, then numerically integrate Newton's Equations



2013 Nobel Prize in Chemistry Awarded to Karplus, Levitt, and Warshel

The nightmare: long time scales



Bond vibration

Bond Isomeration

Water dynamics

Helix forms

Fast conformational change

Slow conformational change

1 day of simulation

10^{-15} femto **_MD step_**

10^{-12} pico

10^{-9} nano

10^{-6} micro

10^{-3} milli

10^{0} seconds

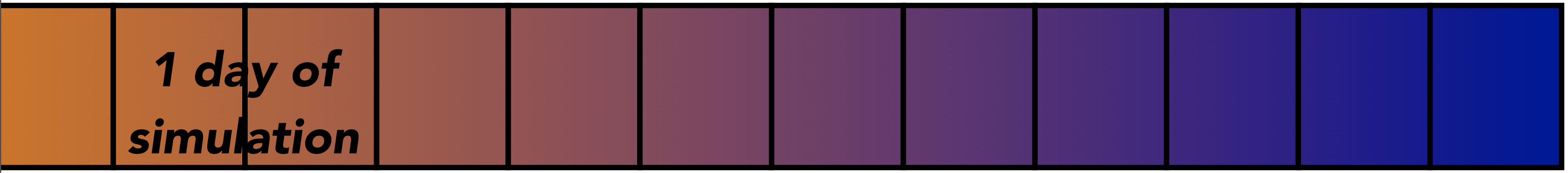
The nightmare: long time scales

Water dynamics

Helix forms

Fast conformational change

Slow conformational change



1 day of simulation

10^{-9} nano

10^{-6} micro

10^{-3} milli

10^{0} seconds

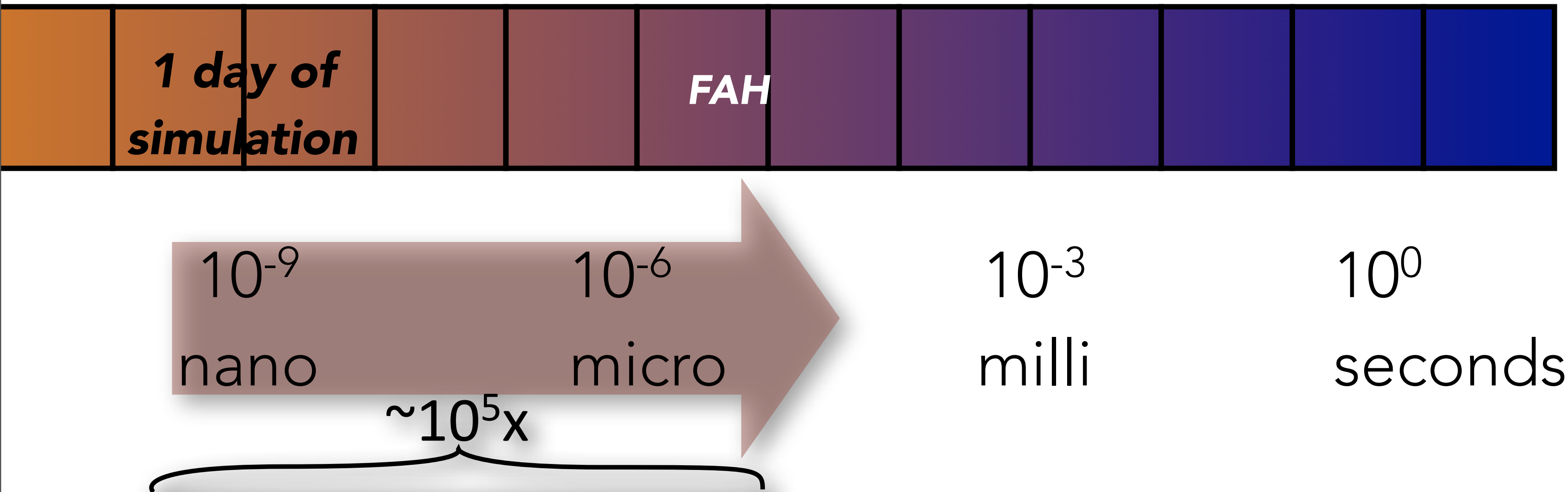
The nightmare: long time scales

Water dynamics

Helix forms

Fast conformational change

Slow conformational change



1 day of simulation

FAH

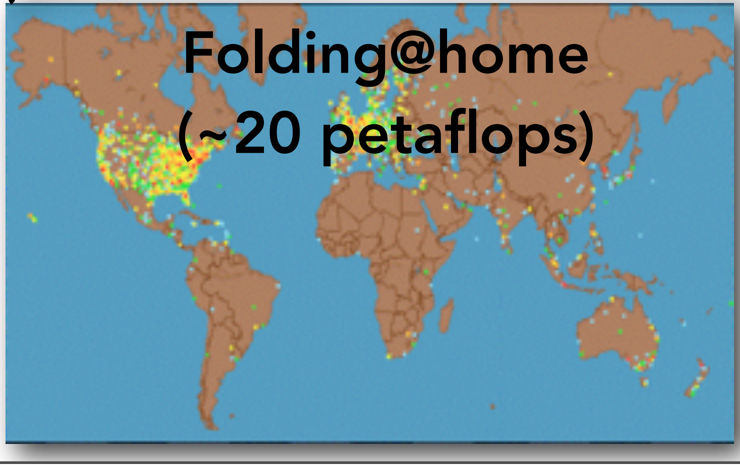
10^{-9} nano

10^{-6} micro

10^{-3} milli

10^{0} seconds

$\sim10^5$x



Folding@home (~20 petaflops)

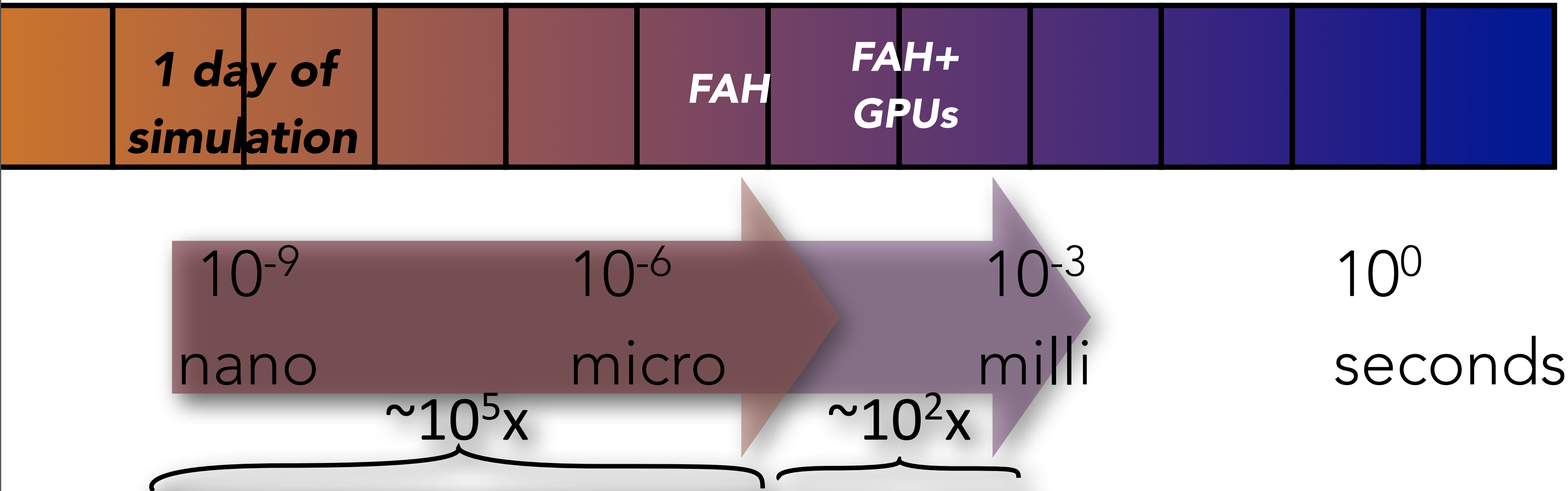
The nightmare: long time scales

Water dynamics

Helix forms

Fast conformational change

Slow conformational change



1 day of simulation

FAH

FAH+ GPUs

10^{-9} nano

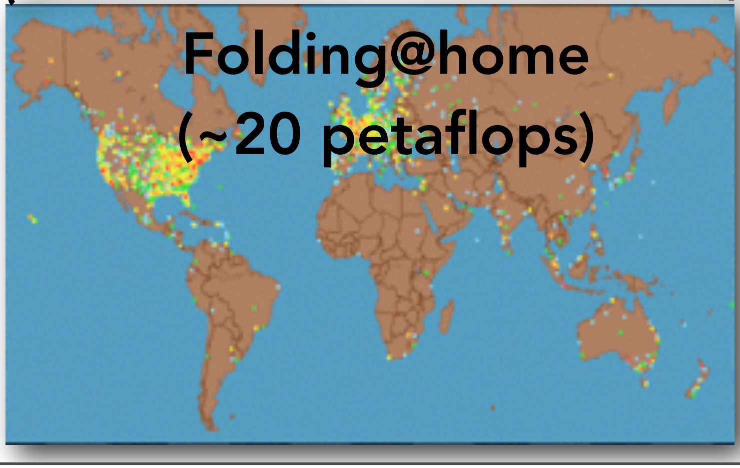
10^{-6} micro

10^{-3} milli

10^{0} seconds

$\sim 10^5$x

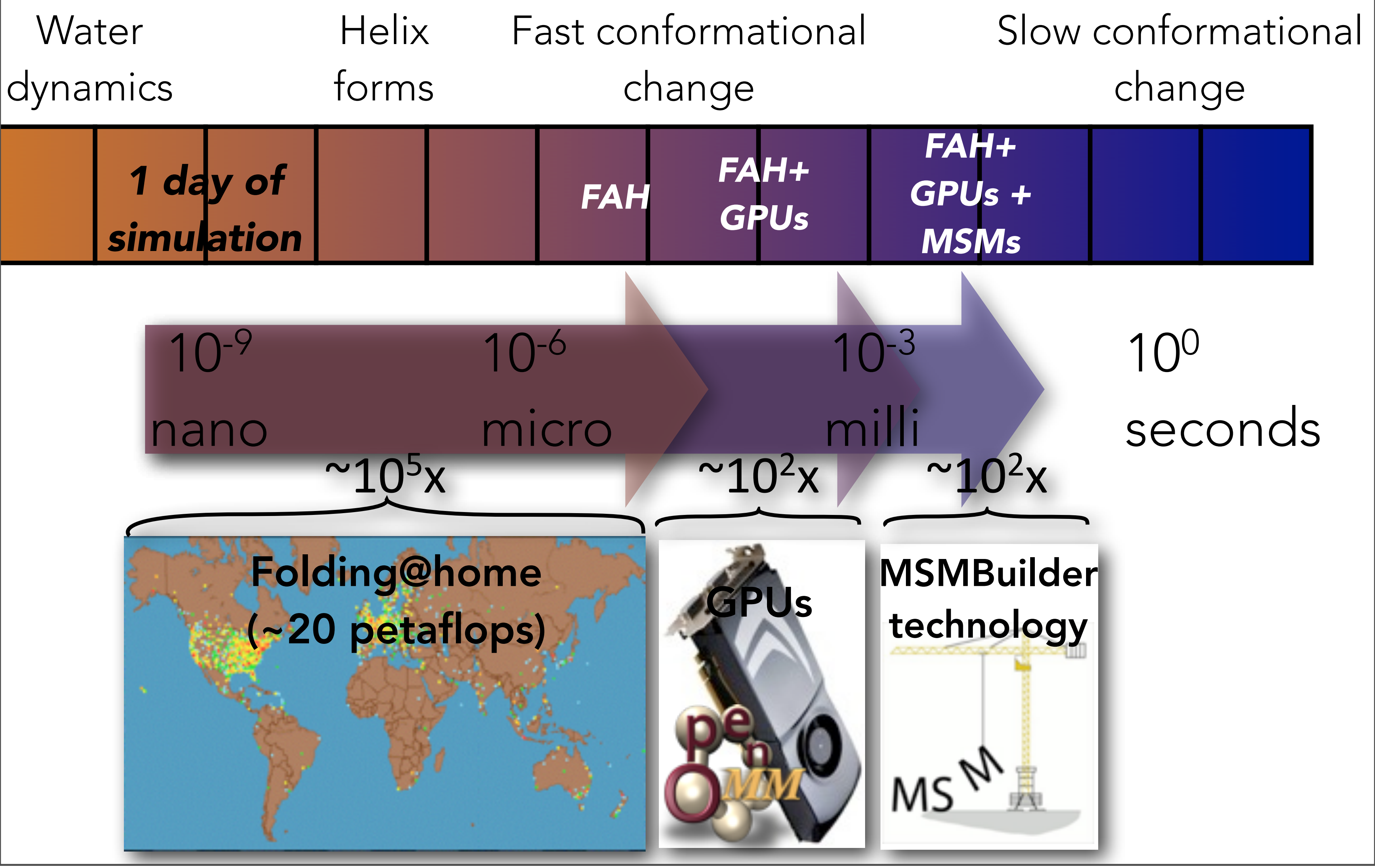
$\sim 10^2$x



Folding@home (~20 petaflops)

GPUs

The nightmare: long time scales



Water dynamics

Helix forms

Fast conformational change

Slow conformational change

1 day of simulation

FAH

FAH+ GPUs

FAH+ GPUs + MSMs

10^{-9} nano

10^{-6} micro

10^{-3} milli

10^{0} seconds

~10^{5}x

~10^{2}x

~10^{2}x

Folding@home (~20 petaflops)

GPUs

MSMBuilder technology

http://simtk.org/home/openmm

OpenMM suite of applications

Fast MD

ΔG calcs

(Chodera Lab)

NEW
Version 1.1
Mar 24 2013
平
The Chinese character is pronounced 'ping' and means 'peace' or 'balance'.

ForceBalance

(Pande Lab)

Odin

ensemble refinement

(Pande Lab)

MSM Accelerator: parallelize

MSM Builder: analyze

MSM Explorer: visualize

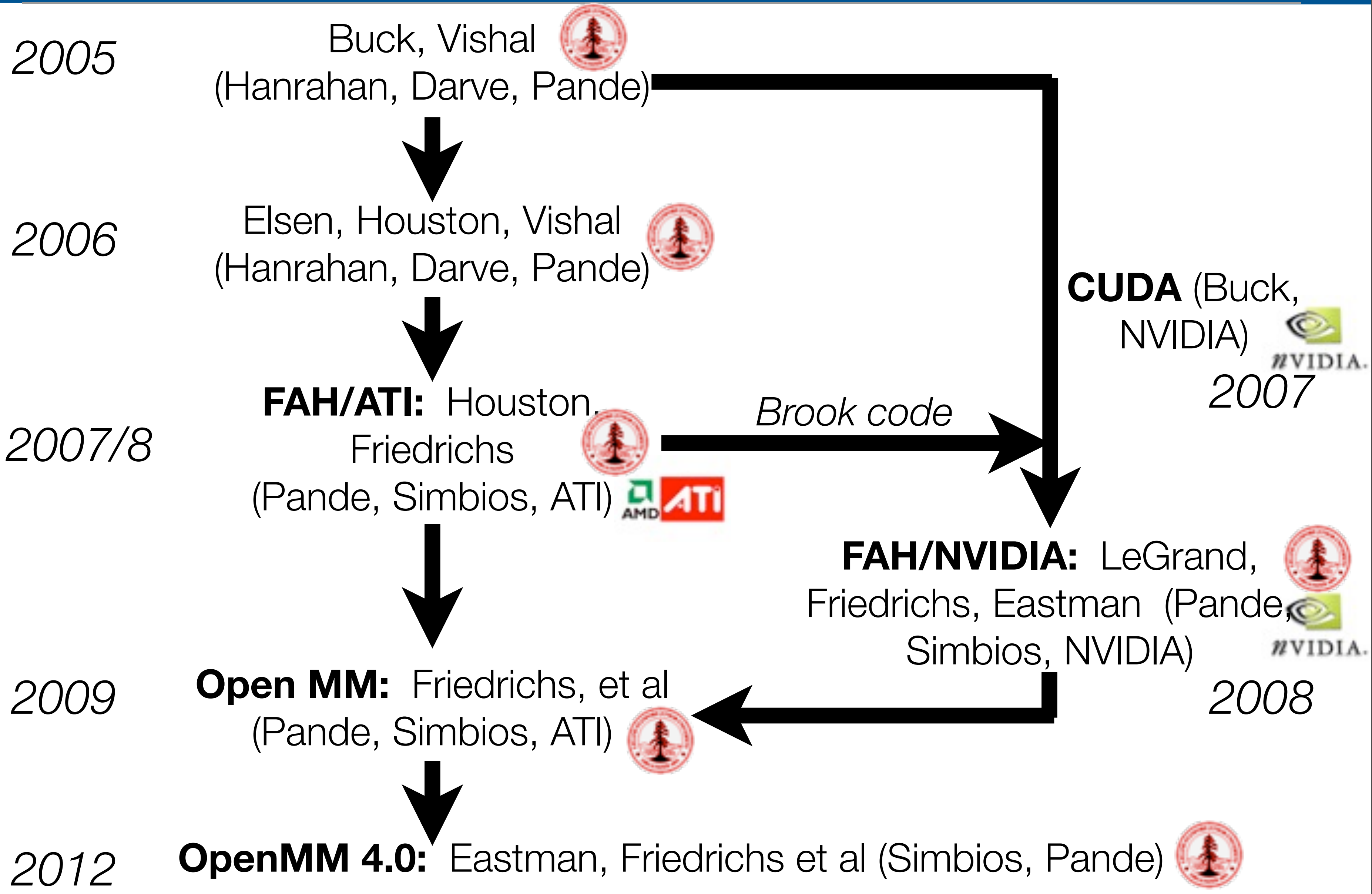
OpenMM

http://simtk.org/home/openmm

= rapid development + rapid execution

OpenMM is an app, API, and library for rapid molecular dynamics.
Easy to modify and incorporate into any code.

History of OpenMM



2005 Buck, Vishal (Hanrahan, Darve, Pande)

2006 Elsen, Houston, Vishal (Hanrahan, Darve, Pande)

2007/8 **FAH/ATI:** Houston, Friedrichs (Pande, Simbios, ATI)

Brook code

CUDA (Buck, NVIDIA) 2007

FAH/NVIDIA: LeGrand, Friedrichs, Eastman (Pande, Simbios, NVIDIA) 2008

2009 **Open MM:** Friedrichs, et al (Pande, Simbios, ATI)

2012 **OpenMM 4.0:** Eastman, Friedrichs et al (Simbios, Pande)

OpenMM: JAC benchmark

	CUDA (GTX Titan)	OpenCL (GTX Titan)	OpenCL (HD 7970)
Implicit hbonds	284	183	120
Implicit hangles	524	324	104
RF 2fs	162	124	83.5
RF 5fs	330	233	90.2
PME 2fs	104	61	49.3
PME 5fs	226	132	63.0

Joint AMBER-CHARMM DHFR Benchmark in ns/day

OpenMM roadmap

- **OpenMM 6**
  - Normal mode analysis script
  - AMOEBA OpenCL implementation
  - Constant pH implementation (JDC)
  - YANK release soon (JDC)
  - test/validate ABSINTH implicit solvent
  - More modeling tools within OpenMM app
- Further development Rosetta force field
- Triclinic boxes
- A more accurate SASA calculation for use with GB models
- Parameterize GB/VI at different temperatures
- CHARMM27 force field
- Thermodynamic ensemble validation tests
- PME for Lennard-Jones

http://wiki.simtk.org/openmm/RoadmapTimeline

Licensing and distribution

- **API & reference BSD license, GPU kernels are LGPL**
  - free & open
  - we want LGPL to have a community owned set of GPU kernels
  - we're looking for collaborations for new features

- **But, please cite us**
  - P. Eastman, M. S. Friedrichs, J. D. Chodera, R. J. Radmer, C. M. Bruns, J. P. Ku, K. A. Beauchamp, T. J. Lane, L.-P. Wang, D. Shukla, T. Tye, M. Houston, T. Stich, C. Klein, M. R. Shirts, and V. S. Pande. OpenMM 4.0: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation. *Journal of Computational and Theoretical Chemistry* **9** 461–469 (2013).

How can we simulate experimentally relevant, long timescales?

The power of
Markov State Models

Comparing two approaches

Comparing two approaches

$15M **ANTON** Specialized hardware from D.E. Shaw can compute 14μs/day

Comparing two approaches

$15M **ANTON** Specialized hardware from D.E. Shaw can compute 14μs/day

$0.3M GPU cluster + **OpenMM +MSMB** can also compute 14μs/day at ~1/50th the cost

Comparing two approaches

50x more powerful =
50x less expensive

$15M **ANTON** Specialized hardware from D.E. Shaw can compute 14µs/day

$0.3M GPU cluster + **OpenMM +MSMB** can also compute 14µs/day at ~1/50th the cost

Comparing two approaches

OpenMM: Over 100ns/day on 24,000 atom JAC

$15M **ANTON** Specialized hardware from D.E. Shaw can compute 14μs/day

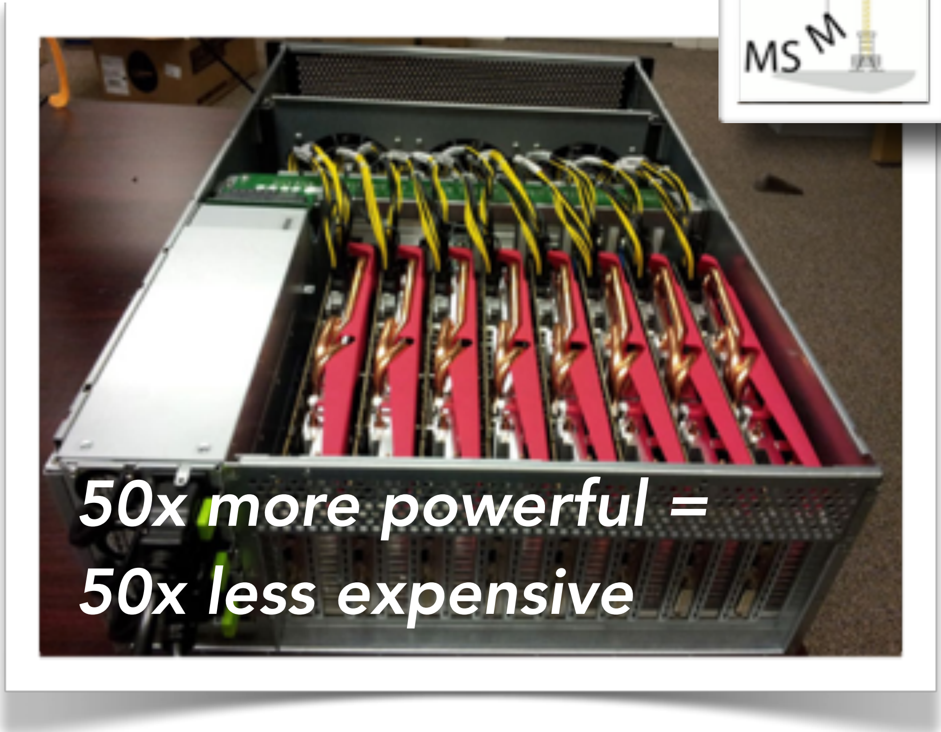
50x more powerful =
50x less expensive

$0.3M GPU cluster + **OpenMM +MSMB** can also compute 14μs/day at ~1/50th the cost

Comparing two approaches

OpenMM: Over 100ns/day on 24,000 atom JAC

$15M **ANTON** Specialized hardware from D.E. Shaw can compute 14μs/day

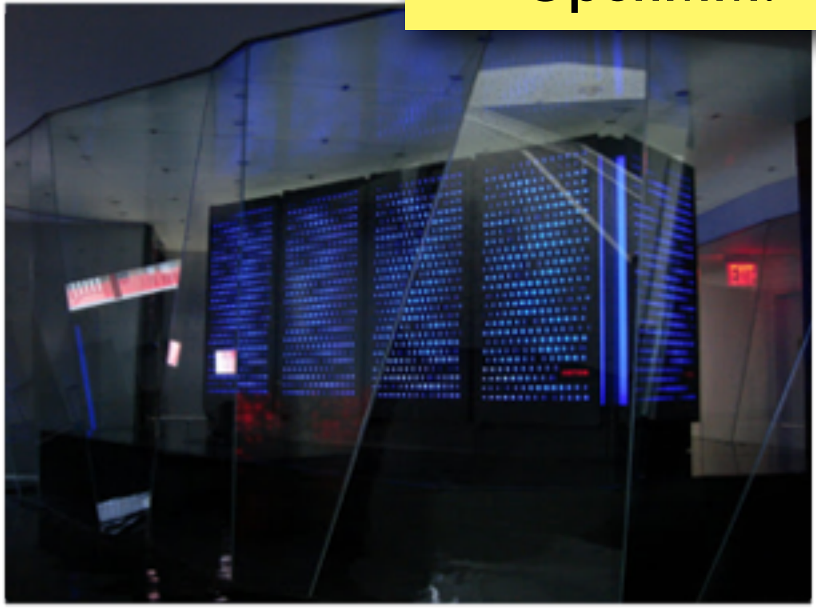

50x more powerful = 50x less expensive

$0.3M GPU cluster + **OpenMM +MSMB** can also compute 14μs/day at ~1/50th the cost

Comparing two approaches

OpenMM: Over 100ns/day on 24,000 atom JAC

MSM Builder: http://msmbuilder.org
OpenMM: http://openmm.org



50x more powerful =
50x less expensive

$15M **ANTON** Specialized hardware from D.E. Shaw can compute 14µs/day

$0.3M GPU cluster + **OpenMM** **+MSMB** can also compute 14µs/day at ~1/50th the cost

Our Goals

- **Build a model which can predict everything**
  - kinetics, thermodynamics, structure
- **Build a model which can yield powerful visualizations**
  - movies of key phenomena
- **Broad applicability**
  - works on many systems
  - easy to use, easily automated

Comparison to other methods

Comparison to other methods

Popular methods

Accelerated MD

Anton

MSM

Metadynamics

Milestoning

Path-based methods

Replica Exchange

Comparison to other methods

Popular methods

Accelerated MD

Anton

MSM

Metadynamics

Milestoning

Path-based methods

Replica Exchange

Our goals

Comparison to other methods

Popular methods

Accelerated MD   Anton   MSM   Metadynamics

Milestoning   Path-based methods   Replica Exchange

Our goals

(1) Fast sampling, including orthogonal degrees of freedom

Comparison to other methods

Popular methods

Accelerated MD

Anton

MSM

~~Metadynamics~~

Milestoning

Path-based methods

Replica Exchange

Our goals

(1) Fast sampling, including orthogonal degrees of freedom

Comparison to other methods

Popular methods

Accelerated MD   Anton   MSM   ~~Metadynamics~~

Milestoning   Path-based methods   Replica Exchange

Our goals

(1) Fast sampling, including orthogonal degrees of freedom

(2) Can discover end points

Comparison to other methods

Popular methods

Accelerated MD

Anton

MSM

~~Metadynamics~~

~~Milestoning~~

Path-based methods

Replica Exchange

Our goals

(1) Fast sampling, including orthogonal degrees of freedom

(2) Can discover end points

Comparison to other methods

Popular methods

Accelerated MD

Anton

MSM

~~Metadynamics~~

~~Milestoning~~

~~Path-based methods~~

Replica Exchange

Our goals

(1) Fast sampling, including orthogonal degrees of freedom

(2) Can discover end points

Comparison to other methods

Popular methods

Accelerated MD Anton MSM ~~Metadynamics~~

~~Milestoning~~ ~~Path-based methods~~ Replica Exchange

Our goals

(1) Fast sampling, including orthogonal degrees of freedom

(2) Can discover end points

(3) Can predict kinetics (& thermodynamics, & structure)

Comparison to other methods

Popular methods

~~Accelerated MD~~ Anton MSM ~~Metadynamics~~

~~Milestoning~~ ~~Path-based methods~~ Replica Exchange

Our goals

(1) Fast sampling, including orthogonal degrees of freedom

(2) Can discover end points

(3) Can predict kinetics (& thermodynamics, & structure)

Comparison to other methods

Popular methods

~~Accelerated MD~~ Anton MSM ~~Metadynamics~~

~~Milestoning~~ ~~Path-based methods~~ ~~Replica Exchange~~

Our goals

(1) Fast sampling, including orthogonal degrees of freedom

(2) Can discover end points

(3) Can predict kinetics (& thermodynamics, & structure)

What are Markov State Models (MSMs)?

MSMs **automatically** build a **Master Equation** with MD simulation, typically with **many short (~μs) trajectories**

$$\frac{dp_i}{dt} = \sum_l [k_{l,i} p_l - k_{i,l} p_i]$$

What are Markov State Models (MSMs)?

MSMs **automatically** build a **Master Equation** with MD simulation, typically with **many short (~μs) trajectories**

$$\frac{dp_i}{dt} = \sum_l \left[k_{l,i} p_l - k_{i,l} p_i \right]$$

probability of being in state *i*

What are Markov State Models (MSMs)?

MSMs **automatically** build a **Master Equation** with MD simulation, typically with **many short (~μs) trajectories**

$$\frac{dp_i}{dt} = \sum_l \left[k_{l,i}p_l - k_{i,l}p_i\right]$$

probability of being in state *i*

rate of change between states

What are Markov State Models (MSMs)?

MSMs **automatically** build a **Master Equation** with MD simulation, typically with **many short (~μs) trajectories**

$$\frac{dp_i}{dt} = \sum_l \left[k_{l,i} p_l - k_{i,l} p_i\right]$$

probability of being in state *i*

rate of change between states

with the goals of:

(1) aiding simulators **reach long timescales** and

(2) **gaining novel insight** from their simulations

What are Markov State Models (MSMs)?

MSMs **automatically** build a **Master Equation** with MD simulation, typically with **many short (~µs) trajectories**

$$\frac{dp_i}{dt} = \sum_l \left[k_{l,i}p_l - k_{i,l}p_i\right]$$

probability of being in state *i*

rate of change between states

with the goals of:

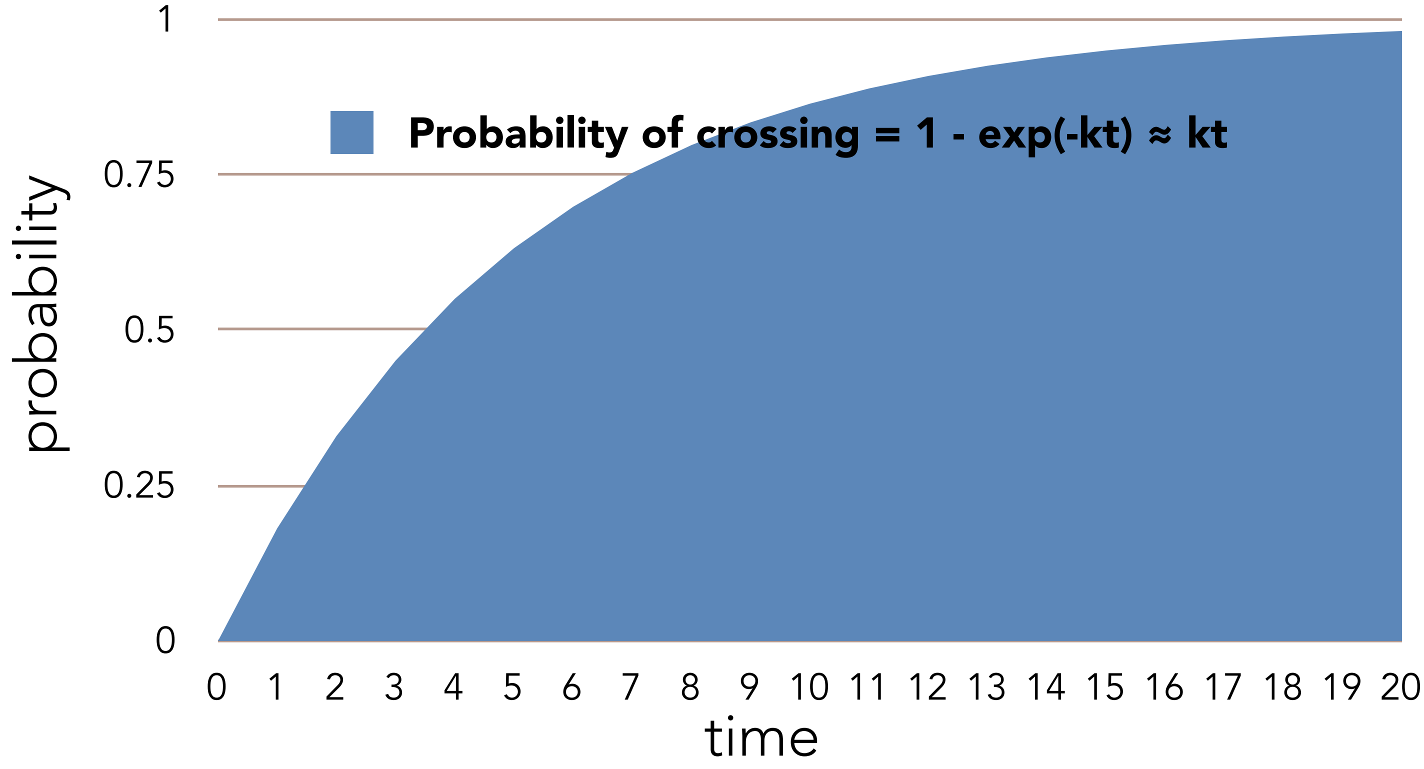
(1) aiding simulators **reach long timescales** and
(2) **gaining novel insight** from their simulations

see the work of: Andersen, Best, Bowman, Caflisch, Chodera, Deuflhard, Dill, Grubmüller, Huang, Hummer, Levy, Noé, Pande, Pitera, Roux, Schütte, Swope, Weber

Short trajectories vs long timescales?

Two state (Single Barrier) Case

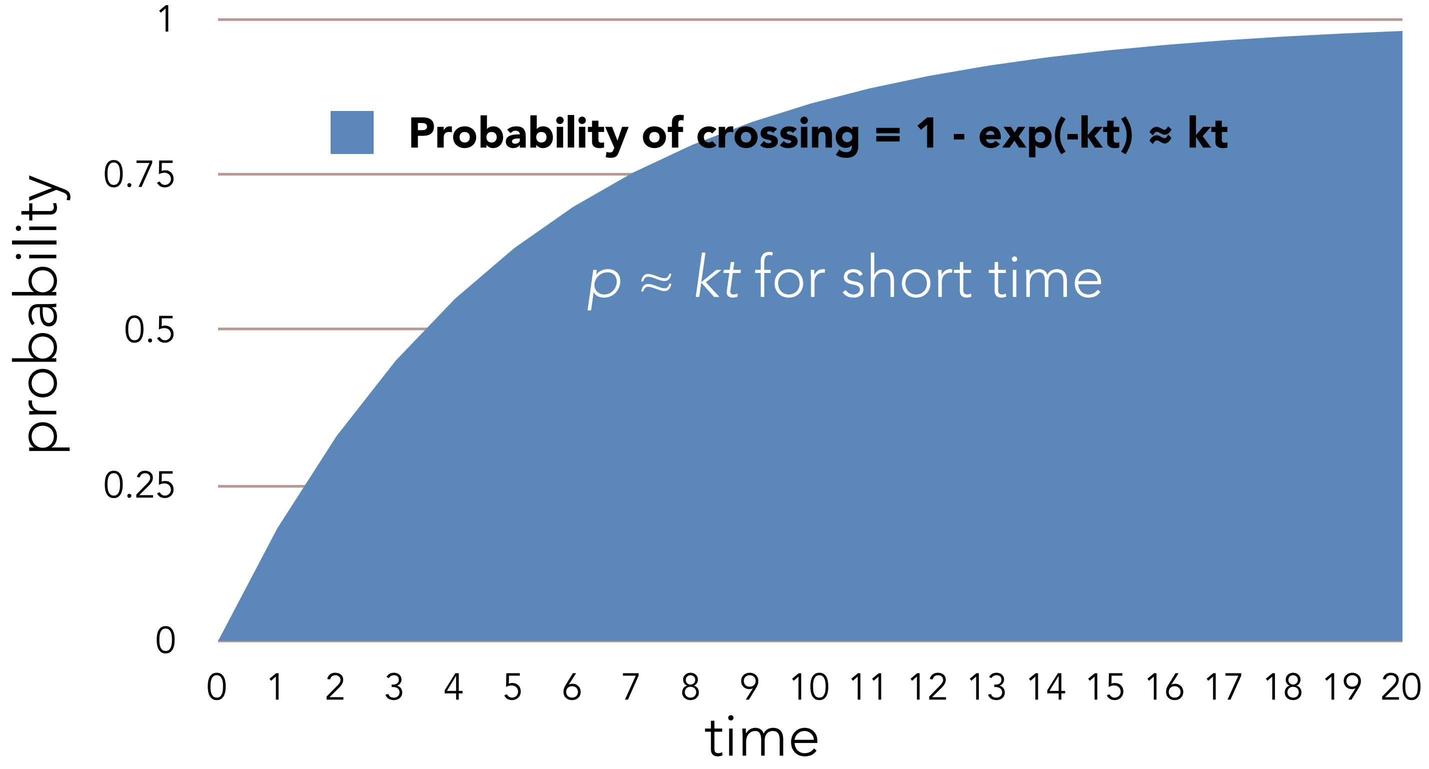
$$A \xrightarrow{k} B$$



Short trajectories vs long timescales?

Two state (Single Barrier) Case

$$A \xrightarrow{k} B$$

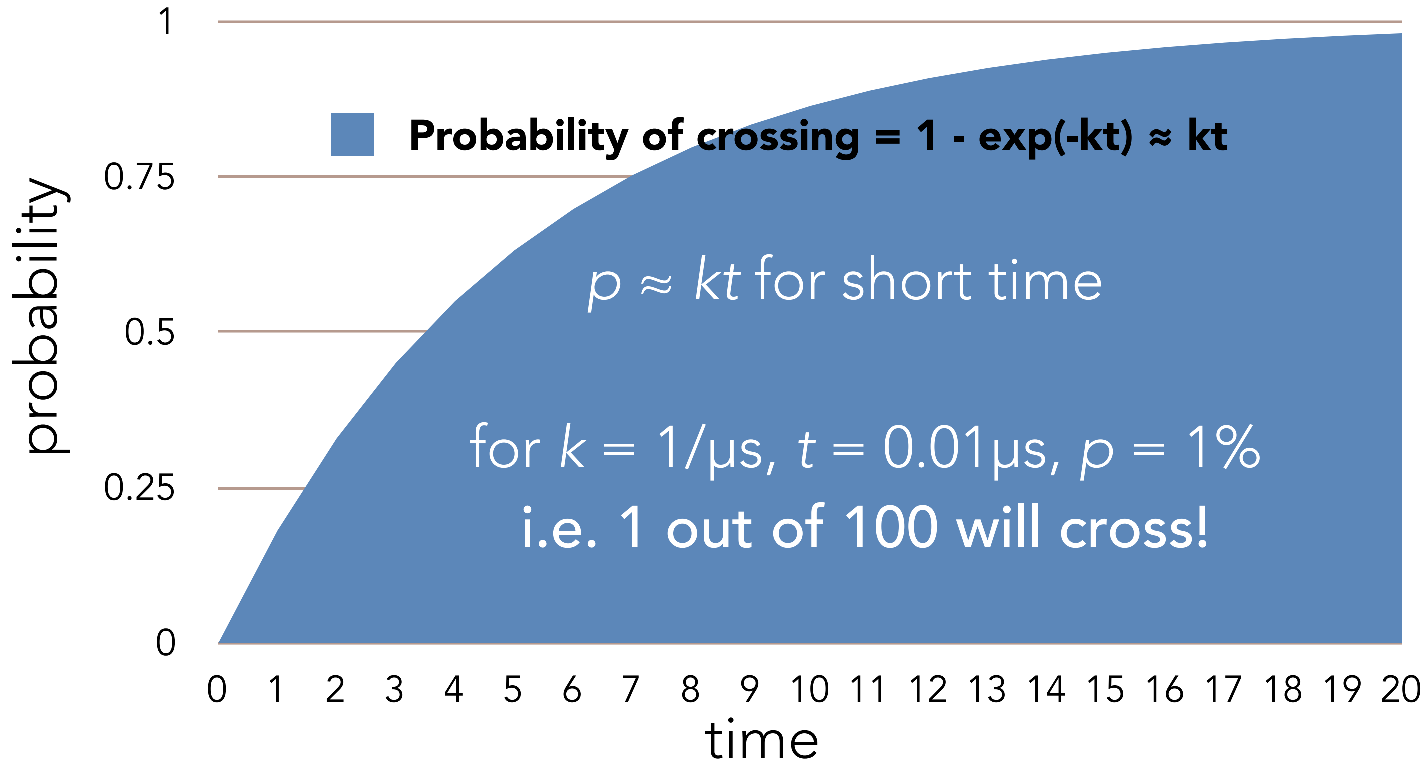


$p \approx kt$ for short time

Short trajectories vs long timescales?

Two state (Single Barrier) Case

$$A \xrightarrow{k} B$$



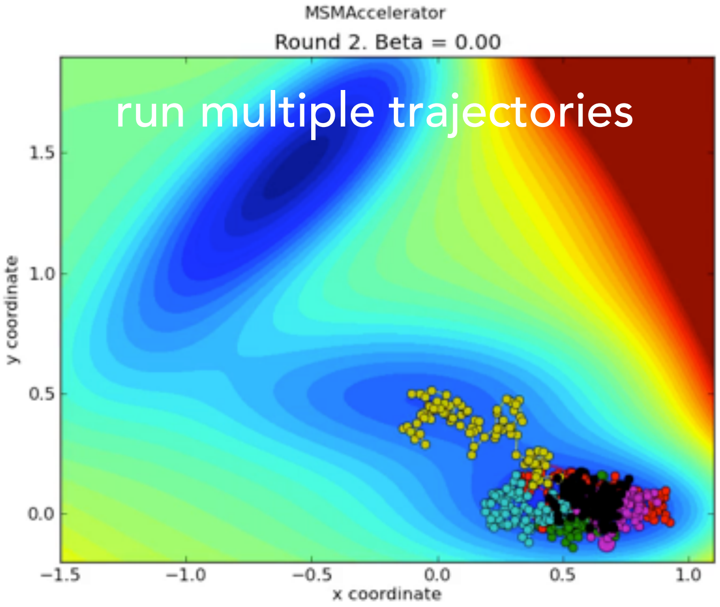
$p \approx kt$ for short time

for $k = 1/\mu s$, $t = 0.01\mu s$, $p = 1\%$

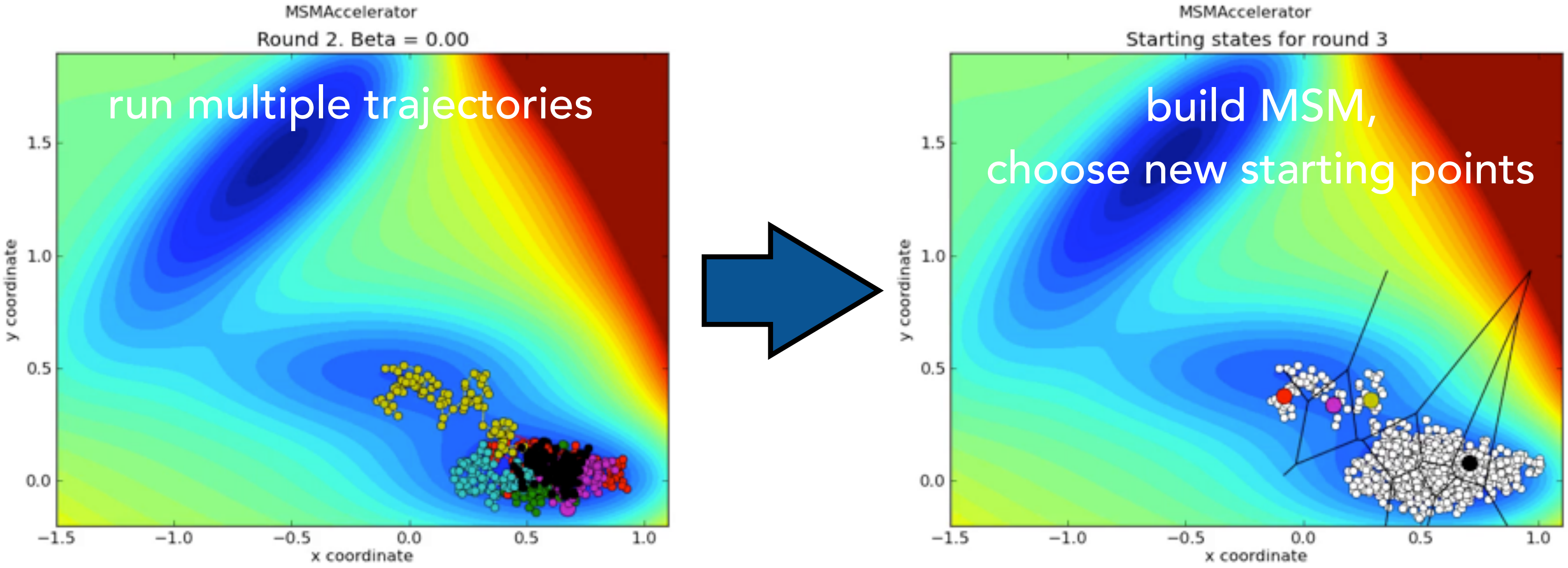
i.e. 1 out of 100 will cross!

Key stages in MSM construction

Key stages in MSM construction



Key stages in MSM construction

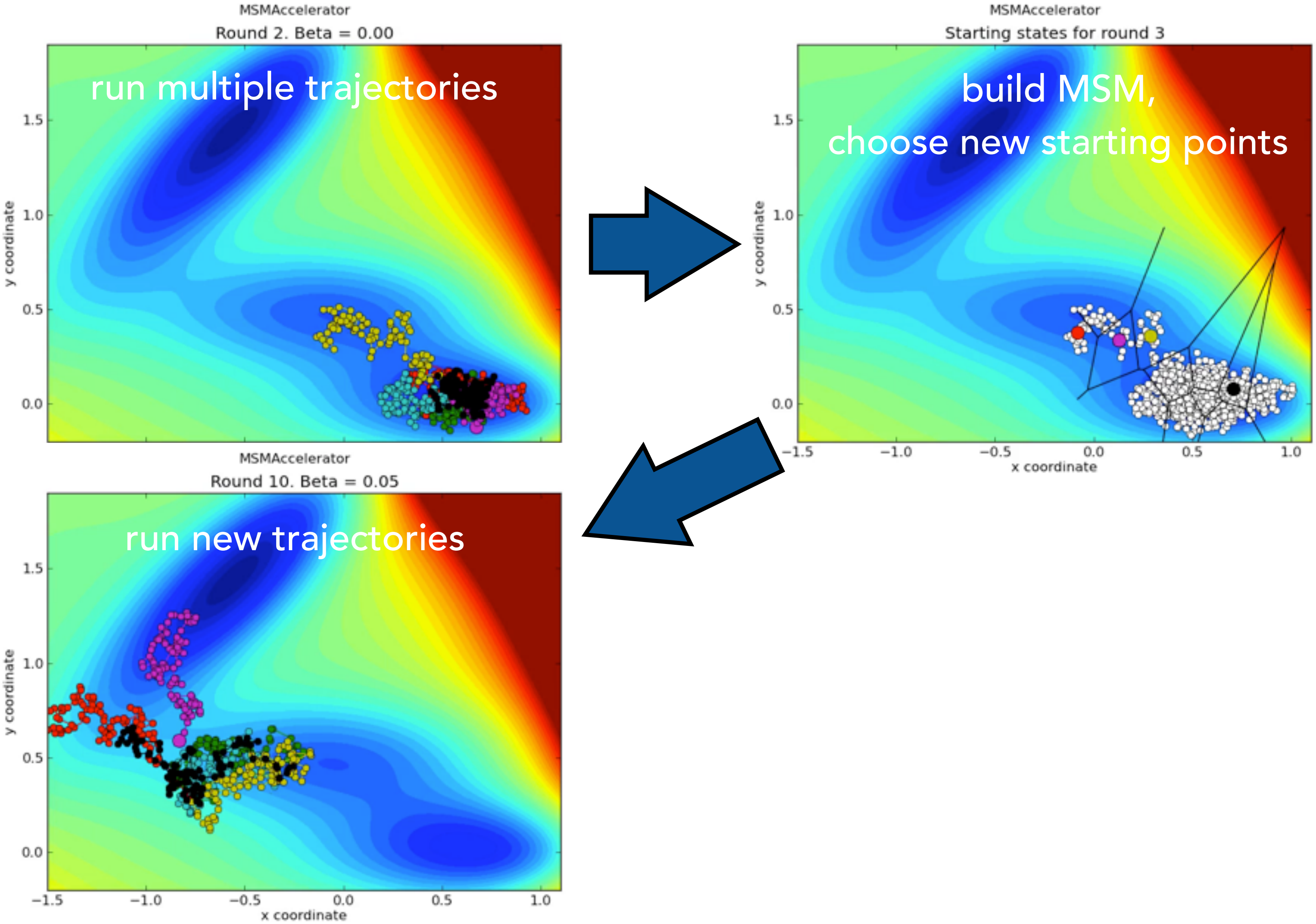


run multiple trajectories

build MSM,

choose new starting points

Key stages in MSM construction



MSMAccelerator

Round 2. Beta = 0.00

run multiple trajectories

MSMAccelerator

Starting states for round 3

build MSM,

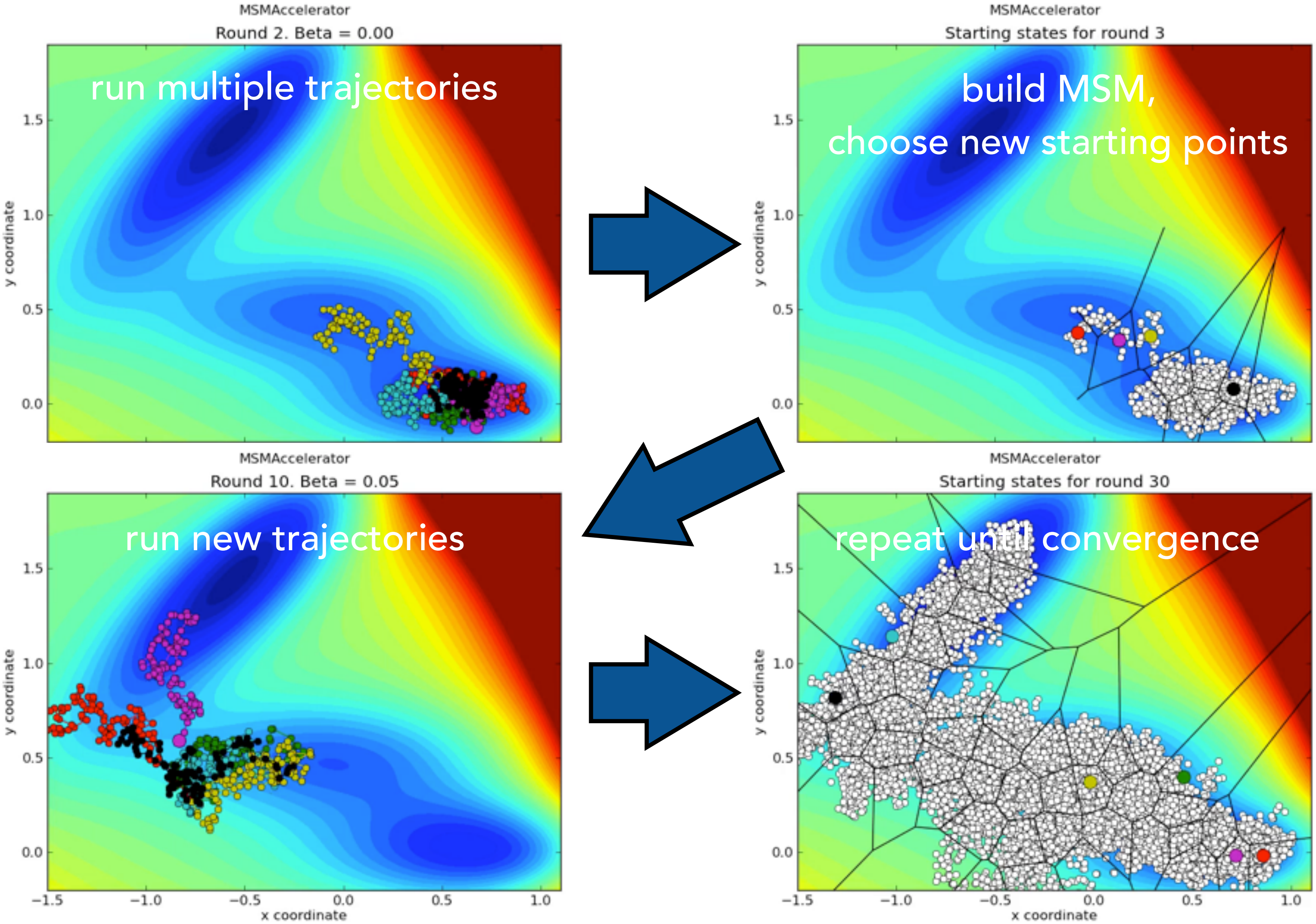
choose new starting points

MSMAccelerator

Round 10. Beta = 0.05

run new trajectories

Key stages in MSM construction



run multiple trajectories

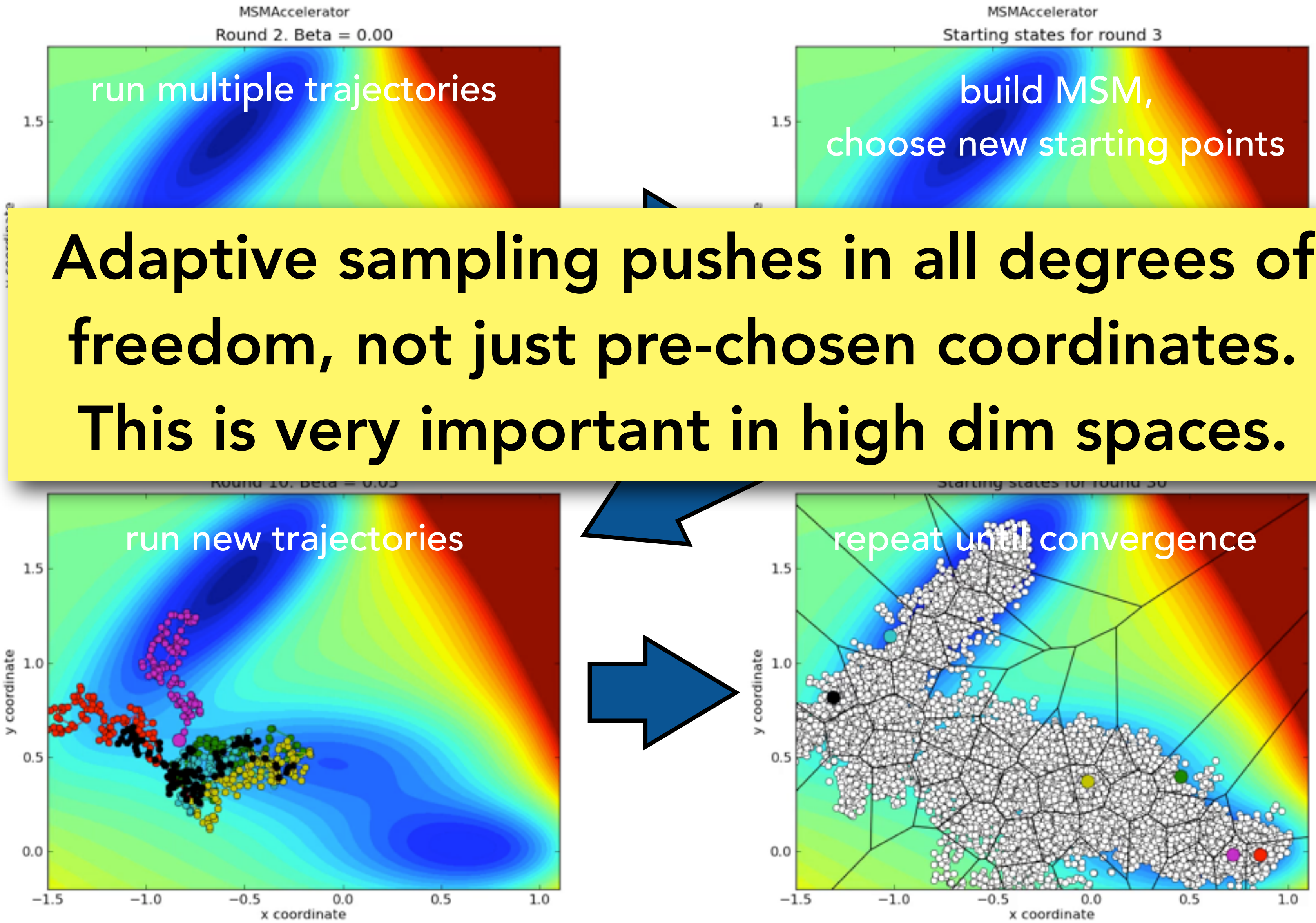
build MSM,

choose new starting points

run new trajectories

repeat until convergence

Key stages in MSM construction



run multiple trajectories

build MSM,
choose new starting points

run new trajectories

repeat until convergence

Adaptive sampling pushes in all degrees of freedom, not just pre-chosen coordinates. This is very important in high dim spaces.

MSM vs long trajectory

(McGibbon, Kiss, Harrigan, Lane, VSP)

(movie by Harrigan, McGibbon)

MSM Adaptive Sampling

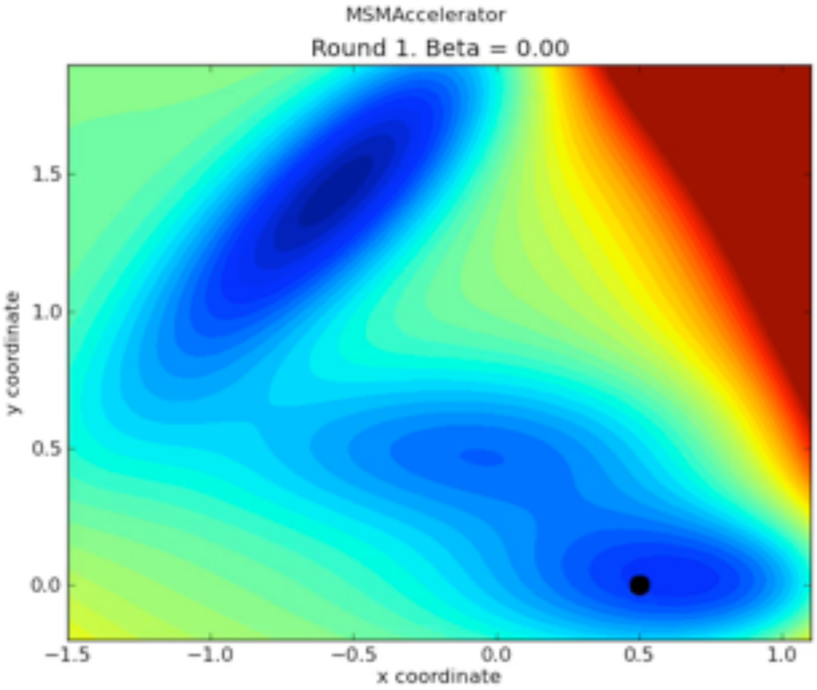
Single long trajectory

MSM vs long trajectory

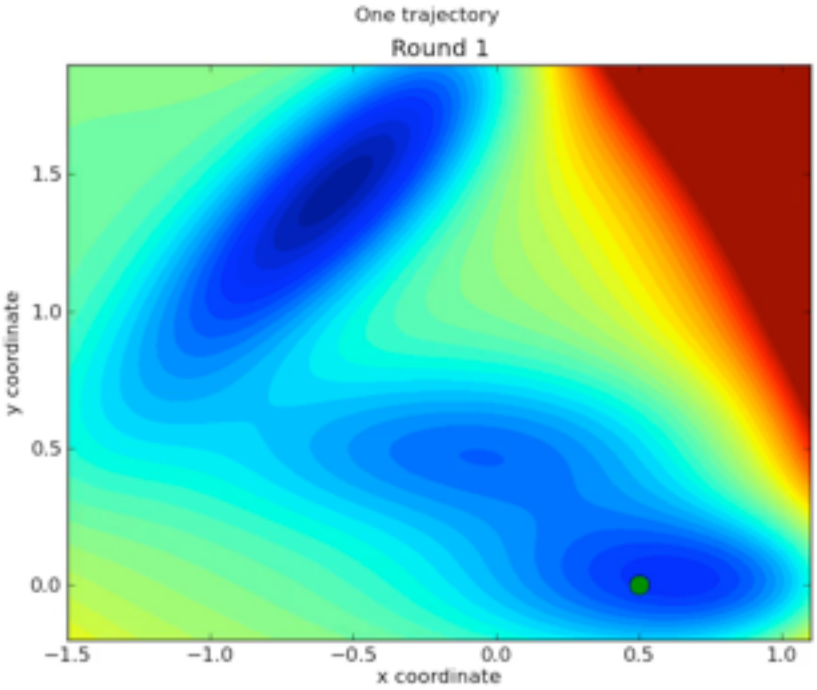
(McGibbon, Kiss, Harrigan, Lane, VSP)

(movie by Harrigan, McGibbon)

MSM Adaptive Sampling



Single long trajectory

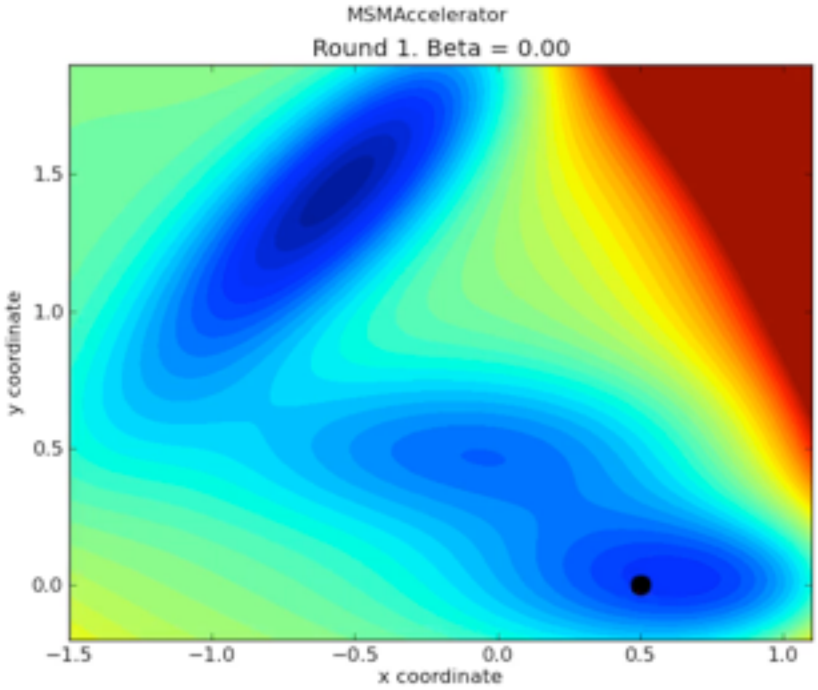


MSM vs long trajectory

(McGibbon, Kiss, Harrigan, Lane, VSP)

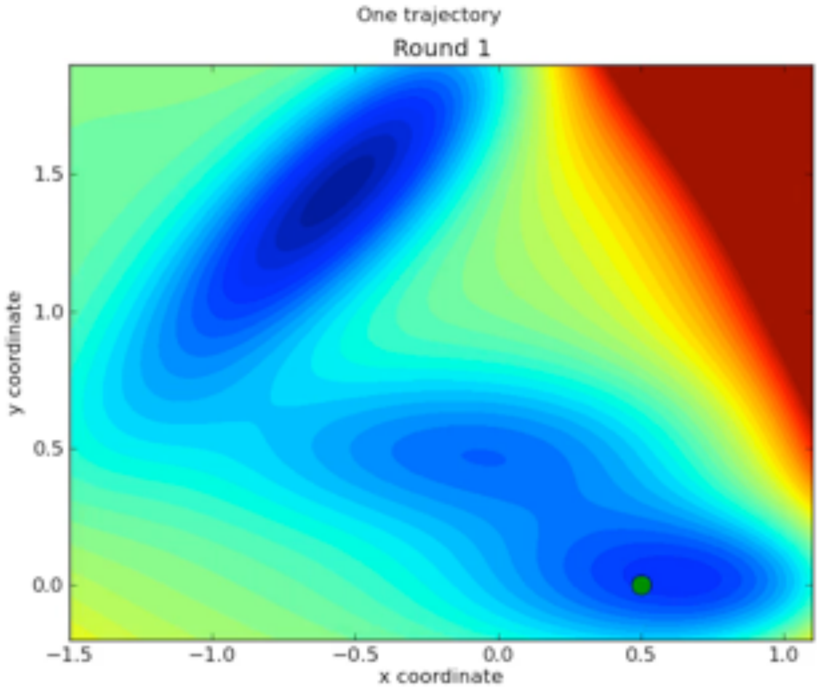
(movie by Harrigan, McGibbon)

MSM Adaptive Sampling



efficient ⊥ sampling, trivial to parallelize

Single long trajectory

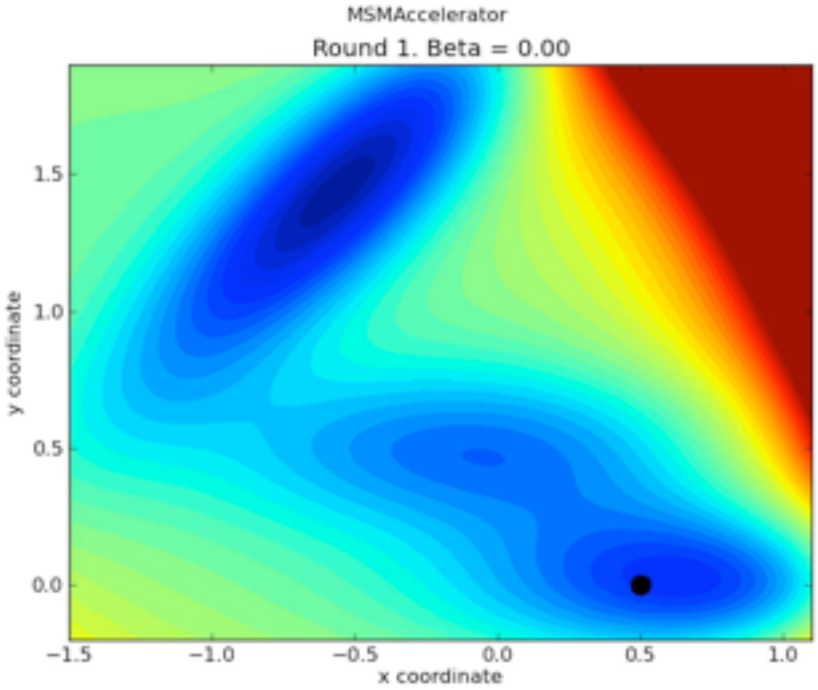


MSM vs long trajectory

(McGibbon, Kiss, Harrigan, Lane, VSP)

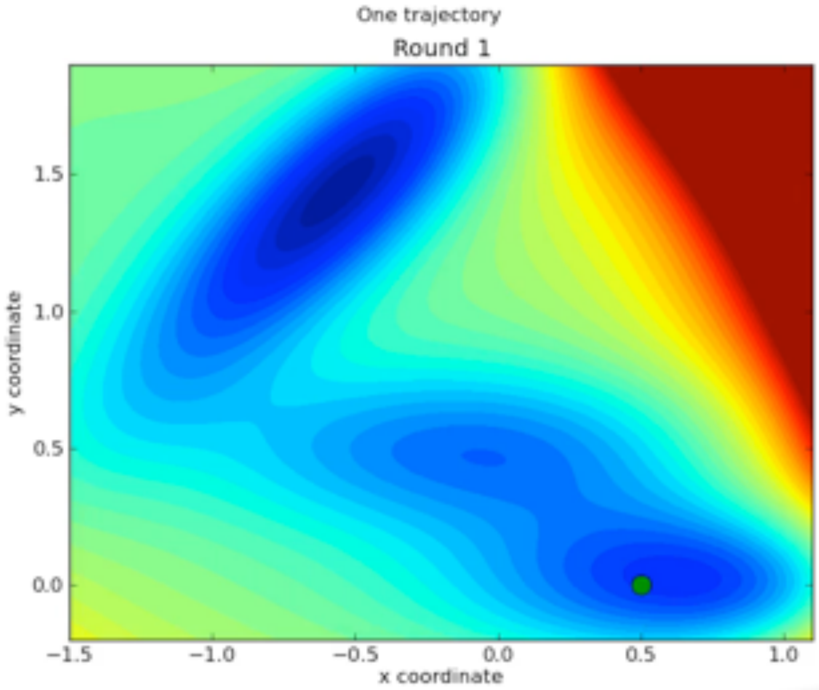
(movie by Harrigan, McGibbon)

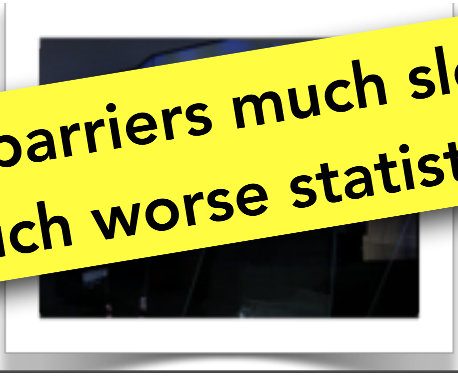
MSM Adaptive Sampling



efficient ⊥ sampling, trivial to parallelize

Single long trajectory

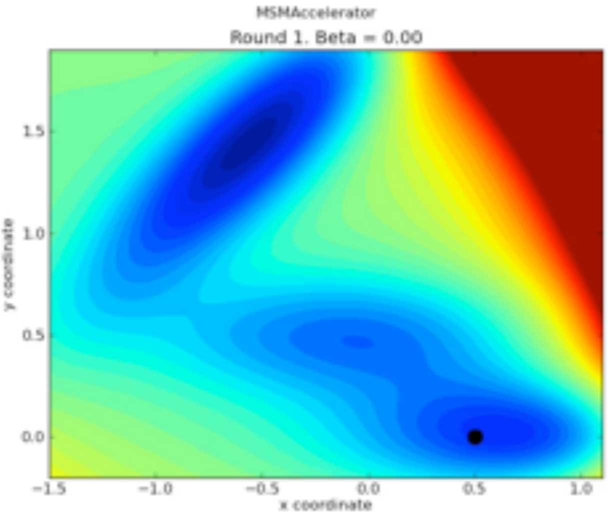




cross barriers much slower, much worse statistics

Comparison to other methods

- **aMD**
  - removes kinetic information
  - speeds on certain degrees of freedom — must know which ones are slow

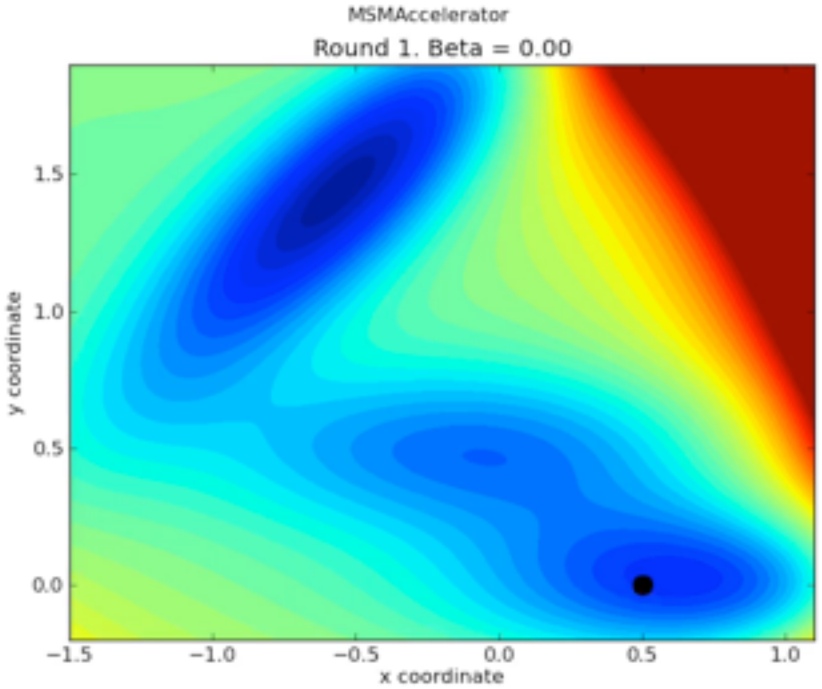


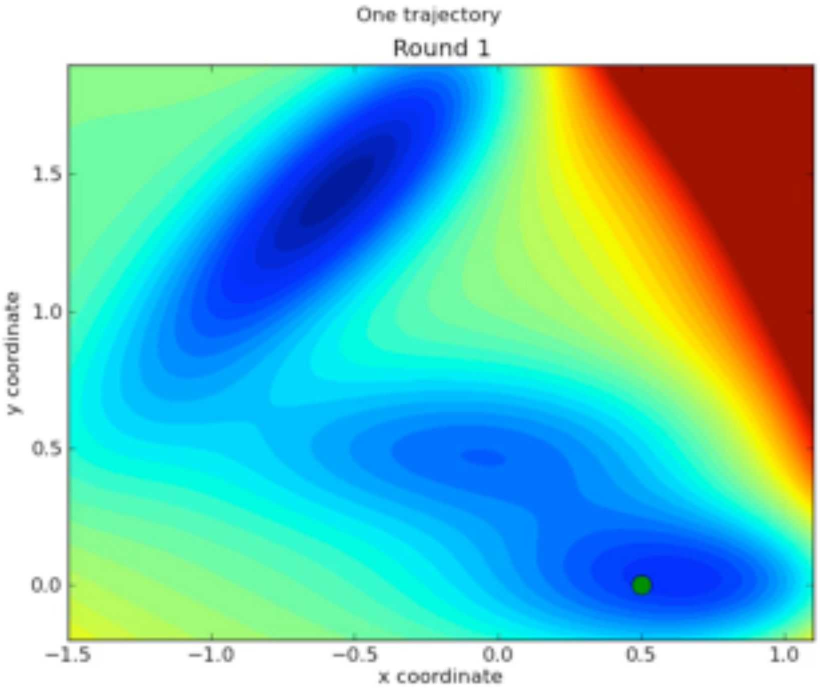
- **Metadynamics**
  - removes kinetic information
  - drives on pre-chosen degrees of freedom, misses key challenge of how to sample orthogonal dofs

- **Replica Exchange**
  - removes kinetic information
  - works best for energy barriers, not ΔG barriers

Comparison to other methods

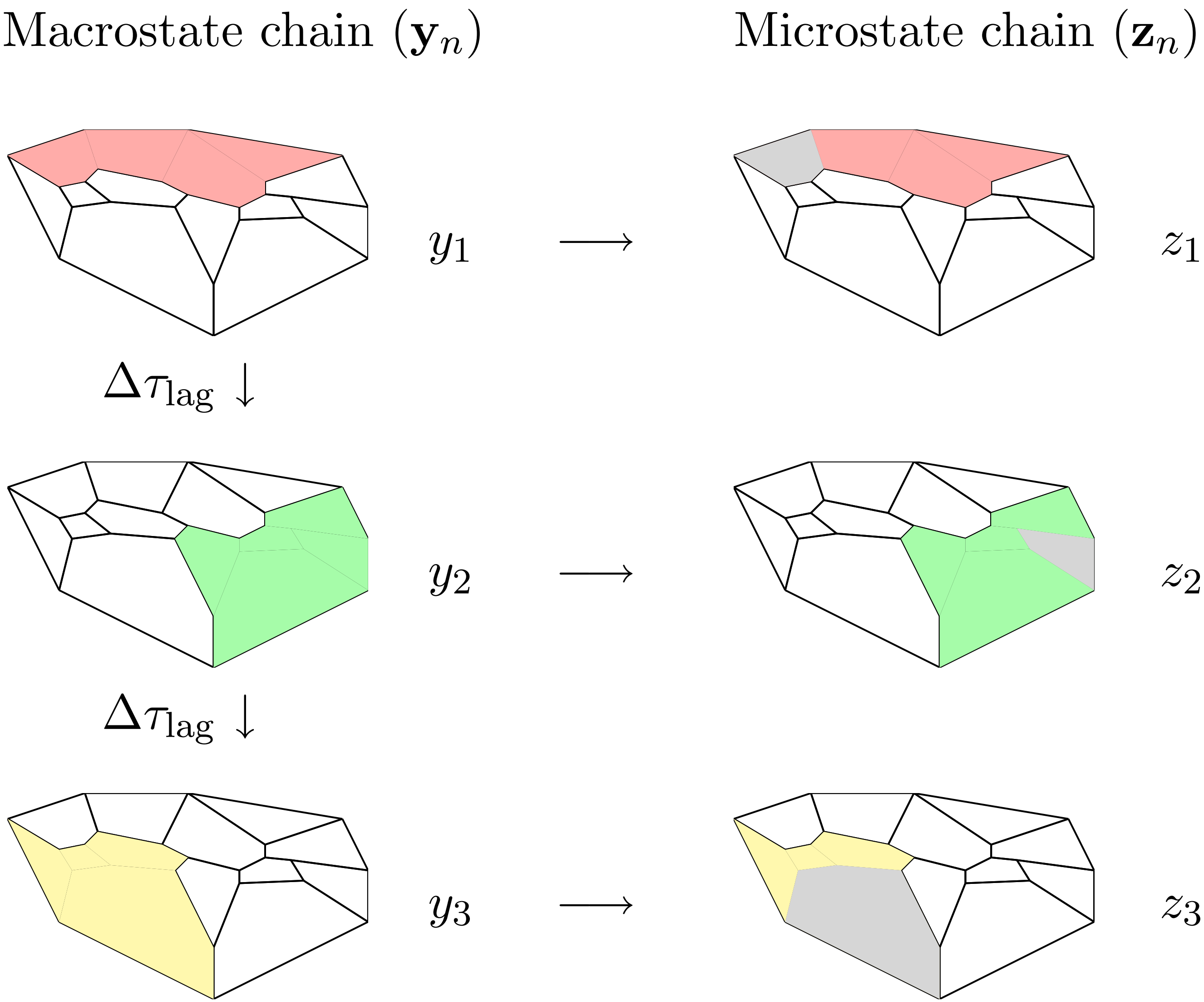
- **Highly parallel MD**
  - still requires the kinetic analysis.
  - many short trajectories are MUCH more efficient
  - very expensive (50x) given throughput: GPU cluster better at many short trajectories





(Bacallado, Chodera, VSP)

Making sense of MSMs: lumping



Macrostate chain $(\mathbf{y}_n)$

Microstate chain $(\mathbf{z}_n)$

$y_1 \longrightarrow z_1$

$\Delta\tau_{\text{lag}} \downarrow$

$y_2 \longrightarrow z_2$

$\Delta\tau_{\text{lag}} \downarrow$

$y_3 \longrightarrow z_3$

(Bacallado, Chodera, VSP)

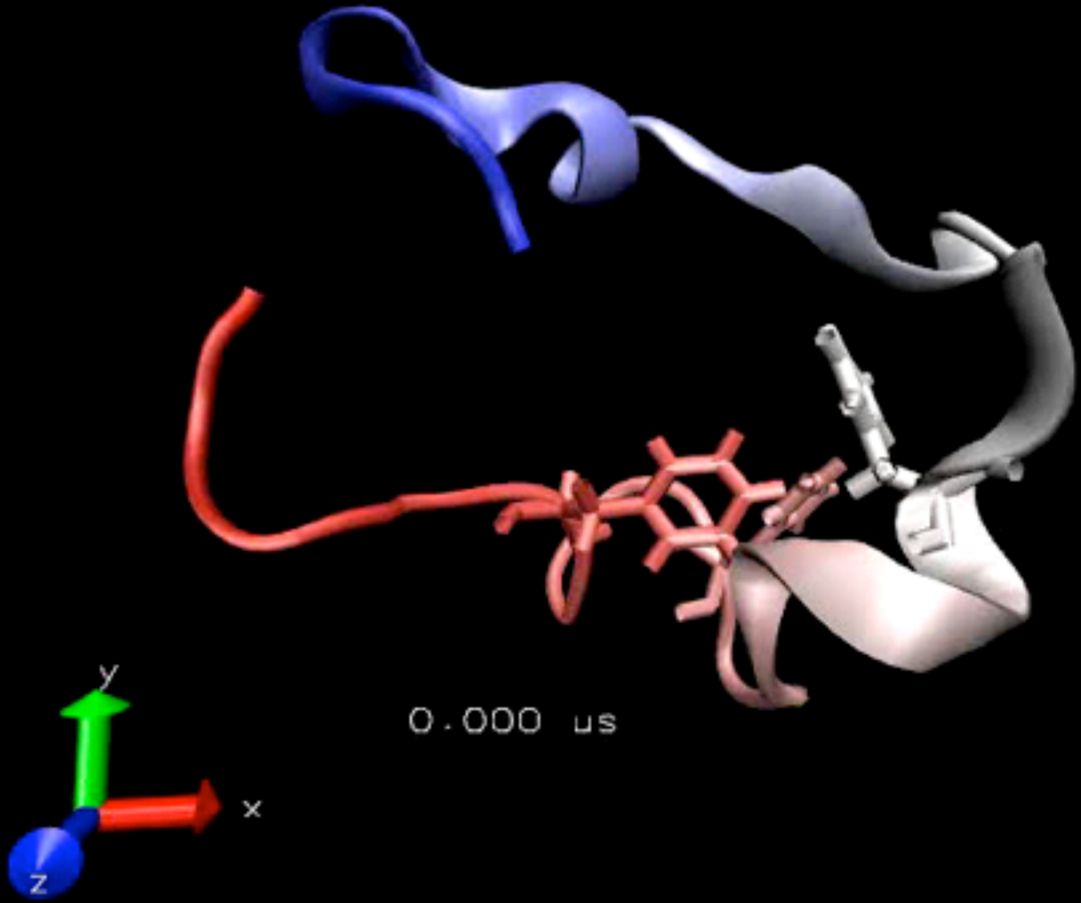
Formalization of lumping

- The model in this case is the lumping $M : Z \rightarrow Y$, a mapping from microstates to macrostates.
- The model is parametrized by the transition probability matrix T, and the local equilibrium distributions for the microstates Θ.
- We can factorize the evidence into two factors:

$$\begin{aligned}
P(\mathbf{z}_n|M) &= \int dT d\Theta P(\mathbf{z}_n|T,\Theta,M)P(T,\Theta|M) \\
&= \int dT d\Theta P(\mathbf{y}_n|T,M)P(T|M)P(\mathbf{z}_n|\mathbf{y}_n,\Theta,M)P(\Theta|M) \\
&= \underbrace{\int dT P(\mathbf{y}_n|T,M)P(T|M)}_{\text{Macrostate Markov chain}} \times \\
&\qquad \underbrace{\int d\Theta P(\mathbf{z}_n|\mathbf{y}_n,\Theta,M)P(\Theta|M)}_{\text{Microstates from equilibrium within macrostates}}
\end{aligned}$$

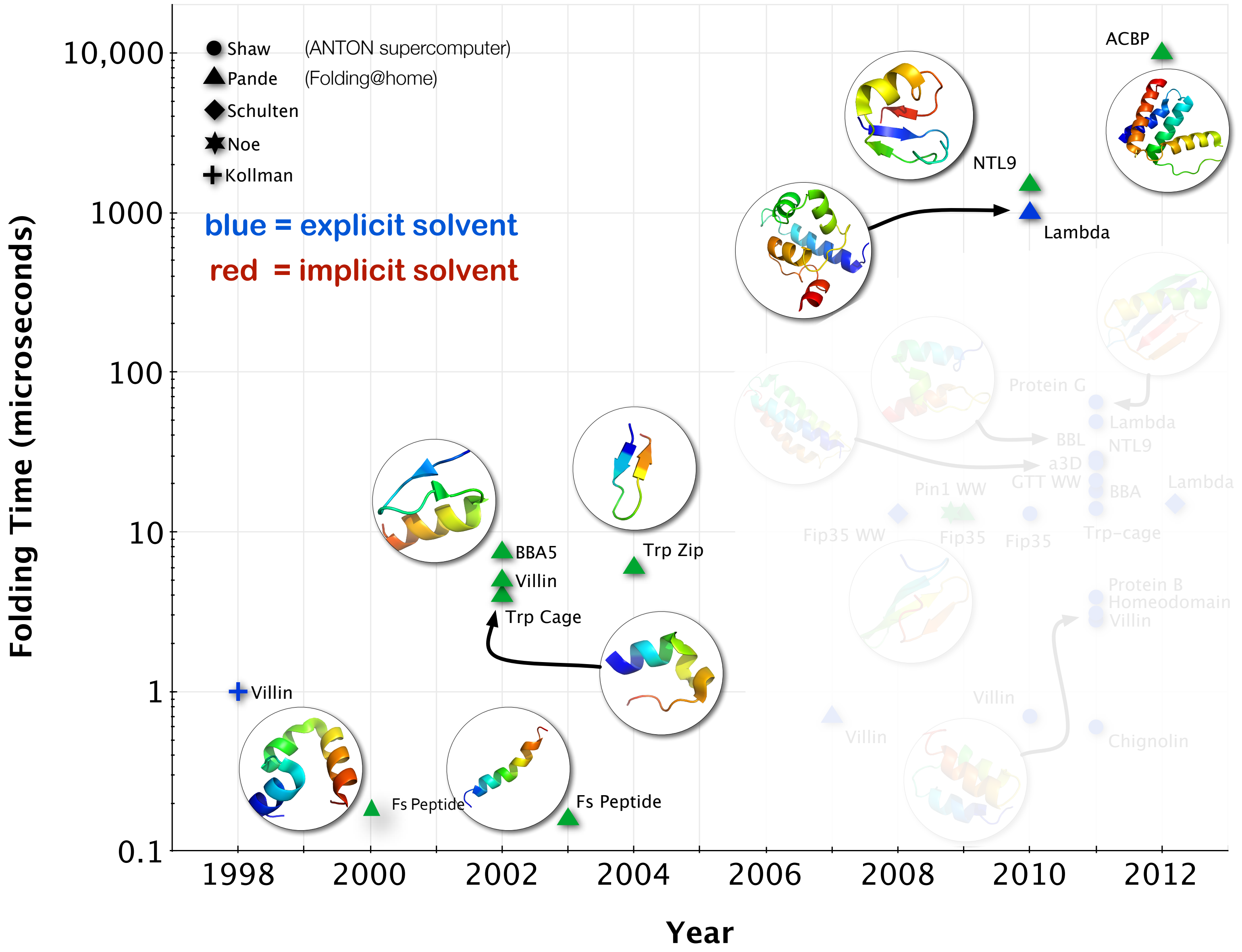
What can MSMs do?

MSMs reach long timescales

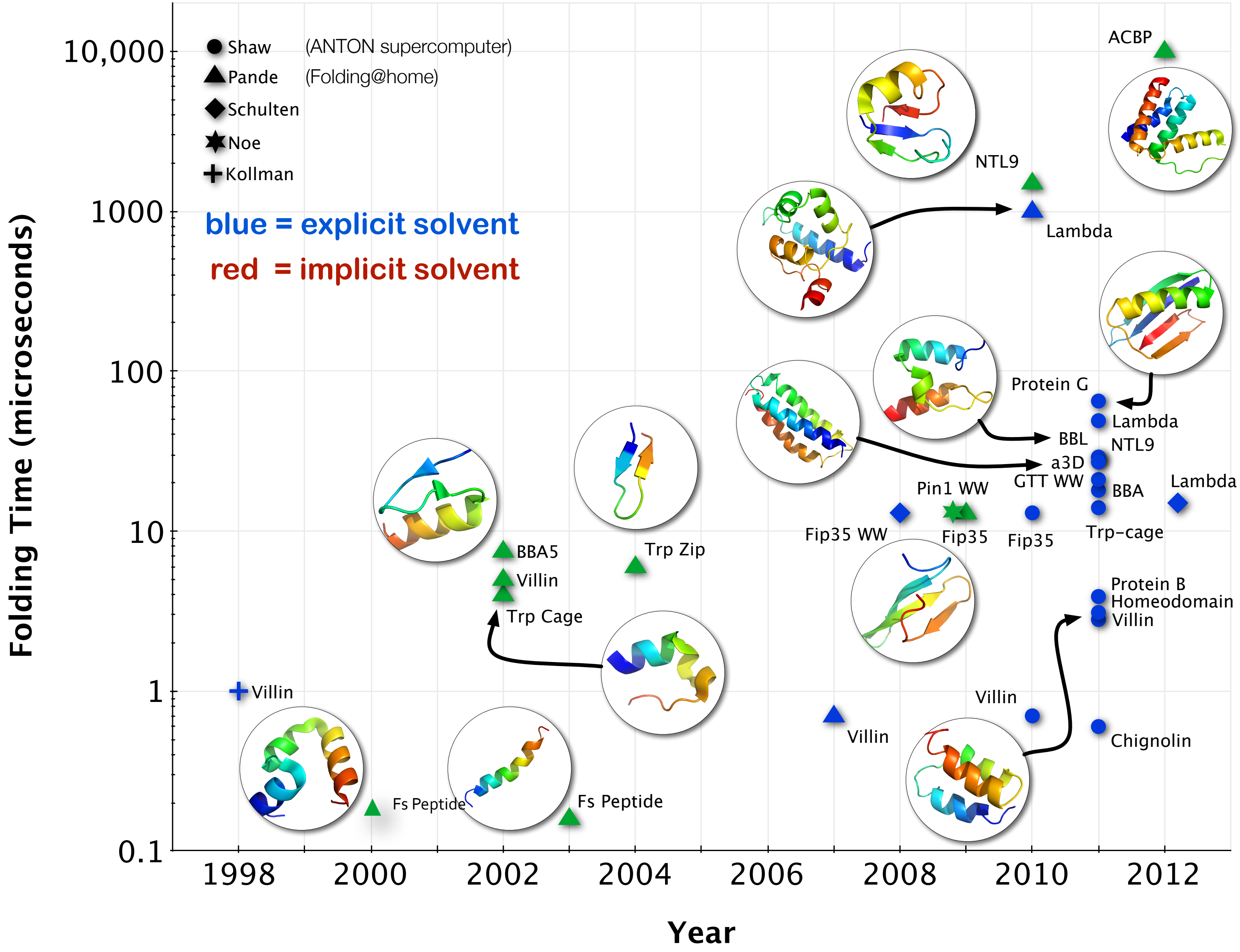


Copernicus: A new paradigm for parallel adaptive molecular dynamics. *Supercomputing 2011* (2011)

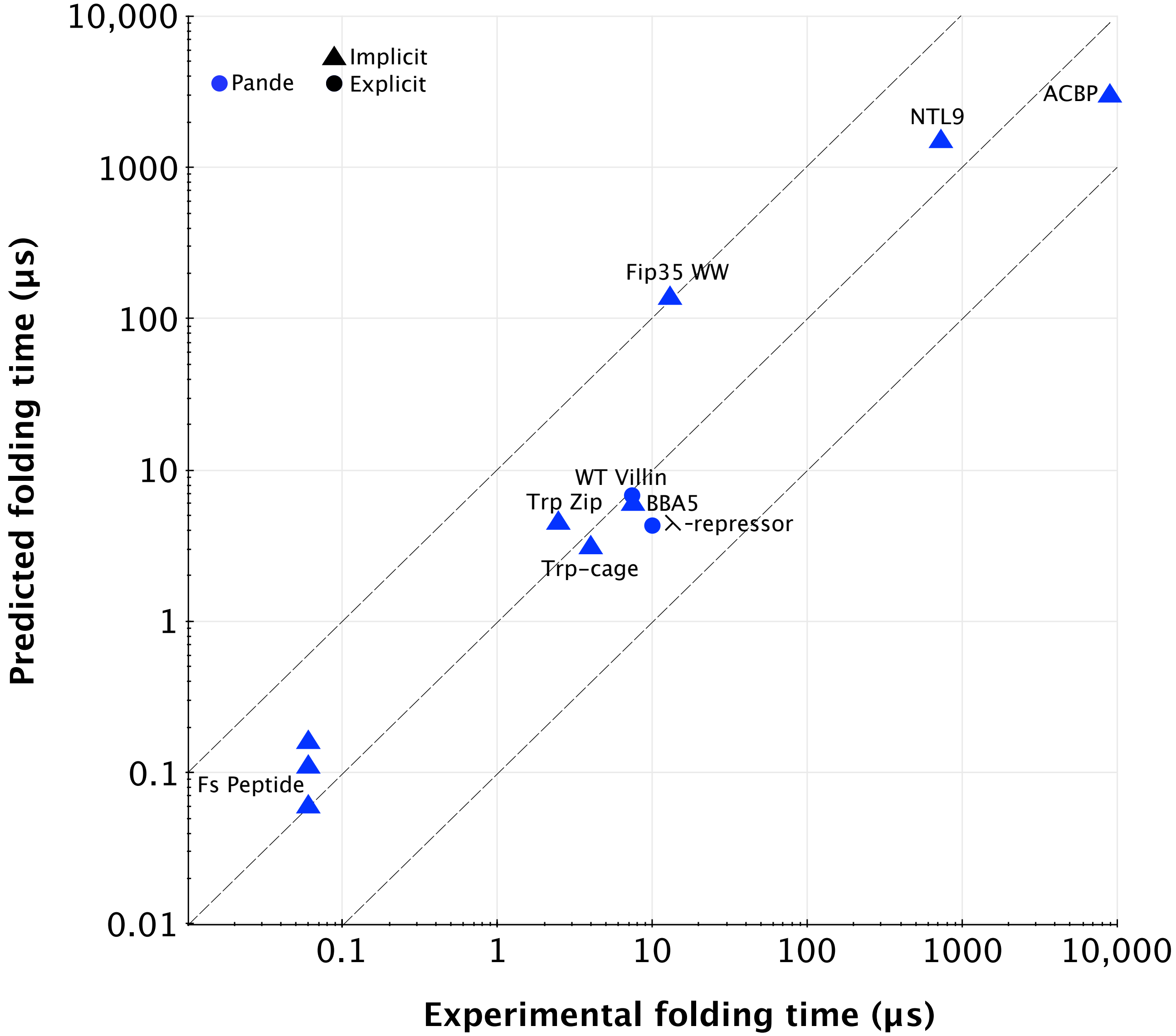
MD simulation has come a long way



MD simulation has come a long way

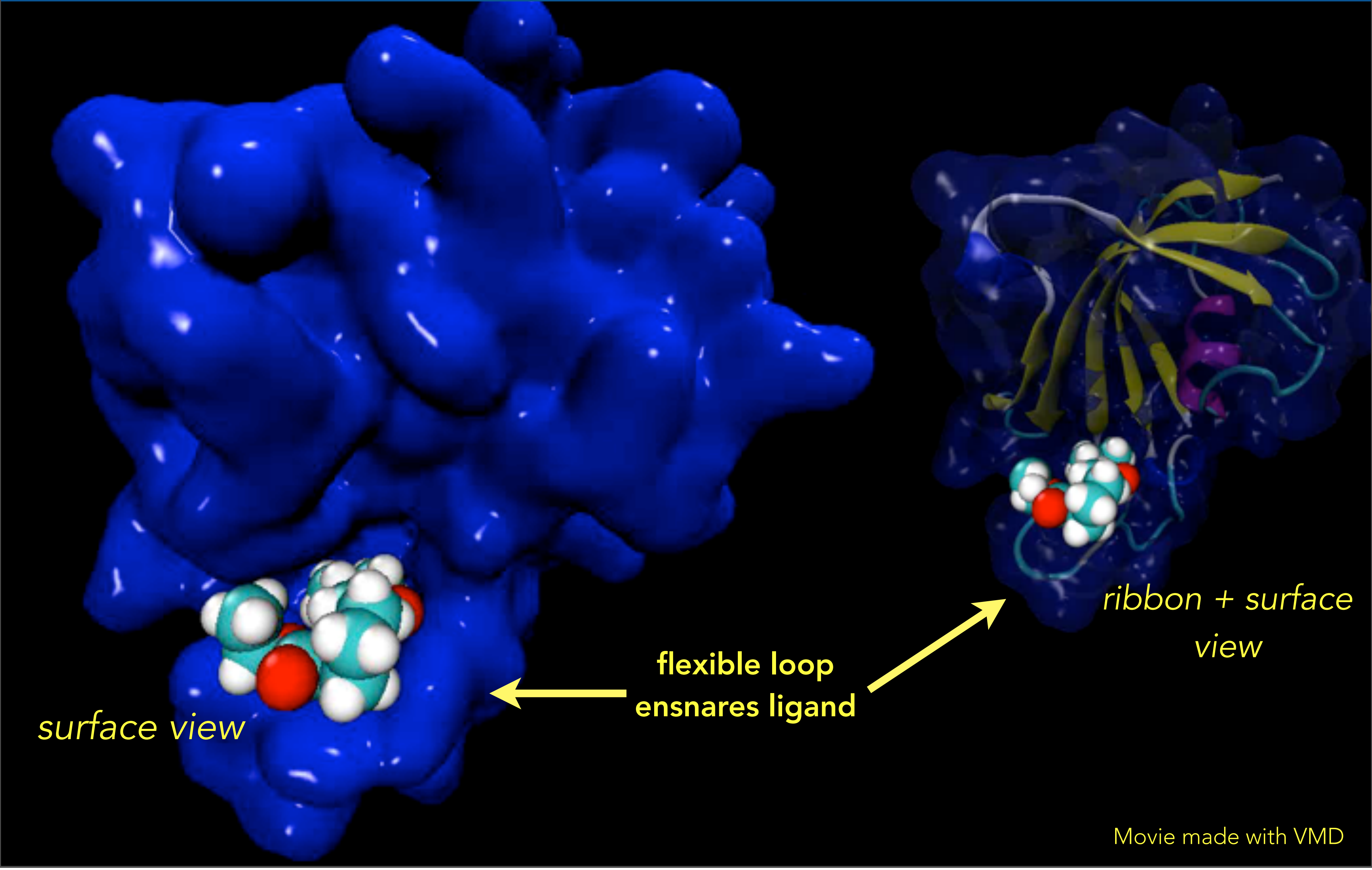


MSMs make quantitative predictions



(Lawrenz, VSP)

MSMs for protein-ligand binding



surface view

flexible loop ensnares ligand

ribbon + surface view

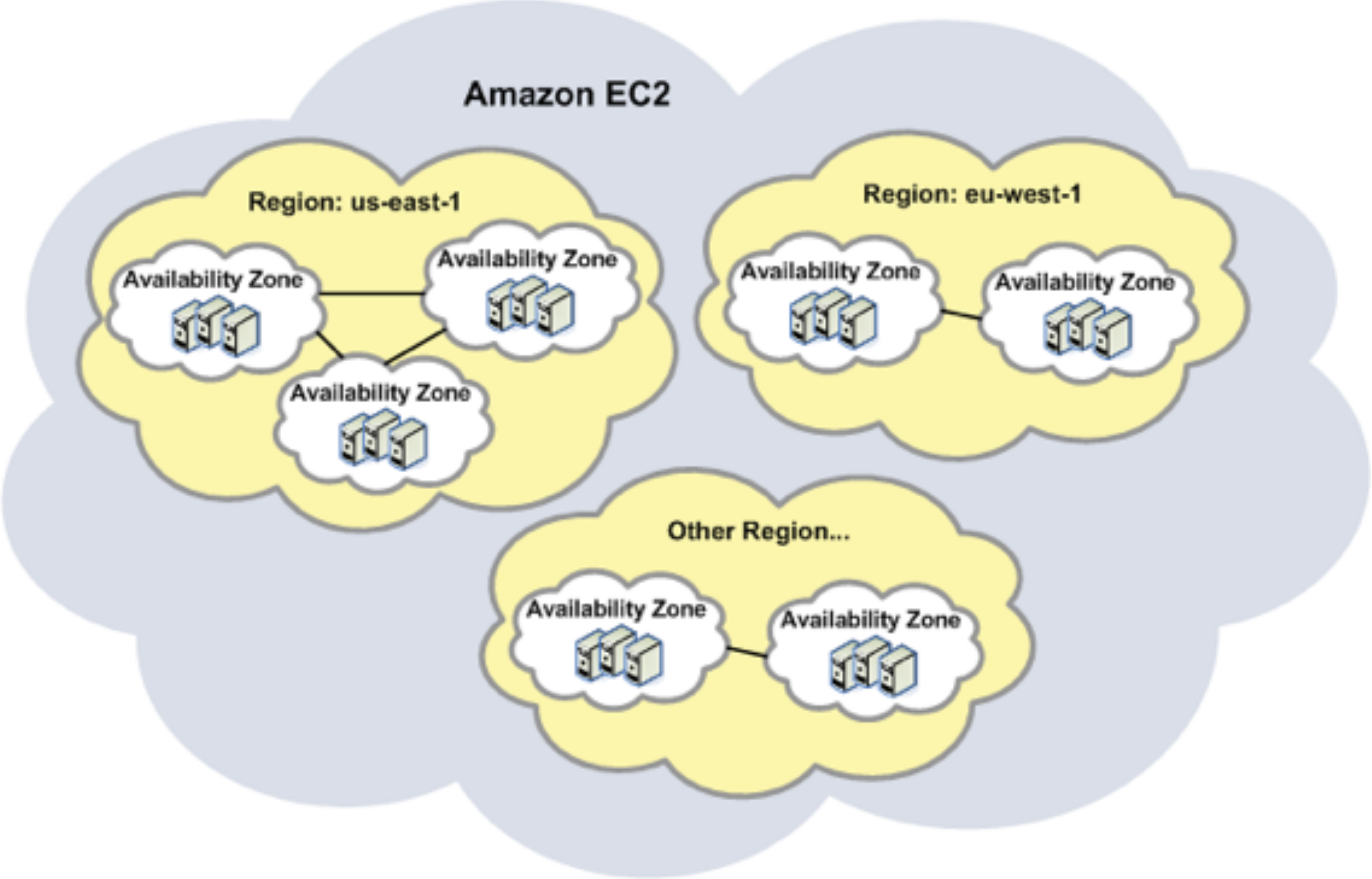
Movie made with VMD

The cloud looks a lot like Folding@home

Large-scale, distributed, heterogeneous, loosely coupled, no common filesystem

The cloud looks a lot like Folding@home

Large-scale, distributed, heterogeneous, loosely coupled, no common filesystem



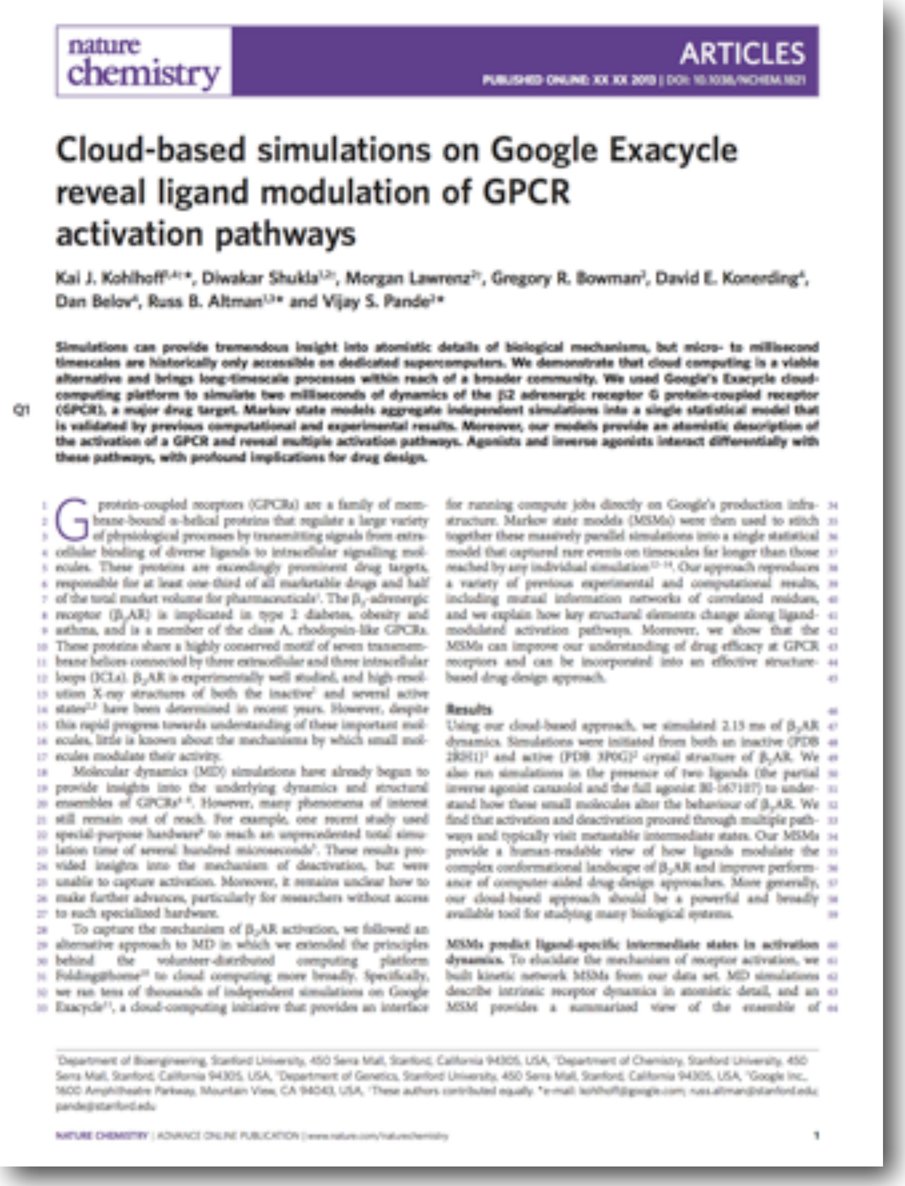
Recent results using Google cloud

nature chemistry

JANUARY 2014 VOL 6 NO 1

www.nature.com/naturechemistry

GPCRS in the Cloud



nature chemistry | ARTICLES

PUBLISHED ONLINE: XX XX 2013 | DOI: 10.1038/NCHEM.1821

Cloud-based simulations on Google Exacycle reveal ligand modulation of GPCR activation pathways

Kai J. Kohlhoff[1,4,†]*, Diwakar Shukla[1,2,†], Morgan Lawrenz[2,†], Gregory R. Bowman[2], David E. Konerding[4], Dan Belov[4], Russ B. Altman[1,3]* and Vijay S. Pande[2]*

Q1

Simulations can provide tremendous insight into atomistic details of biological mechanisms, but micro- to millisecond timescales are historically only accessible on dedicated supercomputers. We demonstrate that cloud computing is a viable alternative and brings long-timescale processes within reach of a broader community. We used Google's Exacycle cloud-computing platform to simulate two milliseconds of dynamics of the β2 adrenergic receptor G protein-coupled receptor (GPCR), a major drug target. Markov state models aggregate independent simulations into a single statistical model that is validated by previous computational and experimental results. Moreover, our models provide an atomistic description of the activation of a GPCR and reveal multiple activation pathways. Agonists and inverse agonists interact differentially with these pathways, with profound implications for drug design.

G protein-coupled receptors (GPCRs) are a family of membrane-bound α-helical proteins that regulate a large variety of physiological processes by transmitting signals from extracellular binding of diverse ligands to intracellular signalling molecules. These proteins are exceedingly prominent drug targets, responsible for at least one-third of all marketable drugs and half of the total market volume for pharmaceuticals[1]. The β_2-adrenergic receptor (β_2AR) is implicated in type 2 diabetes, obesity and asthma, and is a member of the class A, rhodopsin-like GPCRs. These proteins share a highly conserved motif of seven transmembrane helices connected by three extracellular and three intracellular loops (ICLs). β_2AR is experimentally well studied, and high-resolution X-ray structures of both the inactive[1] and several active states[2,3] have been determined in recent years. However, despite this rapid progress towards understanding of these important molecules, little is known about the mechanisms by which small molecules modulate their activity.

Molecular dynamics (MD) simulations have already begun to provide insights into the underlying dynamics and structural ensembles of GPCRs[4–8]. However, many phenomena of interest still remain out of reach. For example, one recent study used special-purpose hardware[9] to reach an unprecedented total simulation time of several hundred microseconds[9]. These results provided insights into the mechanism of deactivation, but were unable to capture activation. Moreover, it remains unclear how to make further advances, particularly for researchers without access to such specialized hardware.

To capture the mechanism of β_2AR activation, we followed an alternative approach to MD in which we extended the principles behind the volunteer-distributed computing platform Folding@home[10] to cloud computing more broadly. Specifically, we ran tens of thousands of independent simulations on Google Exacycle[11], a cloud-computing initiative that provides an interface for running compute jobs directly on Google's production infrastructure. Markov state models (MSMs) were then used to stitch together these massively parallel simulations into a single statistical model that captured rare events on timescales far longer than those reached by any individual simulation[12–14]. Our approach reproduces a variety of previous experimental and computational results, including mutual information networks of correlated residues, and we explain how key structural elements change along ligand-modulated activation pathways. Moreover, we show that the MSMs can improve our understanding of drug efficacy at GPCR receptors and can be incorporated into an effective structure-based drug design approach.

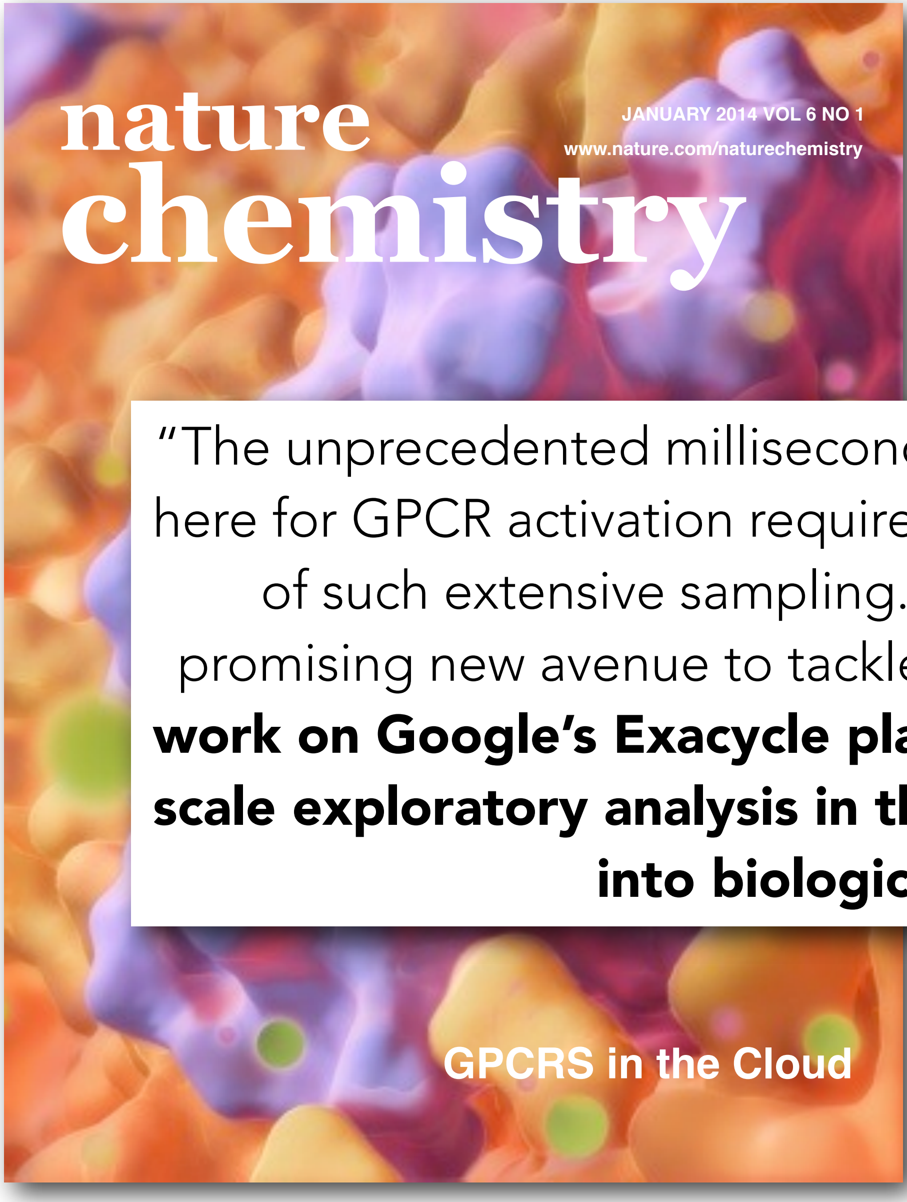
Results

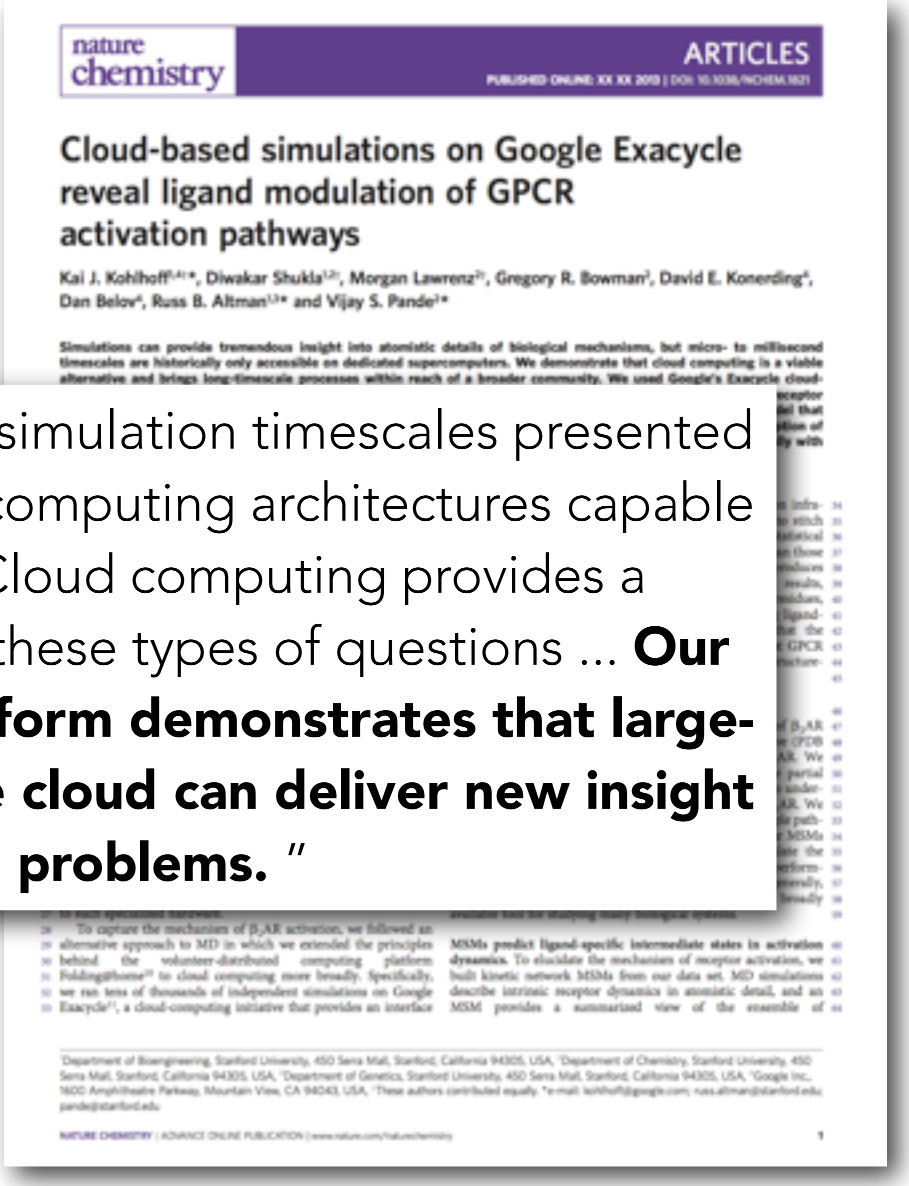
Using our cloud-based approach, we simulated 2.15 ms of β_2AR dynamics. Simulations were initiated from both an inactive (PDB 2RH1)[1] and active (PDB 3P0G)[2] crystal structure of β_2AR. We also ran simulations in the presence of two ligands (the partial inverse agonist carazolol and the full agonist BI-167107) to understand how these small molecules alter the behaviour of β_2AR. We find that activation and deactivation proceed through multiple pathways and typically visit metastable intermediate states. Our MSMs provide a human-readable view of how ligands modulate the complex conformational landscape of β_2AR and improve performance of computer-aided drug design approaches. More generally, our cloud-based approach should be a powerful and broadly available tool for studying many biological systems.

MSMs predict ligand-specific intermediate states in activation dynamics. To elucidate the mechanism of receptor activation, we built kinetic network MSMs from our data set. MD simulations describe intrinsic receptor dynamics in atomistic detail, and an MSM provides a summarized view of the ensemble of

[1] Department of Bioengineering, Stanford University, 450 Serra Mall, Stanford, California 94305, USA, [2] Department of Chemistry, Stanford University, 450 Serra Mall, Stanford, California 94305, USA, [3] Department of Genetics, Stanford University, 450 Serra Mall, Stanford, California 94305, USA, [4] Google Inc., 1600 Amphitheatre Parkway, Mountain View, CA 94043, USA, [†] These authors contributed equally. *e-mail: kohlhoff@google.com; russ.altman@stanford.edu; pande@stanford.edu

Recent results using Google cloud




nature chemistry | ARTICLES

PUBLISHED ONLINE: XX XX 2013 | DOI: 10.1038/NCHEM.1821

Cloud-based simulations on Google Exacycle reveal ligand modulation of GPCR activation pathways

Kai J. Kohlhoff[1,4†]*, Diwakar Shukla[1,2†], Morgan Lawrenz[2†], Gregory R. Bowman[2], David E. Konerding[4], Dan Belov[4], Russ B. Altman[1,3]* and Vijay S. Pande[2]*

Simulations can provide tremendous insight into atomistic details of biological mechanisms, but micro- to millisecond timescales are historically only accessible on dedicated supercomputers. We demonstrate that cloud computing is a viable alternative and brings long-timescale processes within reach of a broader community. We used Google's Exacycle cloud-

to such specialized hardware.

To capture the mechanism of β2AR activation, we followed an alternative approach to MD in which we extended the principles behind the volunteer-distributed computing platform Folding@home[10] to cloud computing more broadly. Specifically, we ran tens of thousands of independent simulations on Google Exacycle[11], a cloud-computing initiative that provides an interface

available tool for studying many biological systems.

MSMs predict ligand-specific intermediate states in activation dynamics. To elucidate the mechanism of receptor activation, we built kinetic network MSMs from our data set. MD simulations describe intrinsic receptor dynamics in atomistic detail, and an MSM provides a summarized view of the ensemble of

[1]Department of Bioengineering, Stanford University, 450 Serra Mall, Stanford, California 94305, USA, [2]Department of Chemistry, Stanford University, 450 Serra Mall, Stanford, California 94305, USA, [3]Department of Genetics, Stanford University, 450 Serra Mall, Stanford, California 94305, USA, [4]Google Inc., 1600 Amphitheatre Parkway, Mountain View, CA 94043, USA, †These authors contributed equally. *e-mail: kohlhoff@google.com; russ.altman@stanford.edu; pande@stanford.edu

NATURE CHEMISTRY | ADVANCE ONLINE PUBLICATION | www.nature.com/naturechemistry 1

"The unprecedented millisecond simulation timescales presented here for GPCR activation require computing architectures capable of such extensive sampling. Cloud computing provides a promising new avenue to tackle these types of questions ... **Our work on Google's Exacycle platform demonstrates that large-scale exploratory analysis in the cloud can deliver new insight into biological problems.** "