

# Molecular Dynamics

## Day 3

Ben Leimkuhler

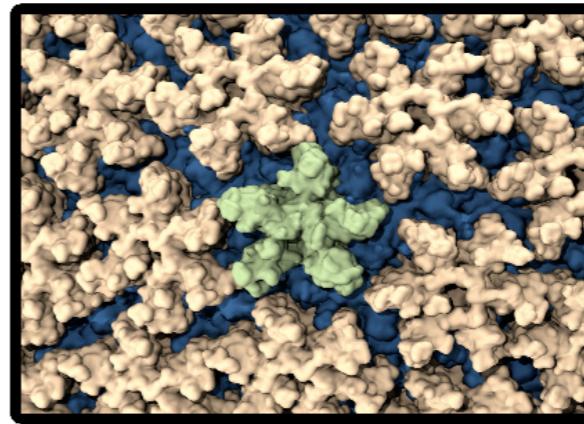
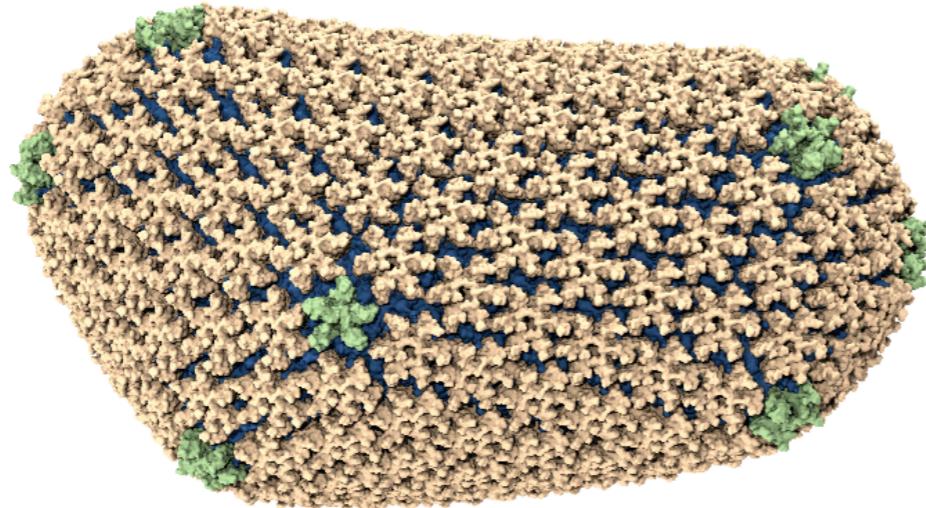
**ensemble perspective  
stochastic differential equations  
numerical methods and error  
Euler-Maruyama and Leimkuhler-Matthews  
Langevin methods by splitting**

Peking 2018

# HIV-1 Virus Capsid, 2014 (protective shield around virus)

Zhao, et al.: Mature HIV-1 capsid structure by cryoelectron microscopy and all-atom molecular dynamics. *Nature* 497 , 643–646 (2013).

*stochastic MD, Langevin piston, r-RESPA, constraints*



Molecular dynamics 64M atoms (including surroundings)  
A few ns of simulation

- **Chaotic, nonlinear dynamics, ever expanding scale**
- **Key questions are of a *probabilistic* nature**

# **Ensemble Perspective**

# Probability Density

$$\rho : \mathcal{D} \subset \mathbb{R}^{2d} \rightarrow \mathbb{R}$$

$$\rho \geq 0 \quad \int_{\mathcal{D}} \rho(z) dz = 1$$

**Probability  
of a set**

$$\Pr(\mathcal{A}) = \int_{\mathcal{A}} \rho(z) dz$$

**Average of  
a function  
wrt density**

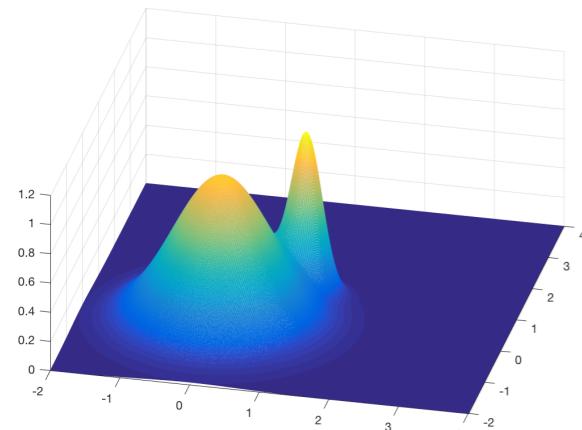
$$\text{Av}_{\rho}(g) = \int_{\mathcal{D}} g(z) \rho(z) dz$$

↑  
**“observable”**

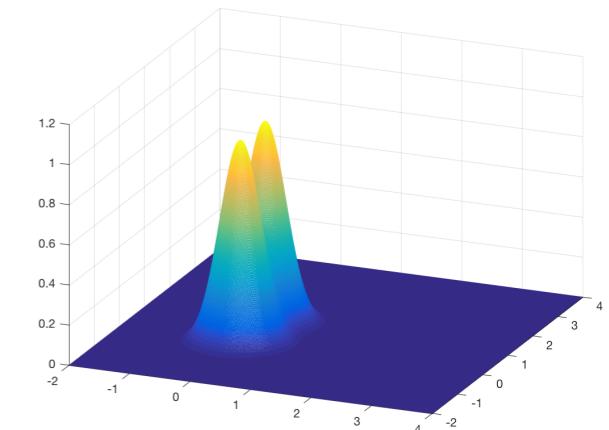
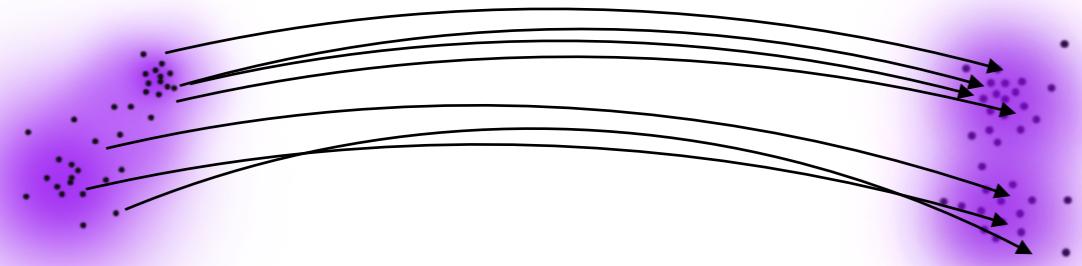
# Evolution of density

$$\dot{z} = f(z)$$

vector field



$$\rho(z, 0)$$



$$\rho(z, t)$$



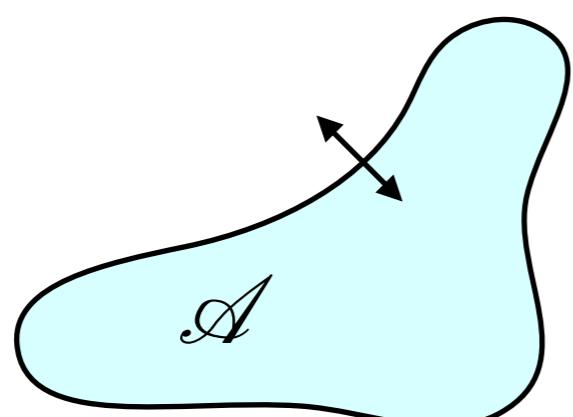
Arthur's Seat



# Continuity Equation

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (f\rho)$$

$$\nu(t) = \int_{\mathcal{A}} \rho(z, t) dz \quad \text{fraction of states in } \mathcal{A}$$



$$\begin{aligned}\nu'(t) &= \int_{\mathcal{A}} \frac{\partial \rho(z, t)}{\partial t} dz \\ &= - \int_{\partial \mathcal{A}} (f\rho) \cdot n d\sigma \\ &= - \int_{\mathcal{A}} \operatorname{div}(f\rho) dz\end{aligned}$$

Transportation of a distribution

# Dual or Adjoint Operator

$$\int_{\mathcal{D}} \varphi(z) \mathcal{L}_f \psi(z) dz = \int_{\mathcal{D}} [\mathcal{L}_f^\dagger \varphi(z)] \psi(z) dz$$

$\mathcal{L}_f^\dagger$  = adjoint wrt inner product of  $L^2(\mathcal{D})$

$$\mathcal{L}_f = f(z) \cdot \nabla_z \quad \mathcal{L}_f^\dagger g = -\nabla \cdot (fg)$$

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (f\rho) = \mathcal{L}_f^\dagger \rho$$

# Fixed Points and Equilibria

$$\mathcal{L}_f g = 0$$

$$\Phi_t g = g \quad \Rightarrow g(z(t)) \equiv g(z(0))$$

$g$  is a first integral

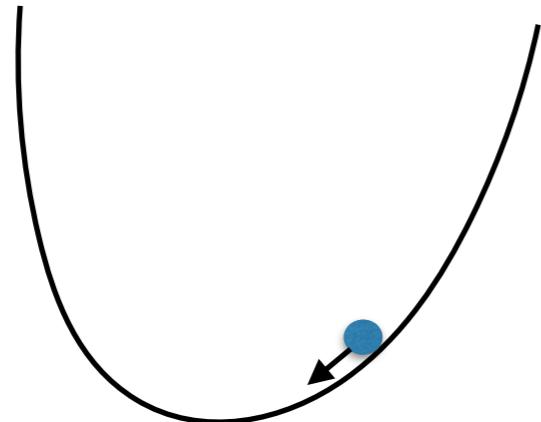
$$\mathcal{L}_f^\dagger \rho = 0 \quad \Rightarrow \frac{\partial \rho}{\partial t} = \mathcal{L}_f^\dagger \rho = 0$$

$\rho$  is invariant

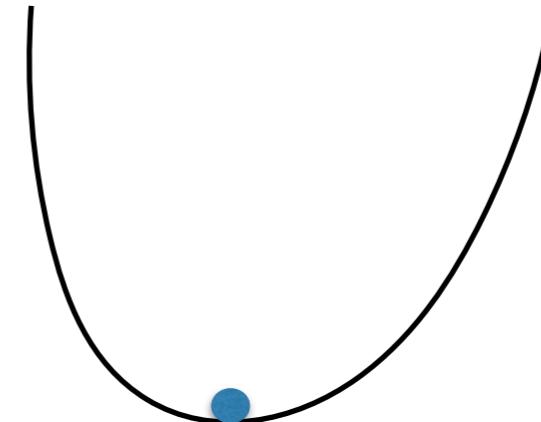
**equilibrium** or **stationary** distribution

# Mechanical vs Statistical Equilibrium

no motion in mechanical equilibrium



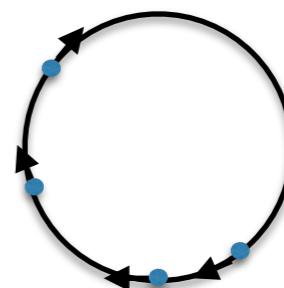
$$\dot{z} = f(z)$$



$$f(z_*) = 0$$

$$\mathcal{L}_f^\dagger \rho = 0$$

$$\begin{aligned}\dot{x} &= u \\ \dot{u} &= -x\end{aligned}$$



plenty of motion in a  
statistical equilibrium state

# Hamiltonian Case

in general:

$$\mathcal{L}_f = f(z) \cdot \nabla_z$$

$$\mathcal{L}_f^\dagger g = -\nabla \cdot (fg)$$

$$\begin{aligned}\mathcal{L}_f g &= (J\nabla H)^T \nabla g \\ &= -\nabla H^T J \nabla g = \{g, H\}\end{aligned}$$

Poisson bracket

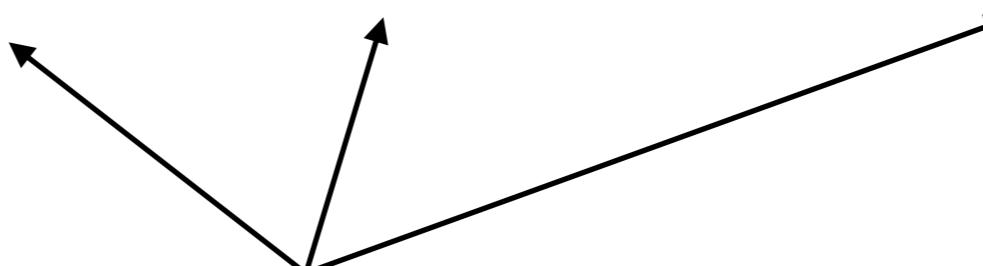
**skew-  
adjoint  
operator**

$$\begin{aligned}\mathcal{L}_f^\dagger g &= -\nabla \cdot (g J \nabla H) \\ &= -\nabla g^T J \nabla H + g \cancel{\nabla \cdot (J \nabla H)} \\ &= -\mathcal{L}_f g\end{aligned}$$

# Hamiltonian Case

$$\mathcal{L}_f^\dagger g = -\mathcal{L}_f g$$

$$\mathcal{L}_f^\dagger \rho = 0 \Rightarrow \mathcal{L}_f \rho = 0$$

$$\rho(z) = \rho(I_1(z), I_2(z), \dots, I_k(z))$$


first integrals

# Two Common Distributions for MD

## microcanonical

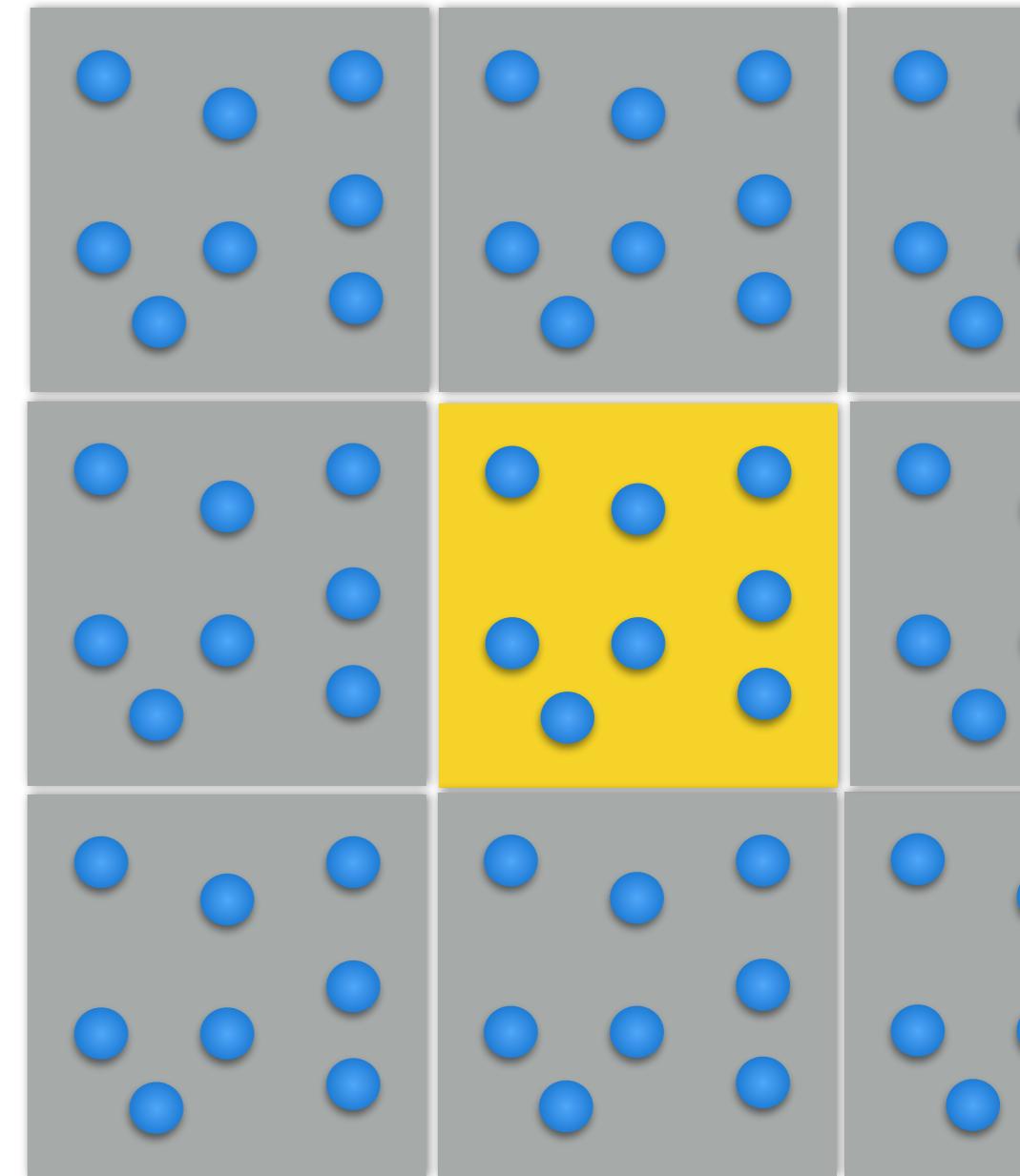
constant particle number N,  
volume V,  
energy E

$$\rho_{mc} \propto \delta[H - E]$$

## canonical

constant particle number N,  
volume V,  
temperature T

$$\rho_\beta \propto \exp(-\beta H) \quad \beta = \frac{1}{k_B T}$$



See e.g. Kinchin...

# chaos → stochasticity

**Microcanonical Ergodicity:** (almost) any pair of phase space points with a given energy  $E$  are linked by a trajectory. We might assume this for MD.

*More precisely:*

$$\forall t \Phi_t S = S \Rightarrow \rho(S) = 0 \text{ or } \rho(S) = 1$$

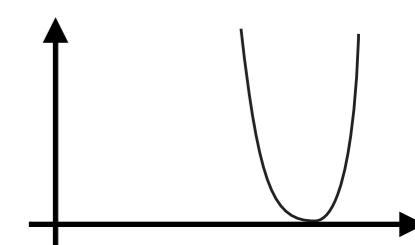
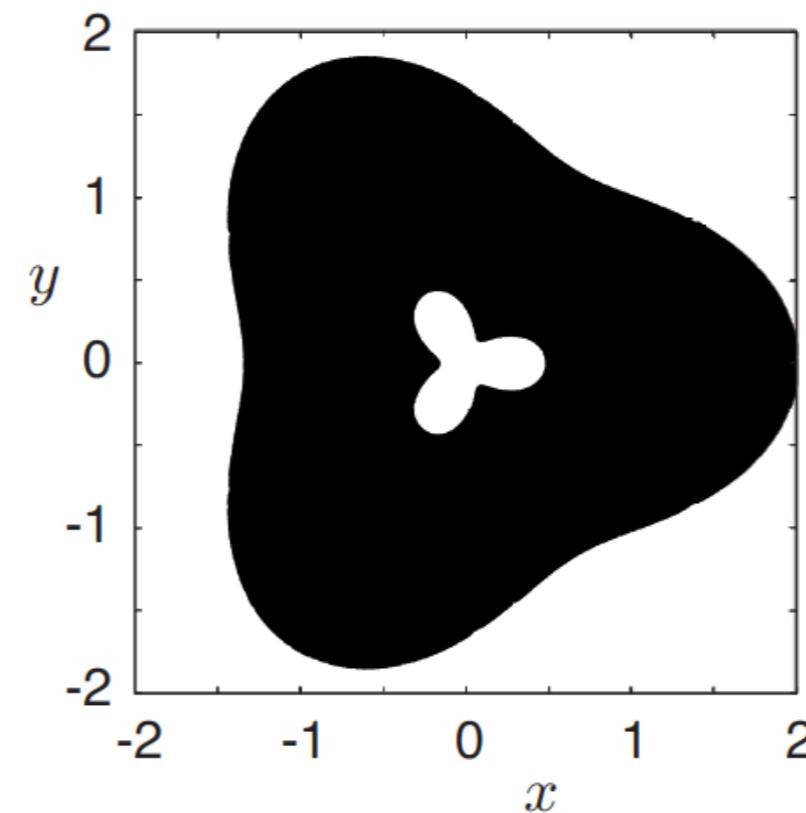
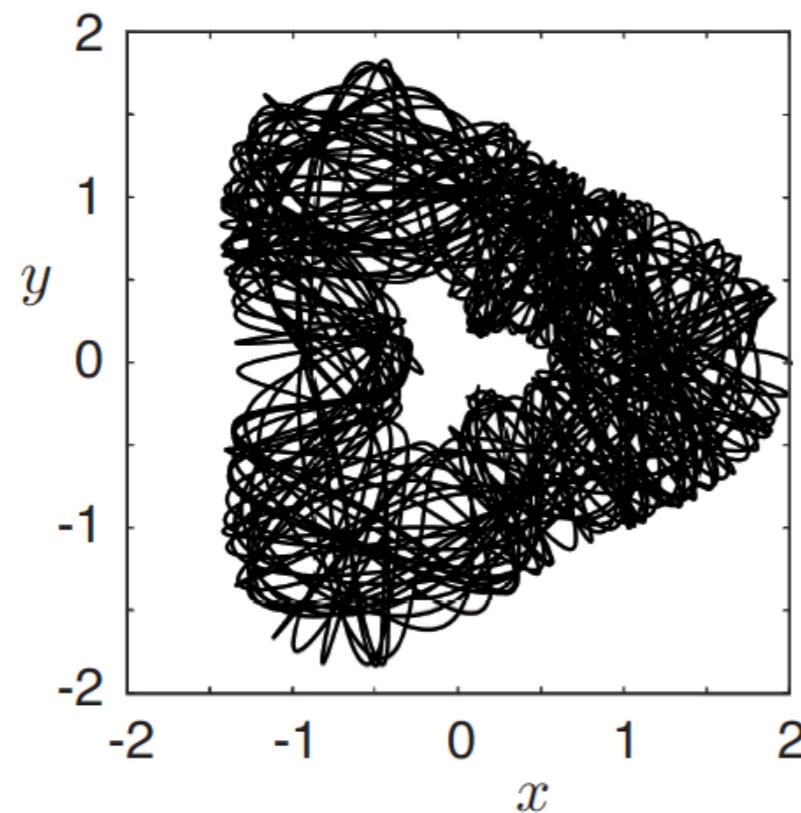
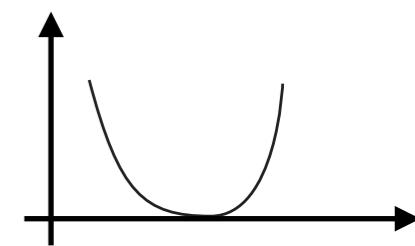
If ergodic, the dynamics generates space-filling curves that map out energy level surfaces.

Such a dynamics can be used to compute **statistics**.

# Invariant Sets

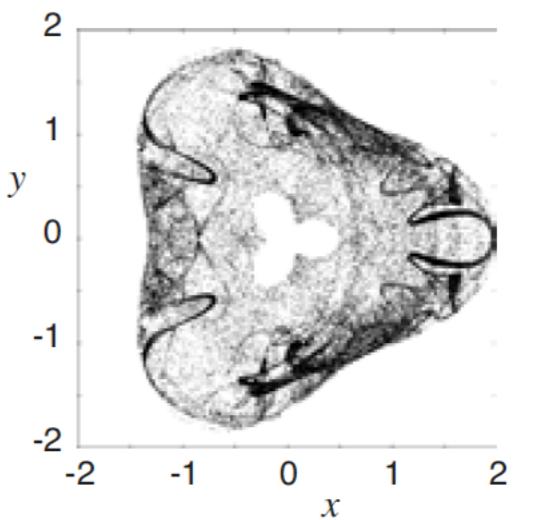
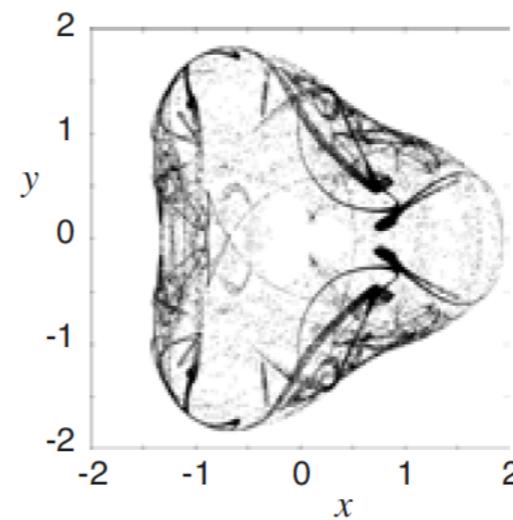
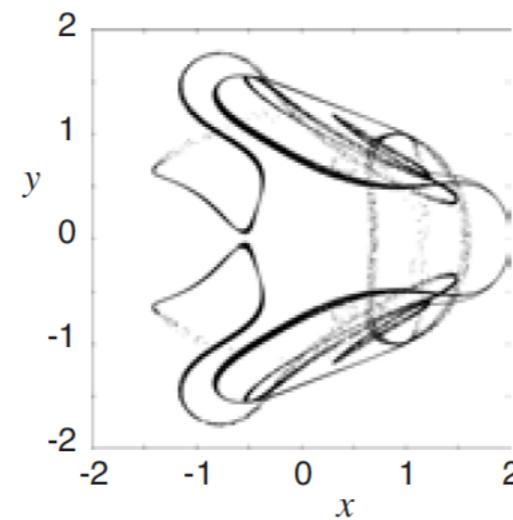
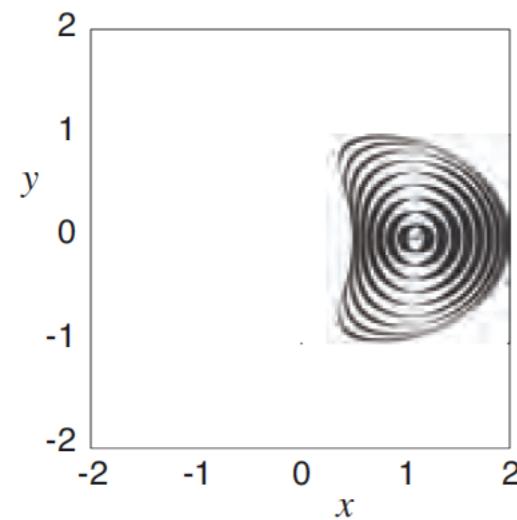
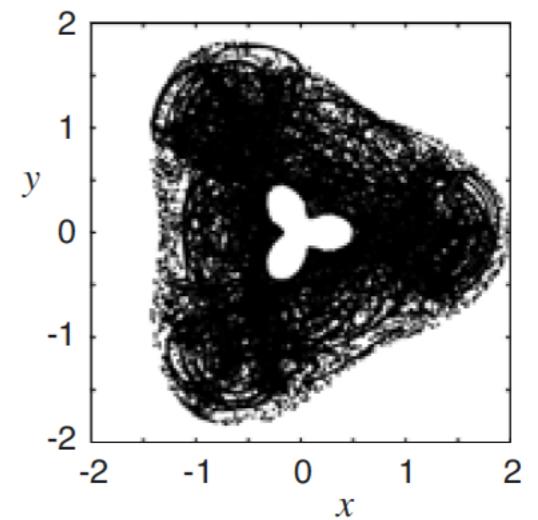
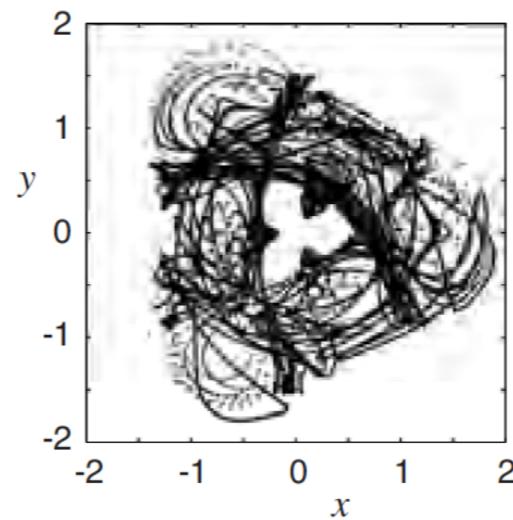
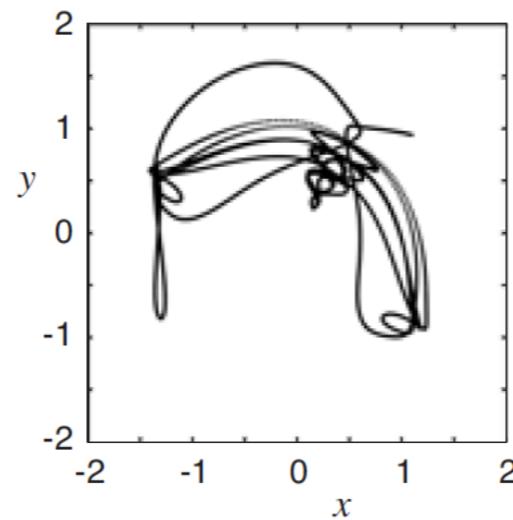
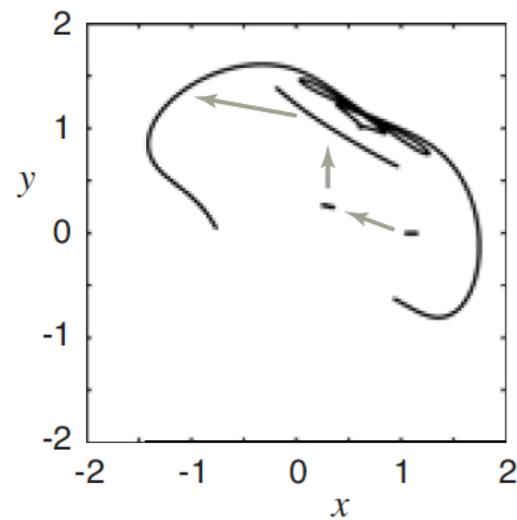
## Anisotropic oscillator

$$E(x, y, \dot{x}, \dot{y}) = \frac{1}{2}(\dot{x}^2 + \dot{y}^2) + \frac{\kappa(c_3)}{2}(r - l(c_3))^2 \quad r = \sqrt{x^2 + y^2}, \quad c_3 = 4c^3 - 3c, \quad \cos \theta = c = \frac{x}{r}$$
$$\kappa(c_3) = \kappa_0(1 - \frac{1}{2}\varepsilon c_3), \quad l(c_3) = l_0(1 + \frac{1}{2}\varepsilon c_3)$$



**Transportation** of set by flow map:  $S \mapsto \Phi_t S$

Even if ergodic, convergence to the invariant set  
may depend heavily on initial condition...



# Statistics from Dynamics

$$\langle f(q, p) \rangle_{\text{microcanonical}} = \frac{1}{\Omega} \int_{\mathbf{R}^{6N}} f(q, p) \rho(q, p) d\omega$$

microcanonical density  $\rho = \delta[H - E]$

if ergodic, **Birkhoff's Theorem** implies

$$\langle f(q, p) \rangle_{\text{microcanonical}} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(q(t), p(t)) dt$$

averaging along a trajectory of energy  $E$

# Sampling the **microcanonical** ensemble

1. Using **Hamiltonian paths** (or symplectic integrators)

**Ergodicity** relies on assumptions regarding the chaotic nature of Hamiltonian dynamics

Suitable when short time dynamical approximation is also required.

2. Alternative: Using constrained **stochastic differential equations** [Faou & Lelievre, *Math. Comp.* 78 (2009), 2047-2074]

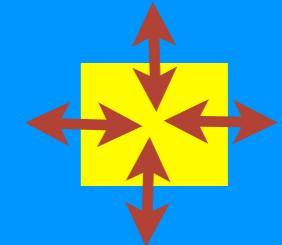
Stochastic perturbations ensure that the full energy surface is ergodically explored.

# **The canonical ensemble**

# Canonical Ensemble

A molecular system is rarely isolated.

We must incorporate the effects of interactions with a larger environment.



**Gibbs distribution:** the states of a small subsystem have an exponential distribution in terms of the energy  $H$  of the isolate.

$$\rho_{\text{can}} \propto e^{-\beta H(q,p)} \quad \beta^{-1} = k_B T$$

# Canonical Ensemble

Gibbs-Boltzmann Entropy:

$$S = -k_B \int \rho \ln \rho d\omega$$

constraints:

$$\int \rho(z) d\omega = 1,$$

probability  
density

$$\int H(z) \rho(z) d\omega = \bar{E}$$

fixed average energy  
(determined by bath)

variational principle

$$V = S/k_B - \lambda \left[ \int \rho(z) d\omega - 1 \right] - \beta \left[ \int H(z) \rho(z) d\omega - \bar{E} \right]$$

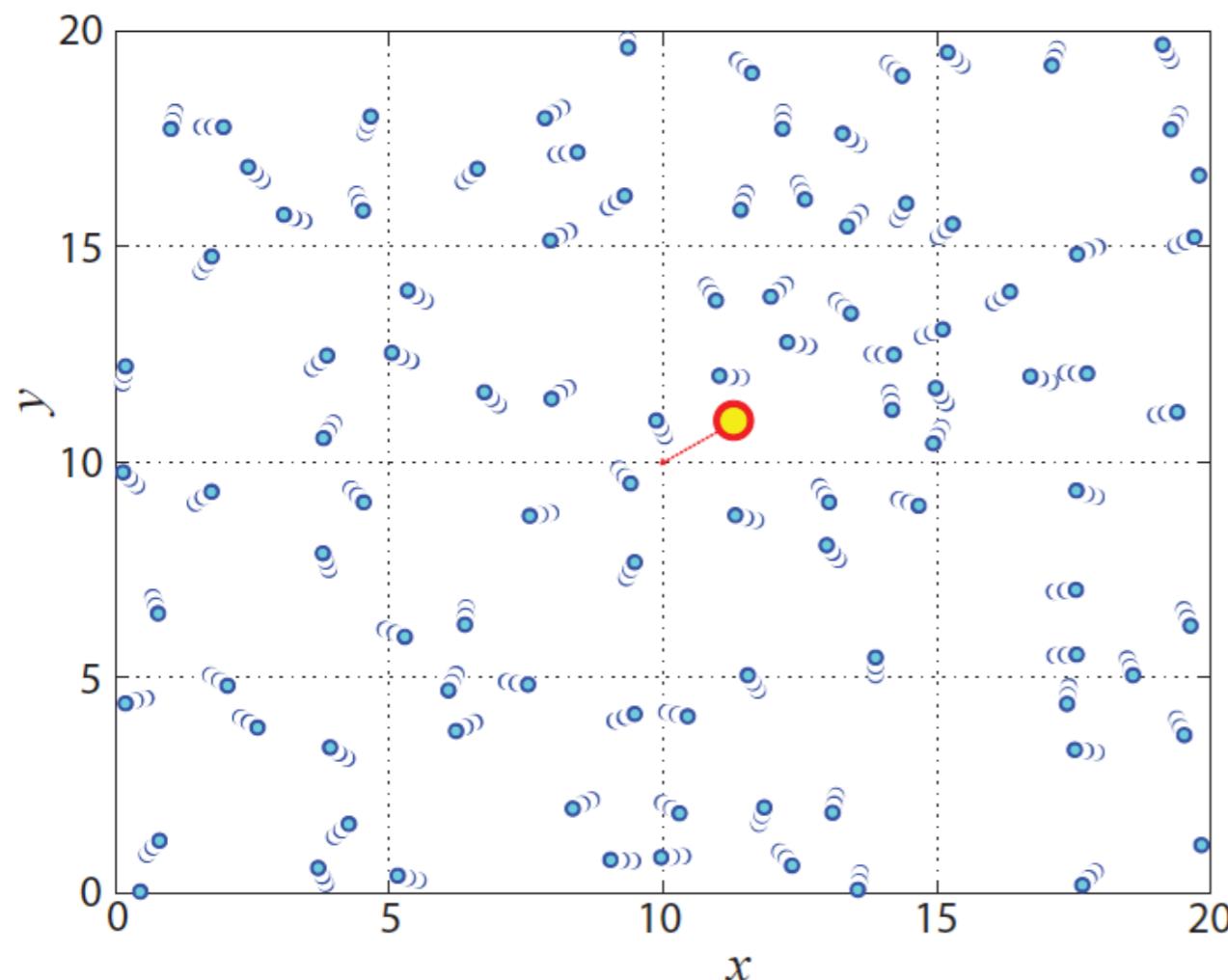
$$\frac{\delta V}{\delta \rho} = 0$$

$$\int \rho(z) d\omega = 1 \quad \Rightarrow \quad \int H(z) \rho(z) d\omega = \bar{E}.$$

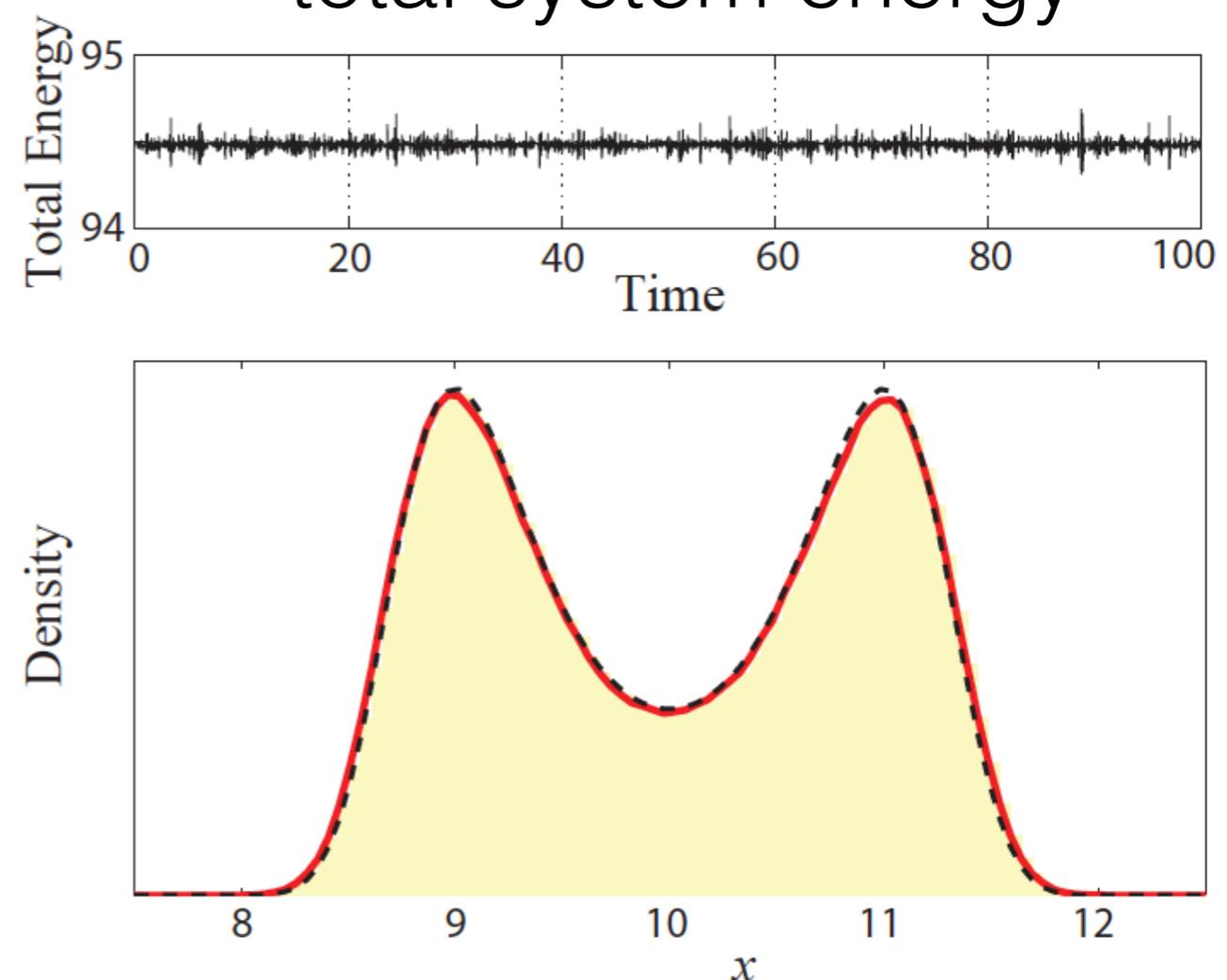
defines  $\beta$

$$\rho_\beta(q, p) = Z_\beta^{-1} e^{-\beta H(q, p)}$$

# 100 atom MD system



total system energy



distribution of  
identified particle

$$\bar{\rho}_\beta(x) \propto e^{-\beta U(x)}$$

# The Sampling Problem

Design methods to calculate

$$\int \varphi(q) d\mu(q)$$

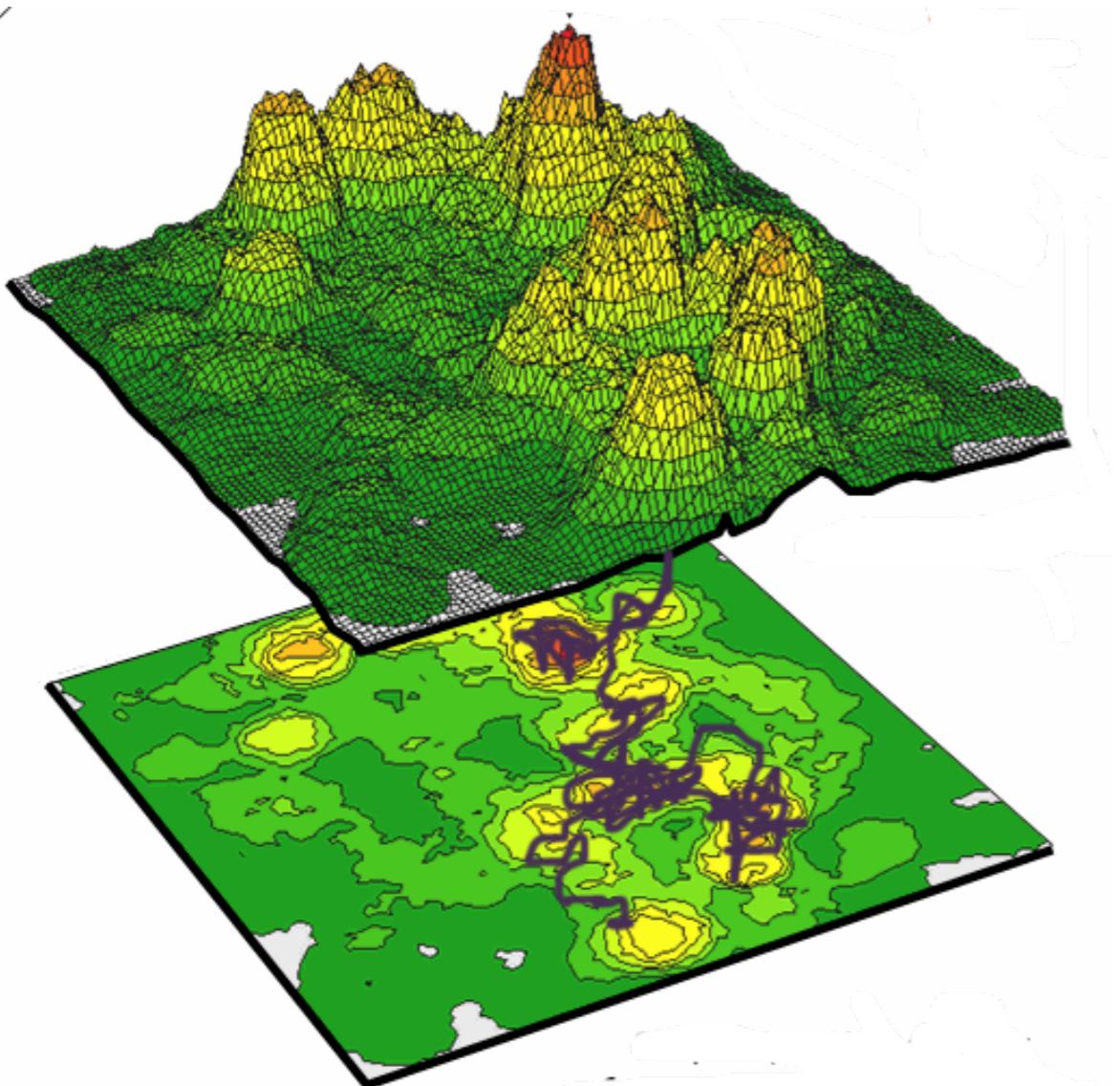
Where:

$\varphi(q)$  observable function

$d\mu(q)$  Gibbs measure     $d\mu(q) \propto e^{-U} dq$

***Integration, but typically in high dimension!***

Options: MCMC, SDE methods



**Dynamical sampling:**  
use paths of a system of  
(stochastic) differential  
equations to explore the  
target distribution.

## Problems:

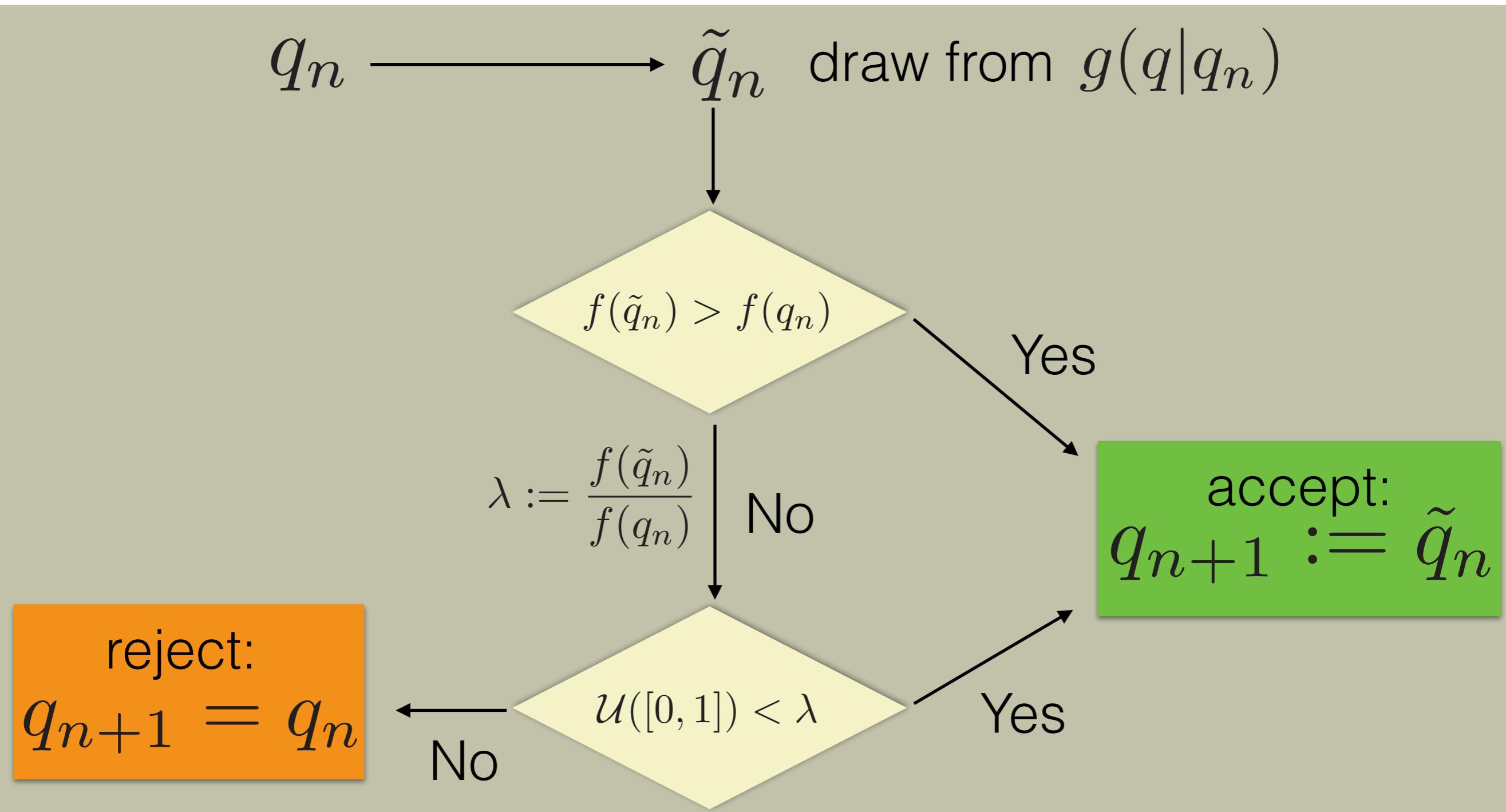
**How to choose the (stochastic)  
dynamical model?**

**Numerical discretization?**

# Metropolis-Hastings Monte Carlo

Target Density:  $\rho(q)$        $f(q) \propto \rho(q)$

Proposal:                           $g(q|q') = g(q'|q)$



# Convergence of Metropolis-Hastings MC

MHMC defines a Markov Chain with transition density

$$\pi(x'|x) = g(x'|x)A(x'|x)$$

proposal    acceptance

$$A(x'|x) = \min \left\{ 1, \frac{f(x')g(x|x')}{f(x)g(x'|x)} \right\}$$

To see that the target distribution is a stationary state, it is enough to check '**detailed balance**' ('**reversibility**')

$$\pi(x'|x)\rho(x) = \pi(x|x')\rho(x')$$

(sufficient but not necessary condition)

# Convergence of Metropolis-Hastings MC

For convergence, we also need to prove that the Markov Chain converges to the target distribution: “**ergodicity**”.

A **finite state Markov Chain** is ergodic (irreducible) if it is **aperiodic** and **positive recurrent**.

These concepts have analogues for **continuous distributions** (through “Harris chains”), but require care. See the book of Meyn and Tweedie.

There are many alternative Markov Chain Monte Carlo methods, such as “**Gibbs Sampling**” which are widely used in data science.

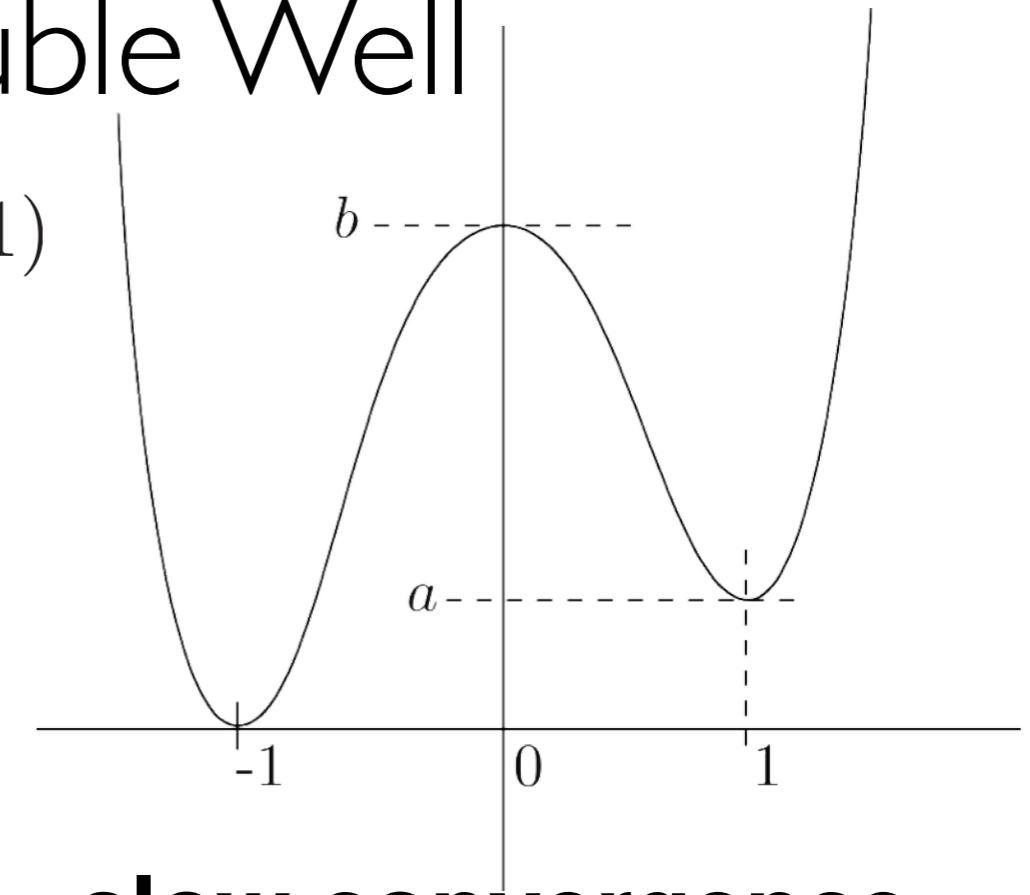
# Example: Metropolis Monte Carlo for the Uneven Double Well

$$\varphi(x) = (b - a/2)(x^2 - 1)^2 + (a/2)(x + 1)$$

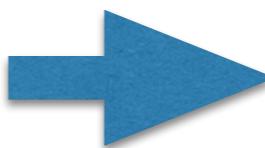
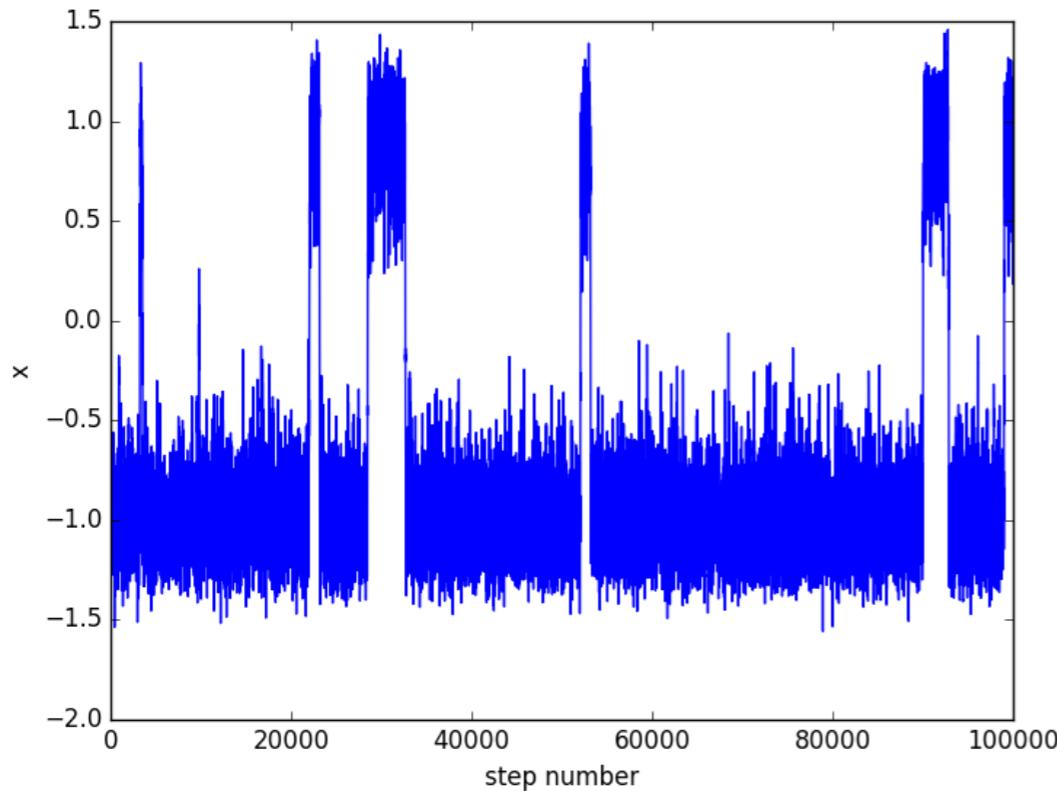
$$f(x) = e^{-\beta\varphi(x)}$$

Gaussian prior:

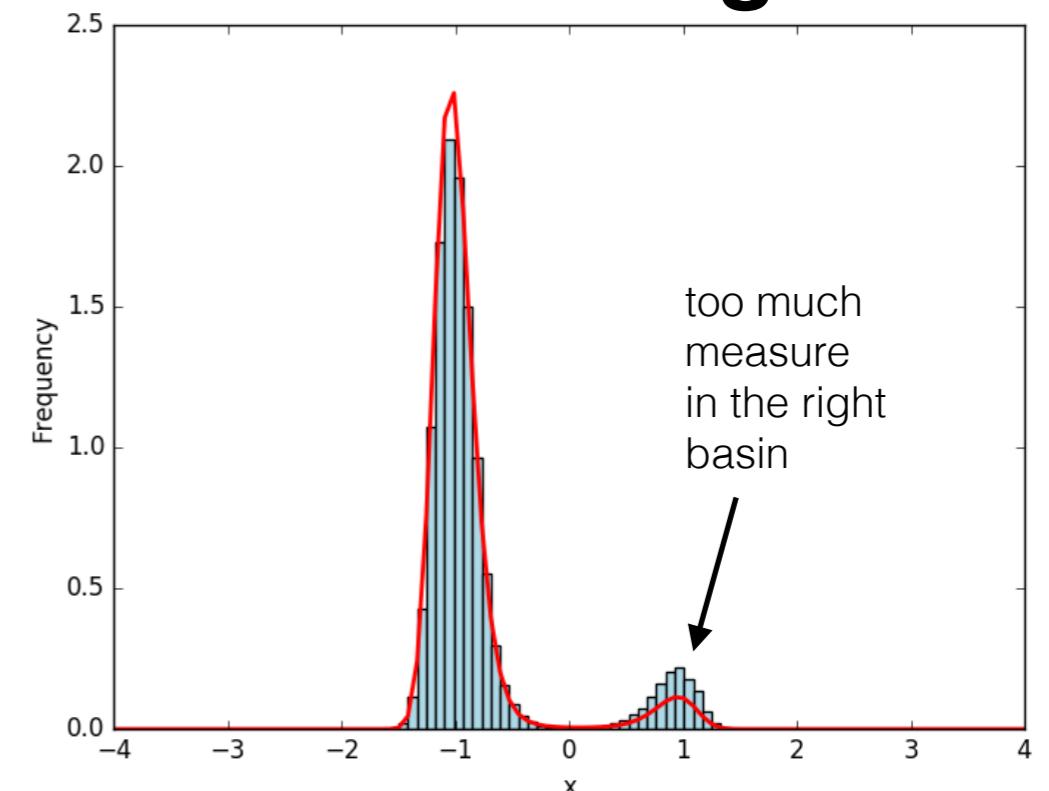
$$g(x|x') = e^{-\frac{(x-x')^2}{2\delta^2}}$$



**“rare events”**



**slow convergence**



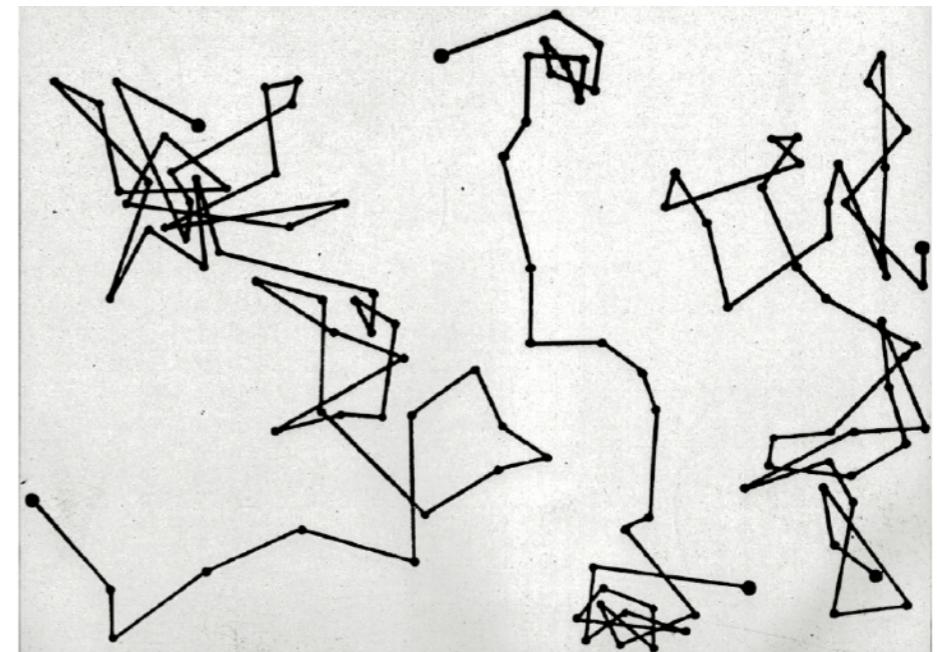
# Why (Stochastic) Dynamical Sampling?

1. SDE/DE methods can make use of **physical (natural) motion** to help select successive steps.
2. Can build on the by now well established principles of SDE/DE discretisation to give **efficient**, convergent approximations.
3. Schemes **can be “Metropolised”** if desired.
4. Mild stochastic perturbations of dynamical systems can **resolve temporally correlated processes**. Especially relevant in **non-equilibrium** settings (e.g. shear flows).
5. Easy to incorporate **constraints** (holonomic, isokinetic, momentum).

# **SDEs and invariant distributions**

# Stochastic Differential Equations

Brownian Motion: dynamics of particles subject to random collisions with a collection of light bodies.



Stochastic differential equations

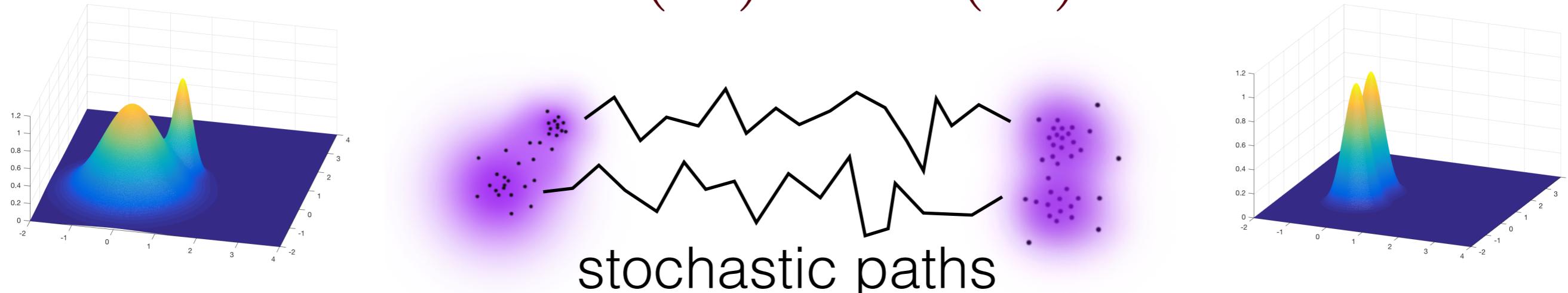
$$dX = a(X)dt + b(X)dW$$

$$W(t) = \widehat{\lim}_{\delta t \rightarrow 0} \sqrt{\delta t} \sum_{k=0}^{\lfloor t/\delta t \rfloor} J_k \quad \begin{aligned} W(t) &\sim \mathcal{N}(0, t) \\ \mathbb{E}(W(t) - W(s))^2 &= |t - s| \end{aligned}$$

# Evolving distribution

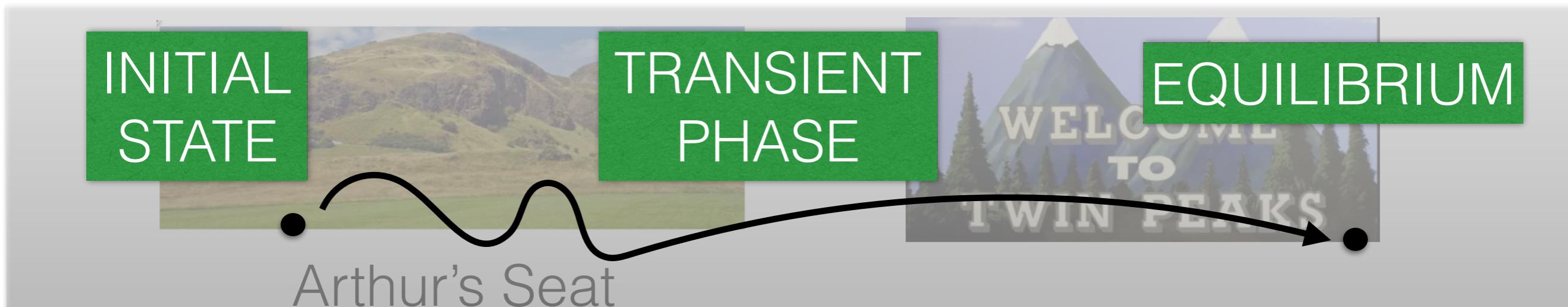
SDEs define stochastic flows for which the distribution evolves in time, in typical cases described by a probability density.

$$dX = a(X)dt + b(X)dW$$



$\rho(X, 0)$

$\rho(X, t)$



# Fokker-Planck Equation

SDE:  $dX = a(X)dt + b(X)dW$

Evolving distribution has density defined by:

$$\frac{\partial \rho}{\partial t} = -\frac{\partial}{\partial x}(a(x)\rho) + \frac{1}{2}\frac{\partial}{\partial x^2}(b(x)^2\rho)$$

Consequence of **Ito's Formula**.

$$d\Phi(X(t)) = \Phi'(X(t))(a(X)dX + b(X)dW) + \frac{1}{2}\Phi''(X(t))b(X)^2dt$$

# Invariant Measures for SDE systems

$$dX = a(X)dt + \sum_{i=1}^k b_i(X)dW_i$$

$$X \in \mathbb{R}^n, \quad a : \mathbb{R}^n \rightarrow \mathbb{R}^n, \quad b_i : \mathbb{R}^n \rightarrow \mathbb{R}^n$$

## Fokker-Planck Equation

$$\frac{\partial \rho}{\partial t} = \mathcal{L}^\dagger \rho$$

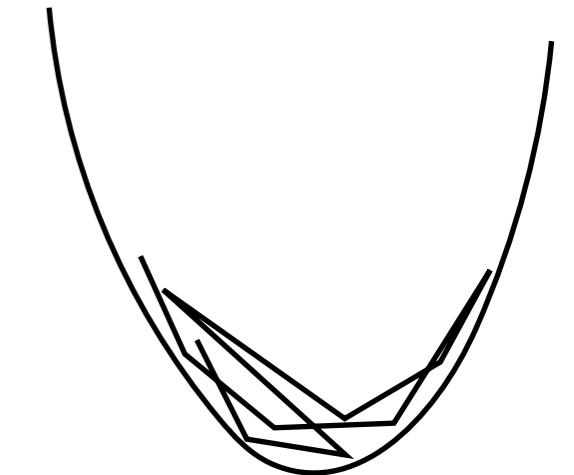
$$\mathcal{L}^\dagger g = -\nabla_x(a(x)g) + \frac{1}{2}\text{tr}\{b(x)^T g''(x)b(x)\}$$

## Stationary (equilibrium) solution

$$\mathcal{L}^\dagger \rho = 0$$

# Brownian Dynamics

*describes a particle diffusing in  
a potential  $U$  at fixed temperature*



$$dx = -\nabla U(x)dt + \sqrt{2}dW_t$$

$$\mathcal{L}^\dagger g = -\nabla \cdot (\nabla U(x)g) + \Delta g$$

$$\mathcal{L}^\dagger e^{-U(x)} = -\nabla \cdot (\nabla U(x)e^{-U(x)}) + \Delta e^{-U(x)} = 0$$

# SDEs and ergodicity

See: book of Pavliotis (Springer), article of Stoltz & Lelievre in Acta Numerica

$$dX_t = a(X_t)dt + b(X_t)dW_t$$

**generator**

$$\mathcal{L} = a(x) \cdot \nabla + b(x)^T b(x) : \nabla^2$$

**ergodic  
wrt to  $\pi$**

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t \varphi(X_t) dt = \int_{\Omega_x} \varphi(x) \pi(x) dx,$$

**transfer  
operator**

$$P_t : \int_{\Omega_x} \varphi(y) P_t(x, dy) \stackrel{\text{def}}{=} \mathbb{E}(\varphi(X_t) | X_0 = x)$$

**hypoelliptic**

$$P_t(x, dy) = p_t(x, y) dy$$

**smooth transition  
density**

# How to prove geometric ergodicity

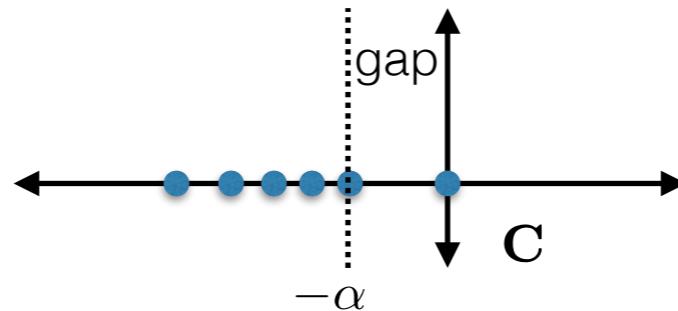
1. Check **parabolic Hörmander condition**
2. Show a **controllability property** - for any tau, there is a stochastic path of time tau linking any two points of the domain
3. Find a **Lyapunov function**

$$\mathcal{L}\mathcal{K} \leq -a\mathcal{K} + b$$

# Geometric Convergence of Brownian Dynamics

*See the book of Pavliotis, Springer, 2015.*

$$C^2(\rho_\beta), \quad \langle f, g \rangle_\beta = \int_{\Omega} f(x)g(x)\rho_\beta(x)dx$$



*Theorem:* Suppose that  $U$  is smooth with periodic boundary conditions  $\Omega = \mathcal{T}^n$  ( $\mathcal{T}^n$  an  $n$ -dimensional torus). Then, wrt  $C^2(\rho_\beta)$  we have:

1.  $\mathcal{L}$  is self-adjoint, negative definite, with spectrum consisting of discrete eigenvalues on the half line  $(-\infty, 0]$ .
2.  $\mathcal{L}$  has a *spectral gap*, meaning that, with exception of the zero eigenvalue, the remainder of the spectrum satisfies  $\sigma(\mathcal{L}) \subset (-\infty, -\alpha]$  for some  $\alpha > 0$ .
3. averages with respect to the evolving, time-dependent distribution converge exponentially rapidly to averages with respect to the invariant distribution, and, for any test function  $\phi$ , we have

$$|\bar{\phi}(t) - \text{Av}_{\text{eq.}}(\phi)| \leq K \|\phi\|_\beta e^{-\alpha t},$$

for some positive  $K$  and  $\alpha$ , where  $\bar{\phi}(t)$  denotes the time average of  $\phi$ .

# Averages using SDEs

Want to calculate

$$\int \varphi(q) d\mu(q)$$

## ***Stochastic Differential Equation***

Wiener increment

$$dq = -\nabla U(q)dt + \sqrt{2}dW_t$$

$$\lim_{t \rightarrow \infty} t^{-1} \int_0^t \varphi(q(\tau)) d\tau = \int \varphi(q) d\mu(q) \quad \textit{Ergodic limit}$$

Just need to find a way of computing numerical solutions of SDEs (on long intervals)!

# *Splitting methods for SDEs*

# Discretization of Brownian Dynamics

## Brownian Dynamics

$$dq = -\nabla U(q)dt + \sqrt{2}dW_t$$

Construct discrete stochastic paths  $q_0 \rightarrow q_1 \rightarrow q_2 \dots$

such that  $q_n \approx q(t_n)$  ... **In what sense?**

**strong order  $p$**

expectation taken over Brownian path

$$[\mathbb{E}(\|q_n - q(nh)\|^2)]^{1/2} = O(h^p) \quad nh \leq \tau$$

**weak order  $p$**

$$\varphi \in S$$

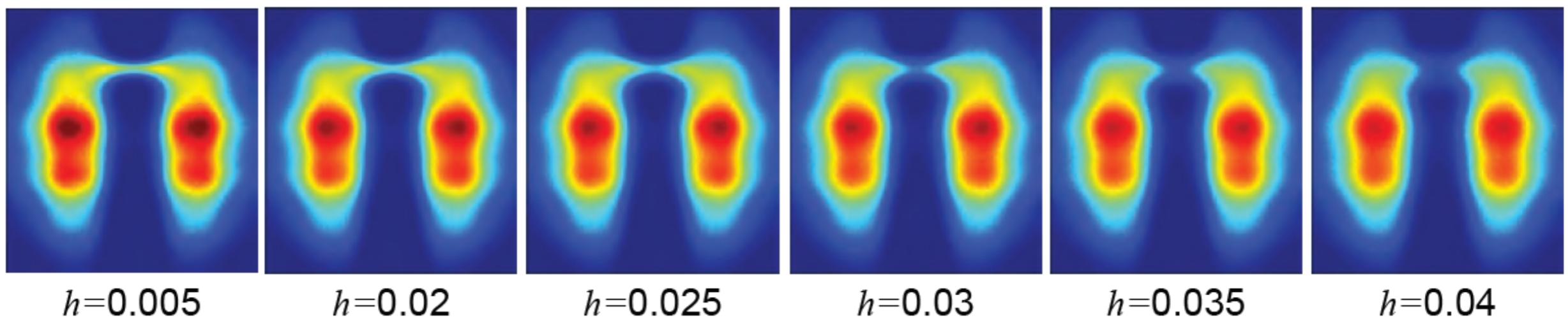
$$|\mathbb{E}\varphi(q_n) - \mathbb{E}\varphi(q(t_n))| = O(h^p) \quad nh \leq \tau$$

**weak  $t \rightarrow \infty$  [ergodic]  
order  $p$**

$$|\lim_{n \rightarrow \infty} \mathbb{E}\varphi(q_n) - \mathbb{E}_{\rho_{\text{eq}}} \varphi(q)| = O(h^p)$$

# Why does discretization bias matter?

stepsize dependent transition corridor



Although it is often stated that sampling error, not discretization bias, dominates the computational challenge in molecular dynamics, in many cases small stepsizes are used to control bias.

**Goal:** maximize the discretization stepsize while retaining high accuracy with respect to the invariant distribution.

# Discrete Paths

## Euler-Maruyama Method

$$q_{n+1} = q_n + hF(q_n) + \sqrt{2h}R_n$$

$R_n \sim \mathcal{N}(0, 1)$

## Leimkuhler-Matthews Method

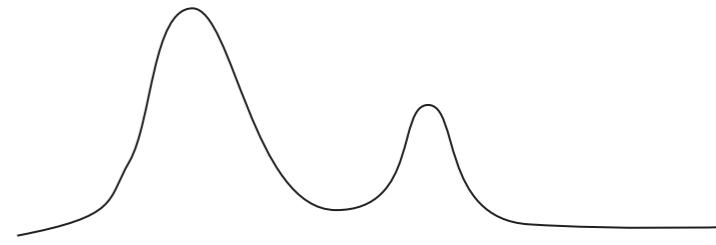
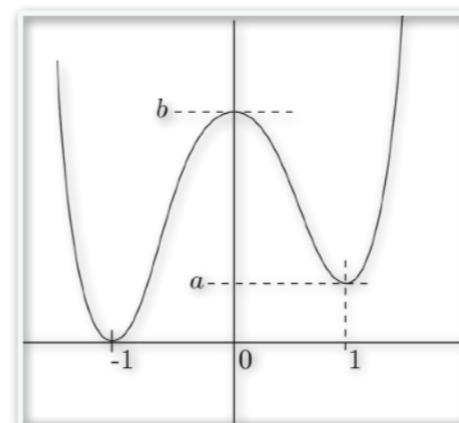
$$q_{n+1} = q_n + hF(q_n) + \sqrt{h/2}(R_n + R_{n-1})$$

[L. & Matthews, AMRX, 2013]

[L., Matthews & Stoltz, IMA J. Num. Anal., 2015]

[L., Matthews & Tretyakov, Proc Roy Soc A, 2014]

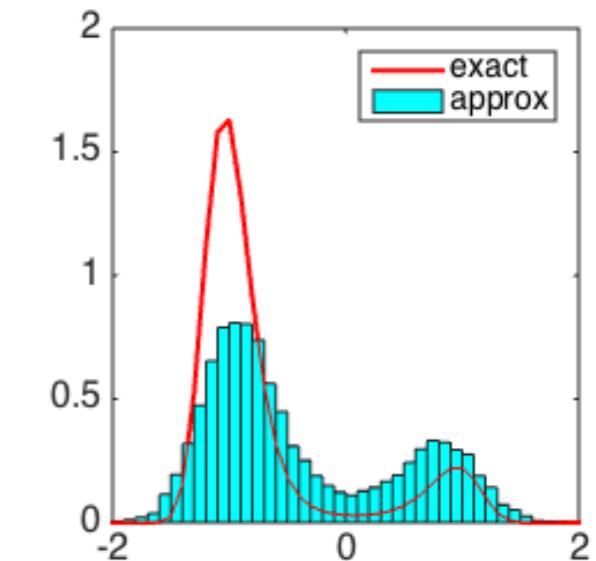
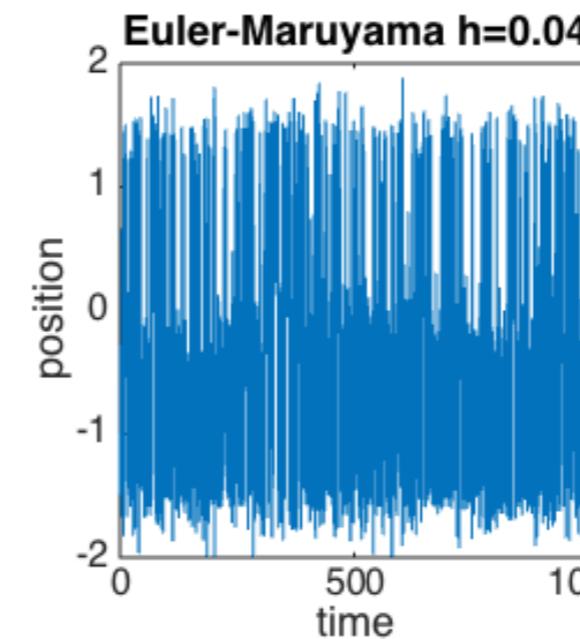
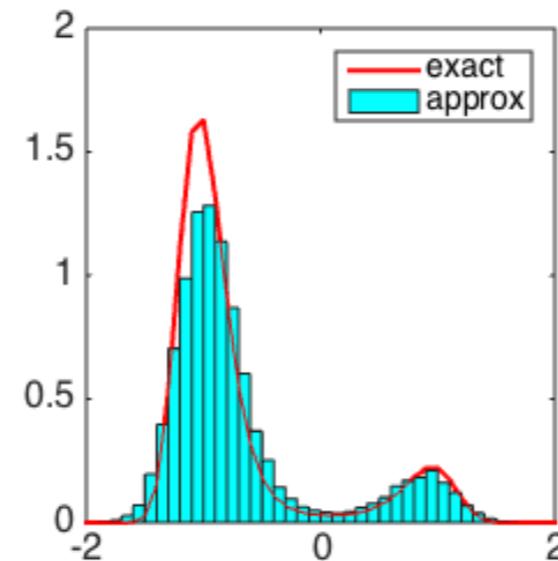
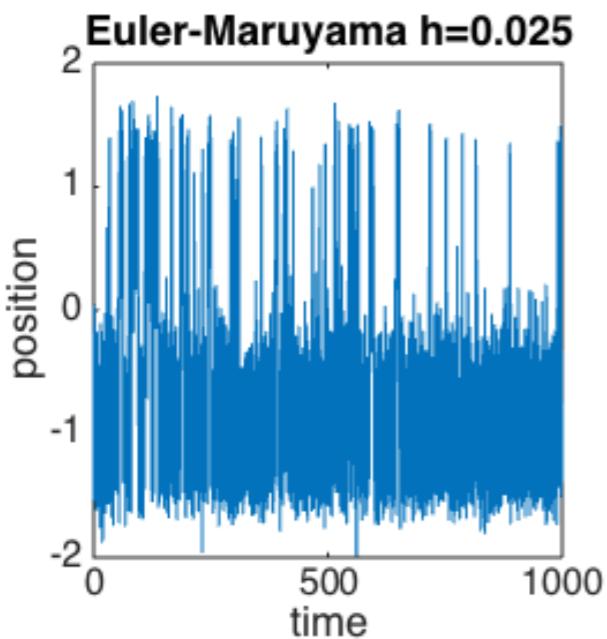
# Uneven Double Well



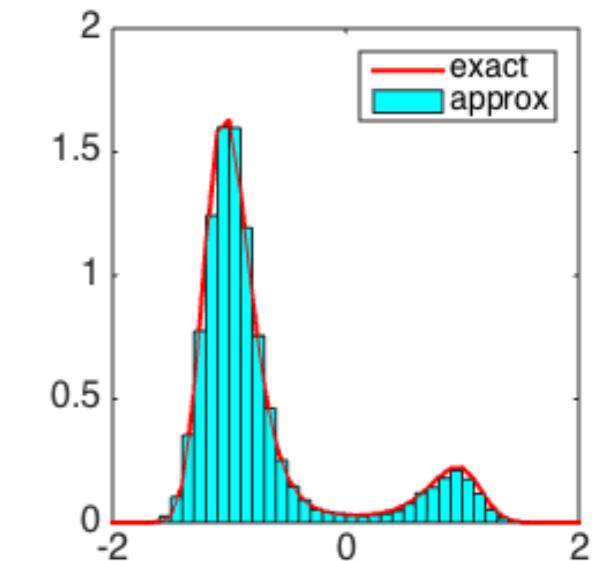
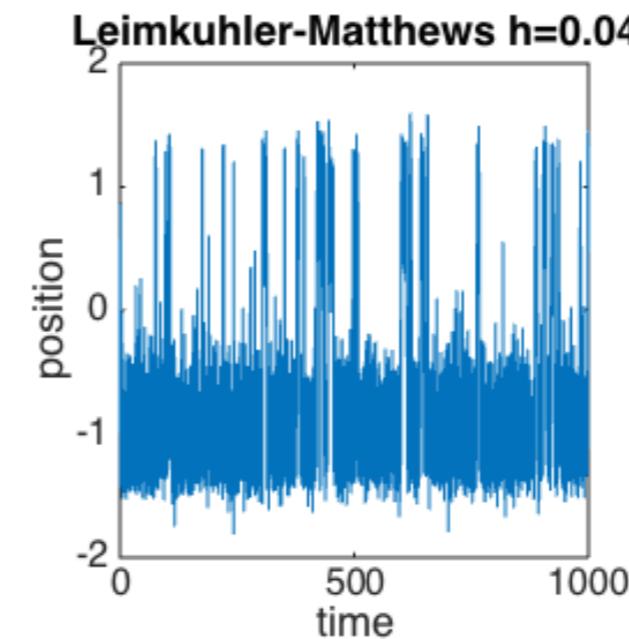
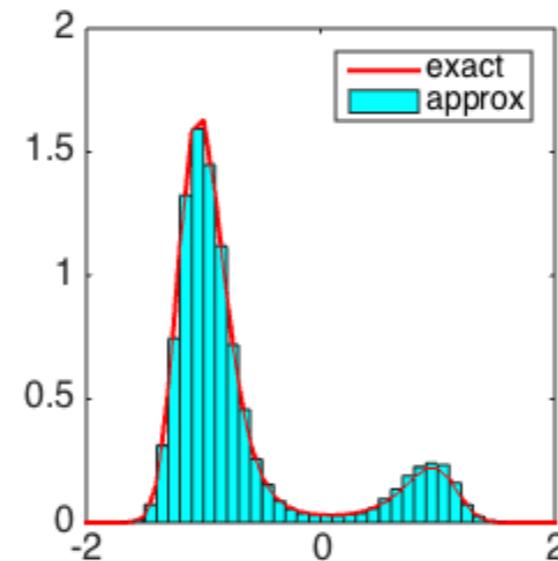
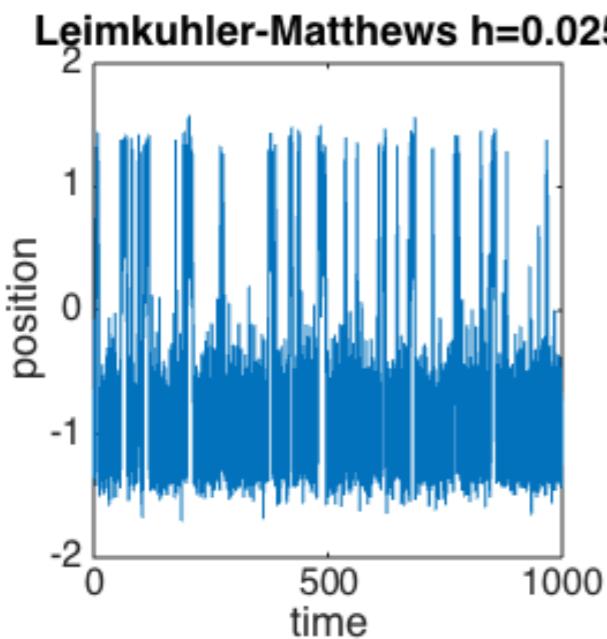
bimodal distribution

small stepsize

**E-M**



**L-M**



$$X_{n+1} = X_n + hF(X_n) + \sqrt{h/2}(R_n + R_{n+1})$$

$$Nh = \tau$$

Theorem (*BL-CM-MT Proc Roy Soc A 2014*)

For the **L-M method**, under suitable conditions,

$$\mathbf{E}\varphi(X_x(\tau)) - \mathbf{E}\varphi(X_N) = C_0(\tau, x)h + C(\tau, x)h^2$$

$$|C_0(\tau, x)| \leq K_0(1 + |x|^\eta)e^{-\lambda_0 \tau}$$

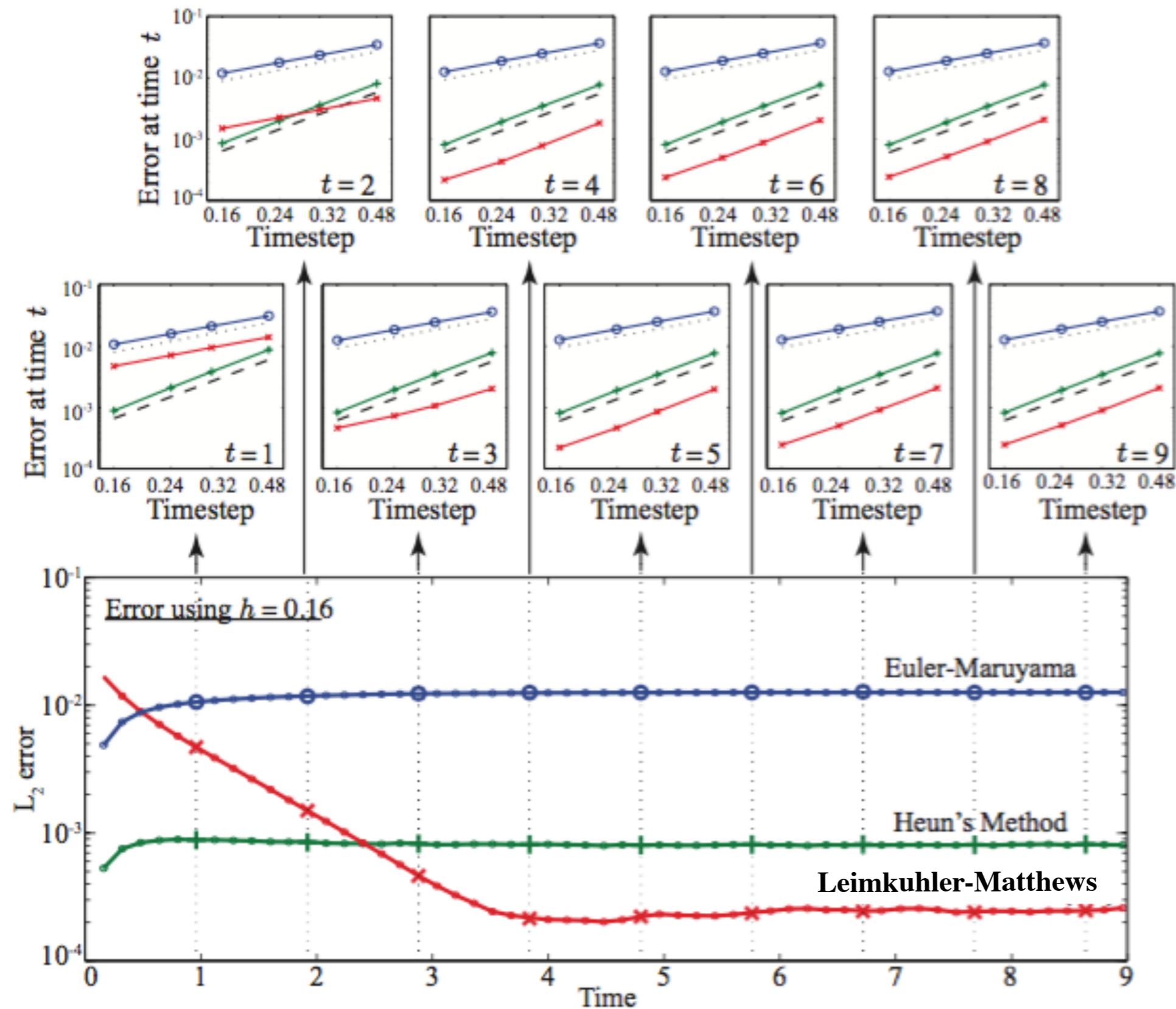
$$\lambda_0, \lambda > 0$$

$$|C(\tau, x)| \leq K(1 + |x|^\eta e^{-\lambda \tau})$$

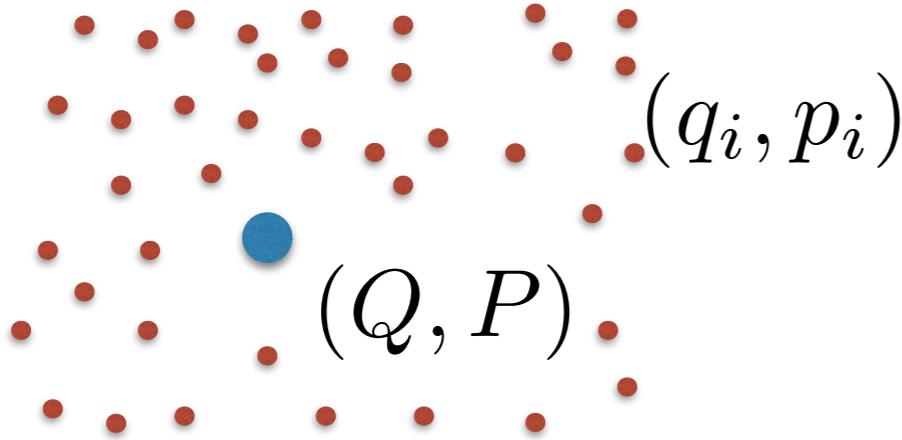
**Weak first order** -> **weak asymptotic second order**

**exponentially fast** in time

# Quartic Oscillator



# Langevin Dynamics



*Harmonic interactions  
between heavy particle  
and light particles*

$$H = \frac{P^2}{2} + \sum_{i=1}^m \frac{p_i^2}{2\mu_i} + U(Q) + \frac{1}{2m} \sum_{i=1}^m (q_i - Q)^2$$

$$\dot{Q} = P,$$

$$\dot{P} = -U'(Q) - m^{-1} \sum_{i=1}^m (Q - q_i),$$

$$\dot{q}_i = p_i / \mu_i,$$

$$\dot{p}_i = (Q - q_i) / m$$

# “Generalized” Langevin Dynamics

$$\begin{aligned}\dot{Q} &= P, \\ \dot{P} &= -U'(Q) - \kappa m^{-1} \sum_{i=1}^m (Q - q_i),\end{aligned}$$

$$= -U'(Q) + f(t) - \int_0^t \psi(t-s) P(s) ds.$$

*An instance of  
Mori-Zwanzig  
reduction*

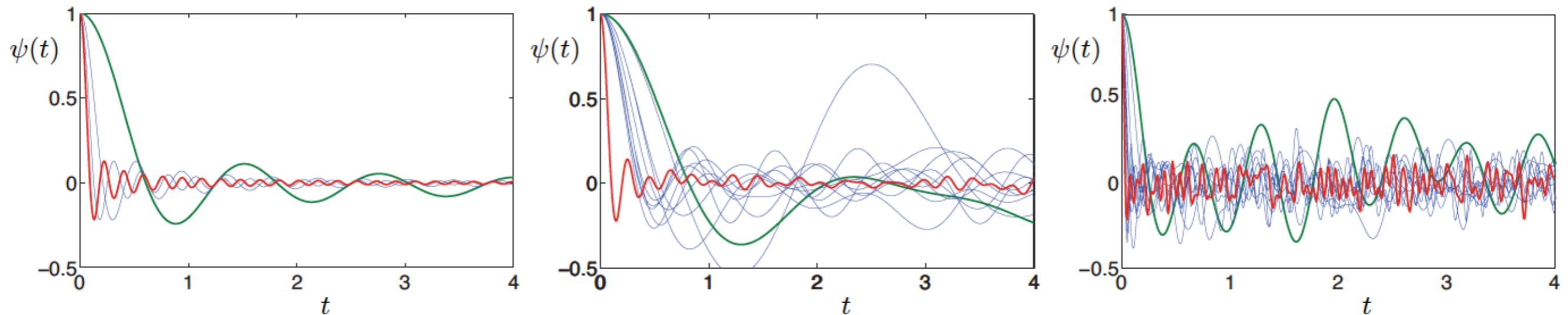
$$f(t) = -m^{-1} \sum_{i=1}^m [\cos(\Omega_i t) q_i(0) + \Omega_i^{-1} \sin(\Omega_i t) p_i(0) / \mu_i - \cos(\Omega_i t) Q(0)],$$

**deterministic model for noise process**

$$\psi(t) = m^{-1} \sum_{i=1}^m \cos(\Omega_i(t)). \quad \text{friction kernel}$$

# Large Bath Limit

**Memory kernels** in various scenarios:

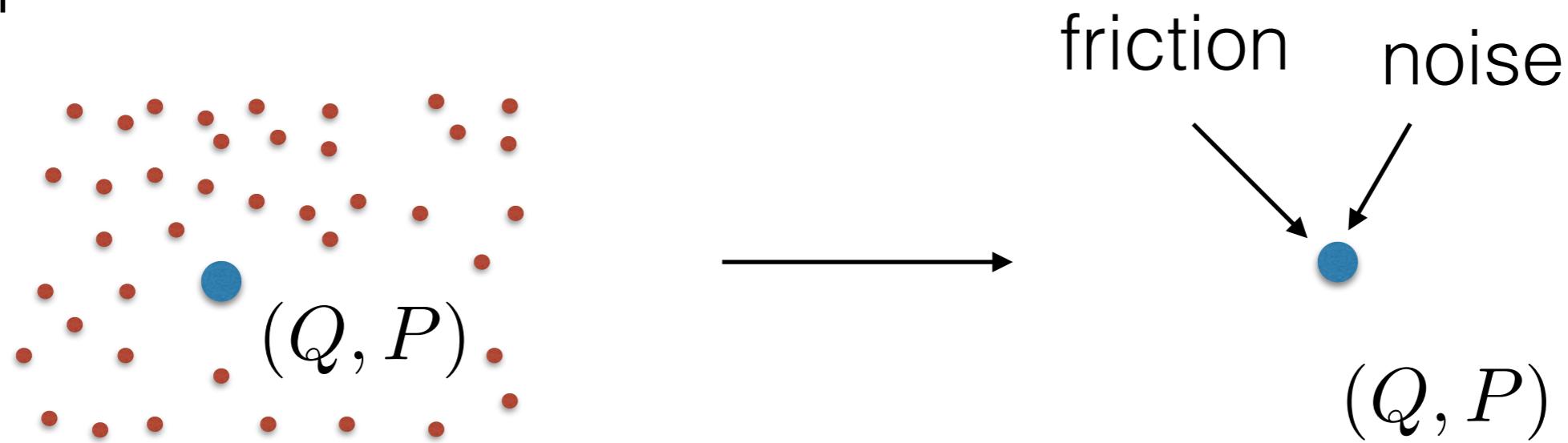


Typically approach delta-function in the limit of large auxiliary system

**Noise process** typically models, in the weak sense, the effect of a Wiener process [See e.g. works of A. Stuart]

# Langevin Dynamics

Simplified model reduction



Limiting SDEs for distinguished particle

$$dQ = P dt$$

$$dP = -\nabla U(Q)dt - \gamma P dt + \sigma dW$$

# Fokker Planck Equation for LD

For a system...

$$dq = M^{-1} p dt,$$

$$dp = -\nabla U(q)dt - \gamma p dt + \sqrt{2\gamma k_B T} M^{1/2} dW.$$

Forward Kolmogorov or Fokker-Planck operator:

$$\mathcal{L}_{LD}^\dagger \rho = -\nabla_q \cdot (M^{-1} p \rho) + \nabla_p \cdot ([\nabla U(q) + \gamma p] \rho) + \gamma k_B T \Delta_M \rho,$$

$$\Delta_M \stackrel{\text{def}}{=} \nabla \cdot (M \nabla) = \sum_{i=1}^{N_c} m_i \frac{\partial^2}{\partial p_i^2}$$

Invariant distribution:

$$\rho(q, p) = e^{-\beta H(q, p)}$$

# Geometric Ergodicity of Langevin Dynamics

**Geometric convergence** relies on establishing:

(1) “**minorization condition**”

(2) “**stability condition**”

Key part of (1): existence of a smooth transition density, usually shown by demonstrating that the operator is **hypoelliptic** by use of a parabolic **Hörmander condition**.

(2) usually shown by obtaining a **Lyapunov function**

# Hypoelliptic Property

$$dX = b_0(X)dt + \sum_{i=1}^k b_i(X)dW_i \in \mathbb{R}^n$$

$\mathcal{L}^\dagger$  is hypoelliptic in  $\Omega$  if the SDE satisfies the parabolic **Hörmander condition** in  $\Omega$

$$\text{Span} (b_{i=1}^k, [b_i, b_j]_{i=1,2,\dots,k, j=0,1,\dots,k}, [b_i, b_j, b_k], \dots) = \mathbb{R}^n$$

*A mixing condition for SDEs.*

**Langevin dynamics:**

$$\partial_{q_i} = \mathbf{e}_i, \quad \partial_{p_i} = \mathbf{e}_{i+n}$$

$$[\partial_{p_i}, p_i \partial_{q_i}] = -\partial_{q_i}$$

# Lyapunov Function

A Lyapunov function for the SDE is defined as a function  $\phi$  satisfying

$$\mathcal{L}\phi \leq -\alpha\phi + \delta$$

for some positive constants  $\alpha, \delta$

For Langevin dynamics, as shown by **[Mattingley, Higham and Stuart, 2002]**, we may use

$$\phi(q, p) = H^l(q, p)$$

We can choose  $l$  to bound a sufficiently large class of observables, giving exponential convergence for that class.

Stronger assumption: bounded  $q$ , e.g. Periodic BCs  
Lyapunov function  $1 + |p|^{2s}$

Decay (“geometric ergodicity”)

$$\|\mathrm{e}^{t\mathcal{L}_\gamma}\|_{\mathcal{B}(\mathcal{H}^1)} \leq K_\gamma \mathrm{e}^{-\lambda_\gamma t},$$

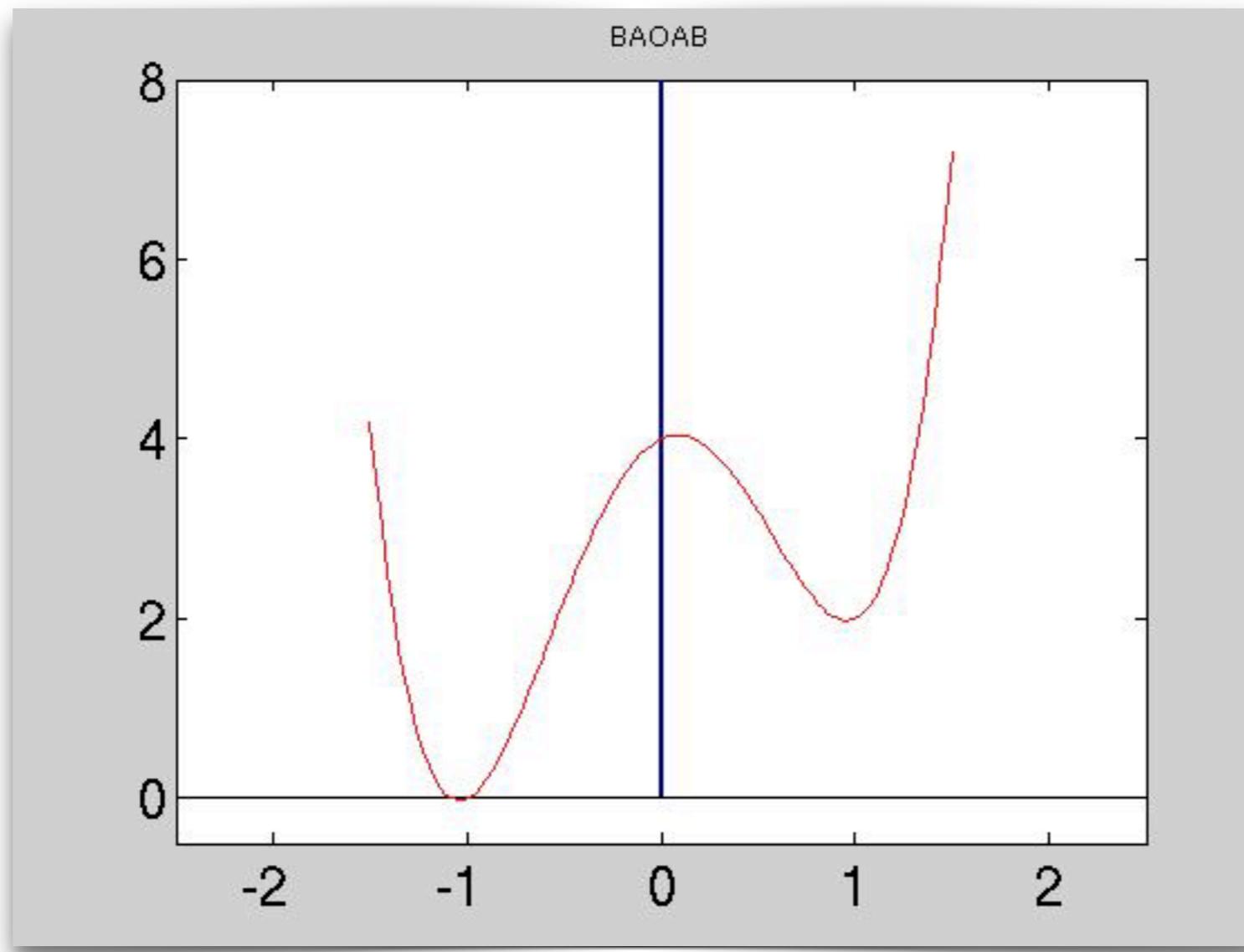
Regularity

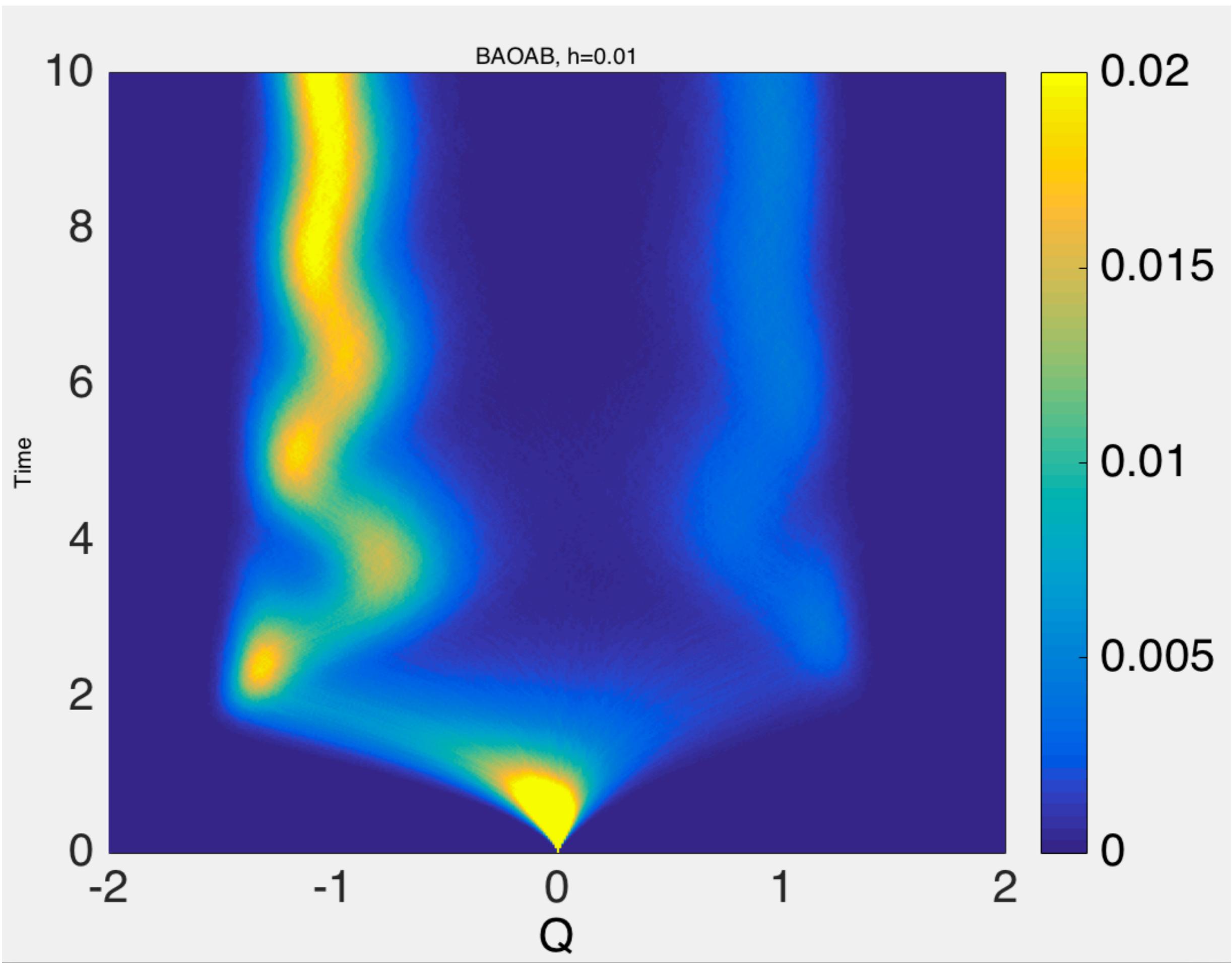
$$\|\mathcal{L}_\gamma^{-1}\|_{\mathcal{B}(\mathcal{H}^1)} \leq \frac{K_\gamma}{\lambda_\gamma}$$

# Langevin Dynamics - Double Well Potential

$$\rho_t = \mathcal{L}^\dagger \rho$$

A. Martinsson





# Splitting Methods

$$\begin{aligned} dq &= pdt \quad (\text{A}) \quad (M=I) \\ dp &= F(q)dt \quad (\text{B}) \quad -\gamma pdt + \sqrt{2\beta^{-1}\gamma} dW \quad (\text{O}) \end{aligned}$$

$$\mathbf{A} : \quad \dot{q} = p, \quad \dot{p} = 0$$

$$\mathbf{B} : \quad \dot{q} = 0, \quad \dot{p} = F(q)$$

$$\mathbf{O} : \quad \dot{q} = 0, \quad \dot{p} = -\gamma p + \sqrt{2\beta^{-1}\gamma} \dot{W}$$

**Time stepsize**  $\delta t$

$$p := e^{-\gamma\delta t} p + \sqrt{(1 - e^{-2\gamma\delta t})\beta^{-1}} \mathcal{N}(0, 1)$$

$$\mathbf{ABO} \quad p := \delta t F(q)$$

$$q := q + \delta t p$$

# Splitting Methods

$$\mathcal{L} = \mathcal{A} + \mathcal{B} + \mathcal{O}$$

$$\mathcal{A} = p^T M^{-1} \nabla_q \quad \mathcal{B} = -\nabla U(q)^T \nabla_p \quad \mathcal{O} = -\gamma p^T M^{-1} \nabla_p + \beta^{-1} \Delta_p$$

Propagator:

$$\mathcal{P}_t = e^{t\mathcal{L}}$$

Splitting Method:

$$\mathcal{P}_t \approx e^{t\mathcal{A}} e^{t\mathcal{B}} e^{t\mathcal{O}}$$

Drift   Kick   Shuffle      “ABO”

“OBA”   “OAB”   “ABOBA”   “OBABO” ...

# In L., Matthews & Stoltz 2015

after Talay 2002, Mattingly 2002,

Bou-Rabee & Owhadi 2010,

Hairer and Mattingley 2010, Kopec 2014

$$\|\mathcal{P}_{\delta t}^n\|_* \leq K e^{-\lambda n \delta t}$$

discrete propagator

$$\{\psi \mid \psi/(1 + |p|^{2s}) \in L^\infty\}$$

$$\left| (\mathcal{P}_{\delta t}^n f)(q, p) - \int_{\Omega} f d\mu_{\gamma, \delta t} \right| \leq K(1 + |p|^{2s}) e^{-\lambda n \delta t} \|f\|_{L_{1+|p|^{2s}}^\infty}$$

*Uniform in stepsize exponential decay  
i.e., geometric ergodicity of the numerical method*

# Invariant Measure of Numerical Method

The stable equilibrium distribution can be understood as a perturbation of the Gibbs distribution, with density

$$\hat{\rho}_{\beta, \gamma, \delta t} \approx \rho = e^{-\beta H}$$

The error in long-term averages is thus directly related to the quality of this approximation.

**How to calculate the error?**

**Ex:**

$$\mathcal{P}_{\delta t} = e^{\delta t \mathcal{A}} e^{\delta t \mathcal{B}}$$

**BCH:**

$$\mathcal{P}_{\delta t} = e^{\delta t [\mathcal{L} + \delta t \mathcal{L}_1 + \delta t^2 \mathcal{L}_2 + \dots]}$$

**Invariant density:**

$$[\mathcal{L}^\dagger + \delta t \mathcal{L}_1^\dagger + \delta t^2 \mathcal{L}_2^\dagger + \dots] \hat{\rho} = 0$$

**For symmetric splittings:**

$$[\mathcal{L}^\dagger + \delta t^2 \mathcal{L}_2^\dagger + \dots] \hat{\rho} = 0$$

**Proposal:**  $\hat{\rho} = \rho_\beta (1 + \delta t^2 d_2 + \dots)$

# Expansion of the invariant distribution

(Talay-Tubaro expansion<sup>1</sup> in the ergodic limit)

$$[\mathcal{L}^\dagger + \delta t^2 \mathcal{L}_2^\dagger + \dots] e^{-\beta(H + \delta t^2 f_2 + \dots)} = 0$$

Leading order:

$$\mathcal{L}^\dagger(\rho_{\text{can}} f_2) = \beta^{-1} \mathcal{L}_2^\dagger \rho_{\text{can}}$$

**L. & Matthews, AMRX, 2013**

**L., Matthews, & Stoltz, IMA J. Num. Anal. 2015**

- detailed treatment of all 1st and 2nd order splittings
- estimates for the operator inverse discrete inv. measure
- treatment of nonequilibrium (e.g. transport coefficients)

<sup>1</sup>Denis Talay and Luciano Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Analysis and Applications* 4 (1990)

# Operator expansions from Baker-Campbell-Hausdorff:

$$e^{\frac{\delta t}{2}X} e^{\frac{\delta t}{2}Y} e^{\delta tZ} e^{\frac{\delta t}{2}Y} e^{\frac{\delta t}{2}X} = e^{\delta tS},$$

where

$$\begin{aligned} S = & X + Y + Z + \frac{\delta t^2}{12} ([Z, [Z, Y + X]] + [Y, [Y, X]] + [Z, [Y, X]] + [Y, [Z, X]] \\ & - \frac{1}{2} [Y, [Y, Z]] - \frac{1}{2} [X, [X, Z]] - \frac{1}{2} [X, [X, Y]]) + \mathcal{O}(\delta t^4). \end{aligned}$$

e.g. for **BAOAB**:

$$\begin{aligned} \mathcal{L}_2^\dagger \rho_\beta = & \rho_\beta \left[ \frac{\gamma}{4} (\Delta_q U(q) - \beta p^T U''(q)p) \right. \\ & \left. + \frac{\beta}{4} p^T U''(q) \nabla U(q) - \frac{\beta}{12} p \cdot \nabla_q p^T U''(q)p \right]. \end{aligned}$$

# Two level expansion

For each of **ABOBA** and **BAOAB**, we find the first terms of the two-level expansion

$$\hat{\rho} = \exp(-\beta[H + \delta t^2(f_{2,0} + f_{2,1}\varepsilon + O(\varepsilon^2)) + O(\delta t^4)])$$

$$\varepsilon = 1/\gamma$$

$$f_{2,0} \equiv f_{2,0}^{\text{BAOAB}} = \frac{1}{8} (p^T U''(x) p - \beta^{-1} \Delta U(x)),$$

$$f_{2,1} \equiv f_{2,1}^{\text{BAOAB}} = \frac{1}{24} \beta^{-1} p^T \nabla_x \Delta_x U(x) - \frac{1}{72} p^T \nabla_x p^T U''(x) p,$$

$$f_{2,2} \equiv f_{2,2}^{\text{BAOAB}} = \frac{1}{296} p^T \nabla_x p^T \nabla_x p^T U''(x) p - \frac{1}{48} \nabla U(x) \cdot \nabla_x p^T U''(x) p.$$

# Configurational Sampling

Configurational sampling means computation of the marginal distribution in positions, or **averages of  $q$ -dependent quantities.**

In the typical case the order of accuracy of the configurational distribution of the numerical method is the same as that of the phase space distribution.

But in the **high friction limit**, in computing the marginal distribution it is possible to eliminate the leading term.

# Configurational Sampling

Integrate out with respect to momenta...and  
discover a **surprise**:

*Proposition:*

The marginal (configurational) distribution of the  
**BAOAB** method has an expansion of the form

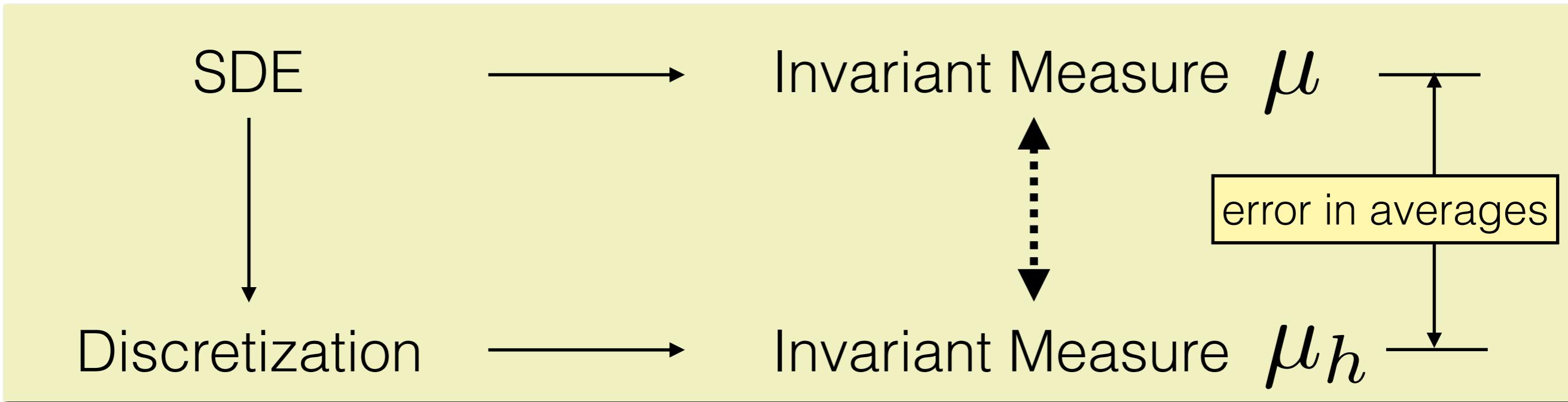
$$\bar{\rho} = e^{-\beta U} (1 + O(\delta t^2 \varepsilon^2) + O(\delta t^4))$$

In the high friction limit: 4th order, and with just one  
force evaluation per timestep.

# The Magic Cancellation

$$\begin{aligned} & \int \exp \left( -\beta \left[ \frac{1}{2} p^T (I + \frac{\delta t^2}{4} U'') p + U - \beta^{-1} \frac{\delta t^2}{8} \Delta U \right] \right) d^N p \\ &= \sqrt{2\pi k_B T / \det \left( I + \frac{\delta t^2}{4} U'' \right)} \exp \left( -\beta \left[ U - \beta^{-1} \frac{\delta t^2}{8} \Delta U \right] \right) \\ & \det A = \exp(\text{trace}(\log(A))) \\ & \propto \exp \left( -\frac{1}{2} \text{trace} \left( \log \left( I + \frac{\delta t^2}{4} U'' \right) \right) \right) \\ & \quad \times \exp \left( -\beta \left[ U - \beta^{-1} \frac{\delta t^2}{8} \Delta U \right] \right). \\ & \quad \propto \exp(-\beta U + O(\delta t^4)) \end{aligned}$$

$$dq = -\nabla U(q)dt + \sqrt{2}dW_t$$



**OBABO**  $\xrightarrow{\gamma \rightarrow \infty}$  **Euler-Maruyama**

$$q_{n+1} = q_n + hF(q_n) + \sqrt{2h}R_n$$

**1st order in**  
 $h = \delta t^2 / 2$

**BAOAB**  $\xrightarrow{\gamma \rightarrow \infty}$  **Leimkuhler-Matthews** **2nd order!**

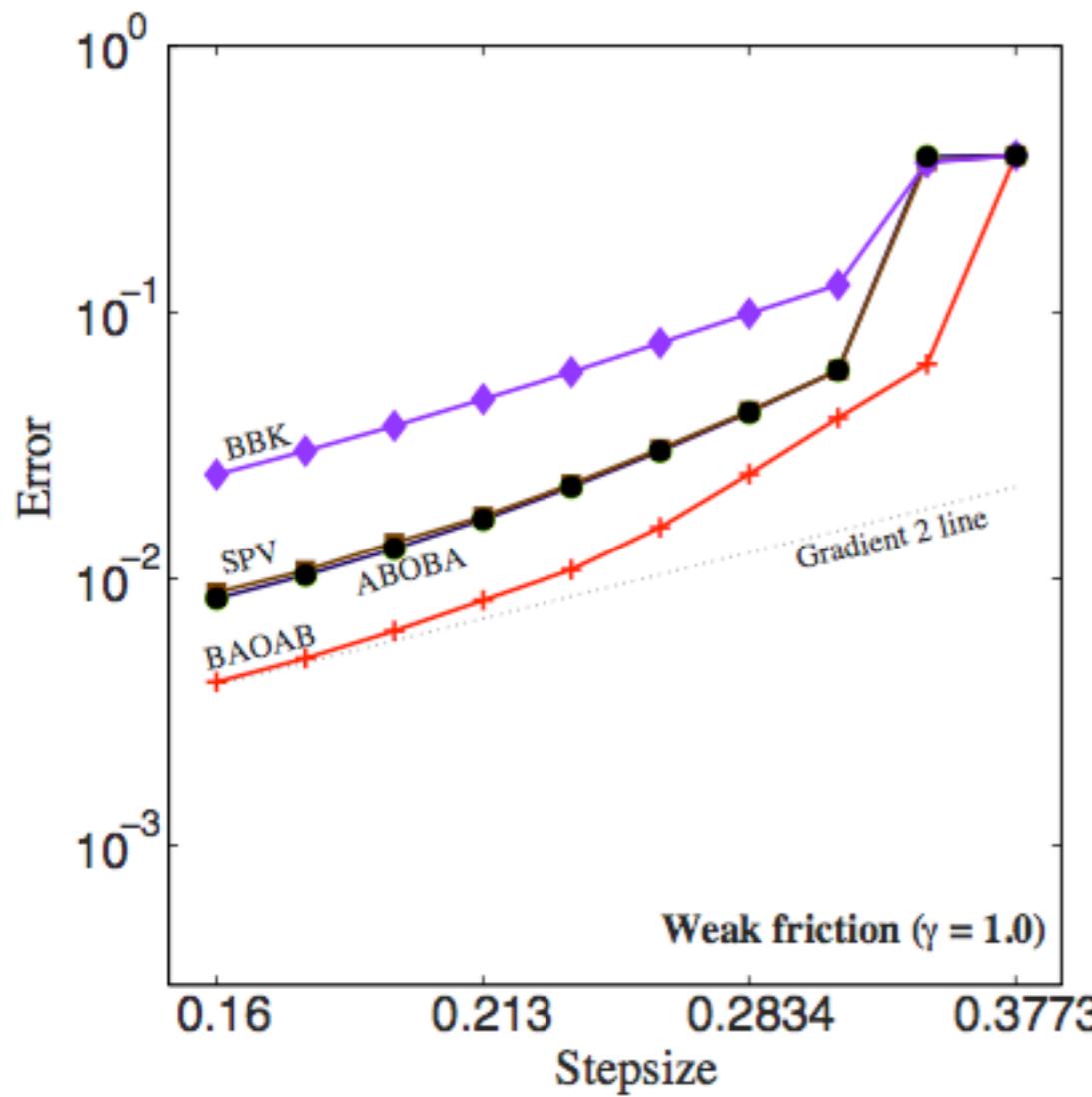
$$q_{n+1} = q_n + hF(q_n) + \sqrt{h/2}(R_n + R_{n-1})$$

# Perturbed Quartic (1D)

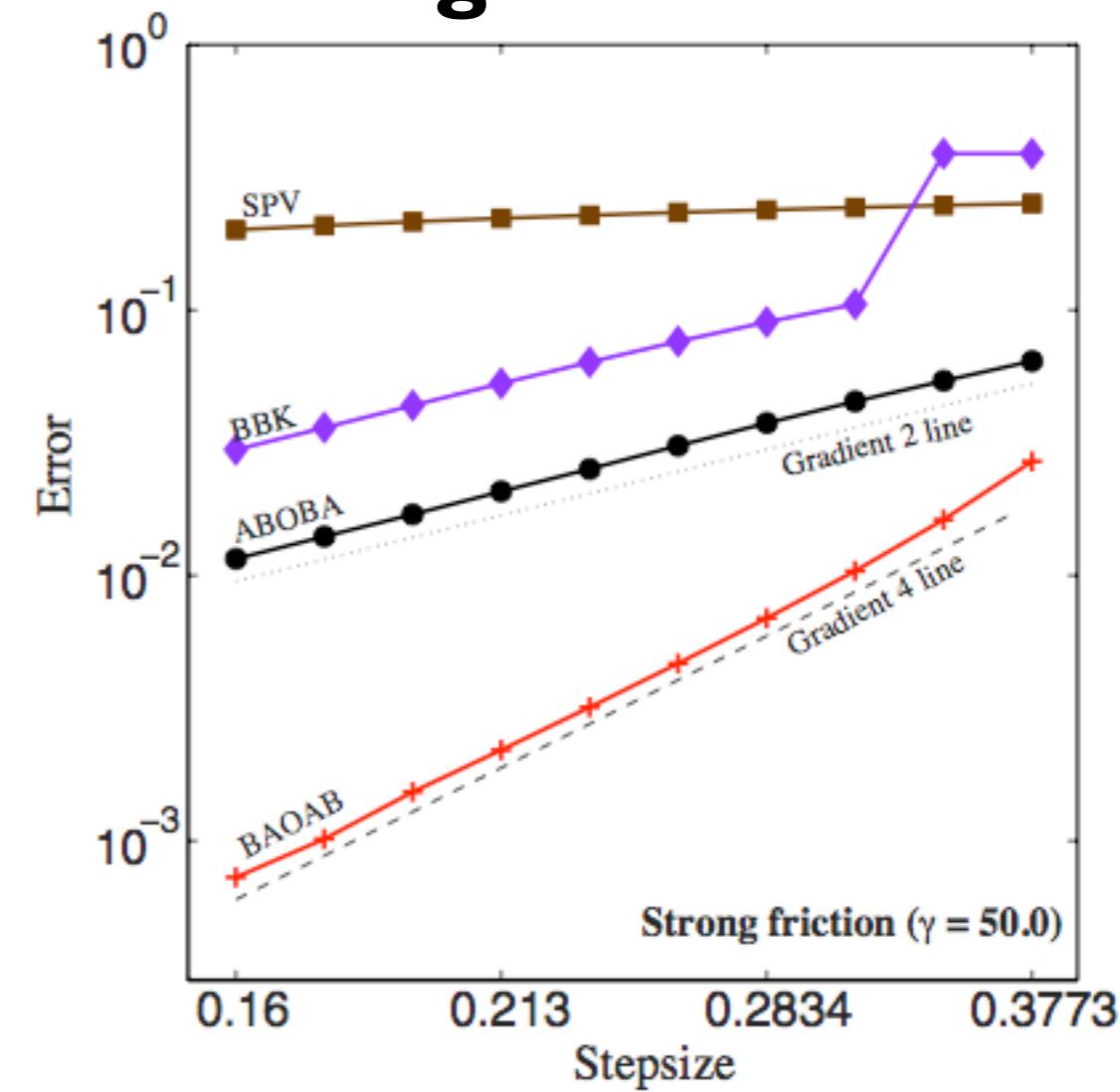
$$U(x) = x^4/4 + \sin(1+5x).$$

Langevin Configurational Dist. Accuracy vs timestep

**Mild Friction**

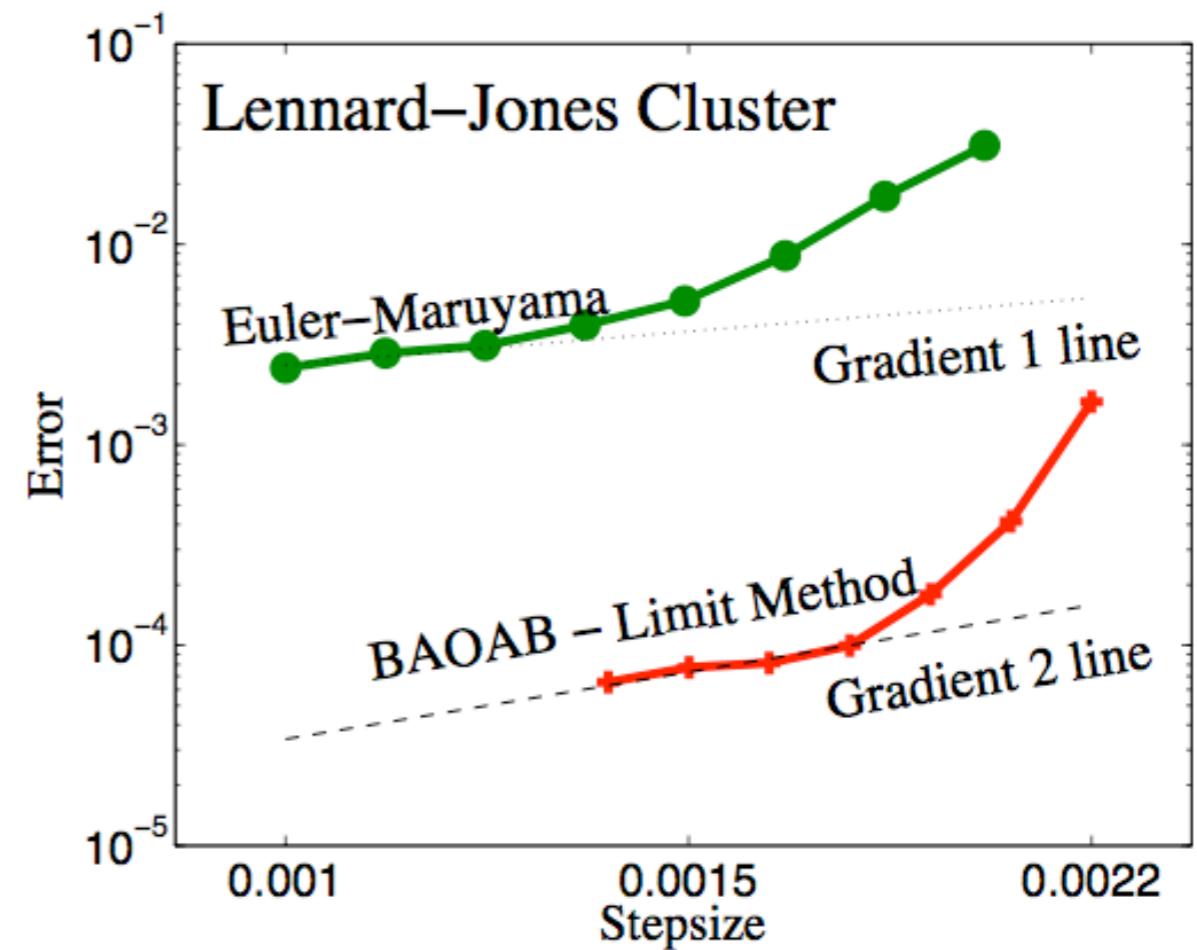
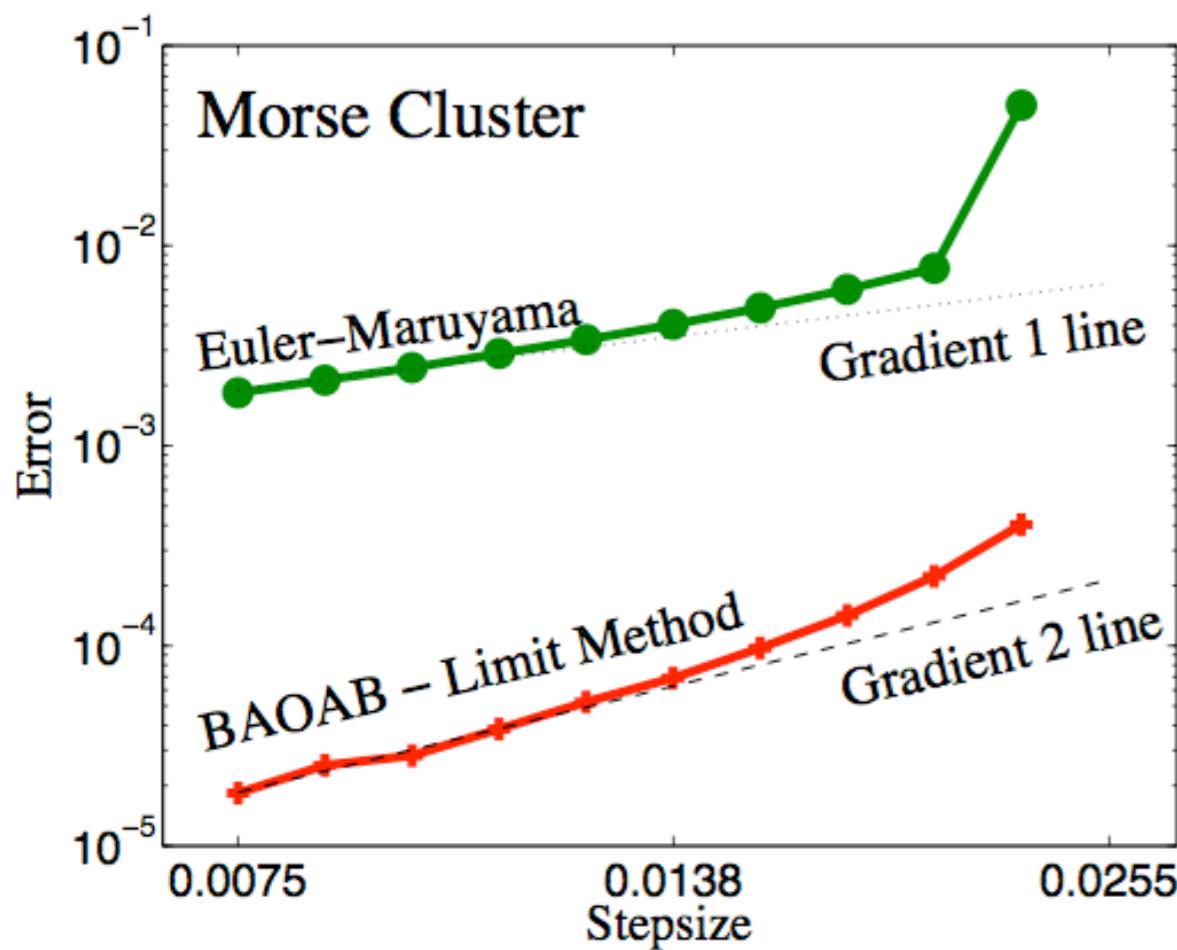
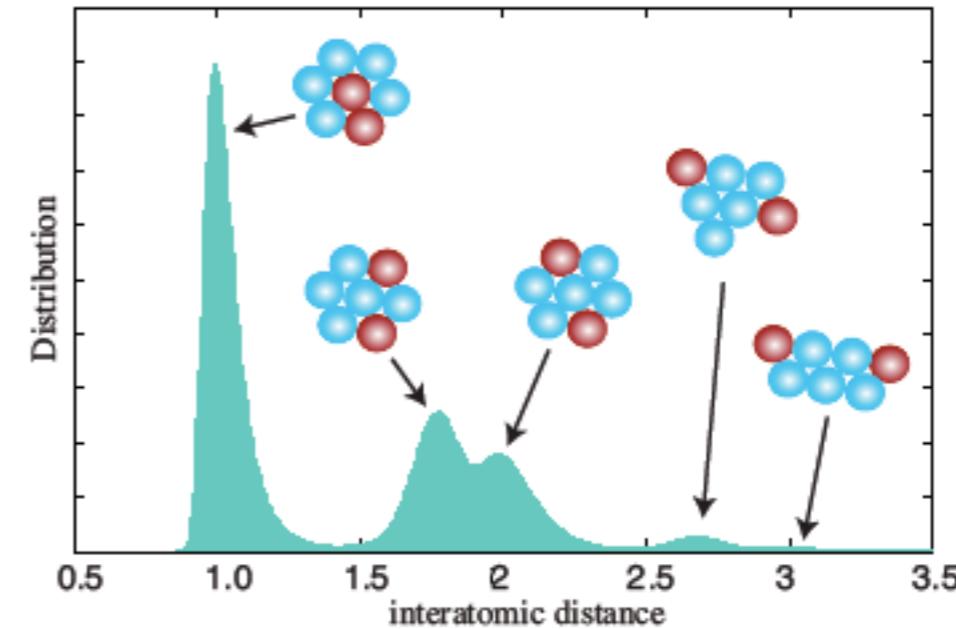


**High Friction**



# Morse and Lennard Jones Clusters

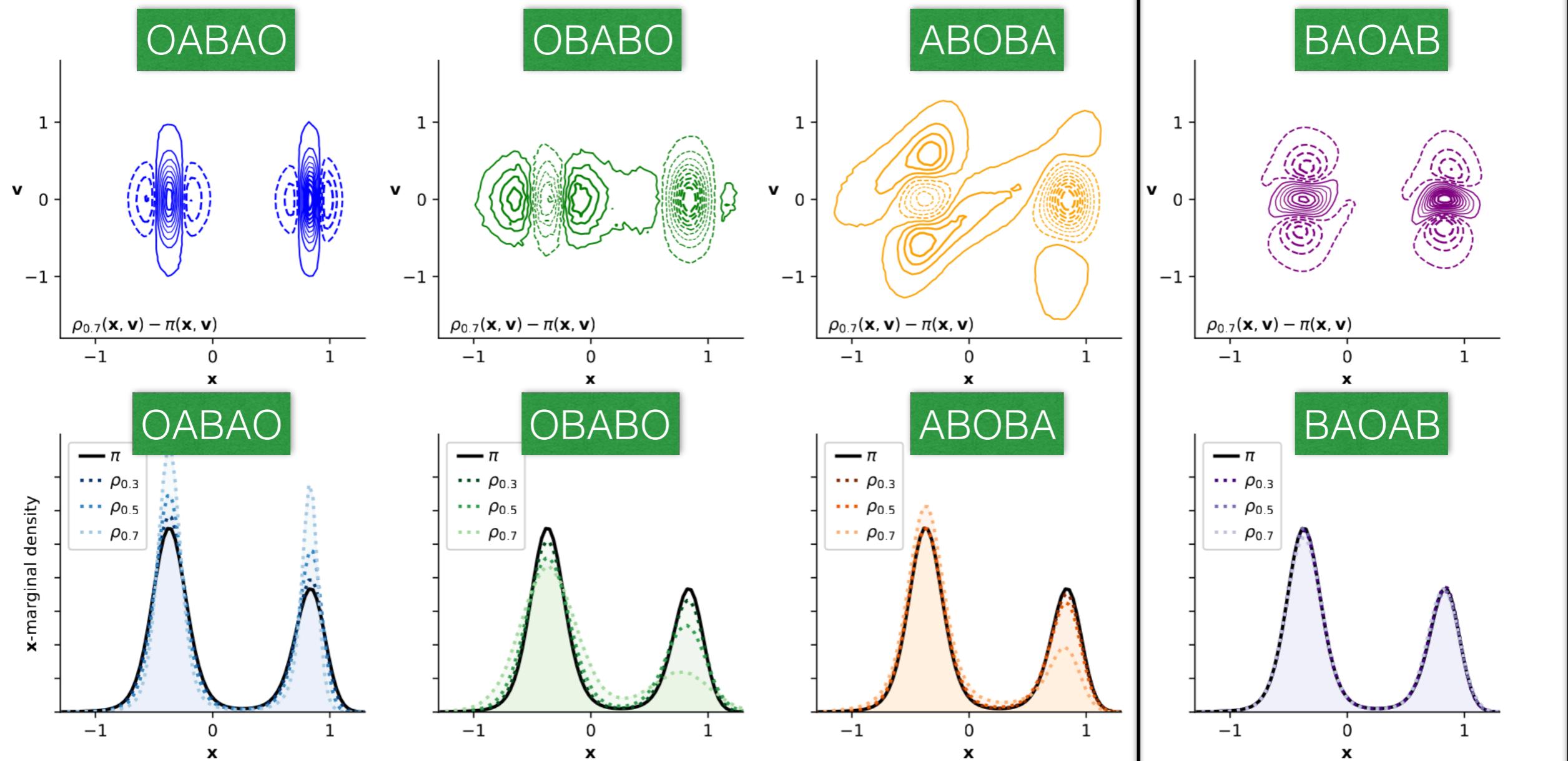
binned radial density



# Quantifying configuration-sampling error in Langevin simulations of complex molecular systems

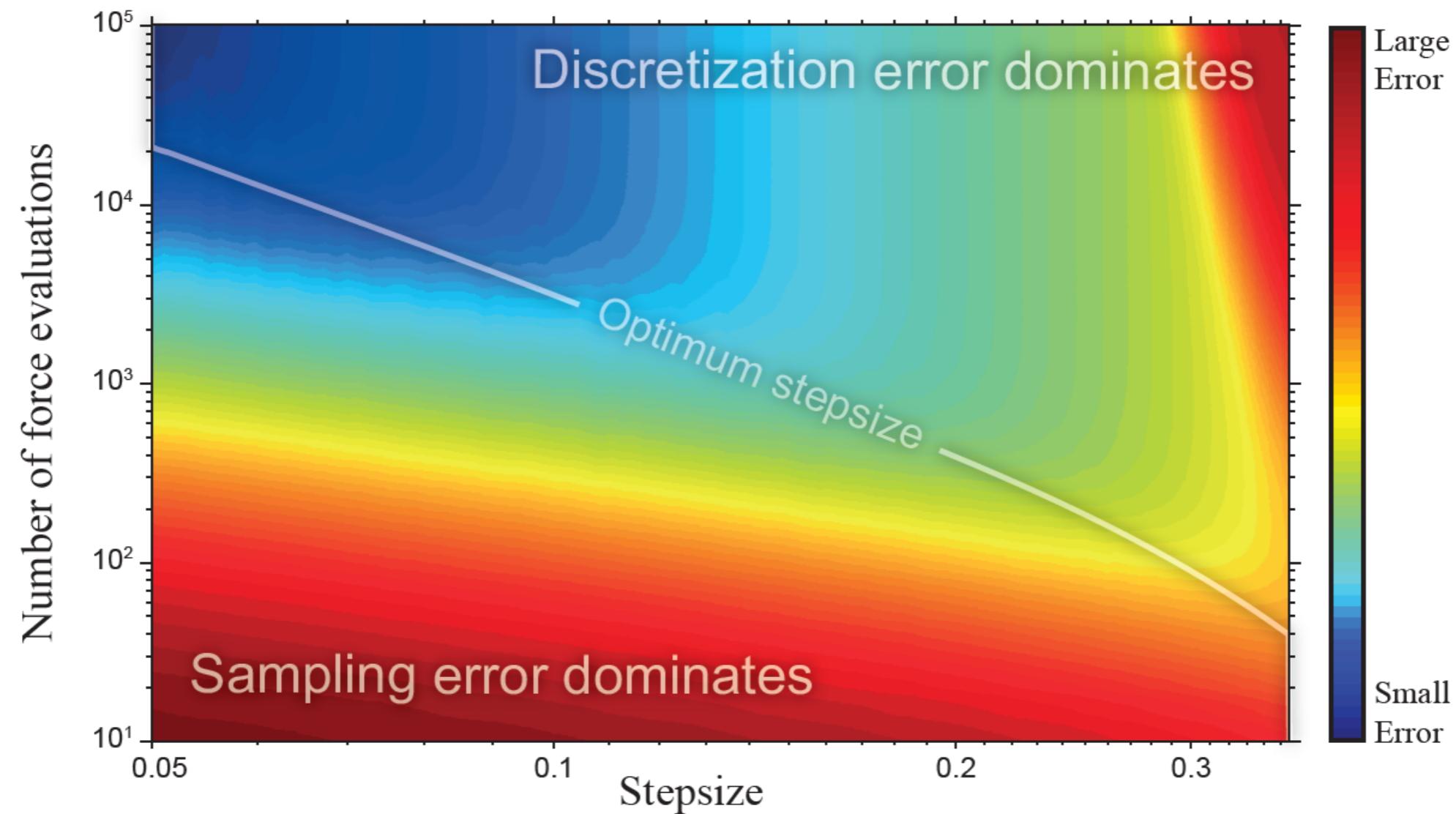
J. Fass, D.A. Sivak , G.E. Crooks , K.A. Beauchamp , B. Leimkuhler, J.D. Chodera, arxiv

## Error in phase space density



## Configurational Marginal

# Convergence and Discretization



**small stepsize:** convergence to accurate stationary state but convergence is slow relative to amount of computational work.

**large stepsize:** more rapid convergence for given computational effort, but the stationary distribution is more severely corrupted.

# Next Up

**Thursday** - Langevin simulations of biomolecules • pushing up stepsize in SDE numerics • multiple timestepping for SDEs • Constraints • g-BAOAB • isokinetic multiple timestepping

**Friday** - thermostats (deterministic and stochastic) • ergodic properties  
• applications - nonequilibrium and in data science