

Sampling Algorithms

“Microcource”

Session 1

Ben Leimkuhler (U of Edinburgh)
Tutorials by: Matthias Sachs (SAMSI) and
Anton Martinssen (U of Edinburgh)

Duke University, Spring 2018

Overview

Lecture 1: Monte Carlo, Hamiltonian systems, splitting methods, microcanonical sampling, SDEs, Euler-Maruyama & Leimkuhler-Matthews, Langevin splitting, superconvergence

Tutorial 1: methods for Langevin (and overdamped) dynamics

Lecture 2: Noisy gradients, then:
increasing the timestep, multiple timestepping,
holonomic constraints, isokinetic constraints & SIN,
ensemble preconditioning

Tutorial 2: enhancing the convergence to equilibrium/
decreasing the integrated autocorrelation time

Molecular Dynamics

2017: article in *Drug Discovery Today*

“Molecular dynamics-driven drug discovery: *leaping forward with confidence*”

“developing a market-approved drug costs a staggering \$2.6 billion”

“Given that the dynamics of the covalent bonds involving hydrogen atoms are not crucial in biological problems, they are usually **constrained** using integration algorithms, such as SHAKE, RATTLE, ... Hence, a time-step value in the range of **1.5–2 fs** is possible and has been shown to be suitable for MD simulations of many biological systems.”

Examples of some **2017** simulations out of around **20000** on google scholar

Ligand binding in the HIV-integrase enzyme

Druggability of membrane bound Rab5

Fission fragment damage in nuclear fuel

Patchy particle systems

Nanoparticle cholesterol metabolism therapeutics

Multiscale protein dynamics in antigen presentation

Bottlebrush copolymers

Modification of membrane structure by lithium

Smectic C semiflexible polymers

Graphene reinforced polymer nanocomposites

Unsaturated Zwitterionic lipids

Identifying the warfarin binding site



Bayesian Inference

- X_N is a sequence be of random variables with density $\pi(\cdot | \theta)$ with parameter $\theta \in \mathbb{R}^n$
- prior distribution $\pi(\theta)$ and observed values x_N of the random variables X_N
- sampling posterior distribution

$$\pi(\theta | x_N) \propto \pi(\theta) \prod_{i=1}^N \pi(x_i | \theta)$$

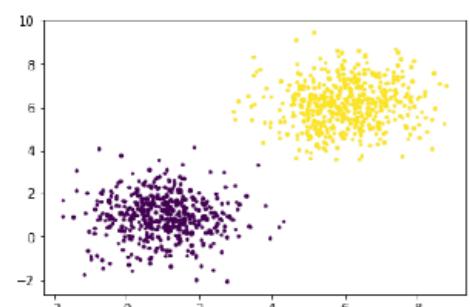
posterior →
 prior →
 likelihood

Examples

- Calculation of credibility regions around an estimate $\hat{\theta} \in A$

$$\mathbb{P}_{\pi(\theta|x_N)}(\theta \in A) = \mathbb{E}_{\pi(\theta|x_N)}[1_A(\theta)]$$

- Classification :



- paired data $x_i = (y_i, z_i)$ - predictor and response (class labels)
- likelihood

$$\pi(x_i | \theta) = \pi(z_i | y_i, \theta)$$

Posterior distribution $\pi(\theta | X) \propto \pi(X | \theta)\pi(\theta)$

Energy $U(\theta) := -\log(\pi(\theta | X)) = -\log(\pi(\theta)) - \log(\pi(X | \theta))$

Sampling

Physicists, chemists and data scientists are interested in calculating

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

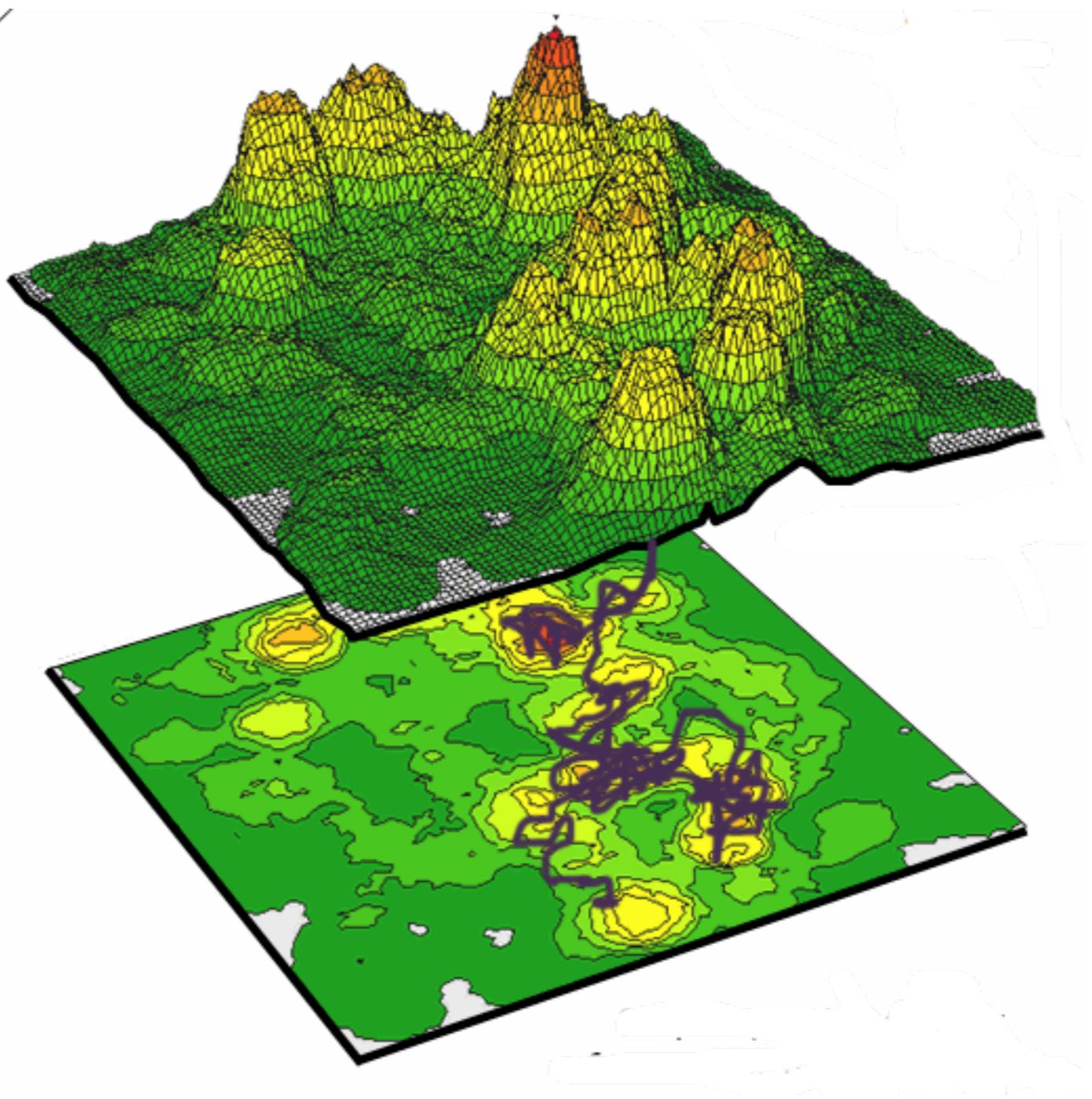
$\varphi(q)$ observable function

$d\mu(q) = \rho(q)dq$ *given density*

just integration...but with 180M variables!

target
density

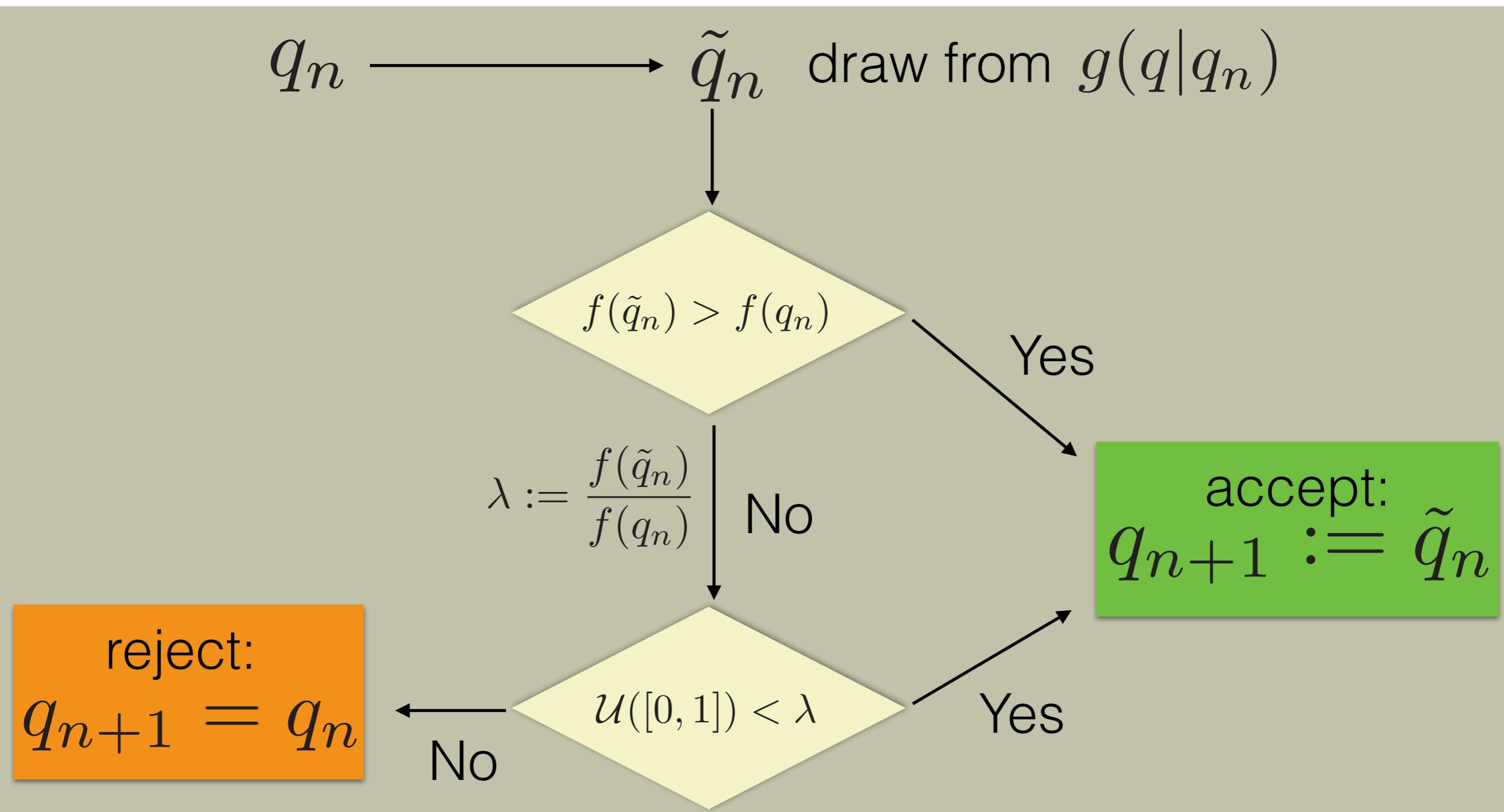
space filling
stochastic
path



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.

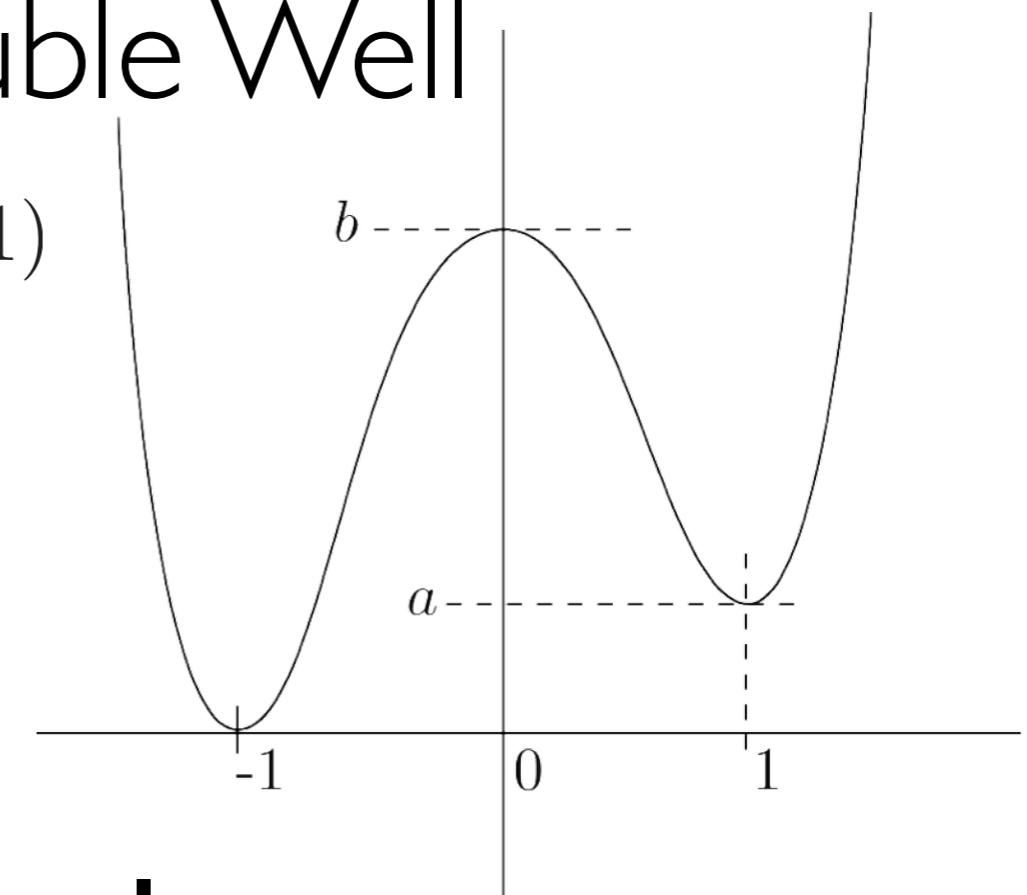
Example: Metropolis Monte Carlo for the Uneven Double Well

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

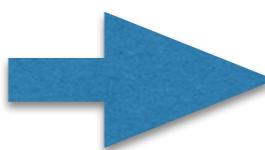
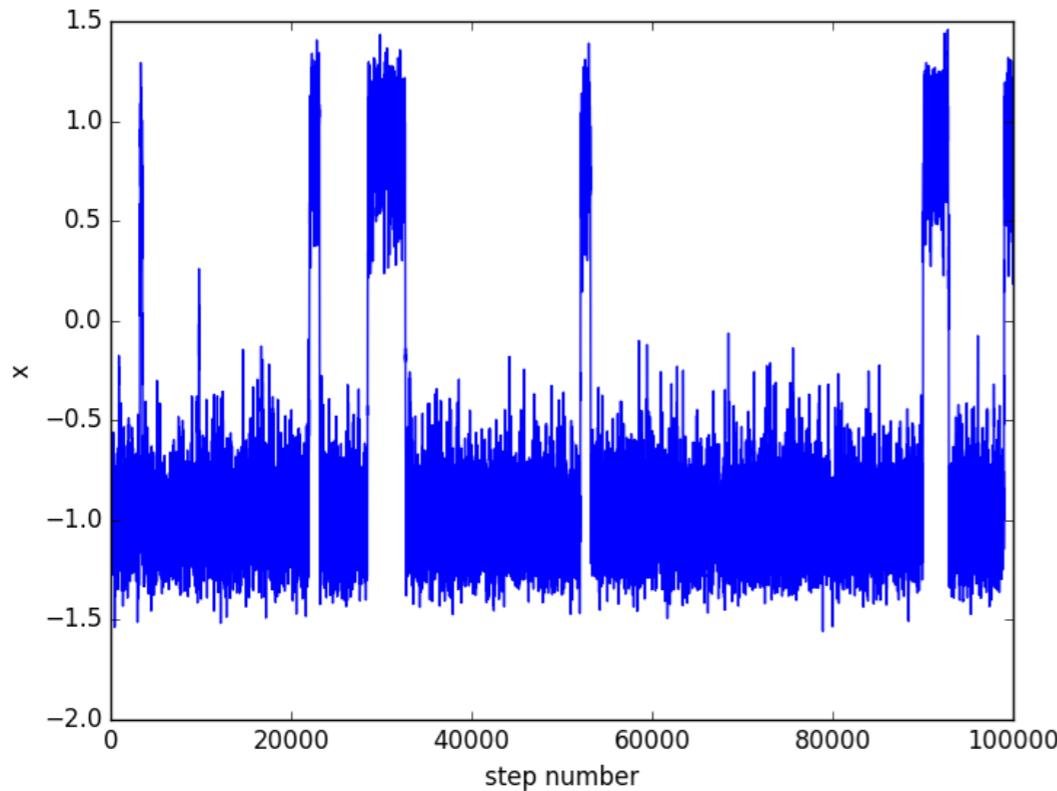
$$f(x) = e^{-\beta\varphi(x)}$$

Gaussian prior:

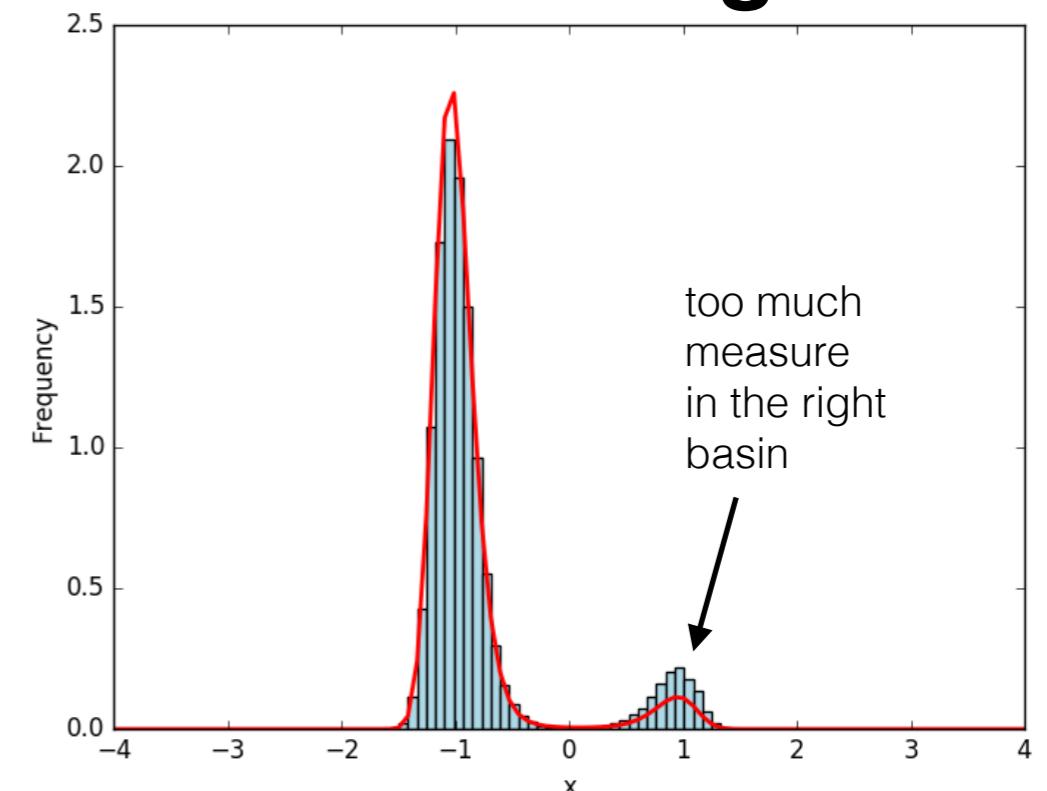
$$g(x|x') = e^{-\frac{(x-x')^2}{2\delta^2}}$$



“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).

*Microcanonical sampling
using Hamiltonian Systems
and Symplectic Methods*

Equations of Motion

Newton's equations (2nd order ODE system)

$$M\ddot{q} = -\nabla U(q)$$

1st order forms

position-velocity

$$\dot{q} = v$$

$$M\dot{v} = -\nabla U(q)$$

position-momentum

$$\dot{q} = M^{-1}p$$

$$\dot{p} = -\nabla U(q)$$

HAMILTONIAN:

$$H(q, p) \equiv p^T M^{-1} p / 2 + U(q)$$

conserved quantity

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

Often 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.

Leapfrog/Verlet Method

$$M(\mathbf{q}_{n+1} - 2\mathbf{q}_n + \mathbf{q}_{n-1}) = -h^2 \nabla U(\mathbf{q}_n)$$

or, in first order form (position-velocity),

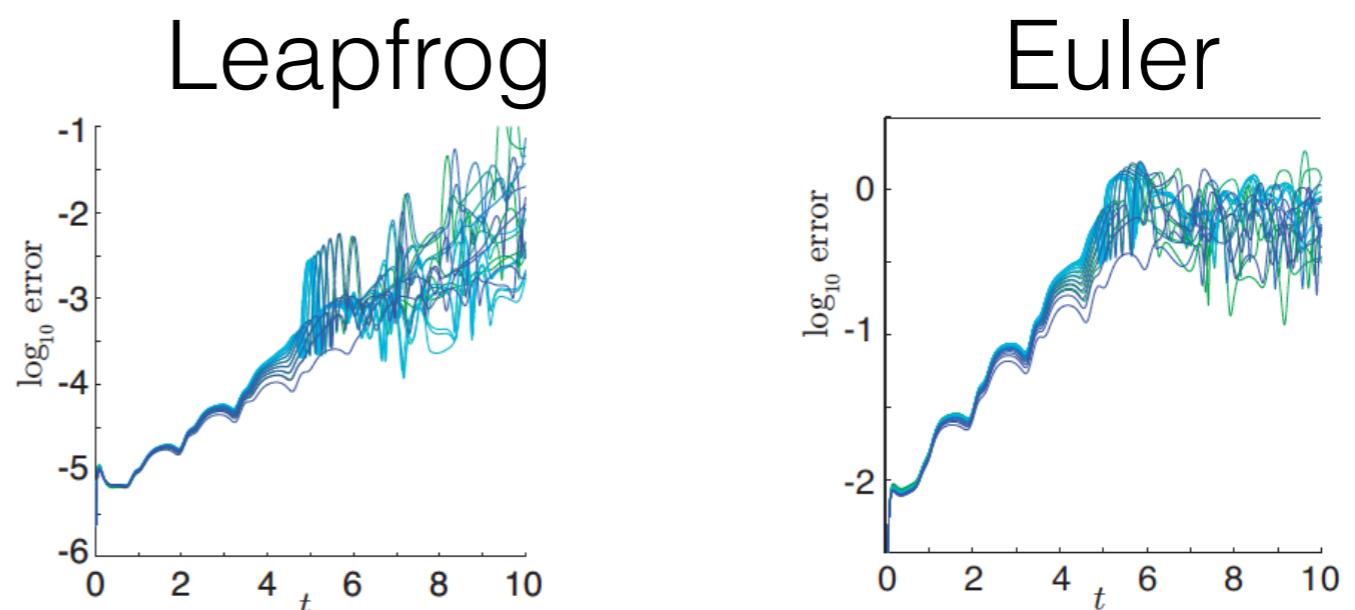
$$\mathbf{v}_{n+1/2} = \mathbf{v}_n + (h/2)M^{-1}\mathbf{F}_n,$$

$$\mathbf{q}_{n+1} = \mathbf{q}_n + h\mathbf{v}_{n+1/2},$$

$$\mathbf{v}_{n+1} = \mathbf{v}_{n+1/2} + (h/2)M^{-1}\mathbf{F}_{n+1}$$

- one force evaluation/timestep
- “2nd order” method

exponential
growth of error



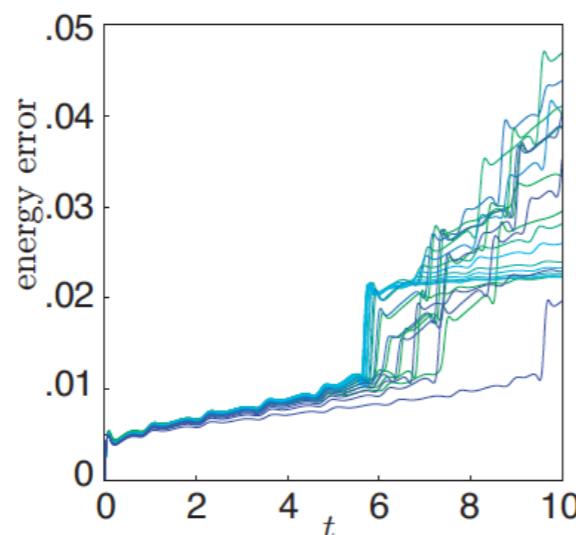
Energy Errors

Recall that the energy is a “conserved quantity” in molecular dynamics. We expect it to be constant along solutions.

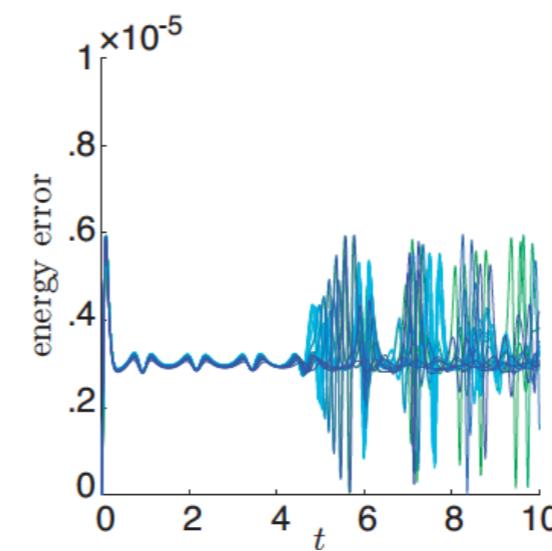
However, in discretization, it is not conserved.

Interestingly the behavior of energy error is very different for different schemes.

Euler



Leapfrog



Symplectic Map

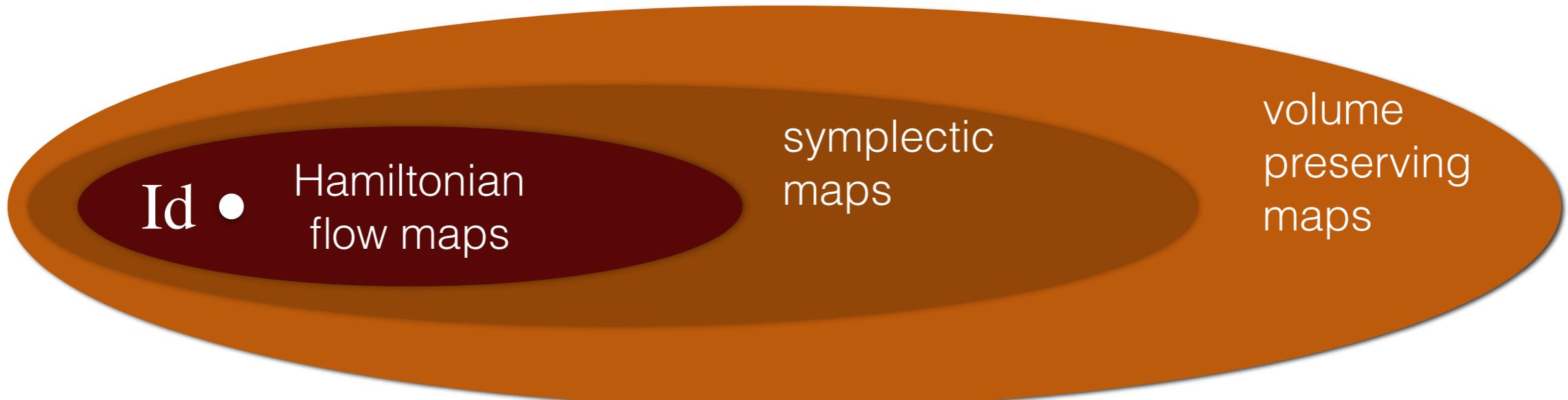
$$\Phi' J \Phi' = J \quad J^T = -J$$

“canonical” choice of J :

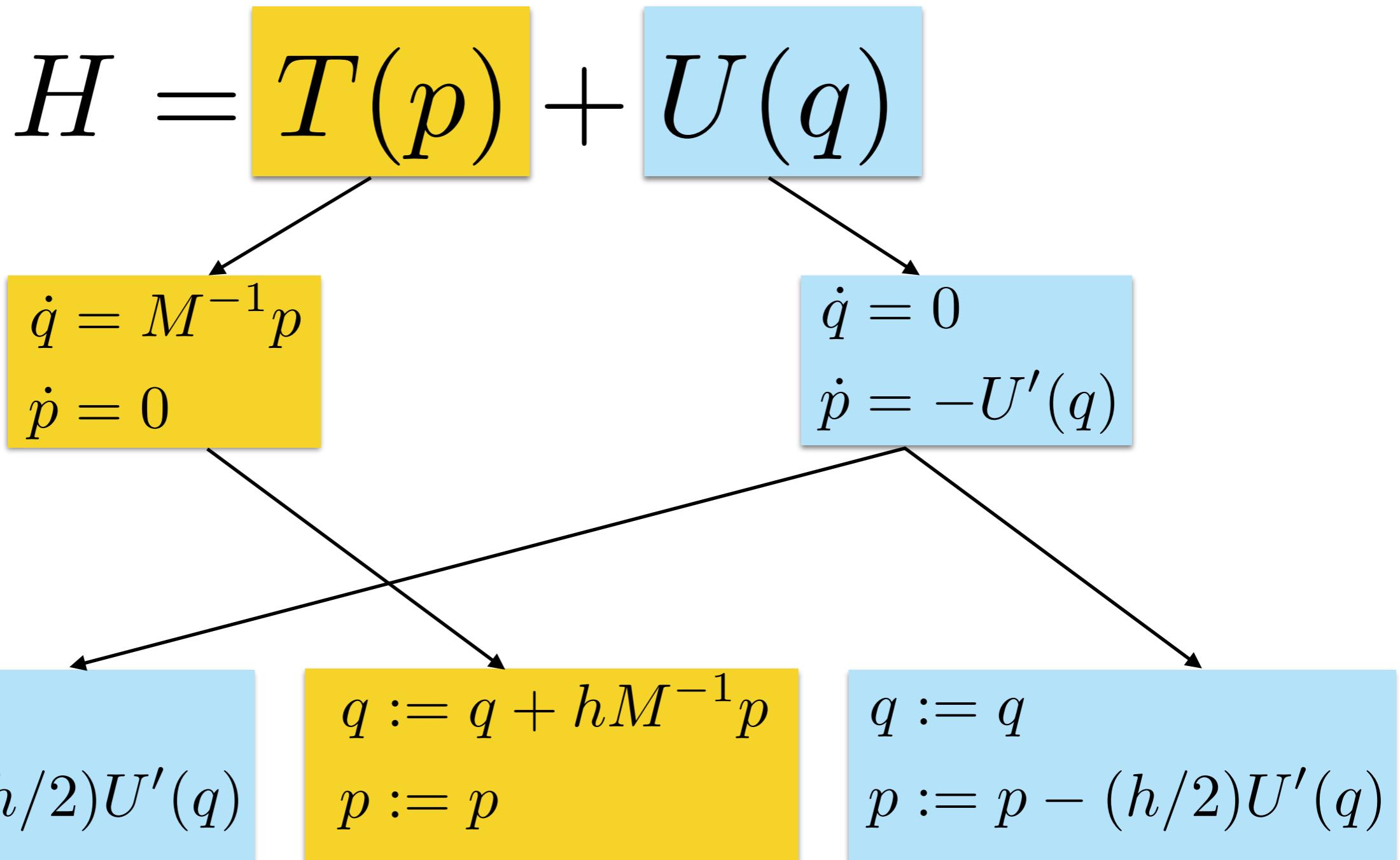
$$J = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}$$

canonical
two-form:

$$\Omega = \sum_{i=1}^d dq_i \wedge dp_i \quad \Omega(\xi, \eta) = \xi^T J \eta$$



Leapfrog/Verlet is a “splitting method”



Second order approximation
Symplectic

A Theorem about Symplectic Integrators

[Benettin and Giorgilli] *Given a symplectic numerical method Ψ_h , there is a modified Hamiltonian whose flow map $\tilde{\Phi}_h$ is exponentially close.*

[Benettin & Giorgilli 1994] On the Hamiltonian interpolation of near-to-the identity symplectic mappings with application to symplectic integration algorithms, J Stat Phys, 74, 1117-1143, 1994.

Lie Derivative

$$\frac{d}{dt} \begin{bmatrix} q \\ p \end{bmatrix} = \begin{bmatrix} \nabla_p H \\ -\nabla_q H \end{bmatrix} \quad \dot{z} = f(z)$$

$$g : \mathbb{R}^{2d} \rightarrow \mathbb{R} \quad \mathcal{L}_f g = f \cdot \nabla g$$

$$g(z(t)) = g(z(0)) + t \left. \mathcal{L}_f g \right|_{z=z(0)} + \frac{t^2}{2} \left. \mathcal{L}_f^2 g \right|_{z=z(0)} + \dots$$

$$\Phi_t g := g \circ \Phi_t = e^{t \mathcal{L}_f} g$$

Transportation of functions

$$\dot{z} = f_1(z) + f_2(z)$$

Numerical Method $\Psi_h = e^{h\mathcal{L}_{f_1}} e^{h\mathcal{L}_{f_2}}$

Flow Map $\Phi_h = e^{h\mathcal{L}_{f_1+f_2}} = e^{h(\mathcal{L}_{f_1} + \mathcal{L}_{f_2})}$

$$\Phi_h = 1 + h(\mathcal{L}_{f_1} + \mathcal{L}_{f_2}) + \frac{h^2}{2}(\mathcal{L}_{f_1} + \mathcal{L}_{f_2})^2 + \dots$$

$$\Psi_h = (1 + h\mathcal{L}_{f_1} + \frac{h^2}{2}\mathcal{L}_{f_1}^2 + \dots)(1 + h\mathcal{L}_{f_2} + \frac{h^2}{2}\mathcal{L}_{f_2}^2 + \dots)$$

$$\Psi_h = \Phi_h + \frac{h^2}{2}[\mathcal{L}_{f_1}, \mathcal{L}_{f_2}] + O(h^3)$$

local error

The Modified Equations

$$\Psi_h = e^{h\mathcal{L}_{\tilde{f}}}$$

$$\tilde{f} = f + hf^{(1)} + h^2f^{(2)} + \dots$$

$$e^{h\mathcal{L}_{\tilde{f}}} = 1 + h\mathcal{L}_{\tilde{f}} + \frac{h^2}{2}\mathcal{L}_{\tilde{f}}^2 + \dots$$

$$= 1 + h(\mathcal{L}_f + h\mathcal{L}_{f^{(1)}} + \dots) + \frac{h^2}{2}\mathcal{L}_f^2 + O(h^3)$$

$$\mathcal{L}_{f^{(1)}} + \frac{1}{2}\mathcal{L}_f^2 = \mathcal{L}_{f_1}\mathcal{L}_{f_2} + \mathcal{L}_{f_1}^2 + \mathcal{L}_{f_2}^2$$

$$\mathcal{L}_{f^{(1)}} = \frac{1}{2}[\mathcal{L}_{f_1}, \mathcal{L}_{f_2}] \quad f^{(1)} = \frac{1}{2}[f_1, f_2]$$

similarly, we may construct expressions for $f^{(2)}, f^{(3)}, \dots$

Baker-Campbell-Hausdorff Expansion

$$\log(\exp X \exp Y) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} \sum_{\substack{r_1+s_1>0 \\ \vdots \\ r_n+s_n>0}} \frac{[X^{r_1} Y^{s_1} X^{r_2} Y^{s_2} \dots X^{r_n} Y^{s_n}]}{\sum_{i=1}^n (r_i + s_i) \cdot \prod_{i=1}^n r_i! s_i!}$$

$$[X^{r_1} Y^{s_1} \dots X^{r_n} Y^{s_n}] = \underbrace{[X, [X, \dots [X,}_{r_1} \underbrace{[Y, [Y, \dots [Y,}_{s_1} \dots \underbrace{[X, [X, \dots [X,}_{r_n} \underbrace{[Y, [Y, \dots Y]] \dots]}_{s_n}]$$

$$\begin{aligned} Z(X, Y) &= \log(\exp X \exp Y) \\ &= X + Y + \frac{1}{2}[X, Y] + \frac{1}{12} ([X, [X, Y]] + [Y, [Y, X]]) \\ &\quad - \frac{1}{24} [Y, [X, [X, Y]]] \\ &\quad - \frac{1}{720} ([Y, [Y, [Y, [Y, X]]]] + [X, [X, [X, [X, Y]]]]) \\ &\quad + \frac{1}{360} ([X, [Y, [Y, [Y, X]]]] + [Y, [X, [X, [X, Y]]]]) \\ &\quad + \frac{1}{120} ([Y, [X, [Y, [X, Y]]]] + [X, [Y, [X, [Y, X]]]]) + \dots \end{aligned}$$

Verlet Example

$$H = T(p) + U(q)$$

Ansatz:

$$e^{(h/2)U} e^{hT} e^{(h/2)U} = e^{h\tilde{H}_h}$$

$$\tilde{H}_h = H + \frac{h^2}{12} \left(\{T, T, U\} - \frac{1}{2} \{U, U, T\} \right) + O(h^4)$$

$$\{A, B, C\} = \{A, \{B, C\}\} \quad \{A, B\} = \nabla A^T J \nabla B$$

Problem: Series expansion does not actually converge

Solution: Define an optimal truncation index

Energy Conservation of Symplectic Methods

Theorem: Suppose the Hamiltonian H and the truncated modified Hamiltonian \bar{H}_k are globally defined and smooth functions on a convex, compact subset \mathcal{B} of \mathbb{R}^{2d} . Suppose that the exact solution and numerical approximations are contained in \mathcal{B} for h sufficiently small. Then

$$H(z_n) = H(z_0) + O(h^r)$$

for $n = 0, 1, 2, \dots, \nu$, where $\tau = \nu h = O(h^{-k+r})$.

k : truncation index of perturbative series

r : order of the method

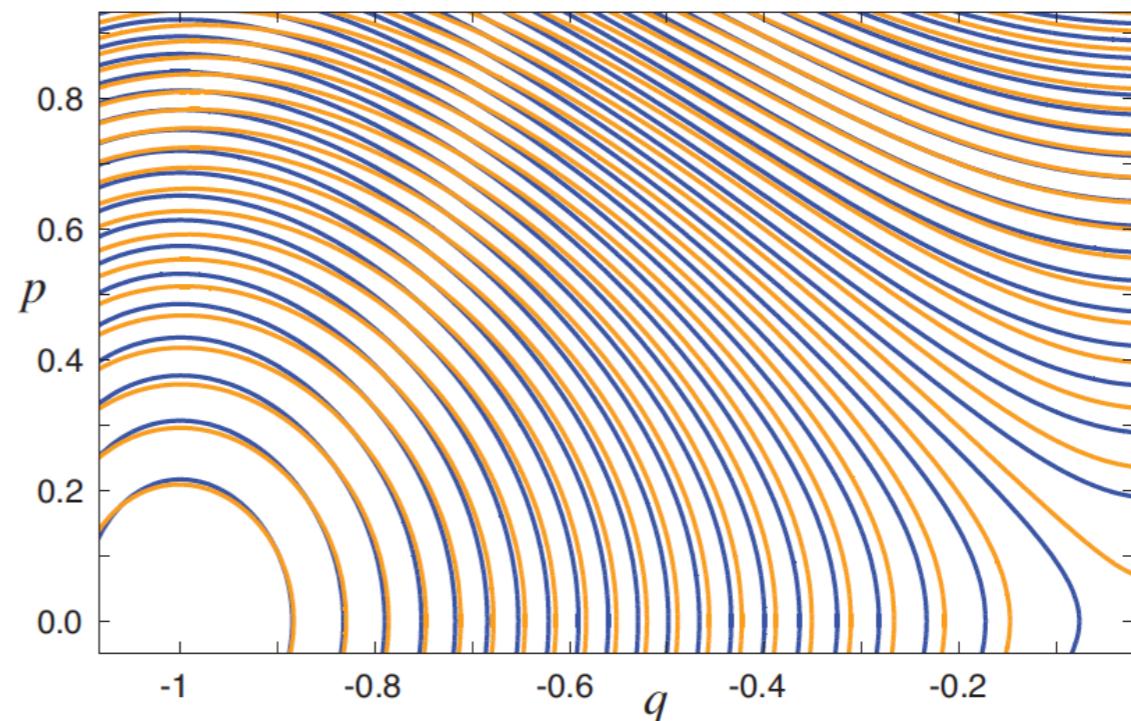
suggests energy conservation on long time intervals
(much longer than would be suggested by the method order)

Exercise

Calculate the terms of the Hamiltonian perturbation through 4th order in the stepsize for the Verlet method applied $H=p^2/2 + U(q)$ (1dof):

$$\begin{aligned}\tilde{H}_h = H &+ \frac{h^2}{24}(2p^T U'' p - (U')^2) \\ &+ h^4 \left(\frac{1}{720} p^4 U'''' - \frac{1}{120} p^2 U' U''' - \frac{1}{240} (U')^2 U'' - \frac{1}{60} p^2 ((U'')^2 + U' U''') \right) \\ &+ \mathcal{O}(h^6)\end{aligned}$$

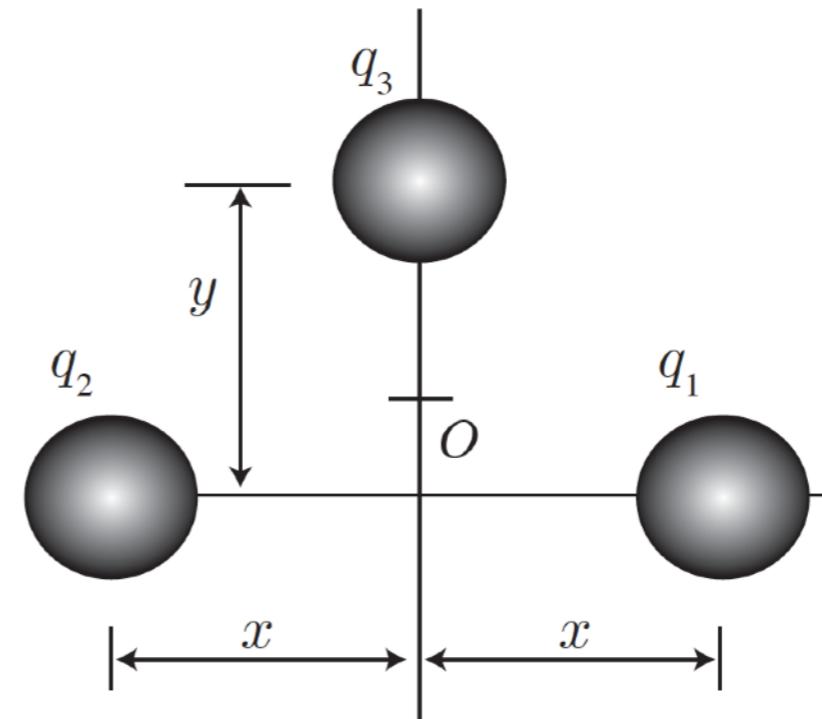
isoenergies:



MD Examples I: Trimer

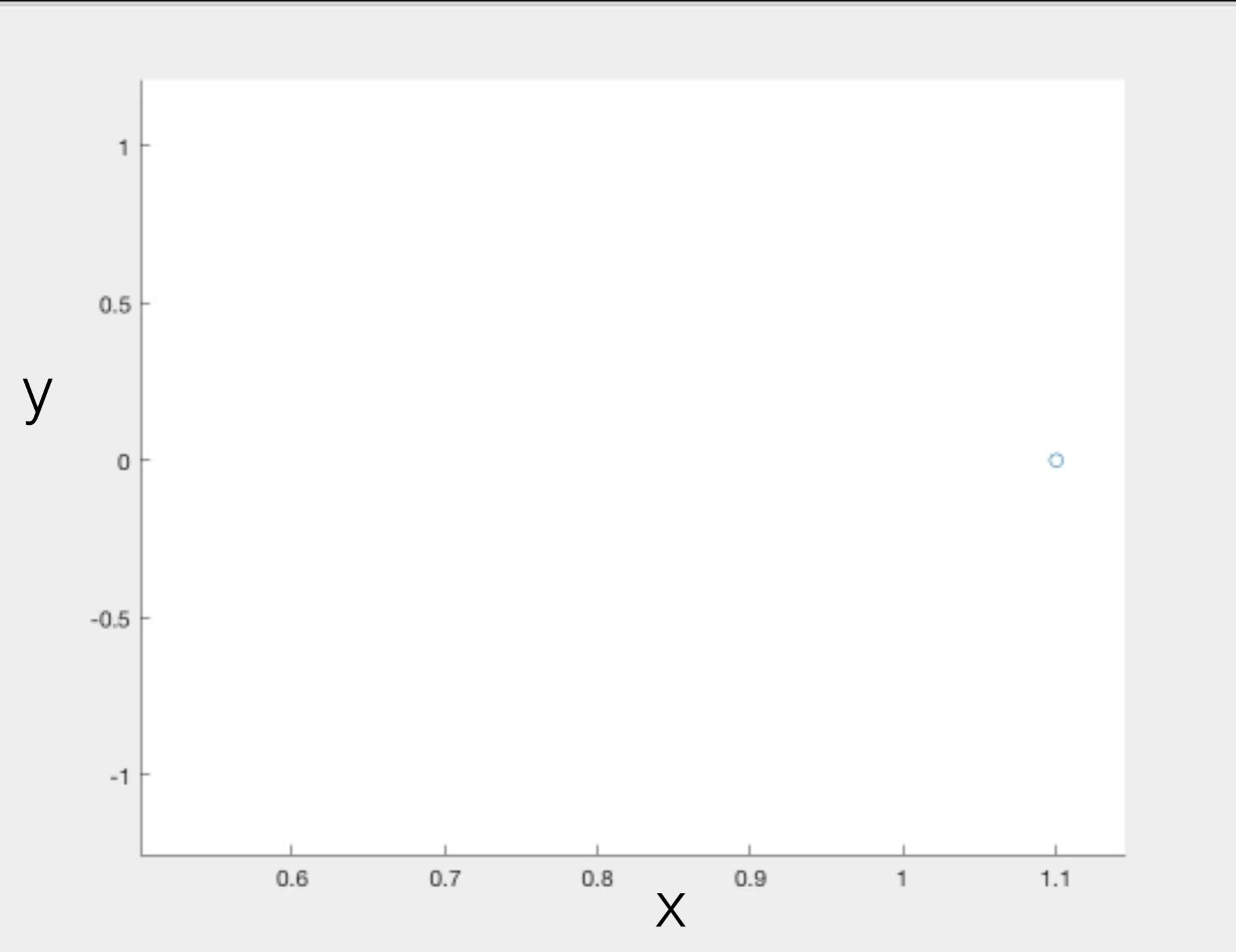
3 atoms moving in the plane
interactions governed
by Lennard-Jones
potentials

isosceles configuration

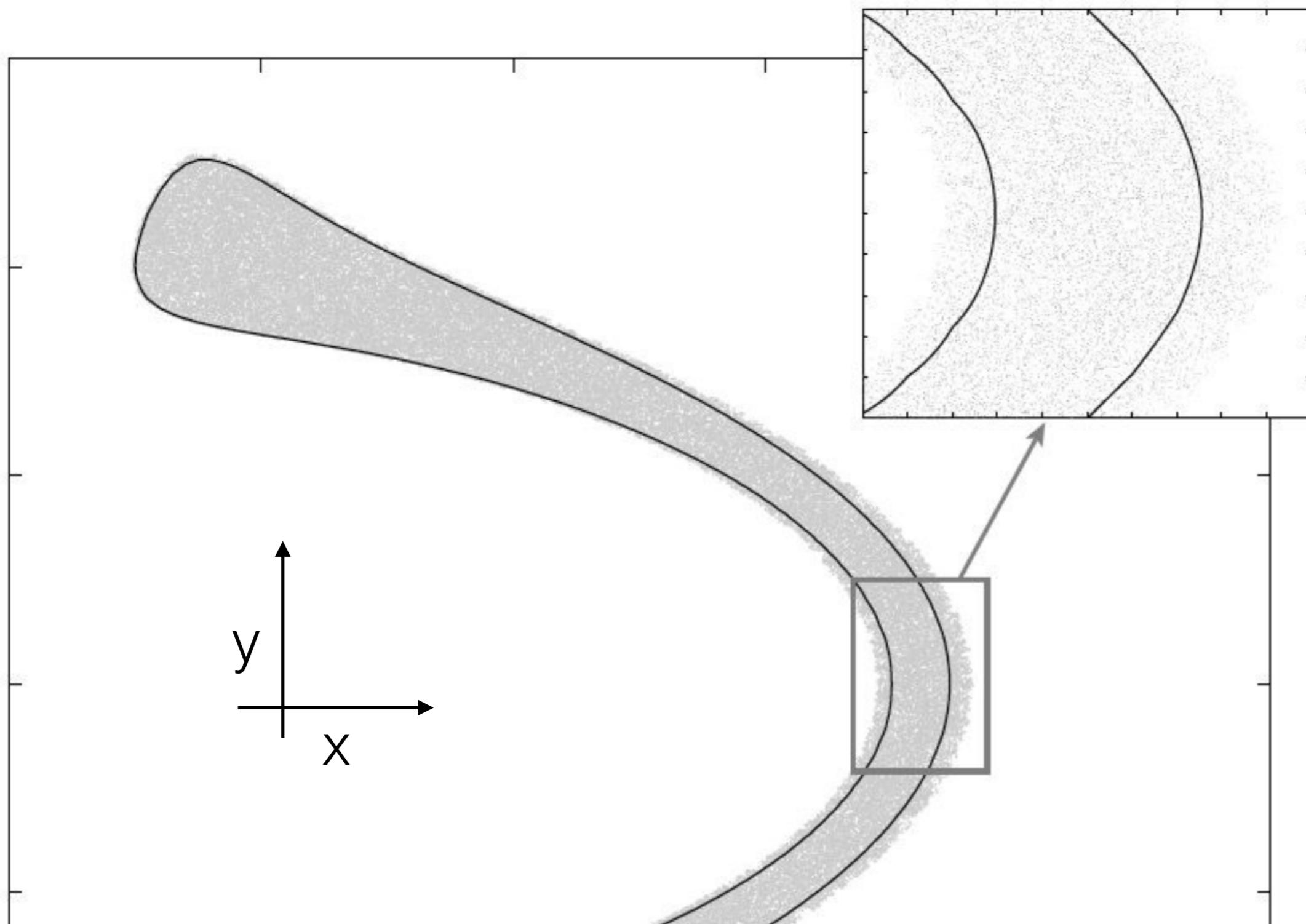


$$E = \dot{x}^2 + \frac{\dot{y}^2}{3} + 2\hat{\varphi}_{\text{LJ}}(\sqrt{x^2 + y^2}) + \hat{\varphi}_{\text{LJ}}(2x)$$

$$\hat{\varphi}_{\text{LJ}}(r) = r^{-12} - r^{-6}$$



Trimer — discretization error



ergodicity and sampling

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:

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

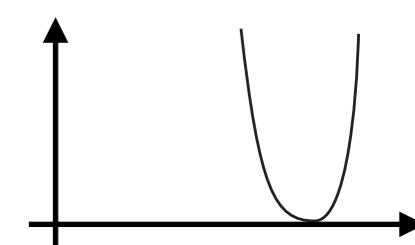
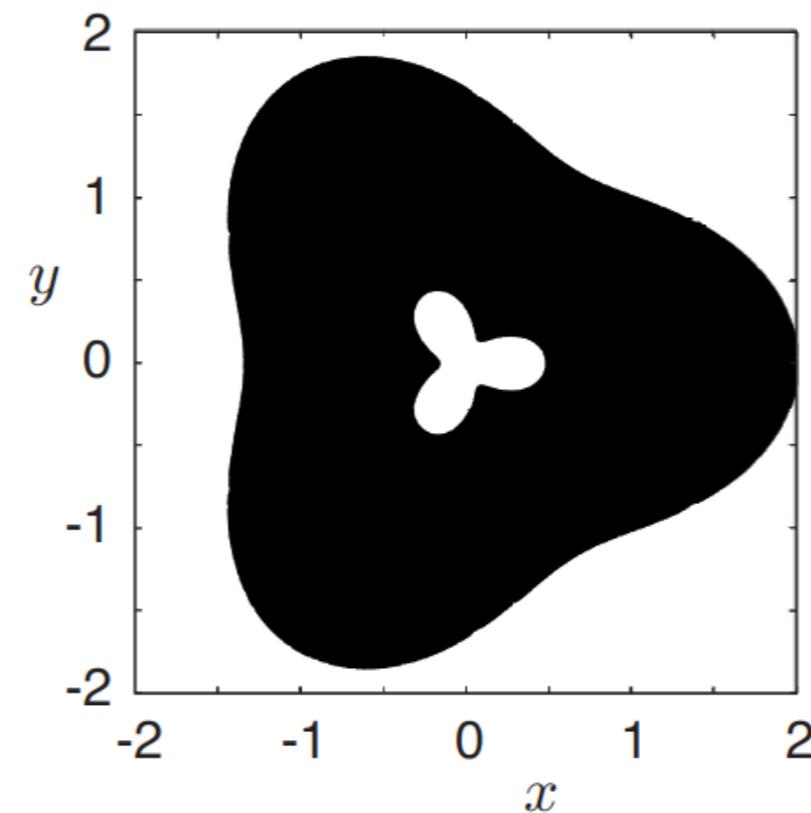
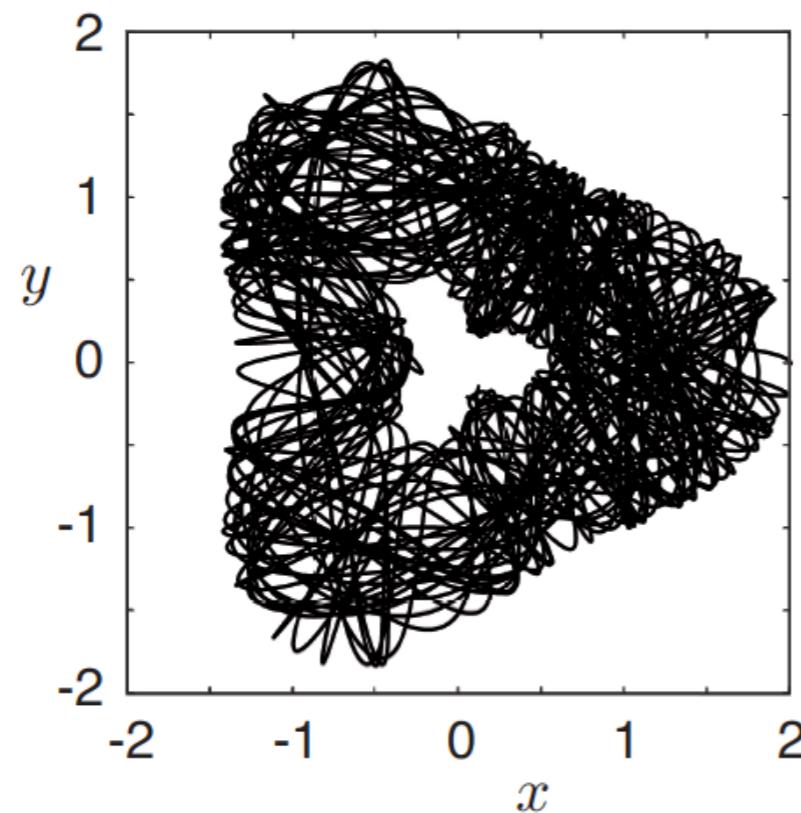
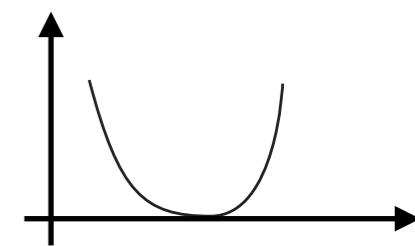
If ergodic, the dynamics generates space-filling curves that map out energy level surfaces.

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

Invariant Sets

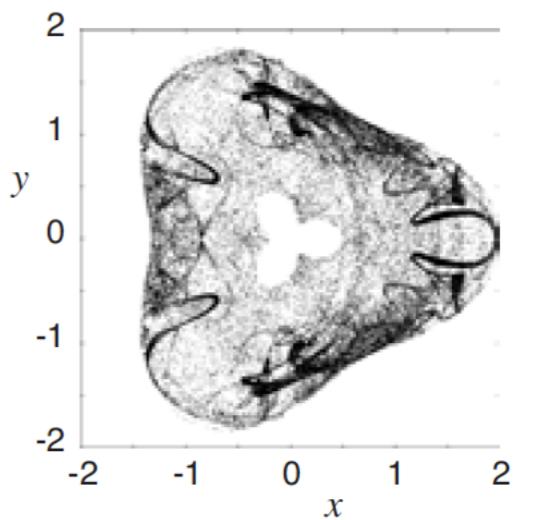
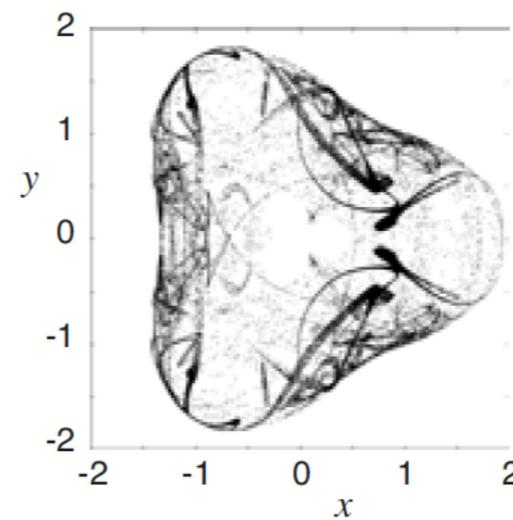
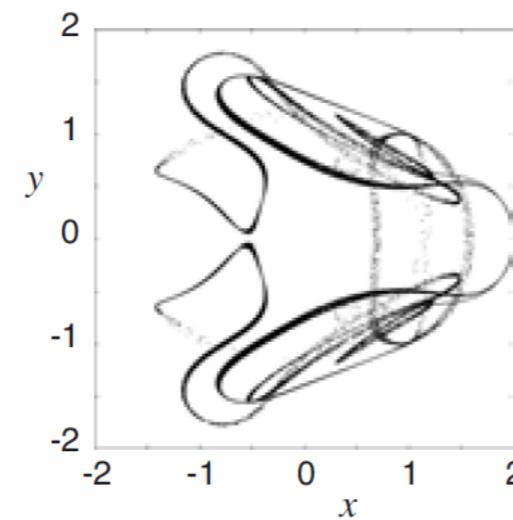
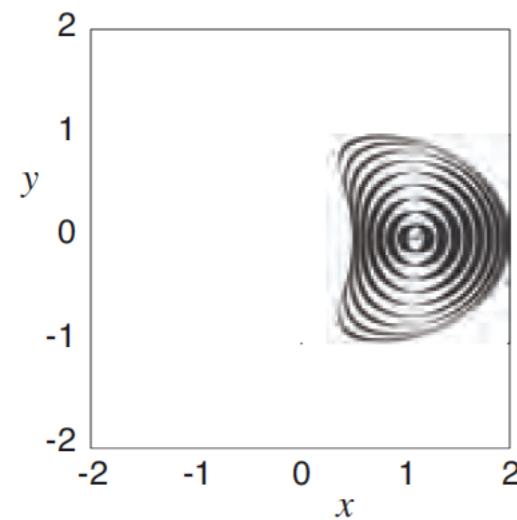
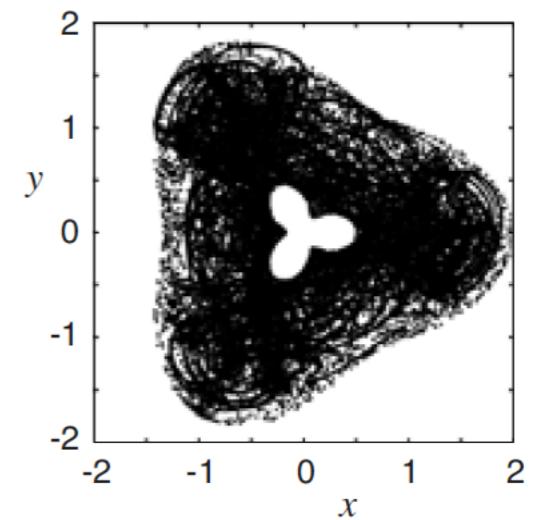
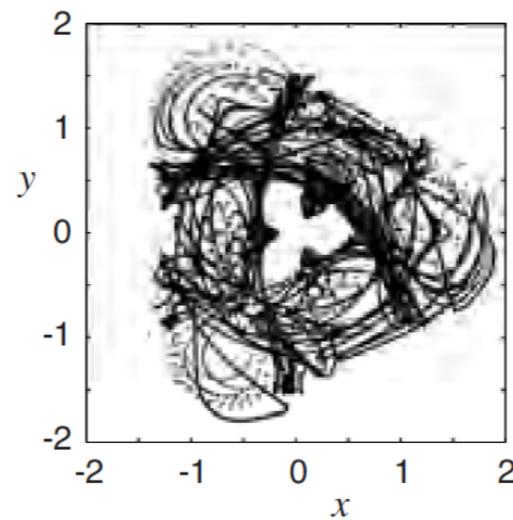
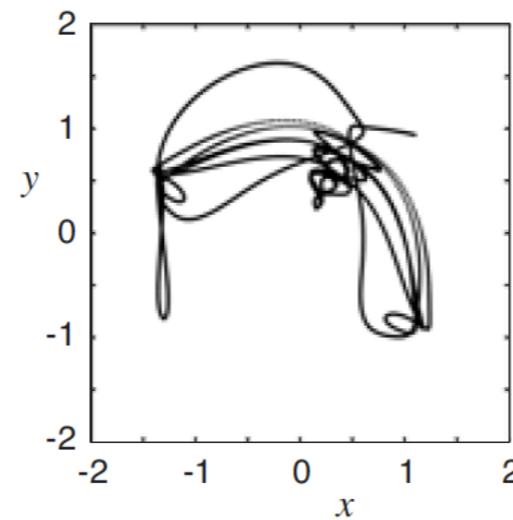
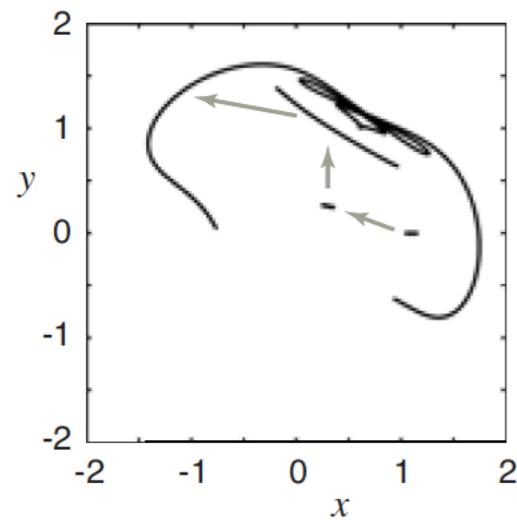
Anisotropic oscillator

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



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

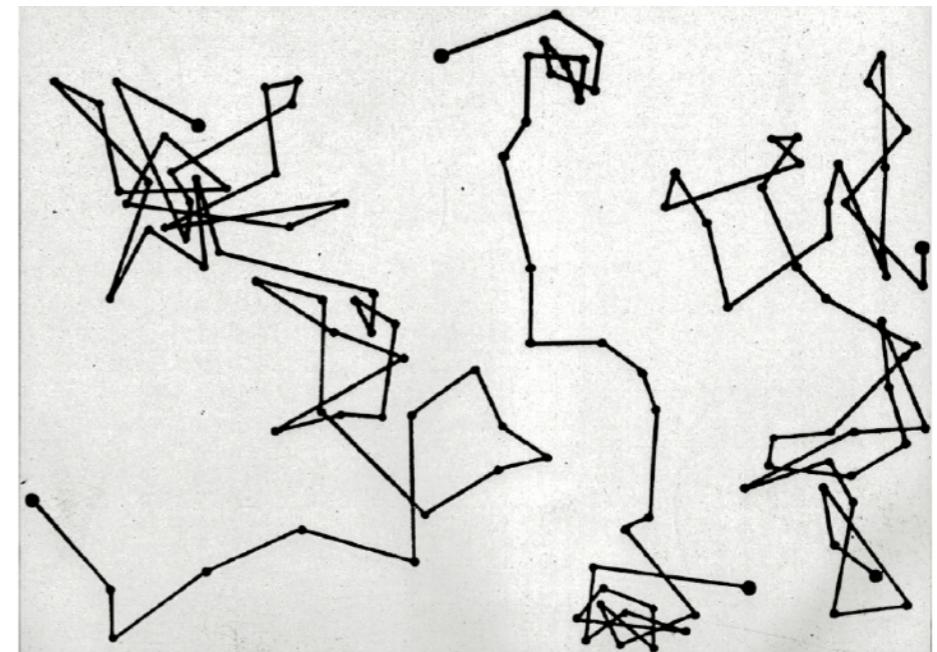
Even if ergodic, convergence to the invariant set
may depend heavily on initial condition...



Sampling using SDEs

Stochastic Differential Equations

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



Stochastic differential equations

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

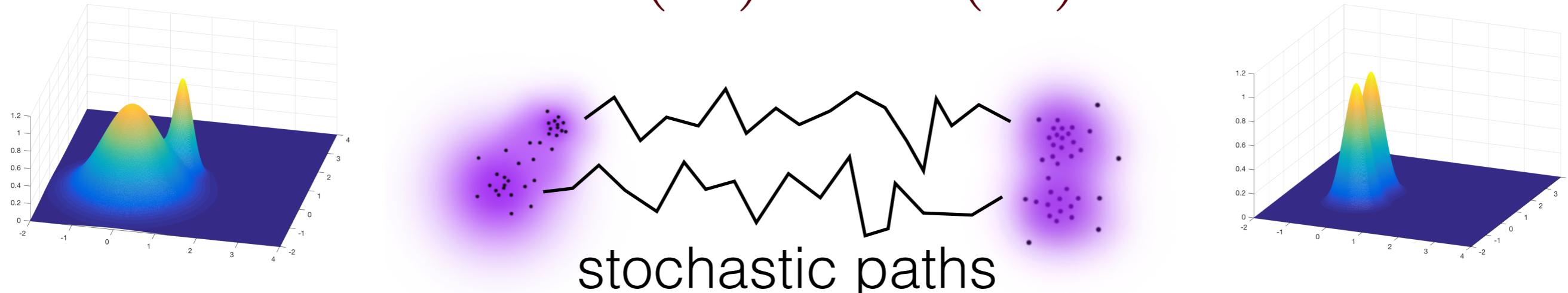
$$W(t) = \widehat{\lim}_{\delta t \rightarrow 0} \sqrt{\delta t} \sum_{k=0}^{\lfloor t/\delta t \rfloor} J_k$$

$W(t) \sim \mathcal{N}(0, t)$
 $\mathbb{E}(W(t) - W(s))^2 = |t - s|$

Evolving distribution

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

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



INITIAL STATE

TRANSIENT PHASE

EQUILIBRIUM

Arthur's Seat

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

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

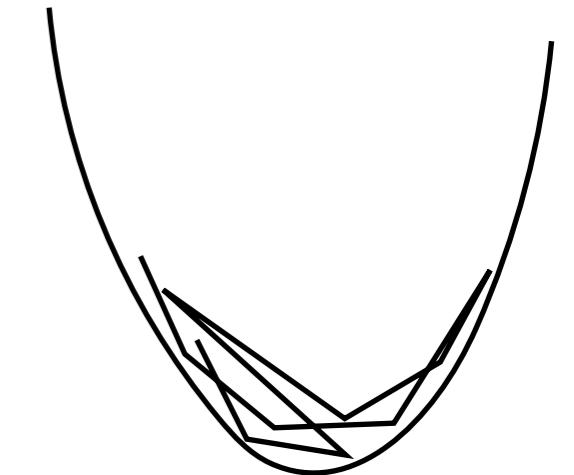
$$\mathcal{L}^\dagger g = -\nabla_x(a(x)g) + \frac{1}{2}\text{tr}\{b(x)^T g''(x)b(x)\}$$

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 π**

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

**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

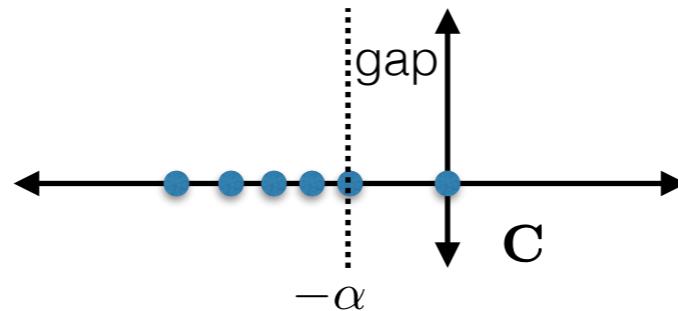
1. Check **parabolic Hörmander condition**
2. Show a **controllability property** - for any tau, there is a stochastic path of time tau linking any two points of the domain
3. Find a **Lyapunov function**

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

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 ϕ , we have

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

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

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

$$\varphi \in S$$

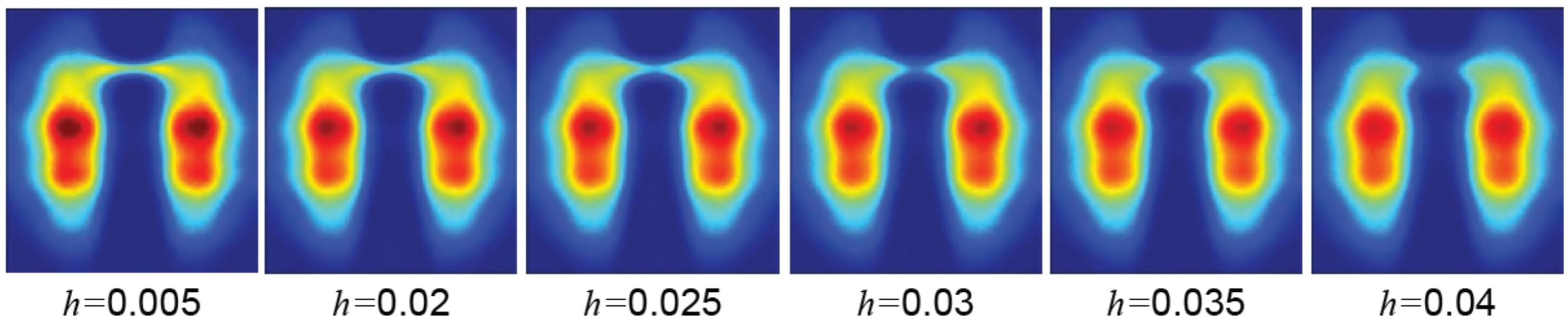
$$|\mathbb{E}\varphi(q_n) - \mathbb{E}\varphi(q(t_n))| = O(h^p) \quad nh \leq \tau$$

**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

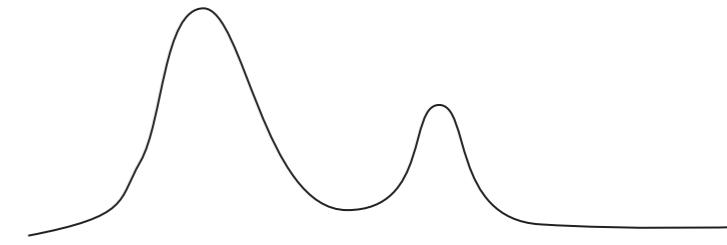
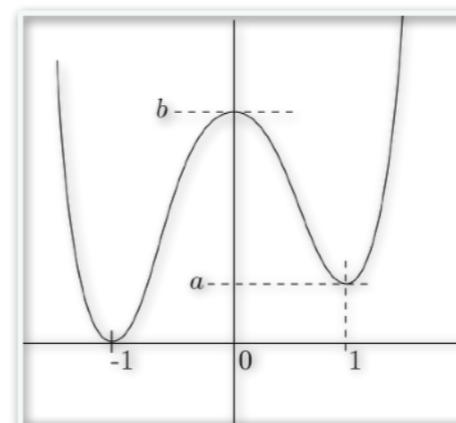
$$q_{n+1} = q_n + hF(q_n) + \sqrt{h/2}(R_n + R_{n-1})$$

[L. & Matthews, AMRX, 2013]

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

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

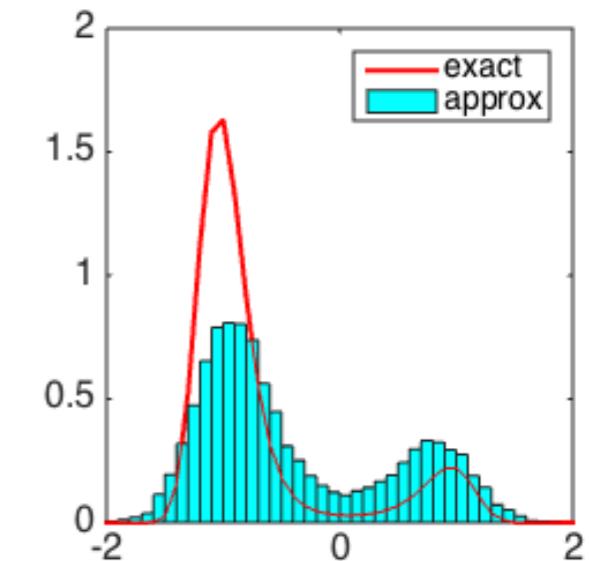
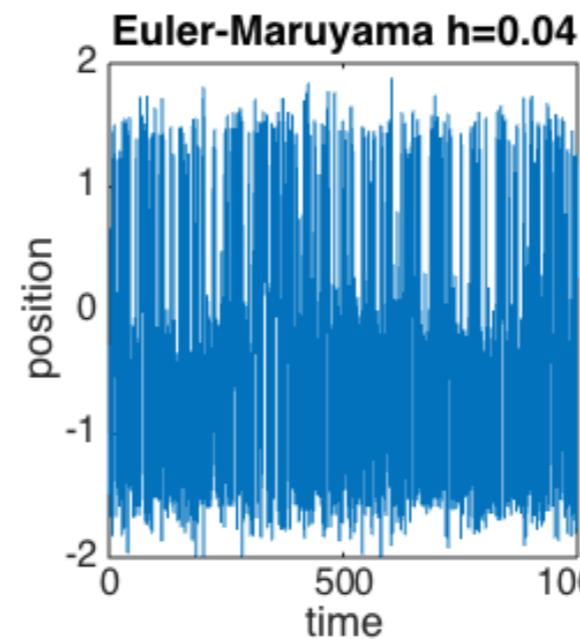
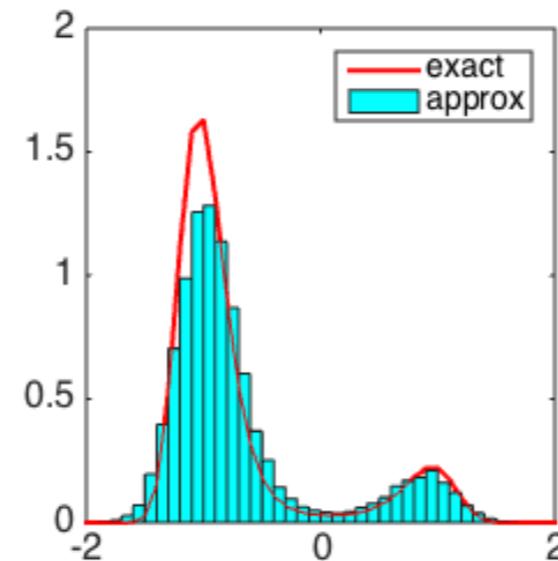
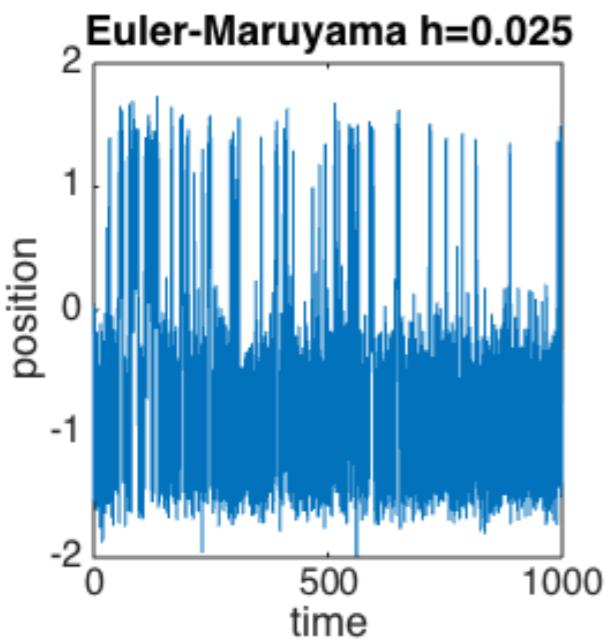
Uneven Double Well



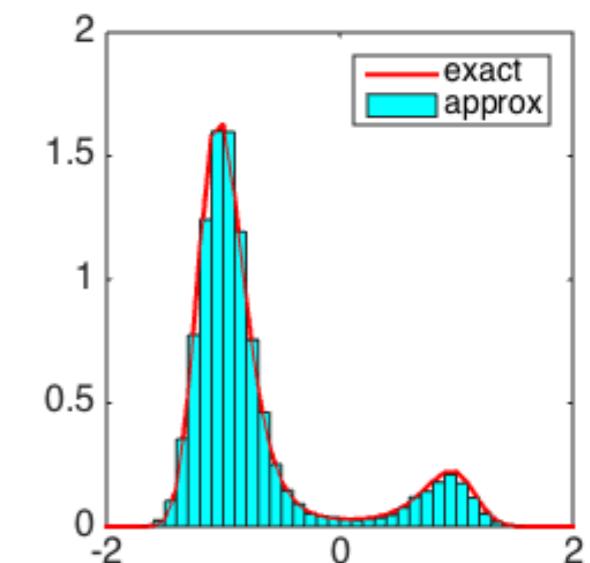
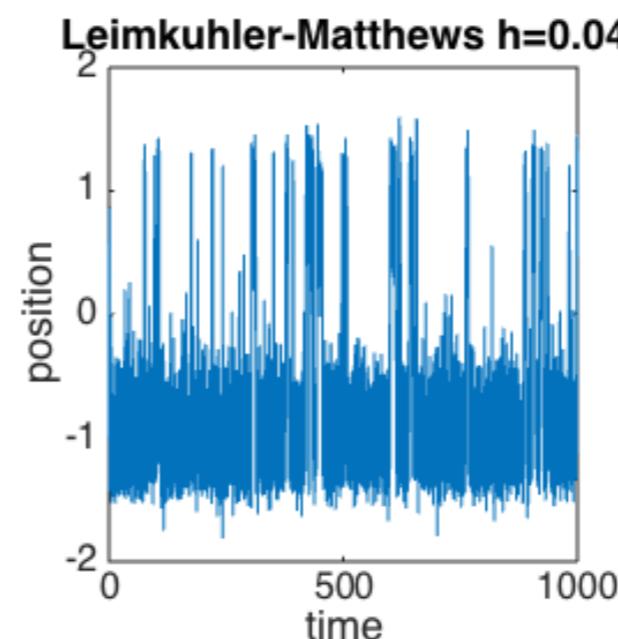
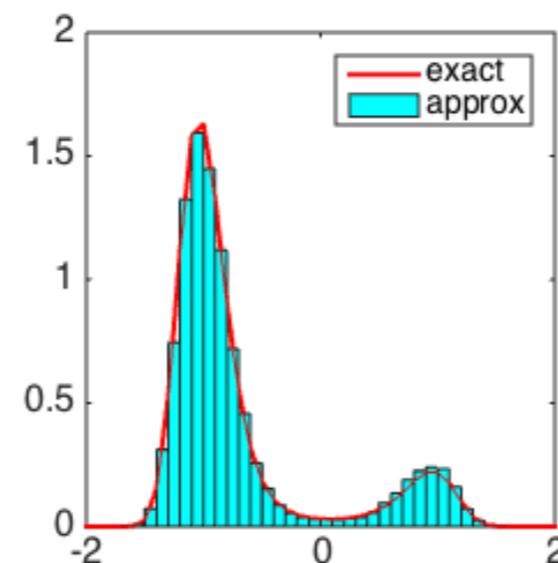
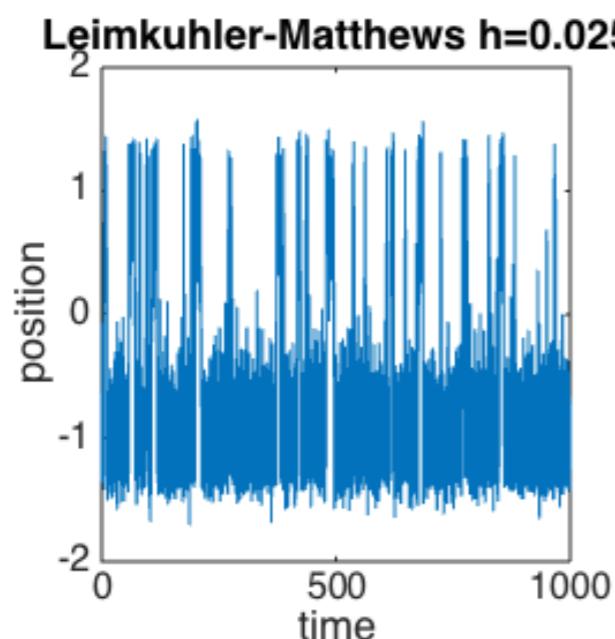
bimodal distribution

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,

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

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

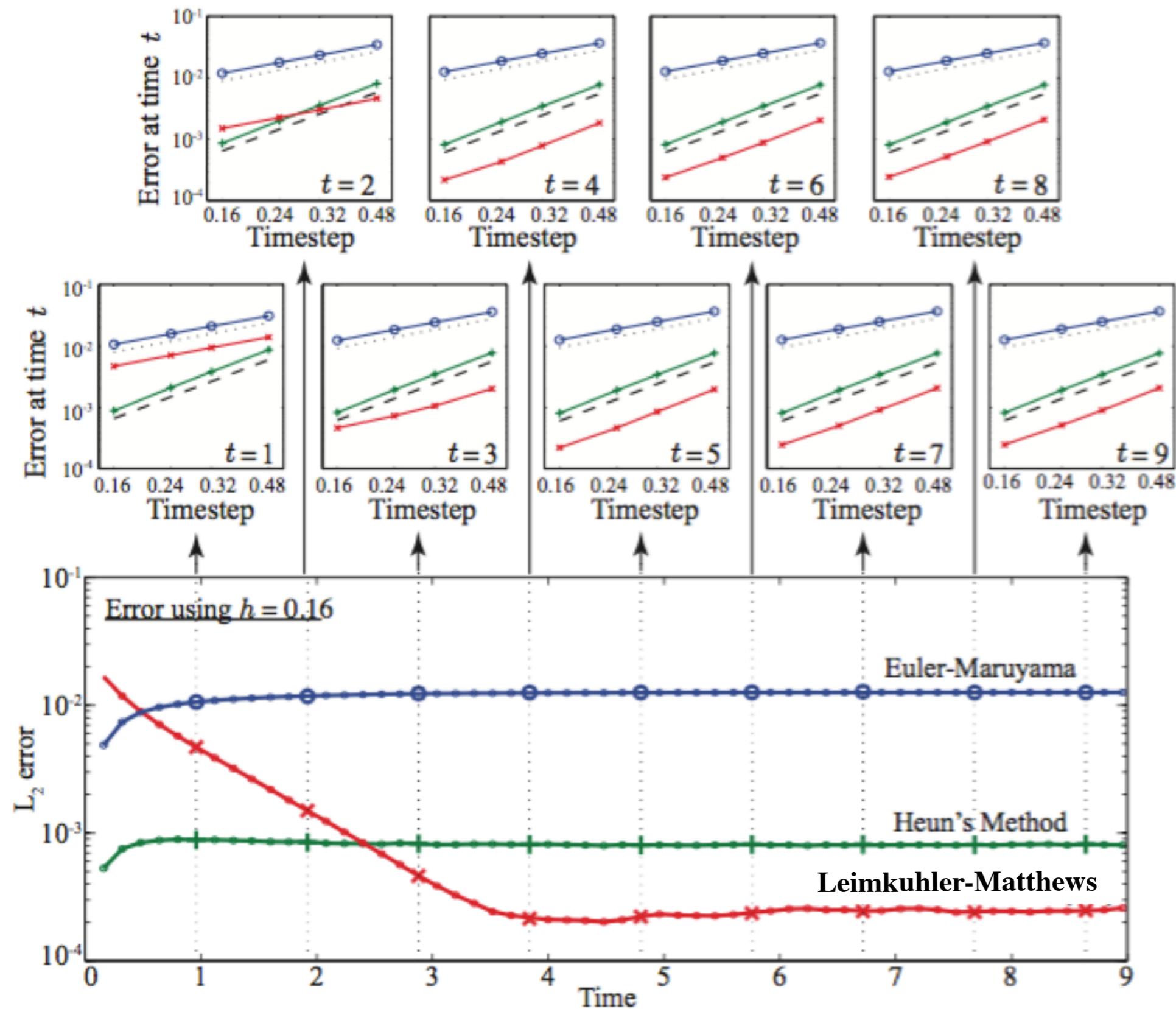
$$\lambda_0, \lambda > 0$$

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

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

exponentially fast in time

Quartic Oscillator



Langevin Dynamics

$$dq = M^{-1} pdt$$

$$dp = F(q)dt - \gamma M^{-1} pdt + \sqrt{2\beta^{-1}\gamma} dW$$

Newton's Equations

Stochastic Perturbation

With Periodic Boundary Conditions
and smooth potential, ergodic sampling
of the canonical distribution with density

$$\rho_{\text{can}} \propto e^{-\beta(p^T M^{-1} p / 2 + U(q))}$$

Mattingly, Stuart, Higham, Hairer, Nier...

Lyapunov Function

A Lyapunov function for the SDE is defined as a function ϕ satisfying

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

for some positive constants α, δ

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”)

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

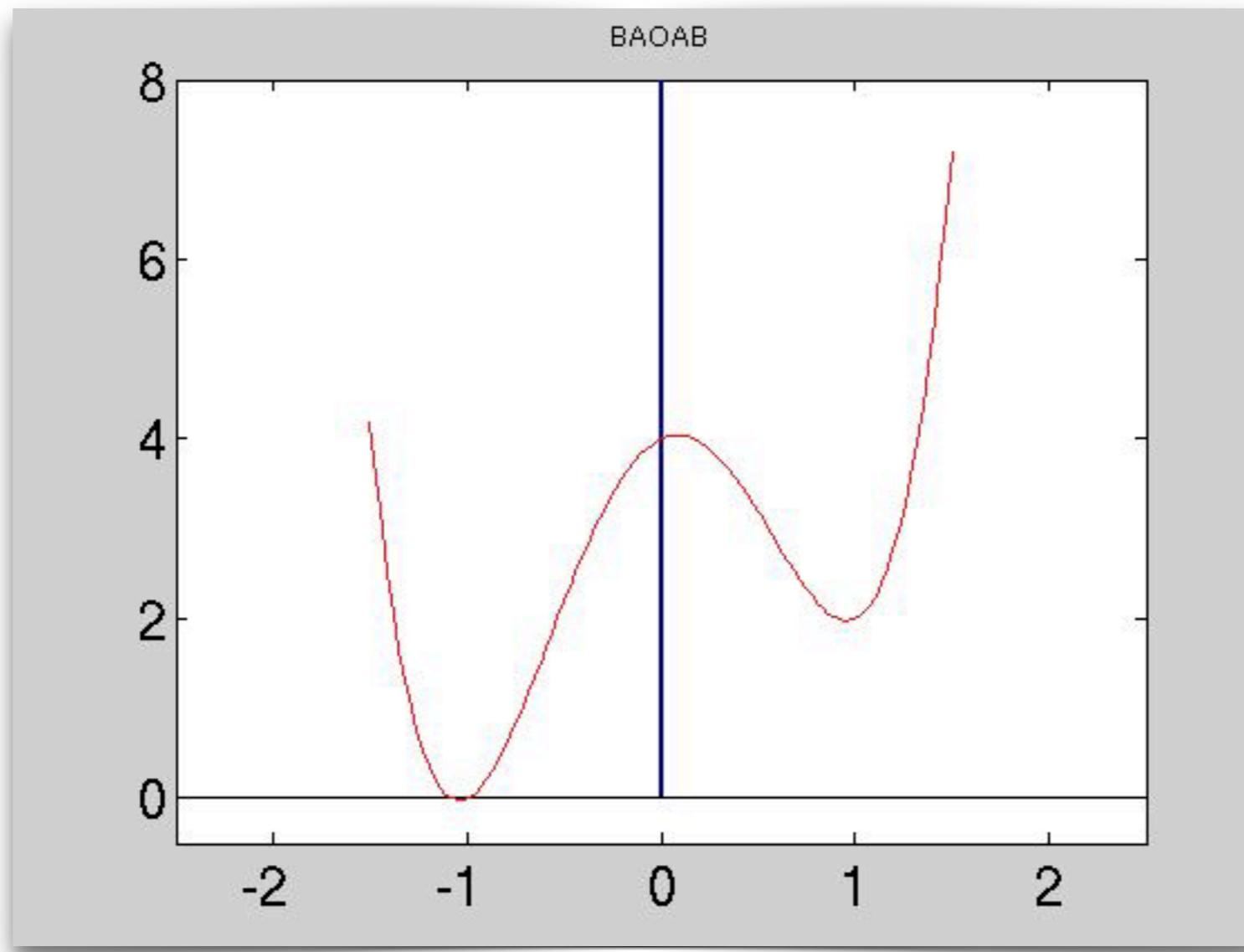
Regularity

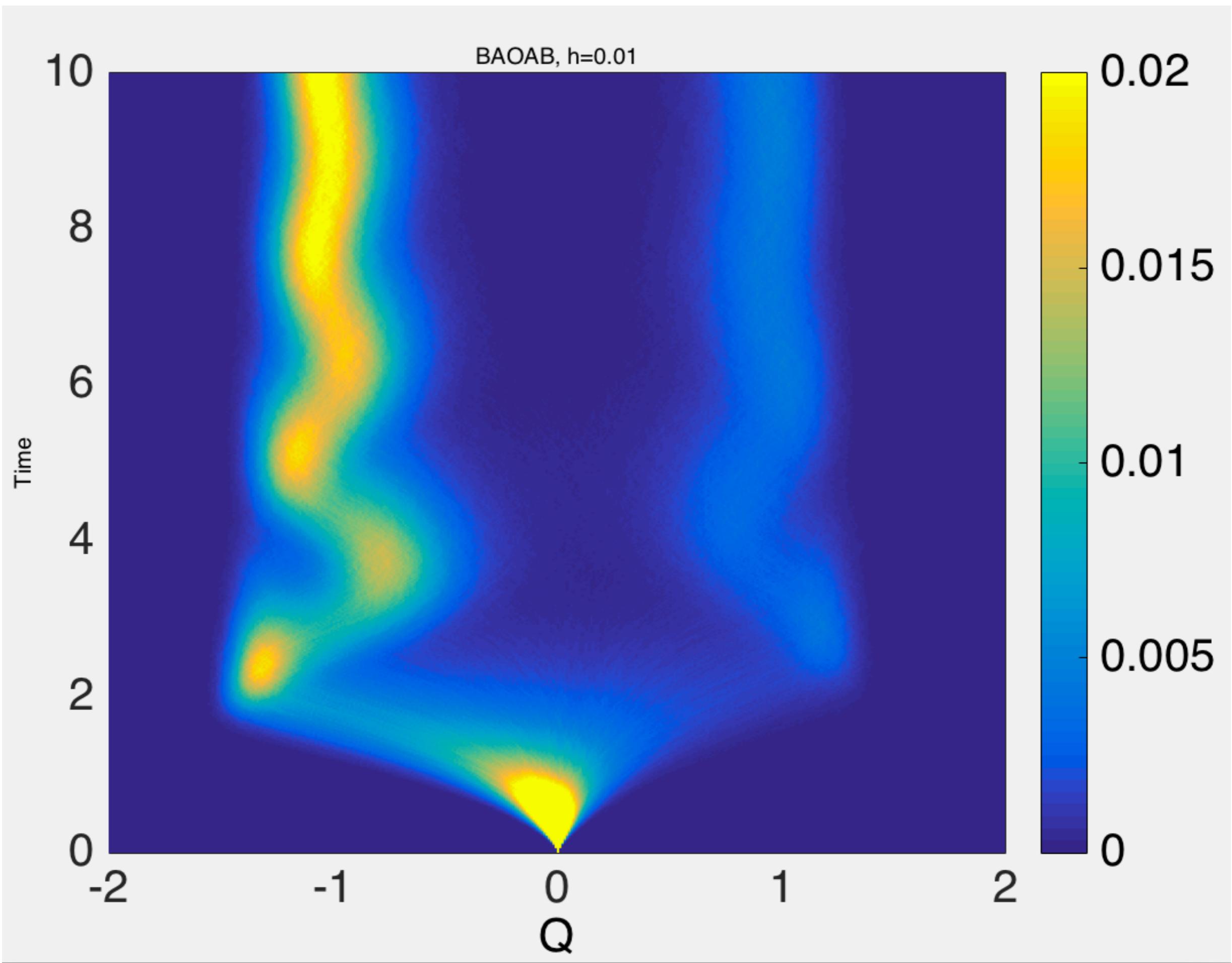
$$\|\mathcal{L}_\gamma^{-1}\|_{\mathcal{B}(\mathcal{H}^1)} \leq \frac{K_\gamma}{\lambda_\gamma}$$

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 δt

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

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

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

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 expansion¹ 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)

¹Denis 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

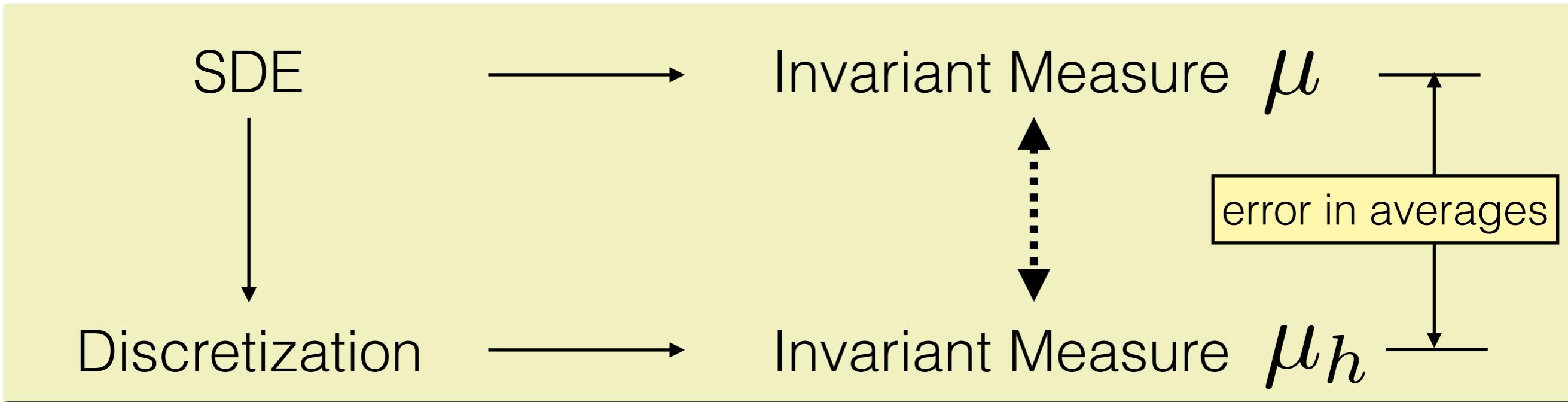
$$\bar{\rho} = e^{-\beta U} (1 + O(\delta t^2 \varepsilon^2) + O(\delta t^4))$$

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!**

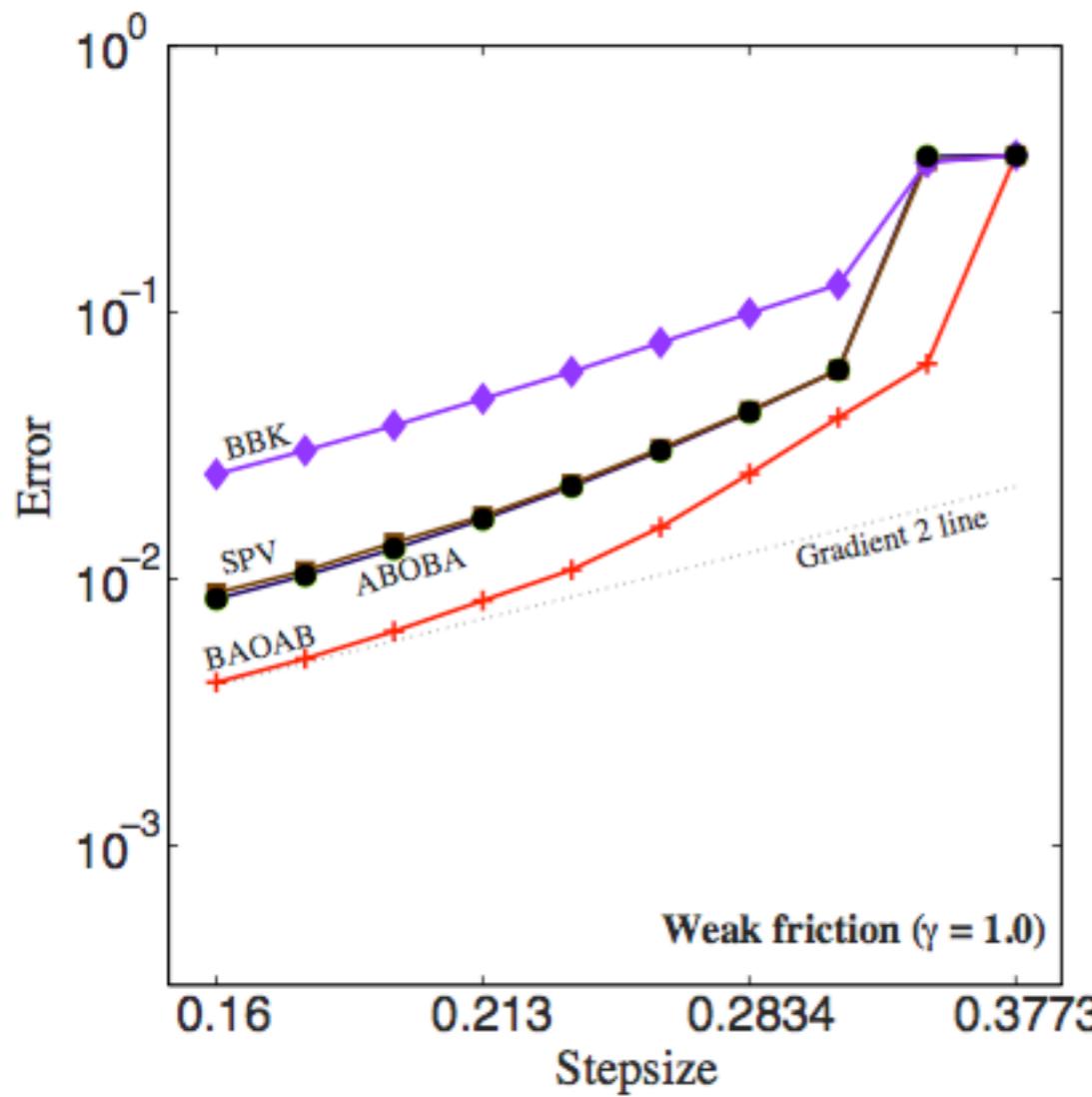
$$q_{n+1} = q_n + hF(q_n) + \sqrt{h/2}(R_n + R_{n-1})$$

Perturbed Quartic (1D)

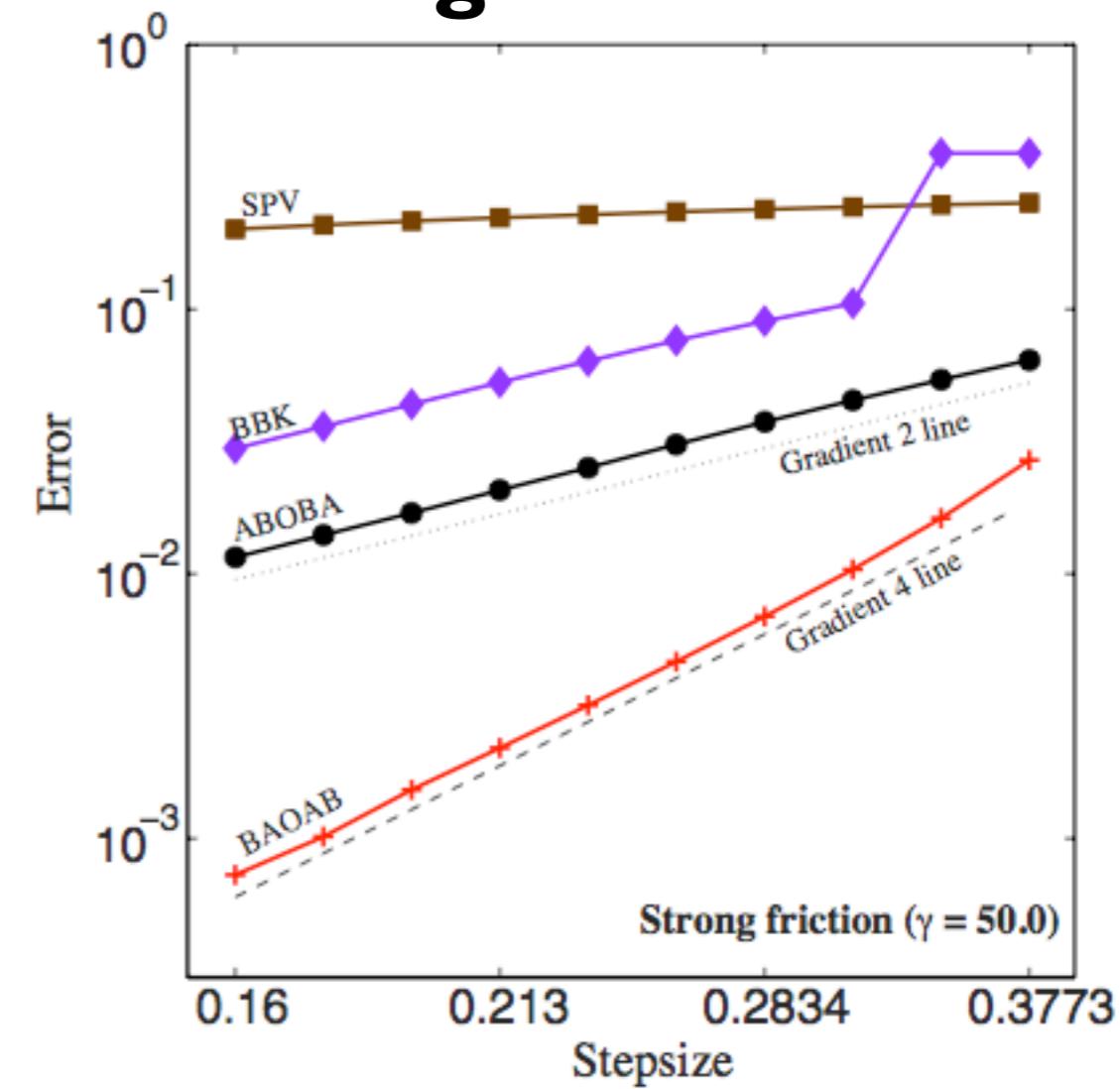
$$U(x) = x^4/4 + \sin(1+5x).$$

Langevin Configurational Dist. Accuracy vs timestep

Mild Friction

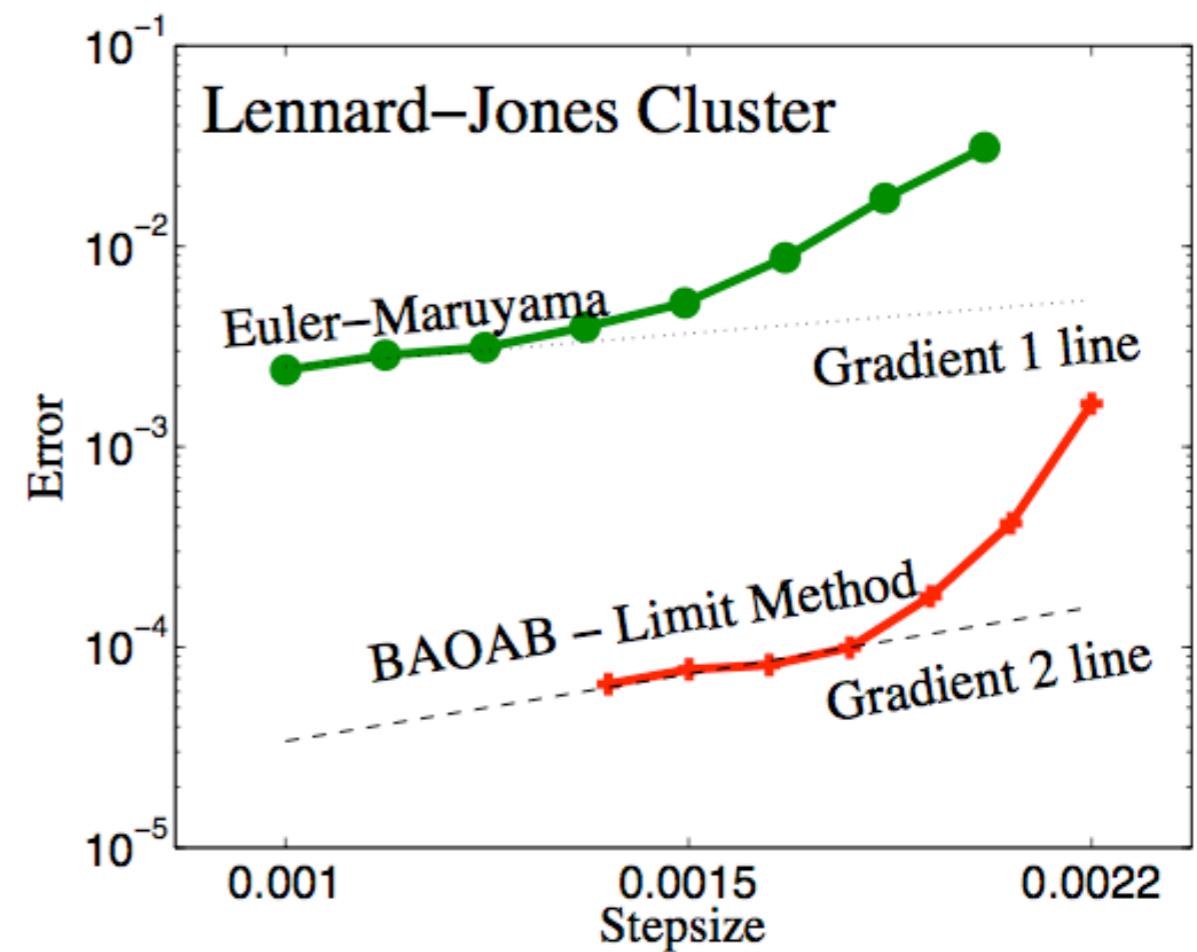
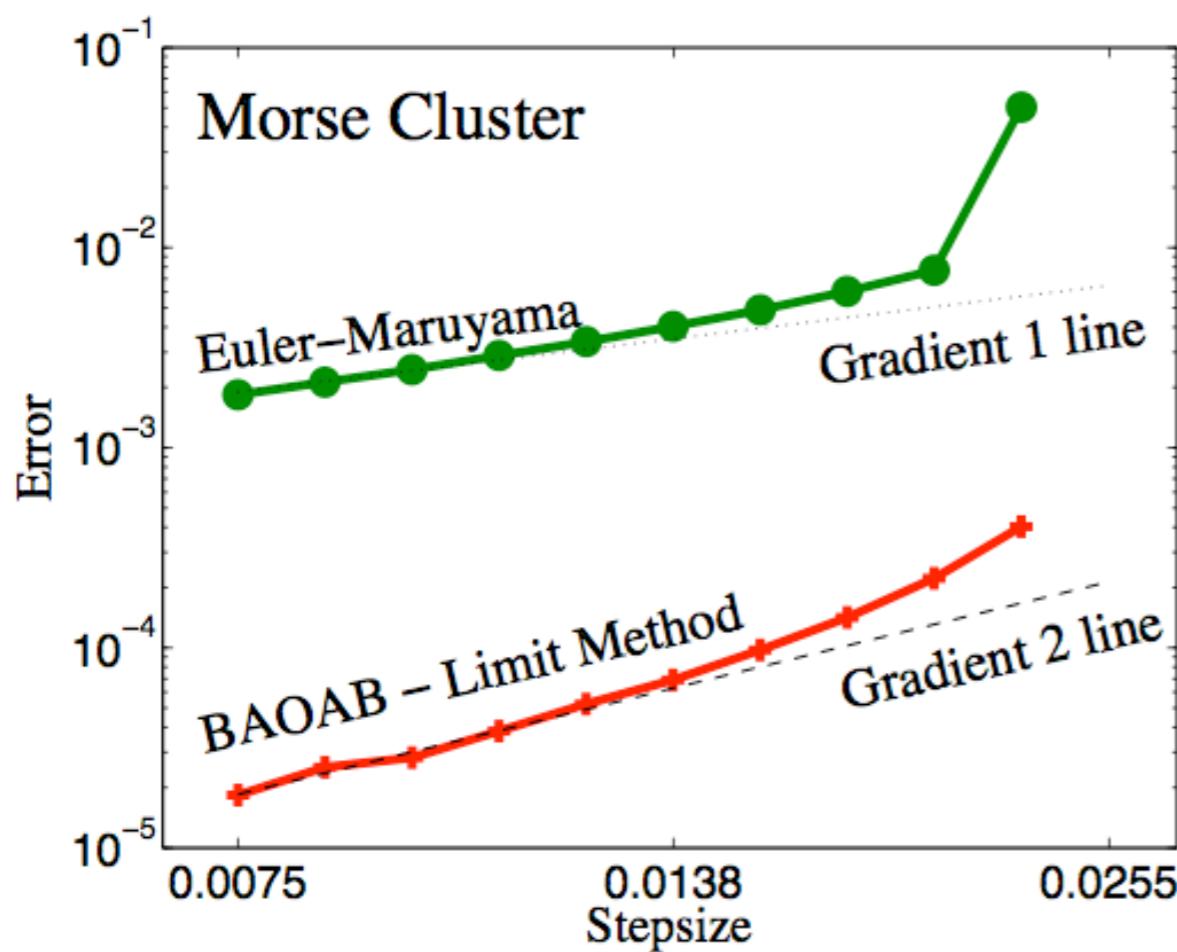
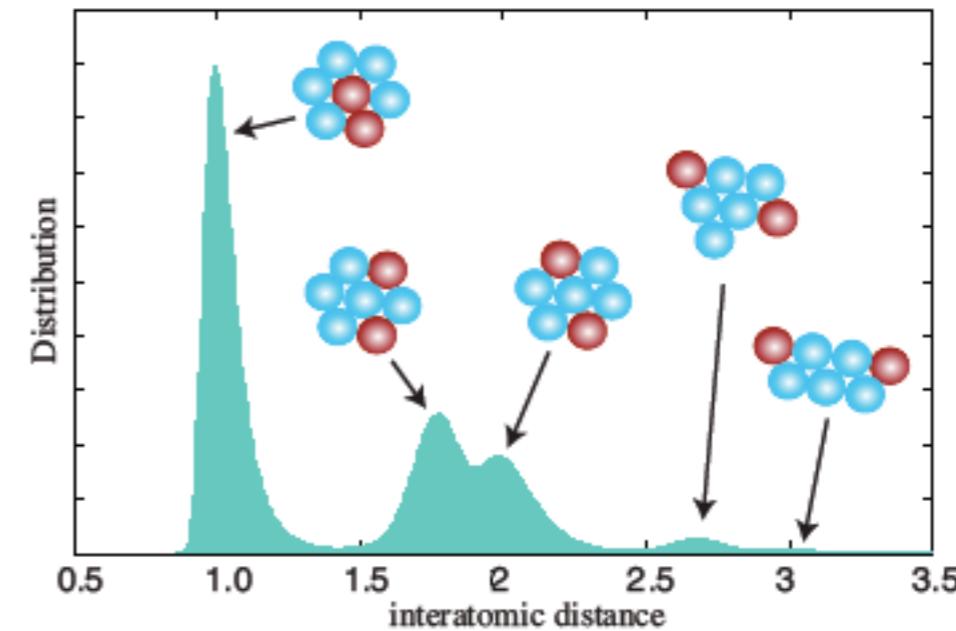


High Friction



Morse and Lennard Jones Clusters

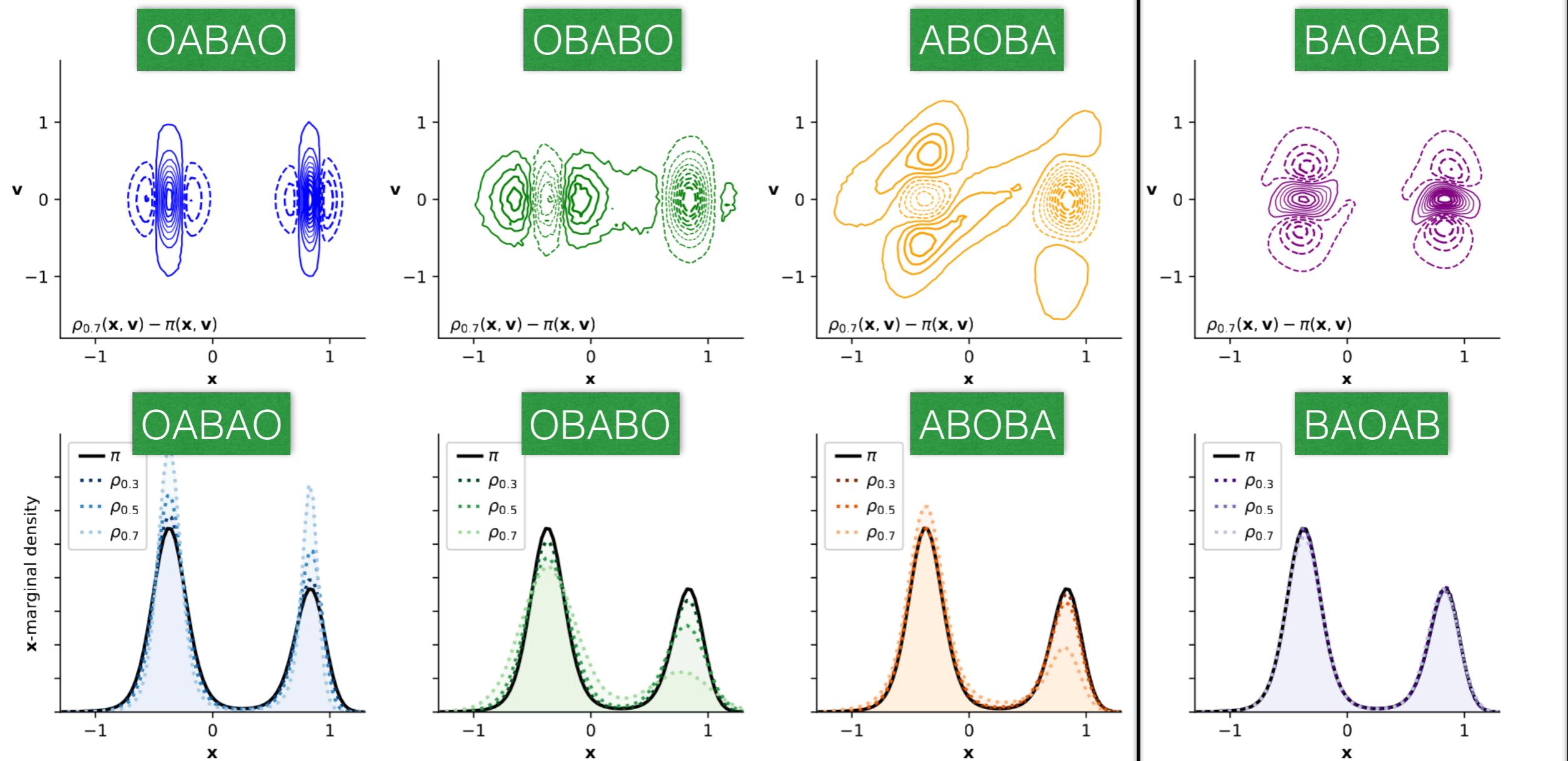
binned radial density



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

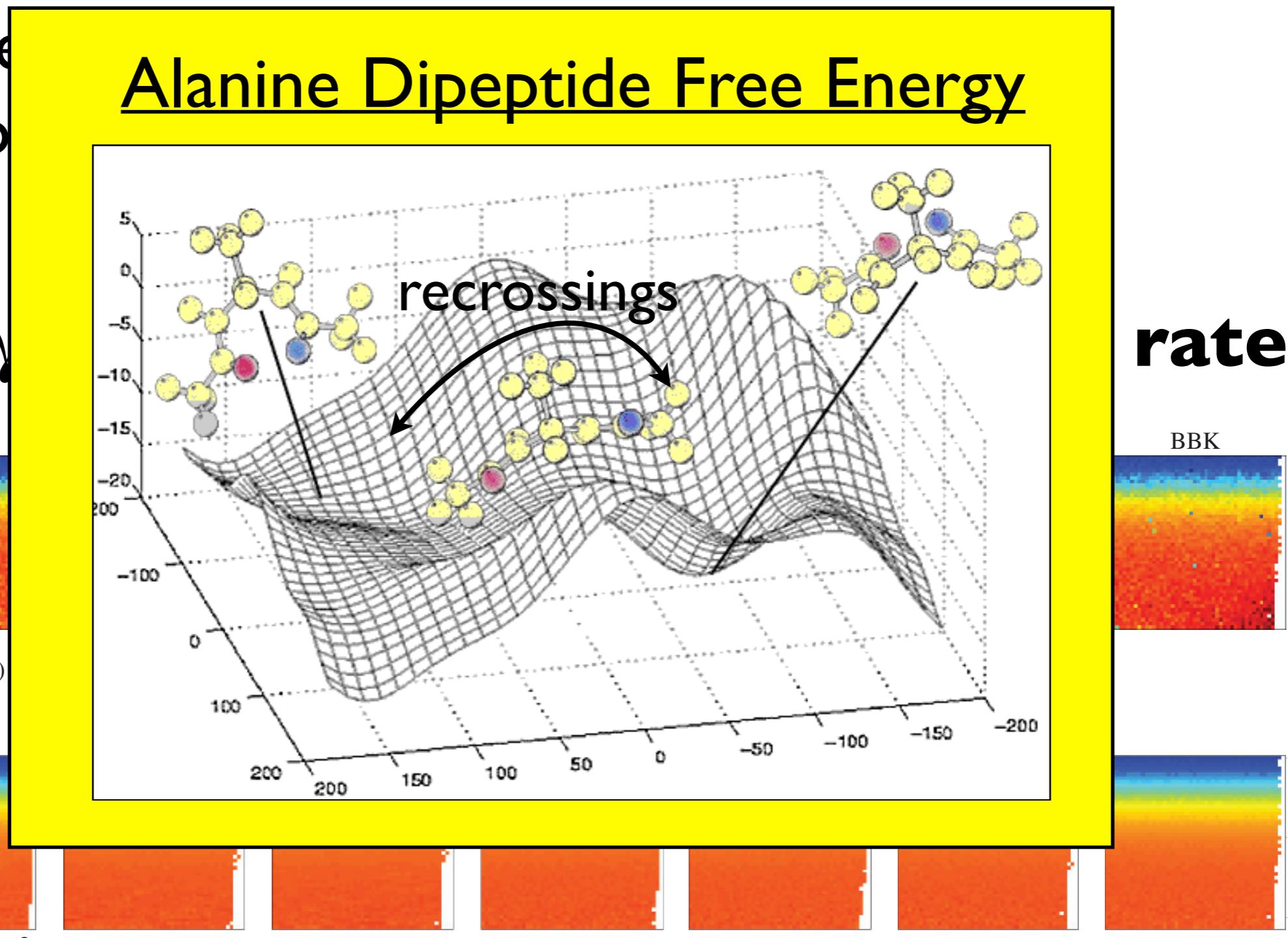
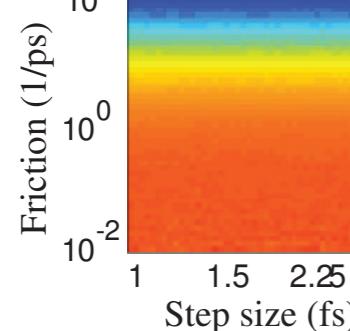
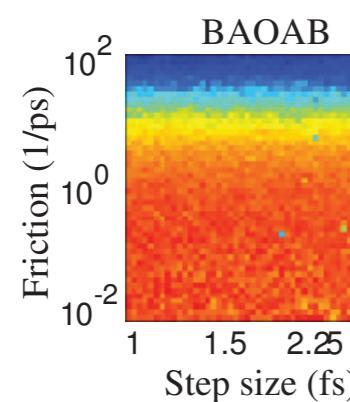


Configurational Marginal

Diffusion vs Accuracy

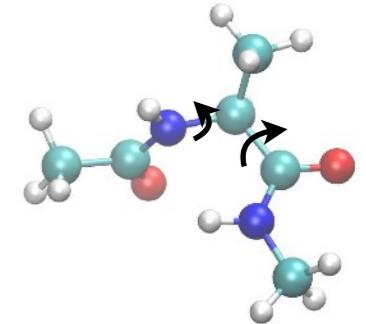
With BAOAB, we would imagine it to be better to
use larger
 \Rightarrow 4th order

But...

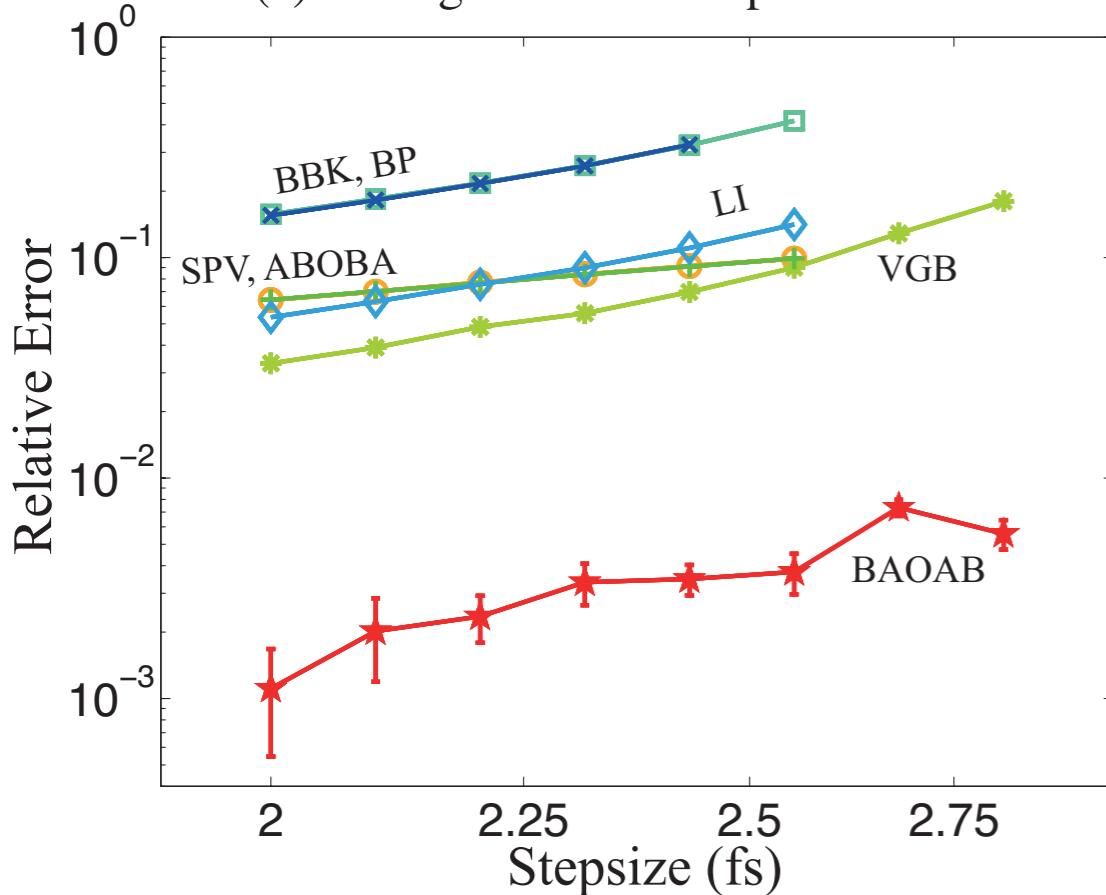


Solvated Systems

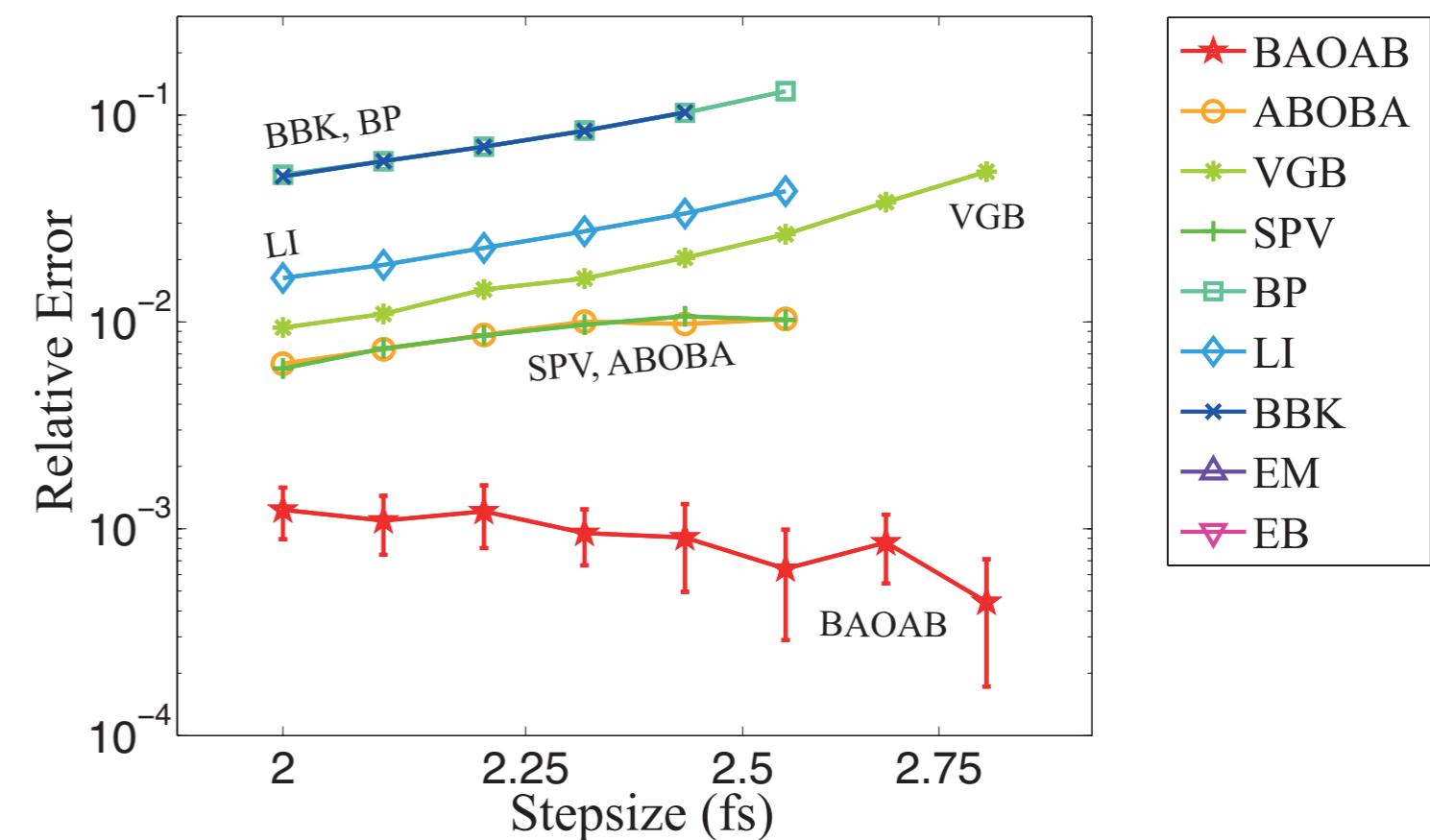
for alanine dipeptide in flexible water



(a) Configurational Temperature Error



(b) Average Total Potential Energy Error



BAOAB much better than alternatives for relevant
(i.e. *configurational*) quantities...up to the
“stability threshold” and ... **At moderate friction!**

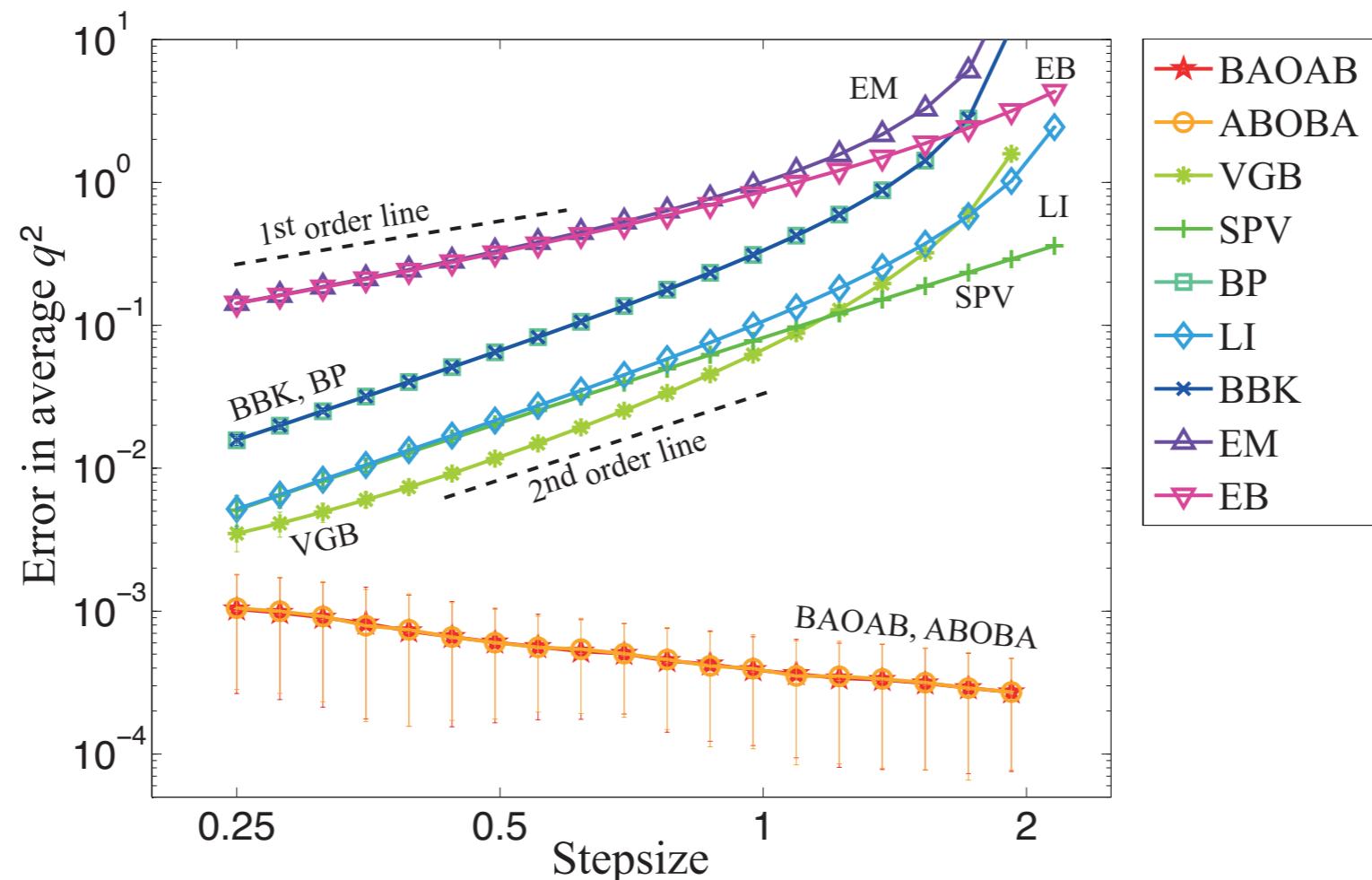
If we are not in the high friction limit, why is **BAOAB** so much better?

Harmonic Oscillator: $H = \frac{p^2}{2} + \frac{\omega^2 q^2}{2}$

ABOBA and **BAOAB** are ***exact*** for average potential

Scheme	$\langle q^2 \rangle$	Scheme	$\langle q^2 \rangle$
Exact	$K^{-1}\beta^{-1}$	SPV	$K^{-1}\beta^{-1} \left(\gamma \delta t \frac{1-e^{-2\gamma\delta t}}{2(1-e^{-\gamma\delta t})^2} \right)$
BAOAB	$K^{-1}\beta^{-1}$	LI	$K^{-1}\beta^{-1} - \frac{\delta t^2}{12M\beta} + O(\delta t^4)$
ABOBA	$K^{-1}\beta^{-1}$	VGB	$K^{-1}\beta^{-1} + \frac{\gamma^2 M - 2K}{24M\beta K} \delta t^2 + O(\delta t^4)$
BBK	$K^{-1}\beta^{-1} \left(1 - \frac{\delta t^2 K}{4M} \right)^{-1}$	EM	$K^{-1}\beta^{-1} \left(1 - \frac{\delta t K}{2\gamma M} \right)^{-1}$
BP	$K^{-1}\beta^{-1} \left(1 - \frac{\delta t^2 K}{4M} \right)^{-1}$	EB	$K^{-1}\beta^{-1} + \frac{\delta t}{2\gamma M\beta} + O(\delta t^2)$

Harmonic Oscillator Configurational Sampling



BAOAB and **ABOBA** both are exact for PE

*But...this is only part of the story since **BAOAB** is much better than **ABOBA** for real molecules...*

Overview

Lecture 1: Monte Carlo, Hamiltonian systems, splitting methods, microcanonical sampling, SDEs, Euler-Maruyama & Leimkuhler-Matthews, Langevin splitting, superconvergence

Tutorial 1: methods for Langevin (and overdamped) dynamics

Lecture 2: Noisy gradients, then:
increasing the timestep, multiple timestepping,
holonomic constraints, isokinetic constraints & SIN,
ensemble preconditioning

Tutorial 2: enhancing the convergence to equilibrium/
decreasing the integrated autocorrelation time