https://github.com/amartinsson/DukeMicroCourse

Sampling Algorithms "Microcourse"

Session 2

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

Langevin Dynamics

$$\mathrm{d}q = M^{-1}p\mathrm{d}t$$

$$\mathrm{d}p = F(q)\mathrm{d}t - \gamma M^{-1}p\mathrm{d}t + \sqrt{2\beta^{-1}\gamma}\mathrm{dW}$$

Newton's Equations *Stochastic Perturbation*

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

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

Mattingly, Stuart, Higham, Hairer, Nier…

Splitting Methods

$$
\begin{aligned}
\mathrm{d}q &= \overbrace{p\mathrm{d}t}^{\mathbf{A}} && (M{=}I)\\
\mathrm{d}p &= \overbrace{F(q)\mathrm{d}t}^{\mathbf{B}} \overbrace{- \gamma p \mathrm{d}t + \sqrt{2\beta^{-1}\gamma}\mathrm{dW}}^{\mathbf{O}}
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{A}: &\quad \dot{q} = p, \quad \dot{p} = 0\\
\mathbf{B}: &\quad \dot{q} = 0, \quad \dot{p} = F(q)\\
\mathbf{O}: &\quad \dot{q} = 0, \quad \dot{p} = -\gamma p + \sqrt{2\beta^{-1}\gamma}\dot{\mathrm{W}}
\end{aligned}
$$

Time stepsize δt

ABO

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

Splitting Methods

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

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

Propagator: $$\mathscr{P}_t = e^{t\mathscr{L}}$$

Splitting Method: $$\mathscr{P}_t \approx e^{t\mathscr{A}} e^{t\mathscr{B}} e^{t\mathscr{O}}$$

Drift Kick Shuffle **“OBA”**

“ABO” “OAB” “ABOBA” “OBABO” …

Expansion of the invariant distribution

(Talay-Tubaro expansion[1] in the ergodic limit)

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

Leading order:

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

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)

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

Two level expansion

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

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

$$\varepsilon = 1/\gamma$$

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

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

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

Configurational Sampling

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

Proposition:

Take Home Message #1

The right order of Langevin building blocks provides substantial improvements in accuracy (low sampling bias).

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

2.1 Noisy gradients

Problem: use stochastic dynamics to accurately sample a distribution with given positive smooth density

$$\rho \propto \exp(-U)$$

in case the force $-\nabla U$ can only be computed approximately

Examples:

Multiscale models
several flavors of hybrid **ab initio MD Methods**
QM/MM methods

…Many applications in **Bayesian Inference & Big Data Analytics**

What to do about the force error?

$$\tilde{F}(x) = -\nabla U(x) + \eta(x)$$

a sampling error… it seems natural to take

$$\eta(x) \sim \mathcal{N}(0, \sigma(x))$$

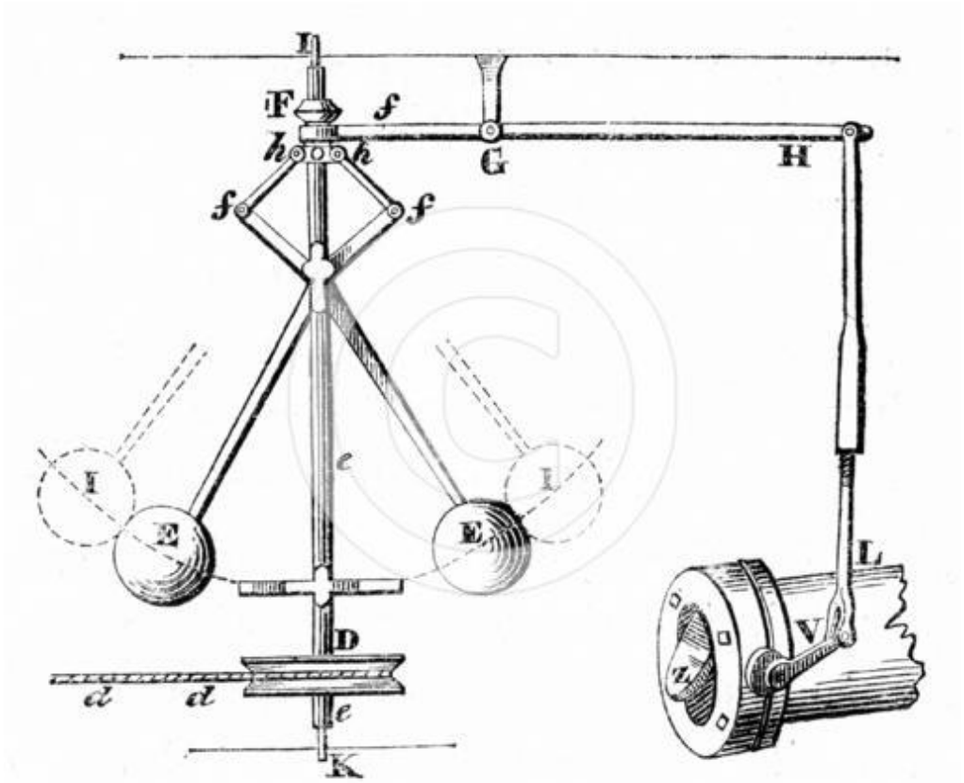
and also, at least in the first stage, to assume $\sigma(x) \approx \sigma$

$$\begin{aligned} x_{n+1} &= x_n + hF(x_n) + h\sigma\tilde{R}_n + \sqrt{2h}R_n \\ &= x_n + hF(x_n) + \sqrt{h}\sqrt{h\sigma^2 + 2}\hat{R}_n \end{aligned}$$

Like Euler-Maruyama discretization of

$$\mathrm{d}x = F(x)\mathrm{d}t + \sqrt{2 + \sigma h}\mathrm{dW}$$

James Watt's Engine



Too fast: balls move to outside, opening valve, releasing steam, reducing pressure, reducing speed

Too slow: balls fall to inside, closing valve, leading to an increase in pressure, increasing speed

Nose-Hoover dynamics - a "Gibbs Governor"

$$\dot{q} = p$$
$$\dot{p} = -\nabla U(q) - \xi p$$
$$\dot{\xi} = p^2 - kT$$

Preserves

$$e^{-\beta[p^2/2+U(q)]} \times e^{-\beta\xi^2/2}$$

Adaptive Langevin, later "SGNHT"

Jones & L. J. Chem. Phys. 2011

Applying Nosé-Hoover Dynamics to a system with noisy gradient corrects the canonical distribution.

Adaptive (Automatic) Langevin

$$\mathrm{d}x = M^{-1}p\mathrm{d}t$$

$$\mathrm{d}p = -\nabla U\mathrm{d}t + \sqrt{h}\sigma\mathrm{d}W - \xi p\mathrm{d}t + \sigma_A\mathrm{d}W_A$$

$$\mathrm{d}\xi = \mu^{-1}[p^T M^{-1}p - n\beta^{-1}]\mathrm{d}t$$

$$\tilde{\rho} = e^{-\beta[p^T M^{-1}p/2 + U(x)]} \times e^{-\beta\mu(\xi-\gamma)^2/2}$$

Shift in auxiliary variable by $\gamma = \dfrac{\beta(h\sigma^2 + \sigma_A^2)}{2\mathrm{Tr}(M)}$

Discretization

[L. & Shang, SISC, 2016]

generator: $\mathscr{L} = \mathscr{L}_{\mathrm{A}} + \mathscr{L}_{\mathrm{B}} + \mathscr{L}_{\mathrm{O}} + \mathscr{L}_{\mathrm{D}}$

$$\mathscr{L}_{\mathrm{A}} = (M^{-1}p) \cdot \nabla_x$$

$$\mathscr{L}_{\mathrm{B}} = -\nabla U(x) \cdot \nabla_p + \frac{h\sigma^2}{2} \Delta_p$$

$$\mathscr{L}_{\mathrm{O}} = -\xi p \cdot \nabla_p + \frac{\sigma_{\mathrm{A}}^2}{2} \Delta_p$$

$$\mathscr{L}_{\mathrm{D}} = G(p) \frac{\partial}{\partial \xi}$$

define related operator by composition, e.g. **BADODAB**

$$e^{h\hat{\mathscr{L}}} = e^{\frac{h}{2}\mathscr{L}_{\mathrm{B}}} e^{\frac{h}{2}\mathscr{L}_{\mathrm{A}}} e^{\frac{h}{2}\mathscr{L}_{\mathrm{D}}} e^{h\mathscr{L}_{\mathrm{O}}} e^{\frac{h}{2}\mathscr{L}_{\mathrm{D}}} e^{\frac{h}{2}\mathscr{L}_{\mathrm{A}}} e^{\frac{h}{2}\mathscr{L}_{\mathrm{B}}}$$

Mimic properties of BAOAB in the high σ_{A} regime!

Superconvergence

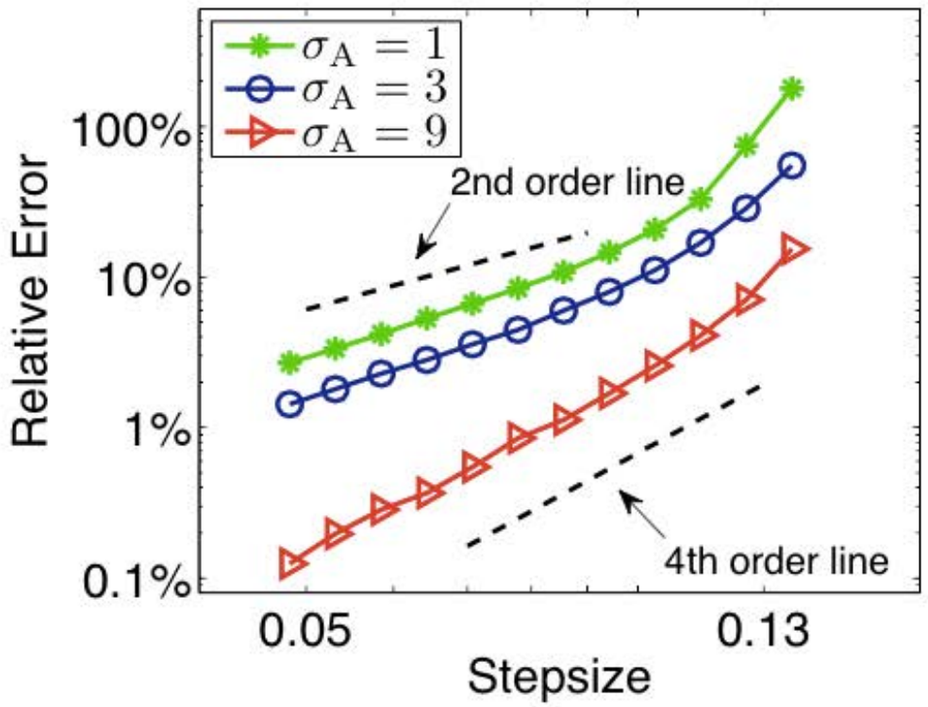
BAOAB, in the high friction limit, gives a ***superconvergence*** property for configurational quantities.

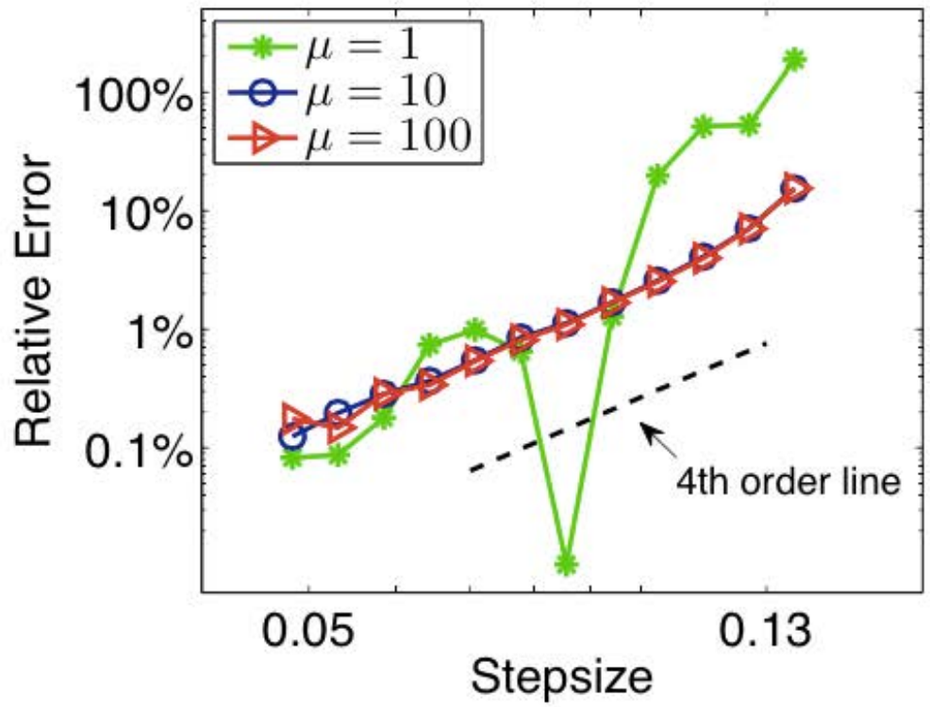
By taking large $\gamma \propto \sigma_A^2$ and $\mu \propto \sigma_A^2$ we can make BADODAB behave like BAOAB in the high friction limit after averaging over the auxiliary variable.

Effectively the extra driving noise implements a projection to the case of Langevin dynamics, ***but large driving noise means large effective friction, restricts phase space exploration.***

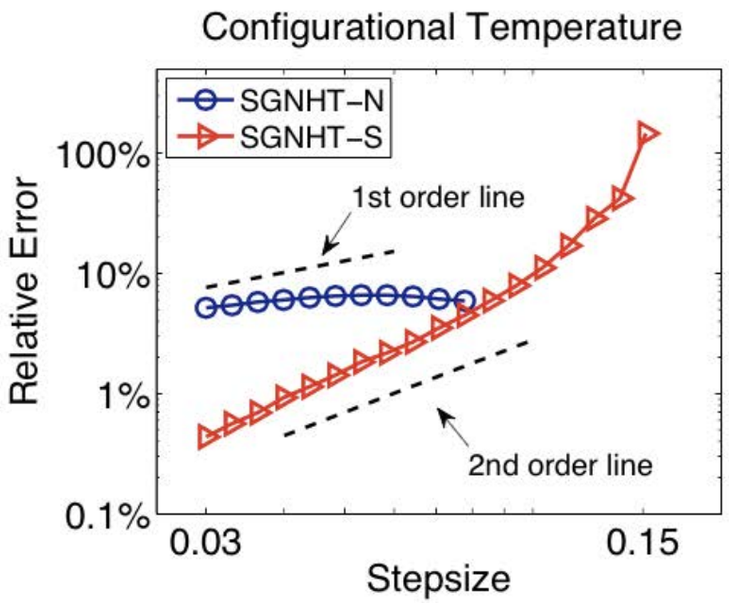
500 LJ particles, clean gradient

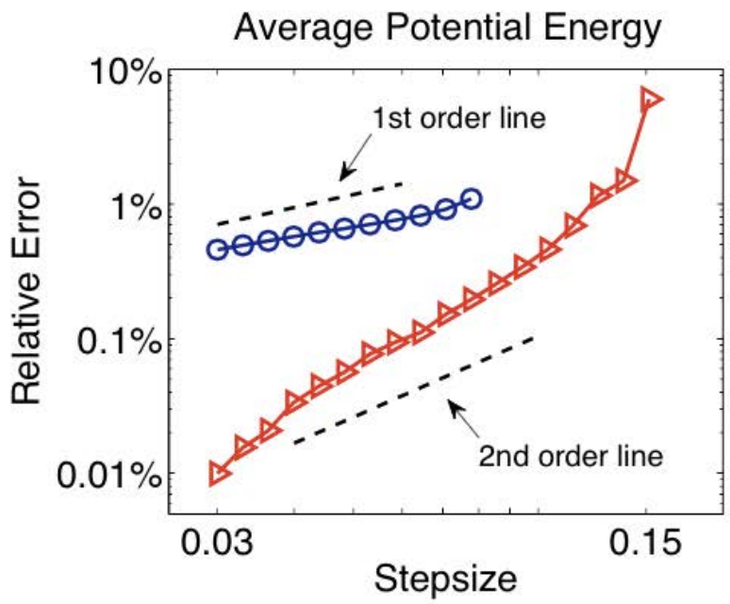
configurational temperature





Comparison with Chen et al. (SGNHT, NIPS 2015)





Bayesian Logistic Regression

$$\pi(y_i|\mathbf{x}_i, \boldsymbol{\beta}) = f(y_i \boldsymbol{\beta}^T \mathbf{x}_i)$$

f: logistic function

data

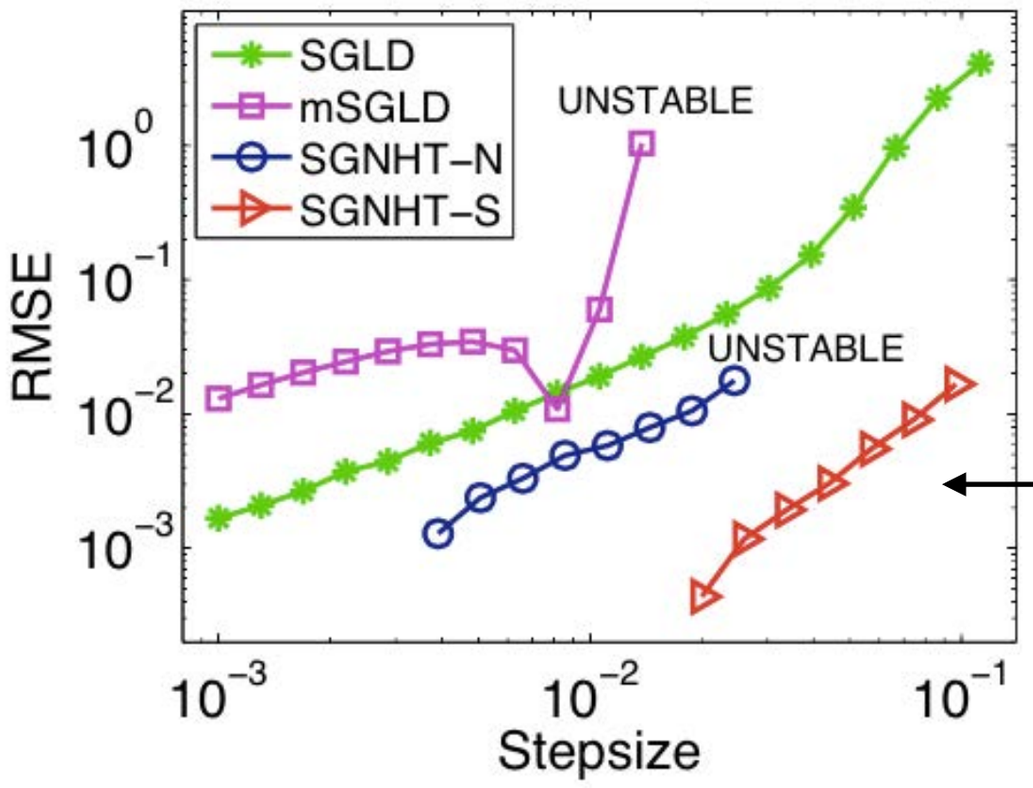
covariates

parameters

posterior parameter distribution

$$\pi(\boldsymbol{\beta}) \propto \exp\left(-\frac{1}{2}\|\boldsymbol{\beta}\|^2\right) \prod_{i=1}^{N} f(y_i \boldsymbol{\beta}^T \mathbf{x}_i)$$

Gaussian prior



superior accuracy for BADODAB splitting

$L^2(\mu)$ Convergence Analysis

[w./Matthias Sachs & Gabriel Stoltz, 2018]

HYPOCOERCIVITY FOR LINEAR KINETIC EQUATIONS CONSERVING MASS,
J. Dolbeault, C. Mouhot, & C. Schmeiser, Trans. AMS 367, 3807-3828, 2015

$$\partial_t f + \mathsf{T}\, f = \mathsf{L}\, f$$

T: anti-symmetric; **L** symmetric part of operator in $L^2(\mu)$
Π : projection operator based on p-marginalization

Define an entropy functional,

$$\mathsf{H}[f] := \frac{1}{2}\,\|f\|^2 + \varepsilon\,\langle \mathsf{A}\, f, f\rangle\,, \quad \text{with } \mathsf{A} := \big(1 + (\mathsf{T}\,\Pi)^*(\mathsf{T}\,\Pi)\big)^{-1}(\mathsf{T}\,\Pi)^*$$

and show it is decreasing along the flow: $\frac{\mathrm{d}}{\mathrm{d}t}\,\mathsf{H}[f] \le -\frac{2\,\kappa}{1+\varepsilon}\,\mathsf{H}[f]\,,$

Consequence: $L^2(\mu)$ hypocoercivity + spectral gap

DMS Convergence Analysis of AdL

$$
\begin{aligned}
\mathrm{d}q_t &= p_t \mathrm{d}t, \\
\mathrm{d}p_t &= \left(-\nabla U(q_t) - \epsilon^{-1}\xi p_t - \gamma p_t\right)\mathrm{d}t + \sqrt{\frac{2\gamma}{\beta}}\mathrm{d}W_t, \\
\mathrm{d}\xi_t &= \epsilon^{-1}\left(|p|^2 - n\beta^{-1}\right)\mathrm{d}t,
\end{aligned}
$$

Assume a Poincaré inequality of the form

$$
\left\|\varphi - \int_\Omega \varphi \mathrm{d}\mu_q\right\|^2_{L^2(\mu_q)} \leq \frac{1}{C}\left\|\nabla\varphi\right\|^2_{L^2(\mu_q)}, \qquad \varphi \in H^1(\mu_q)
$$

Then

$$
\forall t \geq 0, \quad \left\|e^{t\mathscr{L}_{\gamma,\epsilon}}\left(\varphi - \int \varphi \mathrm{d}\mu\right)\right\|_{L^2(\mu)} \leq K_{\epsilon,\gamma} e^{-\kappa_{\epsilon,\gamma} t}\left\|\varphi - \int \varphi \mathrm{d}\mu\right\|_{L^2(\mu)}
$$

with

$$
\kappa_{\epsilon,\gamma} \in O\left(\min(1, \epsilon^{-2}) \cdot \min(\gamma, \gamma^{-1})\right)
$$

CLT for AdL

Bhattacharya et al 1982: *"On the functional central limit theorem…"*

CLT: $\sqrt{t}\,(\mu_t[\varphi] - \mu[\varphi]) \sim \mathcal{N}(0, \sigma_\varphi^2), \text{ as } t \to \infty.$

Take Home Message #2.

Adaptive Langevin dynamics offers a simple method for removing bias in noisy gradient simulations while maintaining high accuracy (low bias).

2.2 The problem of the timestep

Harmonic Oscillator

$$\dot{q} = p$$

$$\dot{p} = -\Omega^2 q$$

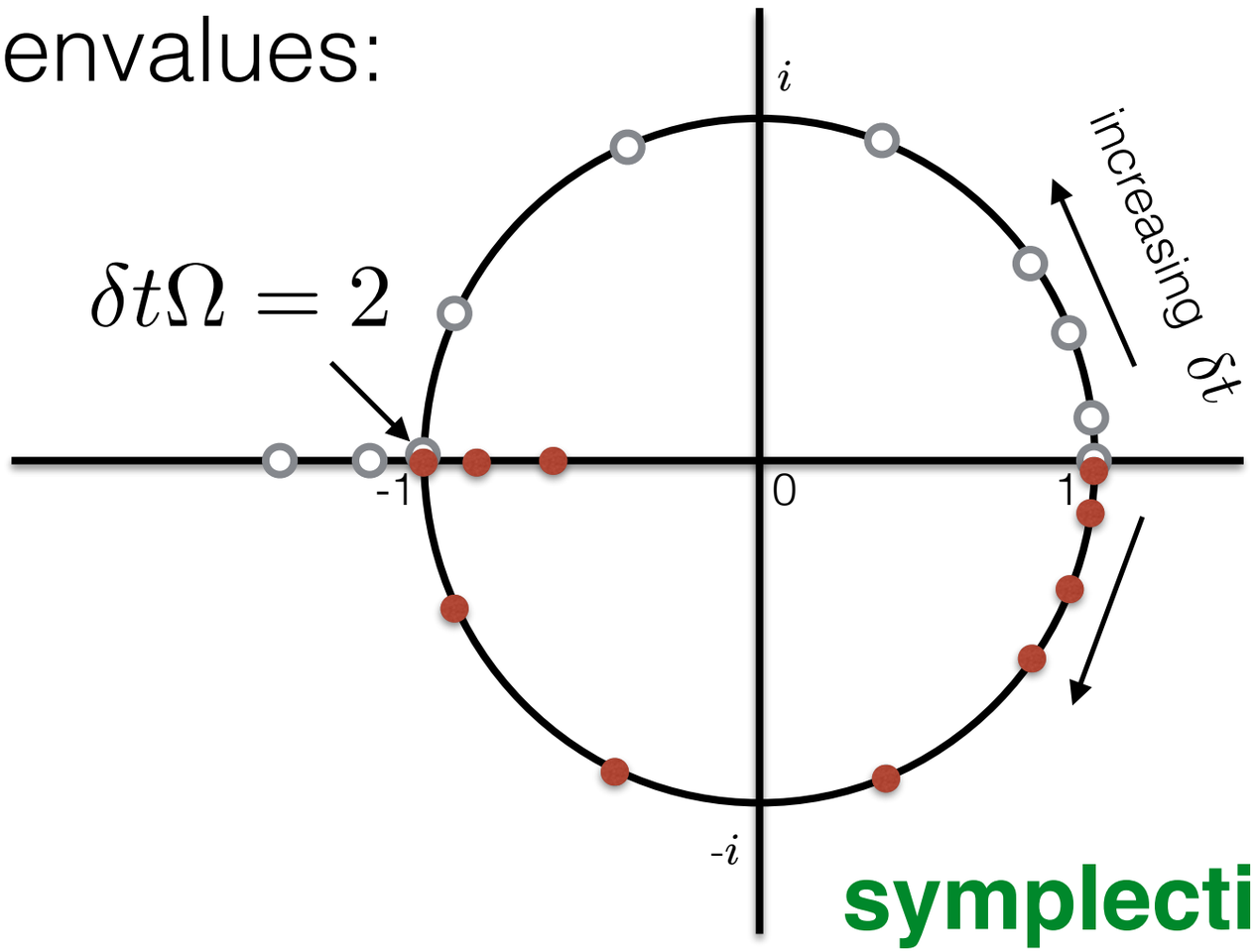
leapfrog/Verlet

$$q_{n+1} = q_n + \delta t p_{n+1/2}$$

$$p_{n+1/2} = p_n - \frac{\delta t}{2}\Omega^2 q_n$$

$$p_{n+1} = p_{n+1/2} - \frac{\delta t}{2}\Omega^2 q_{n+1}$$

eigenvalues:



stable for $\delta t\Omega \leq 2$

symplectic for all δt

but only **useful** for <u>small</u> δt

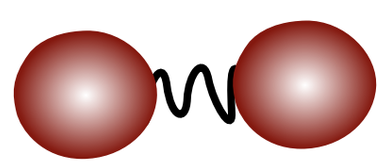
Stepsize Restriction

H.O. $H = \frac{p^2}{2} + \frac{\Omega^2 q^2}{2}$

stability threshold

$$\delta t < 2/\Omega$$

think bonds



in molecular dynamics bonds to H atoms have high frequencies

$$\delta t < 3\text{fs} = 3 \times 10^{-15}\text{s}$$

The Timestep Problem in MD

Find timestepping methods that allow

$$\delta t \gg \delta t_{\mathrm{Verlet}}$$

Three possibilities:

~~Use an **implicit method**~~

Eliminate the motion of fast components

constraints

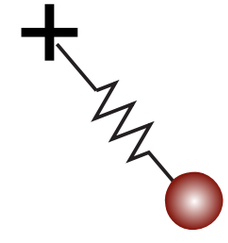
Isolate the stiff terms for efficient treatment

multiple timestepping

Stiff Spring Oscillator

<u>*energy:*</u>

$$H = \frac{|p|^2}{2m} + \frac{k(|q| - L)^2}{2} = E$$

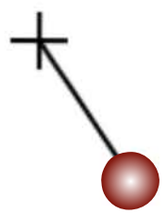


<u>*equations of motion:*</u>

$$\frac{\mathrm{d}q}{\mathrm{d}t} = p/m$$

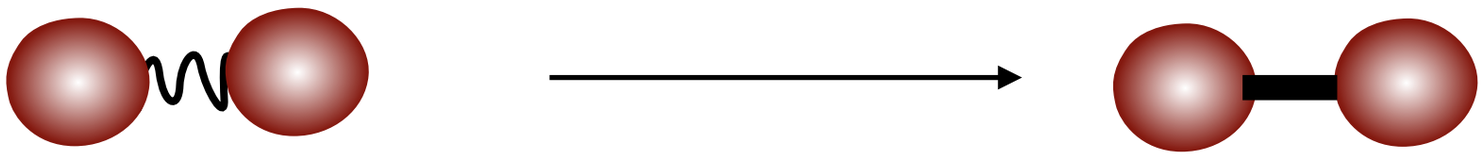
$$\frac{\mathrm{d}p}{\mathrm{d}t} = -k\left(1 - \frac{L}{|q|}\right) q$$

If energy is fixed, then in the limit of large k we must have $|q| \sim L$



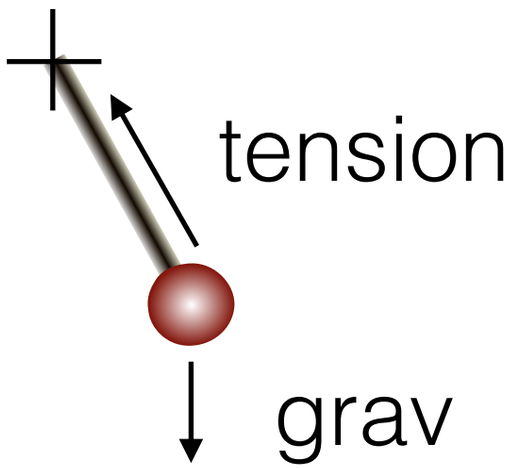
Constrained dynamics: $|q| = L$

Pendulum



stiff harmonic
spring with rest length

rigid rod
(holonomic constraint)



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

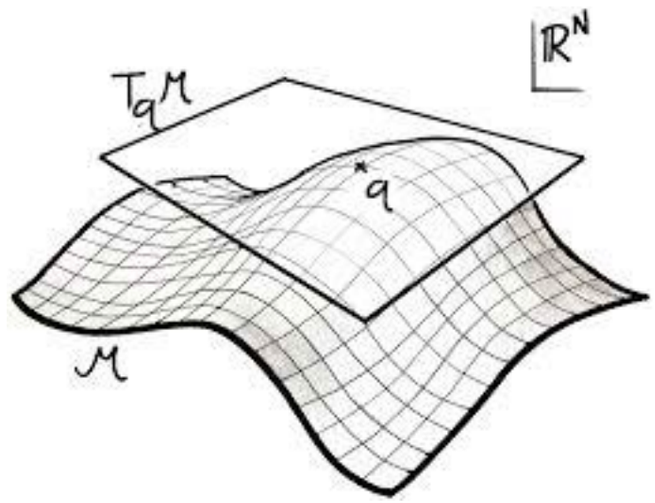
$$\dot{p} = -\begin{bmatrix} 0 \\ g \end{bmatrix} - q\lambda$$

$$0 = q \cdot q - 1$$

Constrained Hamiltonian Systems

$$\frac{\mathrm{d}}{\mathrm{d}t}q = \mathbf{M}^{-1}p$$

$$\frac{\mathrm{d}}{\mathrm{d}t}p = F - \sum_{i=1}^{m} \lambda_i \nabla g_i(q)$$



$$0 = g_j(q), \qquad j = 1, 2, \ldots, m$$

$$0 = \nabla g_j(q)^T \mathbf{M}^{-1} p, \qquad j = 1, 2, \ldots, m$$

“hidden constraints”

“cotangent bundle”

$$T^*\mathcal{M} = \{(q,p)\,|g_j(q) = 0,\, \nabla g_j(q)^T \mathbf{M}^{-1} p = 0, j = 1, 2, \ldots, m\}$$

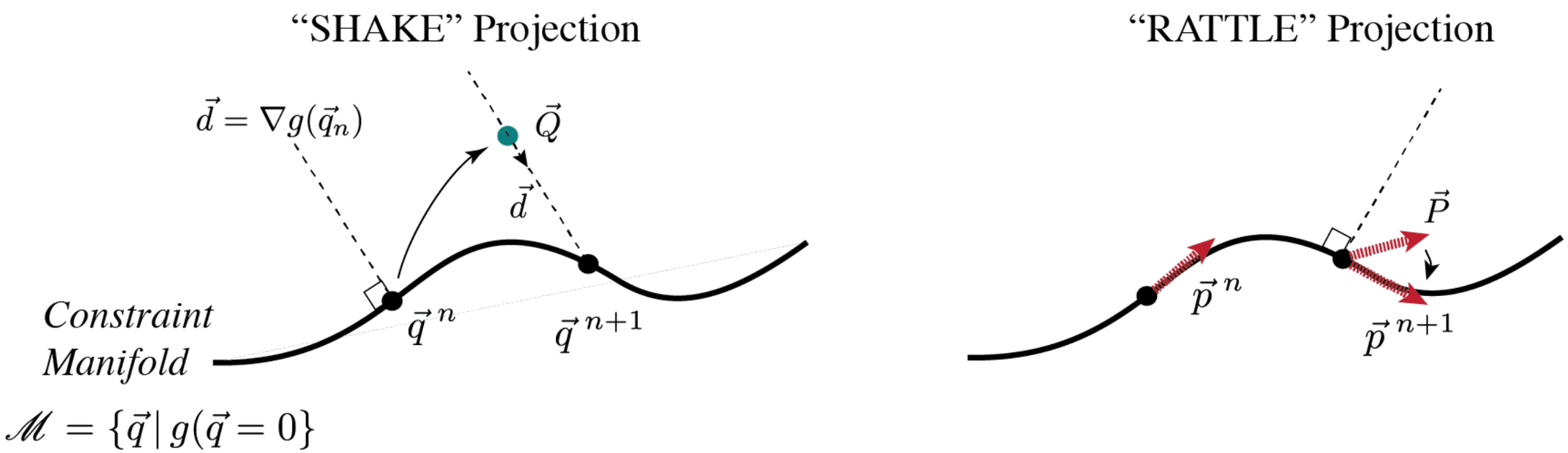
symplectic: dynamics preserves $[\mathrm{d}q \wedge \mathrm{d}p]_{T^*\mathcal{M}}$

SHAKE

[Berendsen, Ciccotti, Ryckaert 1977]

RATTLE

[Andersen 1983]



L. & Skeel, J.Comput. Phys, 1994

SHAKE and **RATTLE** are **symplectic** methods
and are actually the same method (**conjugate** methods).

Constraint preserving methods

$$H(q,p) = |p|^2/2 + U(q)$$

$$\begin{aligned}
\dot{q} &= p \\
\dot{p} &= -\nabla_q U(q) - \lambda \nabla g \\
0 &= g(q) \\
0 &= \nabla g(q) \cdot p
\end{aligned}$$

$$(q,p) \mapsto (Q,P)$$

approximating a step in time of size h

$$\begin{aligned}
Q &:= q + h\tilde{P} \\
\tilde{P} &:= p - h\nabla U(q) - \lambda \nabla g(q) \\
0 &= g(Q)
\end{aligned}$$

$$\begin{aligned}
P &:= \tilde{P} - \mu \nabla g(Q) \\
0 &= \nabla g(Q) \cdot P
\end{aligned}$$

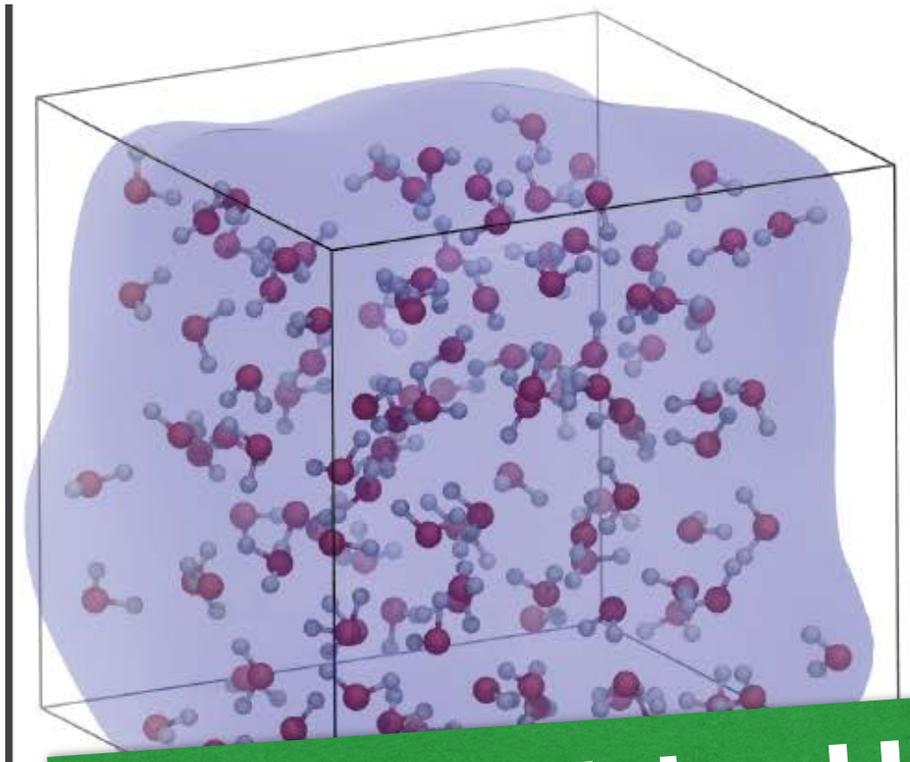
first order

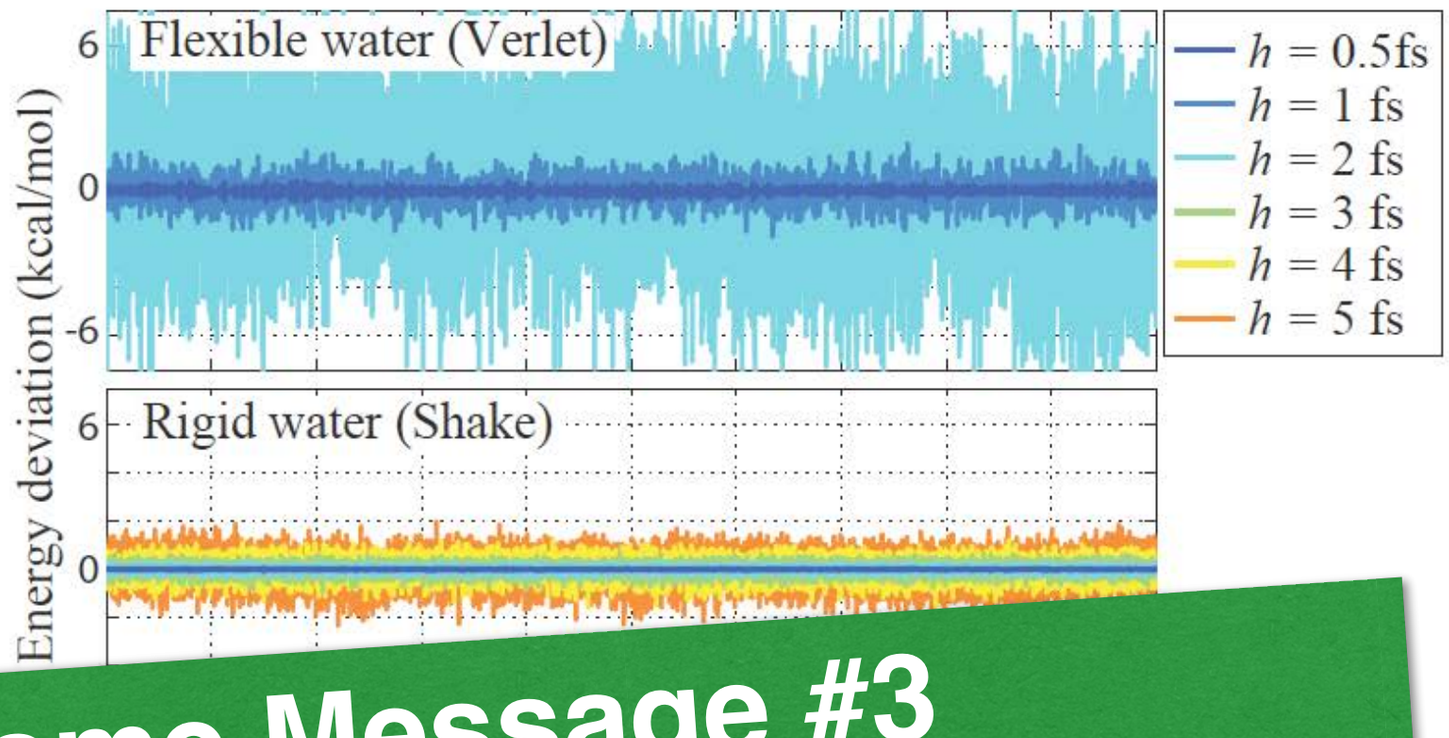
$$\begin{aligned}
Q &:= q + (h/2)\tilde{P} \\
\tilde{P} &:= p - (h/2)\nabla U(q) - \lambda \nabla g(q) \\
0 &= g(Q)
\end{aligned}$$

$$\begin{aligned}
P &:= \tilde{P} - (h/2)\nabla U(Q) - \mu \nabla g(Q) \\
0 &= \nabla g(Q) \cdot P
\end{aligned}$$

2nd order “RATTLE”

Microcanonical Sampling





Take Home Message #3

Constraints don't buy much stability in deterministic MD, but they do improve accuracy!

Ve

1fs with 1 kcal/mol energy accuracy

SHAKE (RATTLE) OK up to as much as

3-4fs with energy accuracy

Multiple timestepping

Multiple Timestepping (RESPA)

Tuckerman, Martyna, Berne '90

also Grubmüller, Heller, Windemuth, Schulten '91

Some forces are ***fast-changing*** (e.g. short-ranged) but **cheap** to compute

Some forces are ***slow-changing*** (e.g. long-ranged) and **costly** to compute

$$\exp\left(\frac{\delta t}{2}\mathcal{L}_{U\text{slow}}\right) \times \left[\exp\left(\frac{\delta t}{2r}\mathcal{L}_{U\text{fast}}\right)\exp\left(\frac{\delta t}{r}\mathcal{L}_{K}\right)\exp\left(\frac{\delta t}{2r}\mathcal{L}_{U\text{fast}}\right)\right]^{r} \times \exp\left(\frac{\delta t}{2}\mathcal{L}_{U\text{slow}}\right)$$

Many fast cheap evaluations
Few slow costly ones

Linear Model Problem

model problem

$$H = \frac{p^2}{2} + \frac{(1+\Omega^2)q^2}{2} \qquad \Omega \gg 1$$

idealized *multiple timestepping (fast solve is exact)*

$$H(q,p) = p^2/2 + U_{\mathrm{S}}(q) + U_{\mathrm{F}}(q)$$

$$\boxed{U_{\mathrm{S}}(q) = q^2/2} \qquad \boxed{U_{\mathrm{F}}(q) = \Omega^2 q^2/2}$$

Kick with U_{S}	Solve $H_{\mathrm{F}} = p^2/2 + U_{\mathrm{F}}(q)$	Kick with U_{S}
$\delta t/2$	δt	$\delta t/2$

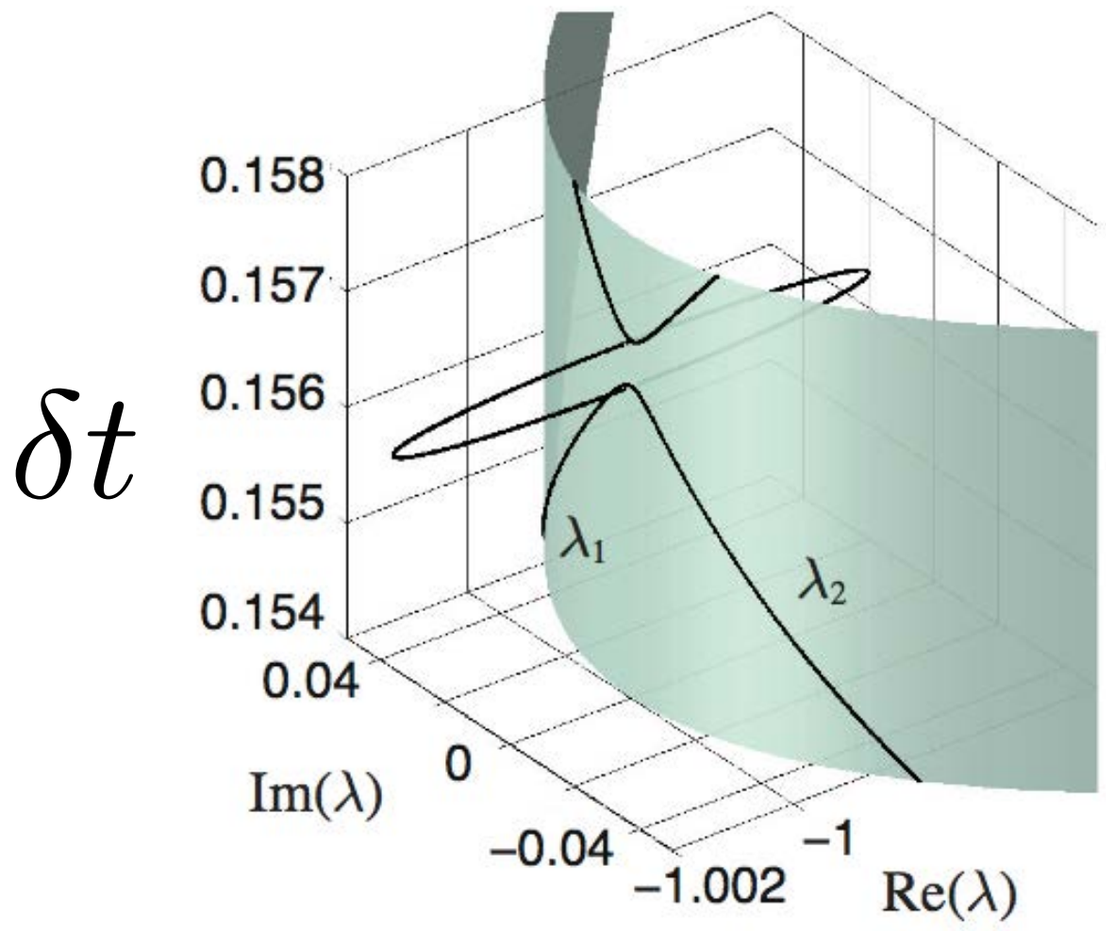
Resonance

Eigenvalues of timestep map:

$$\lambda_1\lambda_2 = 1; \quad \lambda_1 + \lambda_2 = 2\cos(\delta t\Omega) - \frac{\delta}{\Omega}\sin(\delta t\Omega)$$

$$\varepsilon \propto \delta t - \pi/\Omega < 0 \qquad \lambda = -1 \pm (|\varepsilon|/2)^{1/2} + O(\varepsilon)$$

instabilities…



Resonance

For the harmonic model, Verlet introduces a stability restriction

$$\delta t < 2/\Omega$$

Multiple timestepping, in the idealized form described here has a stability restriction of about

$$\delta t < \pi/\Omega$$

i.e., not very dramatic improvement. In practice, we are limited to around **3.5fs**

Mollified Impulse Method

Garcia-Archilla, Sanz-Serna and Skeel 1998

To stabilize multiple timestepping, one idea is to average out over an associated fast dynamics to ``mollify'' the impulse in RESPA

starting from e.g $(q,0)$, solve

$$H^{\mathrm{fast}}(\boldsymbol{q},\boldsymbol{p}) = \boldsymbol{p}^T \boldsymbol{M}^{-1}\boldsymbol{p}/2 + U_{\mathrm{F}}(\boldsymbol{q})$$

to produce a sequence $\theta_0 = q, \theta_1, \theta_2, \ldots, \theta_K$
then define:

$$\mathcal{A}(\boldsymbol{q}) = \frac{1}{K+1}\sum_{i=0}^{K} \phi_i \boldsymbol{\theta}_i(\boldsymbol{q})$$

Advance the step using a *mollified impulse*:

$$\tilde{U}(q) := U_s(\mathcal{A}(q))$$

Pushing the limits

SHAKE-MOLLY *[Izaguirre, Reich, Skeel 1999]* constrains the slow force evaluations to lie exactly at the bond stretch minima (multiple timestepping + constraints).

With the very best **deterministic** schemes, however, we find for a biological molecule:

B

Take Home Message #4

Combining constraints and multiple timestepping can improve accuracy and stability

Stochastic Methods

Introducing random perturbations might seem to complicate the numerical integrator.

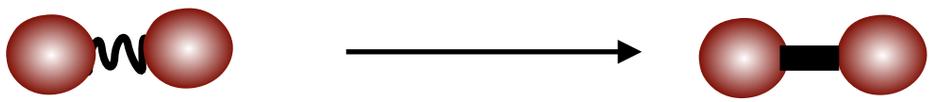
In fact, if done correctly, it is possible to gain in two ways:

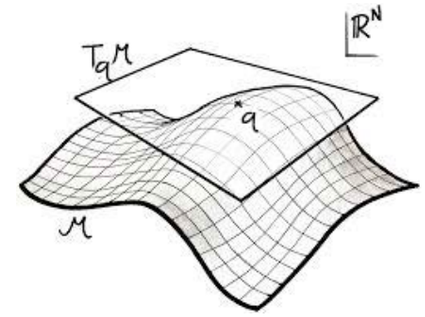
1) a significant **stability improvement** is possible
2) substantially **higher accuracy** is possible by shifting emphasis to the **invariant distribution**

These benefits carry over to **multiple timestepping** and **constraints,** and **the combination can be even more powerful than expected!!**

Constrained Langevin Dynamics

Constrained Langevin Dynamics





$$\frac{\mathrm{d}}{\mathrm{d}t}q = \mathbf{M}^{-1}p$$

$$\frac{\mathrm{d}}{\mathrm{d}t}p = F - \gamma p + \sqrt{2\mathrm{k_B T}\gamma}\mathbf{M}^{1/2}\eta(t) - \sum_{i=1}^{m}\lambda_i \nabla g_i(q\,)$$

Lagrange multipliers

$$0 = g_j(q\,), \qquad j = 1, 2, \ldots, m$$

$$0 = \nabla g_j(q\,)^T \mathbf{M}^{-1}p, \qquad j = 1, 2, \ldots, m$$

$$T^*\mathcal{M} = \{(q,p)\,|g_j(q) = 0,\, \nabla g_j(q\,)^T\mathbf{M}^{-1}p = 0, j = 1, 2, \ldots, m\}$$

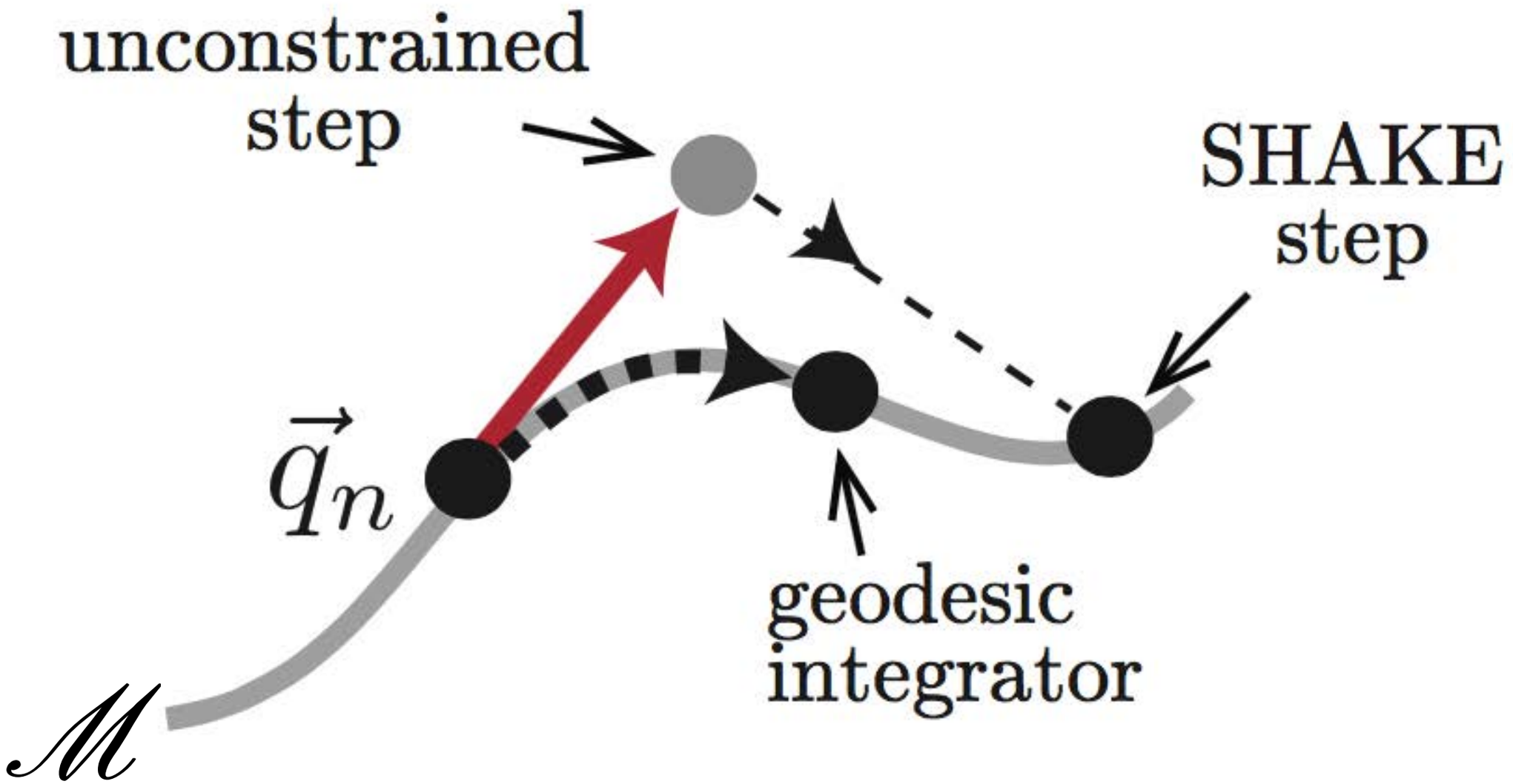
Some technical aspects:

Lelievre, Rousset and Stoltz, Math. Comp., 2010

Geodesic Integrator

B.L. and G. Patrick, J. Nonlin. Sci. 1996 (deterministic)

B.L. and C. Matthews, Proc Roy Soc A 2016 (stochastic)



Geodesic Integrator

B.L. and G. Patrick, J. Nonlin. Sci. 1996 (deterministic)
B.L. and C. Matthews, Proc Roy Soc A 2016 (stochastic)

An alternative to SHAKE discretization

Idea: *preserve the configuration manifold during position moves and the cotangent space during impulse.*

The natural constrained analogue of ***Verlet (BAB)*** *or* ***BAOAB***

g-Verlet: $\Phi_{h,H_{T^*\mathscr{M}}} \approx \Phi^{\mathcal{B}}_{h/2,U_{T^*\mathscr{M}}} \circ \Phi^{\mathcal{A}}_{h,T_{T^*\mathscr{M}}} \circ \Phi^{\mathcal{B}}_{h/2,U_{T^*\mathscr{M}}}$

Combines: **geodesic flow** **projected "kicks"**
projective OU

Constrained OU

$$\mathrm{d}p = -\gamma \mathbf{\Pi} p \mathrm{d}t + \sqrt{2\gamma \mathrm{k_B T}} \mathbf{\Pi} \mathbf{M}^{1/2} \mathrm{dW}$$

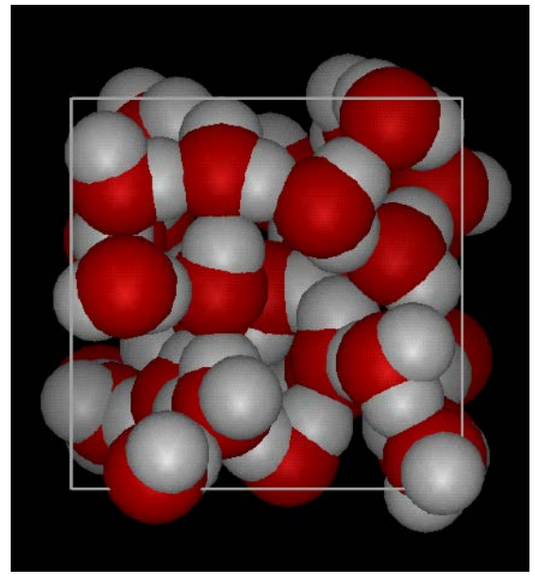
Projector onto the co-tangent space

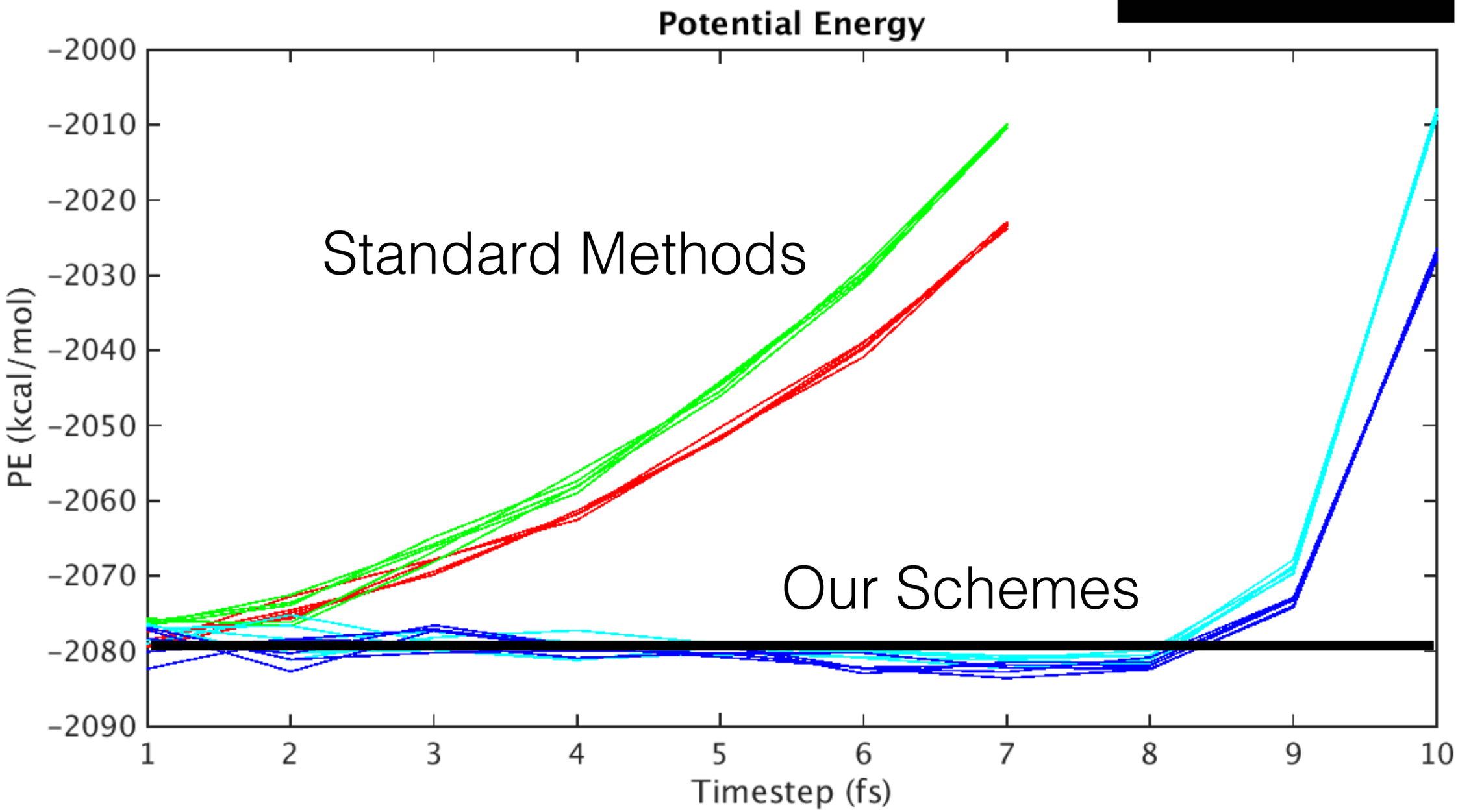
Easily solved using Rodrigues' formula:

$$p\left(t\right) = e^{-\gamma t} p\left(0\right) + \sqrt{\mathrm{k_B T}(1 - e^{-2\gamma t})} \mathbf{\Pi} \mathbf{M}^{1/2} \mathrm{R}(t).$$

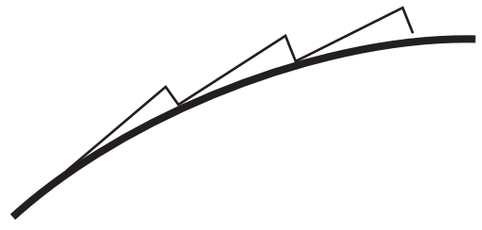
assuming $\mathbf{\Pi} p(0) = p(0)$

For rigid body water, implement the geodesic integrator using
SETTLE: ***exact geodesic steps***





Multiple-Timestepping Implementation



To implement the geodesic integrator, we use a sequence of SHAKE/RATTLE steps for the "**A**" step

These steps do not require re-evaluation of the force field, so each iteration is rel. cheap (in the MD universe).

The **B** and **O** (for Langevin) steps are simple RATTLE projections, so there is no significant added cost for these.

No need for Hessians/normal modes, or tuning/parameterization

Geodesic Integrator

$$\boldsymbol{p} \leftarrow \boldsymbol{p} + \tfrac{\delta t}{2}\boldsymbol{F}(\boldsymbol{q}) + \textstyle\sum_j \mu_j \boldsymbol{G}_j(\boldsymbol{q}) \in T^*_{\boldsymbol{q}}\mathcal{M},$$

Cotangent space projection

For k from 1 to K_r do:

$$(\boldsymbol{q},\boldsymbol{p}) \leftarrow A\left(\boldsymbol{q},\boldsymbol{p},\tfrac{\delta t}{2K_\mathrm{r}}\right)$$

end do

Several Rattle (geodesic flow) Steps

$$\boldsymbol{p} \leftarrow a_2\boldsymbol{p} + b_2\mathbf{M}^{1/2}\boldsymbol{R} + \textstyle\sum_j \mu_j \boldsymbol{G}_j(\boldsymbol{q}) \in T^*_{\boldsymbol{q}}\mathcal{M},$$

O Step

For k from 1 to K_r do:

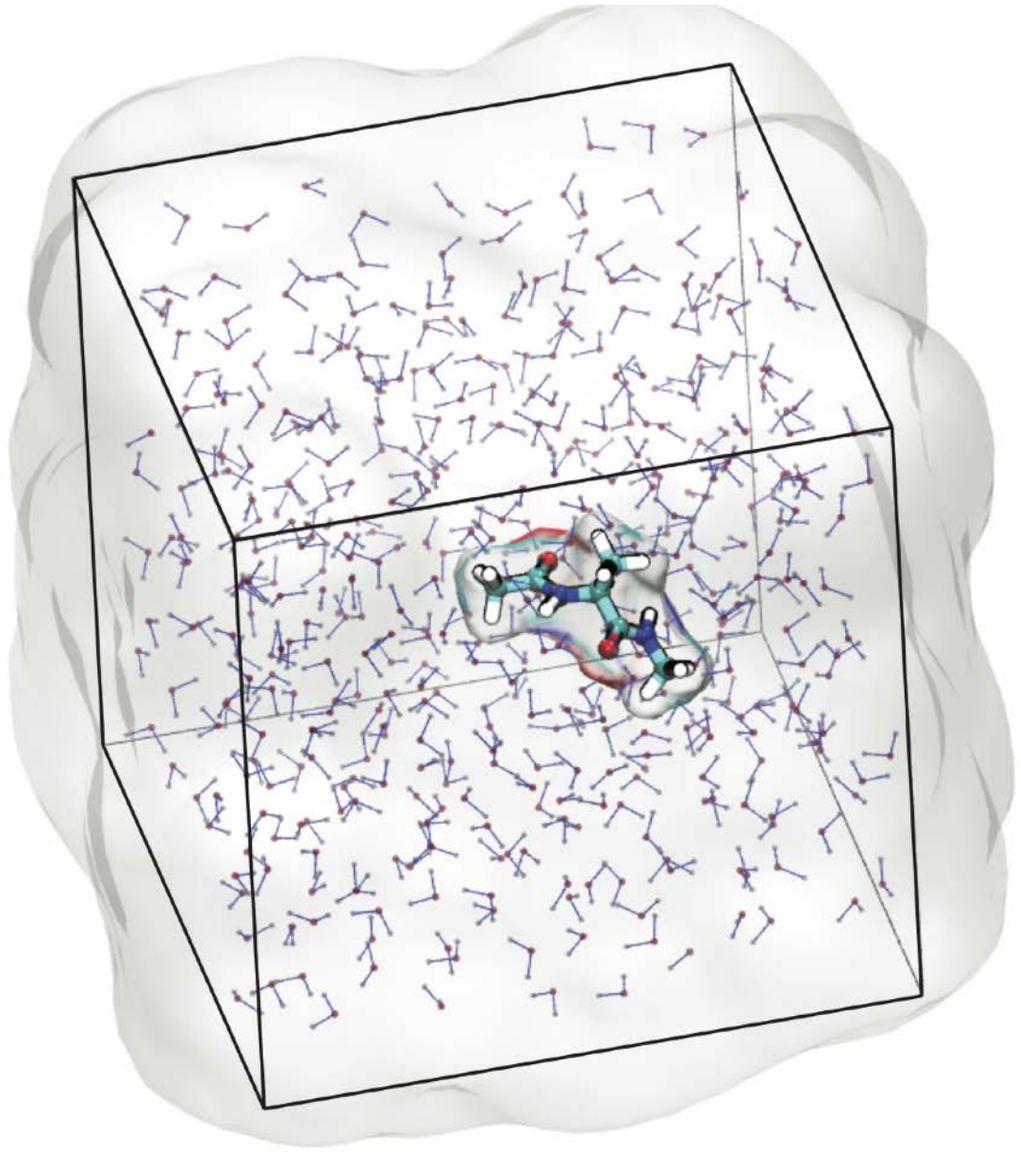
$$(\boldsymbol{q},\boldsymbol{p}) \leftarrow A\left(\boldsymbol{q},\boldsymbol{p},\tfrac{\delta t}{2K_\mathrm{r}}\right)$$

end do

$$\boldsymbol{p} \leftarrow \boldsymbol{p} + \tfrac{\delta t}{2}\boldsymbol{F}(\boldsymbol{q}) + \textstyle\sum_j \mu_j \nabla g_j(\boldsymbol{q}) \in T^*_{\boldsymbol{q}}\mathcal{M},$$

Solute-Solvent Splitting

The object of **bio-MD simulation** is virtually always a protein or nucleic acid fragment + solvent (water) bath.



Once the bonds and selected (H-X-H, X-Y-H) angles of the solute and waters are constrained, **the next fastest modes are due to flexible angle bonds of the solute.**

These would limit the stepsize to around **5fs**, even with the geodesic integrator.

Solute-Solvent Splitting

The obvious solution is to break the interaction forces into three pieces, denoted **PP** (protein-protein), **PS** (protein-solvent), and **SS** (solvent-solvent)

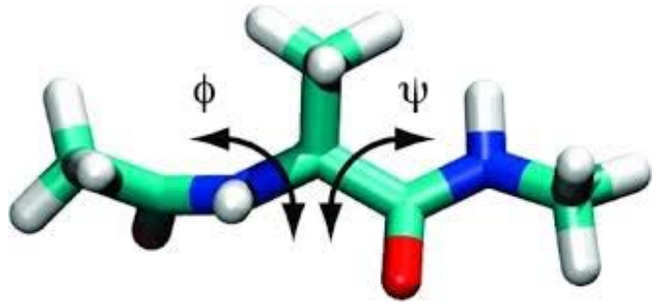
SS dominates the computational cost. **PP, PS** determine the stepsize. Therefore, consider ***stochastic two-level multiple timestepping:***

$$\exp\left(-\frac{h}{2}U_{\mathrm{SS}}(q)\right)\boxed{\exp\left(-\frac{h}{2}[T(p)+U_{\mathrm{PP}}(q)+U_{\mathrm{PS}}(q)]\right)}\exp\left(-\frac{h}{2}U_{\mathrm{SS}}(q)\right)$$

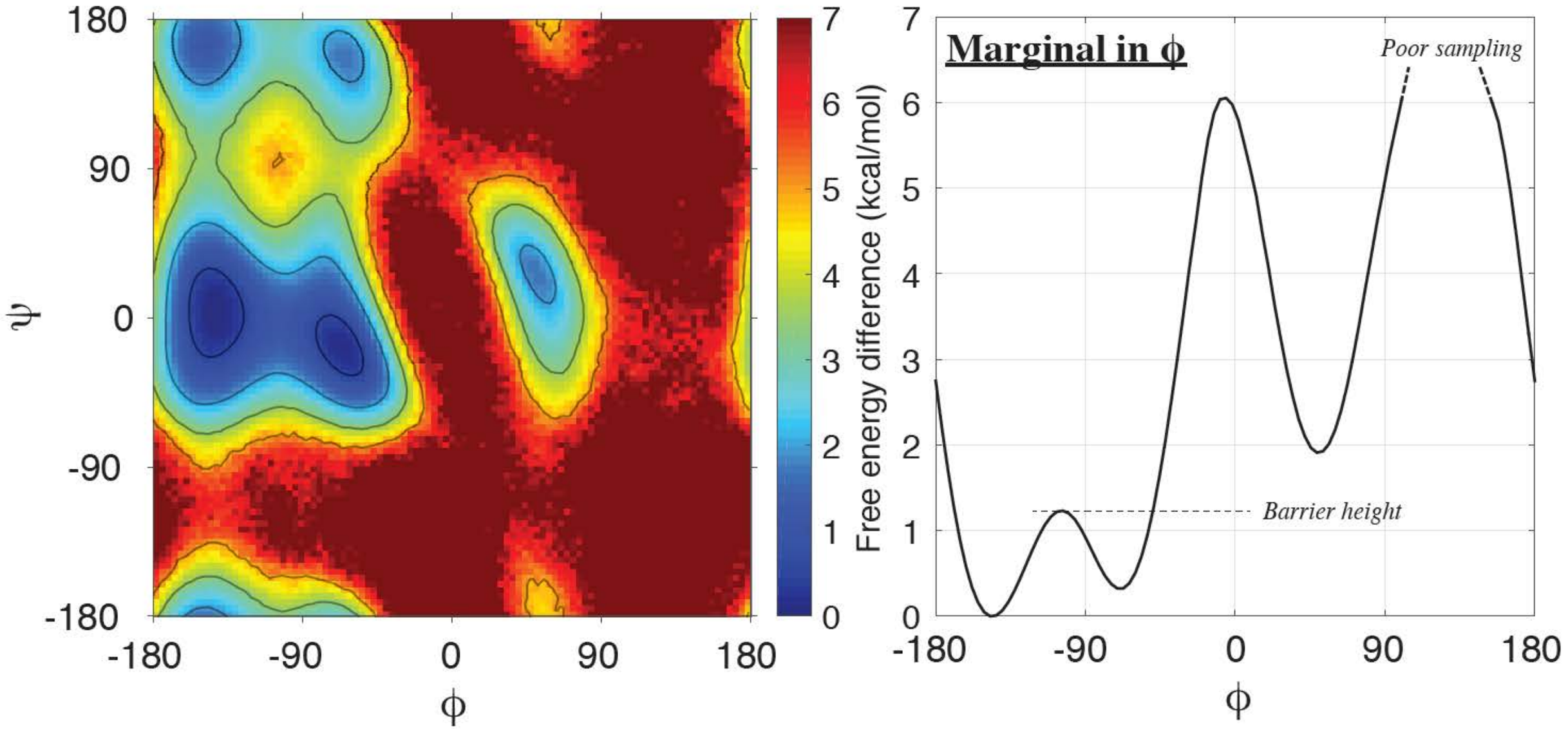
m=2 or m=3 iterations of a geodesic Langevin integrator using several RATTLE substeps

1. **PP+PS** is viewed as a many-body, stochastic system
2. Water motion can be implemented using **SETTLE**.

Alanine Dipeptide, Solvated, 300K, Tinker Code



illustrative FE surface



Comparisons

RATTLE: Standard RATTLE + Langevin

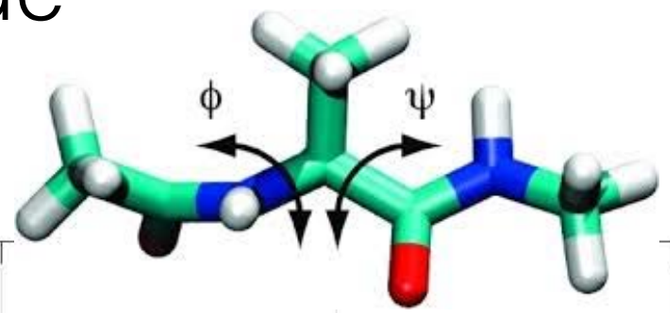
TINKER: Tinker's internal scheme

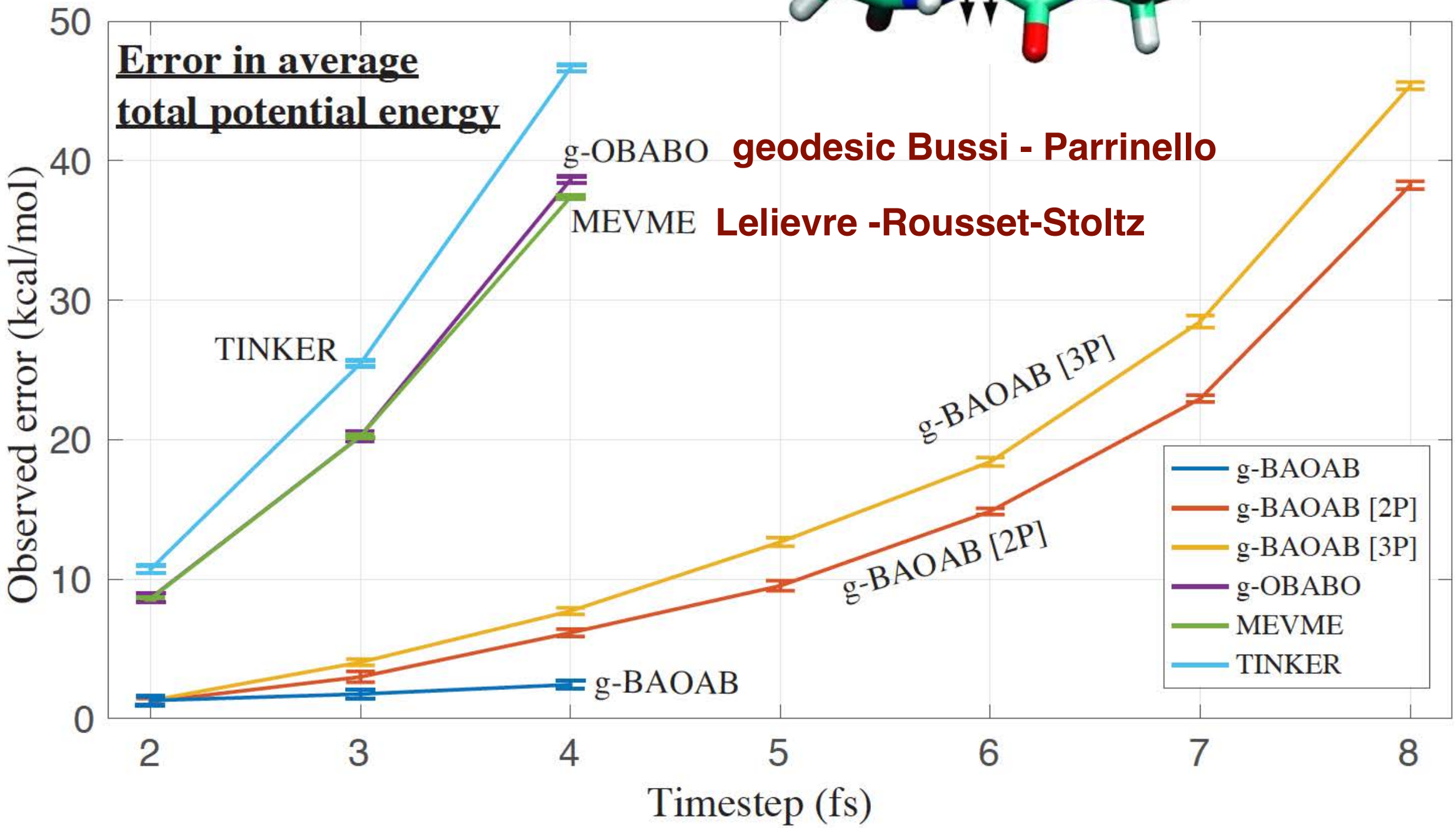
OBABO with constraint projections

MEVME: Scheme of Lelievre, Rousset, Stoltz 2010

MTS-BAOAB: geodesic-MTS method, using 20 steps to compute the geodesics and m=3 (PP+PS) steps per step

Alanine Dipeptide





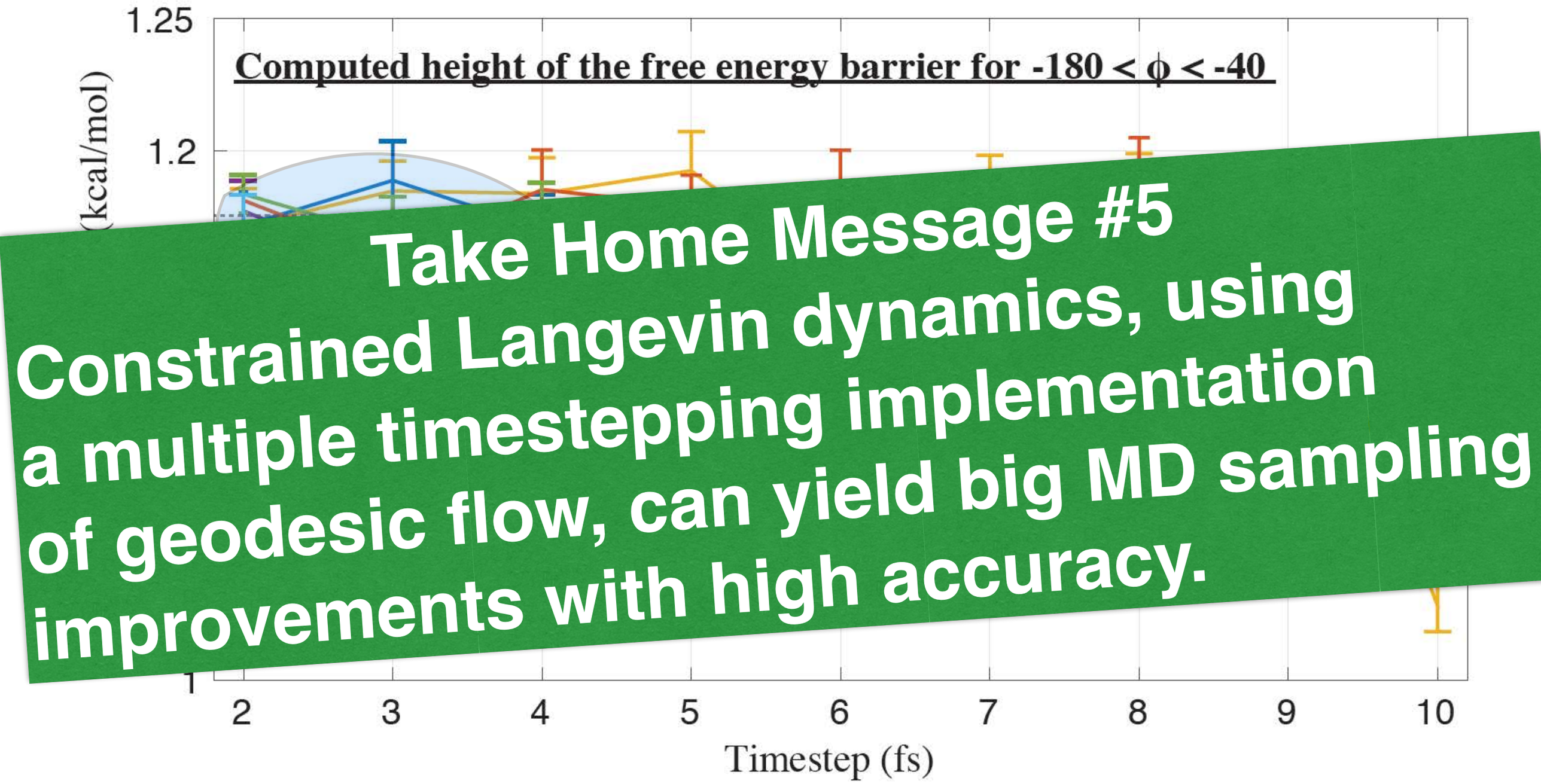
geodesic Bussi - Parrinello

Lelievre -Rousset-Stoltz

WHY do we get this extreme accuracy?

not explained by current BAOAB analysis

Alanine Dipeptide Effective Free Energy Barrier Height
using geodesic solvent-solute splitting



Take Home Message #5

Constrained Langevin dynamics, using a multiple timestepping implementation of geodesic flow, can yield big MD sampling improvements with high accuracy.

Stochastic Backpropagation and Approximate Inference in Deep Generative Models

Danilo J. Rezende, Shakir Mohamed, Daan Wierstra
{danilor, shakir, daanw}@google.com
Google DeepMind, London

Abstract

We marry ideas from deep neural networks and approximate Bayesian inference to derive a generalised class of deep, directed generative models, endowed with a new algorithm for scalable inference and learning. Our algorithm introduces a recognition model to represent an approximate posterior distribution and uses this for optimisation of a variational

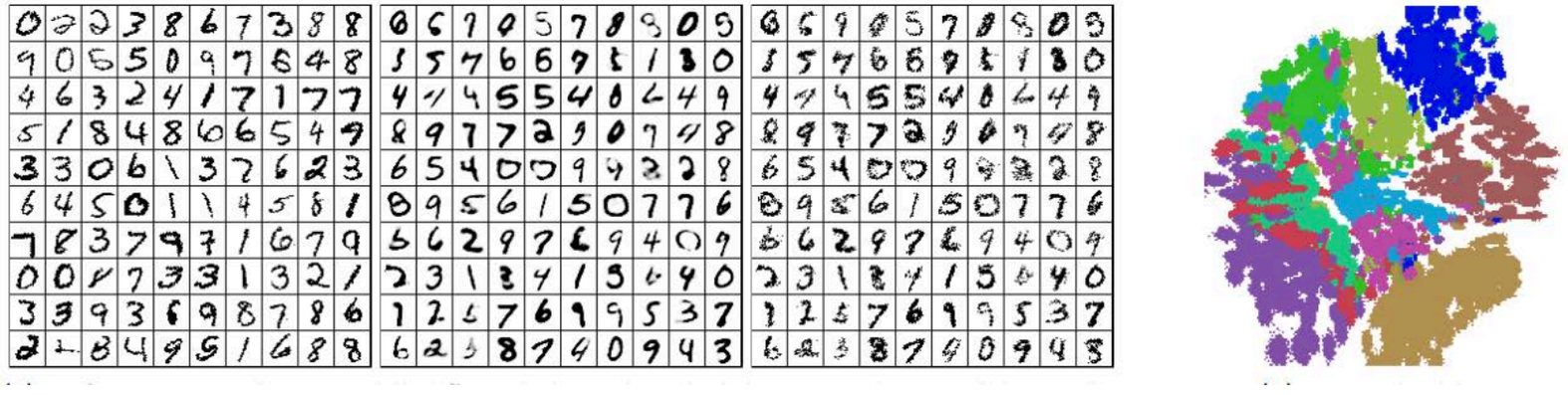
Uria et al., 2014; Gregor et al., 2014) can be easily sampled from, but in most cases, efficient inference algorithms have remained elusive. These efforts, combined with the demand for accurate probabilistic inferences and fast simulation, lead us to seek generative models that are i) *deep*, since hierarchical architectures allow us to capture complex structure in the data, ii) allow for *fast sampling* of fantasy data from the inferred model, and iii) are computationally *tractable and scalable* to high-dimensional data.

high D
inputs

$$U \mapsto Y = [\theta, X] = g(U) = [g_1(U), g_2(U)]$$

parameters

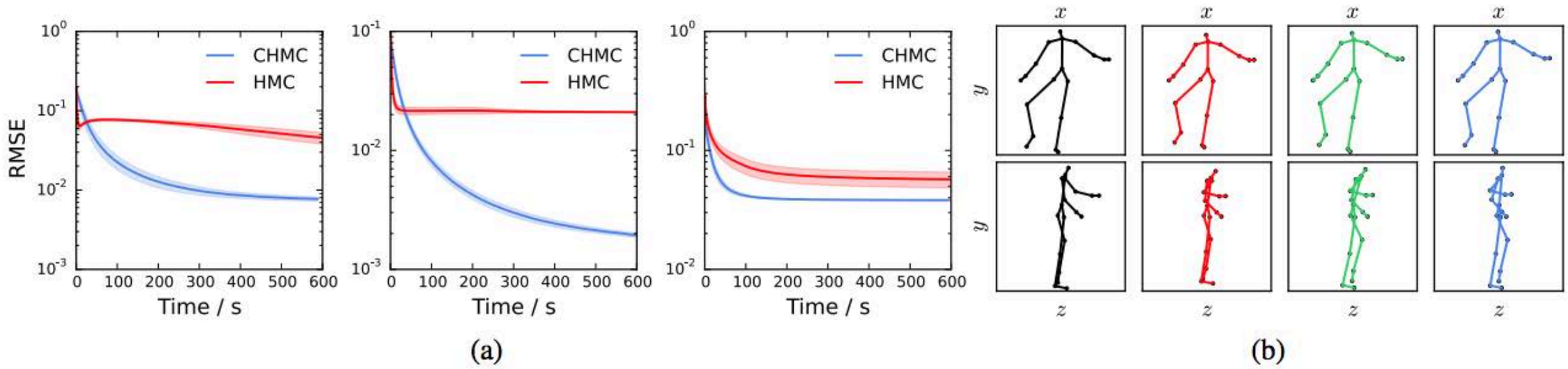
“code” features
low D



Constraints in Statistical Inference

From ***A. Storkey and M. Graham, AISTATS 2017***

Ex. Human Pose Reconstruction



—> Stochastic, constrained model within HMC

"We use a generalisation of the RATTLE scheme to simulate the dynamic. The inner updates of the state to solve for the geodesic motion on the constraint manifold are split into multiple smaller steps. This is a special case of the scheme described in **[L. & Matthews, 2016]** and allows more flexibility in choosing an appropriately small step-size to ensure convergence of the iterative solution of the equations projecting on to the constraint manifold while still allowing a more efficient larger step size for updates to the momentum"

Stochastic multiple timestepping

Langevin-RESPA

L. & Mathews, Molecular Dynamics, Springer 2015

It is natural to think that stochastic perturbations may have a stabilizing influence on multiple timestepping.

Consider the combination of Langevin (stochastic) dynamics with RESPA,

$$\begin{bmatrix} \mathrm{d}q \\ \mathrm{d}p \end{bmatrix} = \underbrace{\begin{bmatrix} p\,\mathrm{d}t \\ -\Omega^2 q\,\mathrm{d}t - \gamma p\,\mathrm{d}t + \sqrt{2\gamma/\beta}\,\mathrm{dW} \end{bmatrix}}_{\text{Fast}} + \underbrace{\begin{bmatrix} 0 \\ -q\,\mathrm{d}t \end{bmatrix}}_{\text{Slow}}$$

Configurational Sampling

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

$$e^{-\beta H} = e^{-\beta T(p)} e^{-\beta U(q)}$$

In MD we typically are interested in **q-dependent quantities**. The Hamiltonian is used to enhance exploration but ultimately we don't much care about momenta.

Question: can we use this freedom, together with stochastic dynamics, to **control resonance in multiple timestepping** while taking advantage of a known force field decomposition?

Fast Dynamics:

$$\boldsymbol{F} = \begin{bmatrix} 0 & 1 \\ -\Omega^2 & -\gamma \end{bmatrix} \qquad \begin{bmatrix} q(t) \\ p(t) \end{bmatrix} = \exp(\boldsymbol{F}t) \begin{bmatrix} q(0) \\ p(0) \end{bmatrix} + \boldsymbol{B}_t \mathbf{R}$$

$$\boldsymbol{B}_t \boldsymbol{B}_t^T = \frac{2\gamma}{\beta} \int_0^t \exp(\boldsymbol{F}(t-s)) \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \exp(\boldsymbol{F}^T(t-s)) \mathrm{d}s.$$

Combine with kicks by the slow force

$$z_{n+1} = X z_n + Y R_n$$

$$\boldsymbol{X} = \begin{bmatrix} 1 & 0 \\ -h & 1 \end{bmatrix} \exp(h\boldsymbol{F}), \qquad \boldsymbol{Y} = \begin{bmatrix} 1 & 0 \\ -h & 1 \end{bmatrix} \boldsymbol{B}_h$$

$$z_{n+1} = X z_n + Y R_n$$

Stability if

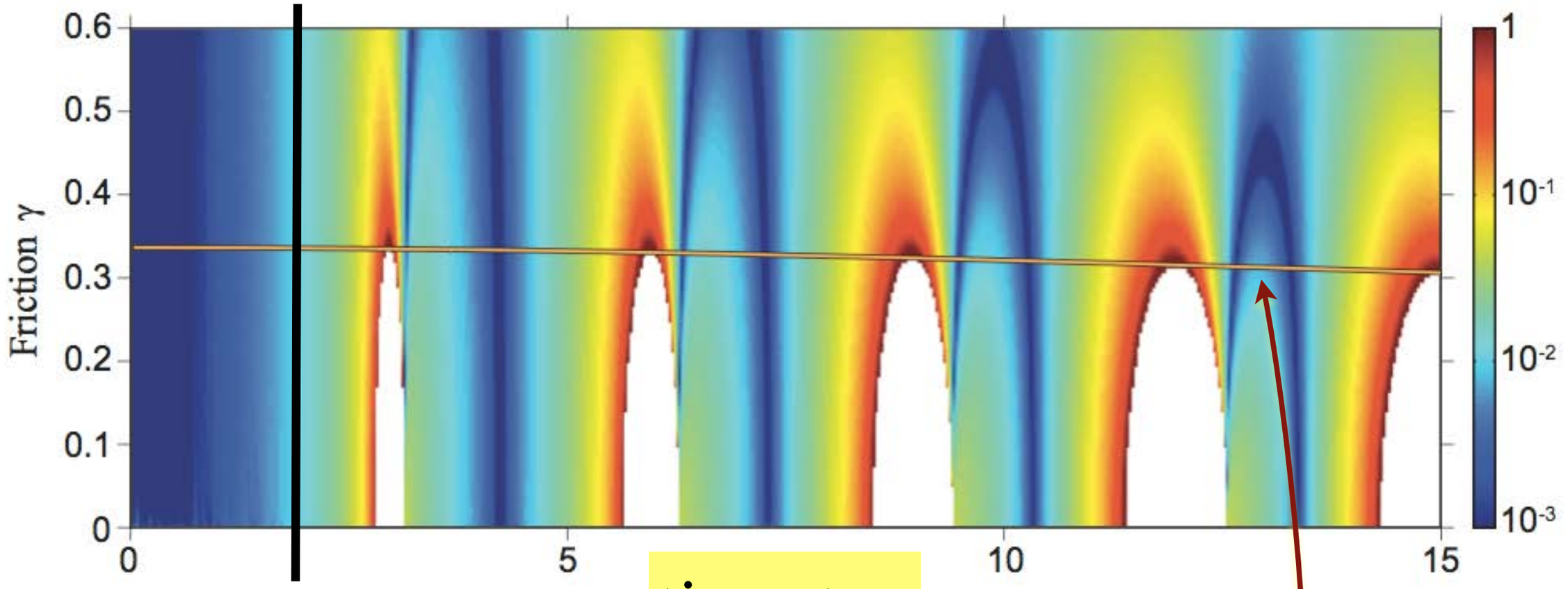
$$\rho(X) \leq 1$$

$$\left| \frac{\eta - h}{\eta} e^{(\eta-\gamma)h/2} + \frac{\eta + h}{\eta} e^{(-\eta-\gamma)h/2} \right| \leq 1 + e^{-\gamma h}$$

$$\frac{h}{\sqrt{4\Omega^2 - \gamma^2}} \leq \sinh\left(\frac{\gamma h}{2}\right)$$

sampling error

noise strength



timestep

Verlet stability threshold

$$\frac{h}{\sqrt{4\Omega^2 - \gamma^2}} \leq \sinh\left(\frac{\gamma h}{2}\right)$$

Langevin can stabilize RESPA but only at high friction

Stochastic Isokinetic Nosé-Hoover (SIN)

L., Margul and Tuckerman, Mol. Phys. 2013

view as 1 dof model

$$\dot{q} = p$$

$$\dot{p} = F(q) - \lambda p$$

$$\dot{\xi}_1 = -\lambda\xi_1 - \xi_1\xi_2$$

$$\dot{\xi}_2 = \xi_1^2 - \mathrm{k}T - \gamma\xi_2 + \sqrt{2\gamma\mathrm{k}T}\eta(t)$$

$$\lambda = \frac{2p\cdot F - \xi_1^2\xi_2}{2K(p,\xi_1)} \qquad \text{to make} \quad K(p,\xi_1) := p\cdot p + \frac{\xi_1^2}{2} = \mathrm{k}T$$

isokinetic constraint

Compressible statistical mechanics

M. E. Tuckerman, Y. Liu, G. Ciccotti, and G. J. Martyna, *J. Chem. Phys.* **115**, 1678 (2001).

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

conservation laws: $C_j(x) = c_j, j = 1, 2, \ldots, k$

phase space compressibility: $\kappa(x) = \mathrm{div} f(x)$

and find: $\kappa = \mathscr{L}_f w$

Then, if ergodic, the partition function is:

$$\Omega(N, V, \beta^{-1}) = \int e^{-w(x)} \Pi_{j=1}^{k} \delta[C_j(x) - c_j] \mathrm{d}x$$

$f(x)$

$$\dot{q} = p$$
$$\dot{p} = F(q) - \lambda p$$
$$\dot{\xi}_1 = -\lambda\xi_1 - \xi_1\xi_2$$
$$\dot{\xi}_2 = \xi_1^2 - \mathrm{k}T$$

OU

$$- \gamma\xi_2 + \sqrt{2\gamma\mathrm{k}T}\eta(t)$$

$$\kappa(x) = \mathrm{div}\, f(x) = -2\lambda - p\partial\lambda/\partial p - \xi_1\partial\lambda/\partial\xi_1 - \xi_2$$
$$= \mathscr{L}_f \left(\beta[U(q) + \xi_2^2/2]\right)$$

∴ ***an*** invariant distribution is given by

$$\rho(x) = e^{-\beta U(x)} e^{-\beta\xi_2^2/2} \delta[K(p, \xi_1) - \mathrm{k}T]$$

$$\rho(x) = e^{-\beta U(x)} e^{-\beta \xi_2^2/2} \delta[K(p, \xi_1) - \mathrm{k}T]$$

Dynamics evolves on the **generalized semi-cylinder** defined by periodic boundary conditions (q), and the isokinetic constraint on (p, ξ_1), with the restriction $\xi_1 > 0$

We need to demonstrate the **Hörmander condition** on the manifold, and find a **Lyapunov function**.

Lyapunov function: 😀

$$\Phi(\xi_2) = 1 + \xi_2^{2s}$$

Ergodicity Property of SIN

$$\boldsymbol{f} = p\partial_q - (q + \lambda p)\partial_p - (\lambda\xi_1 + \xi_2\xi_1)\partial_{\xi_1} + \left[(\xi_1^2 - k_{\mathrm{B}}T) - \gamma\xi_2\right]\partial_{\xi_2},$$

$$\boldsymbol{g} = \sigma\partial_{\xi_2}.$$

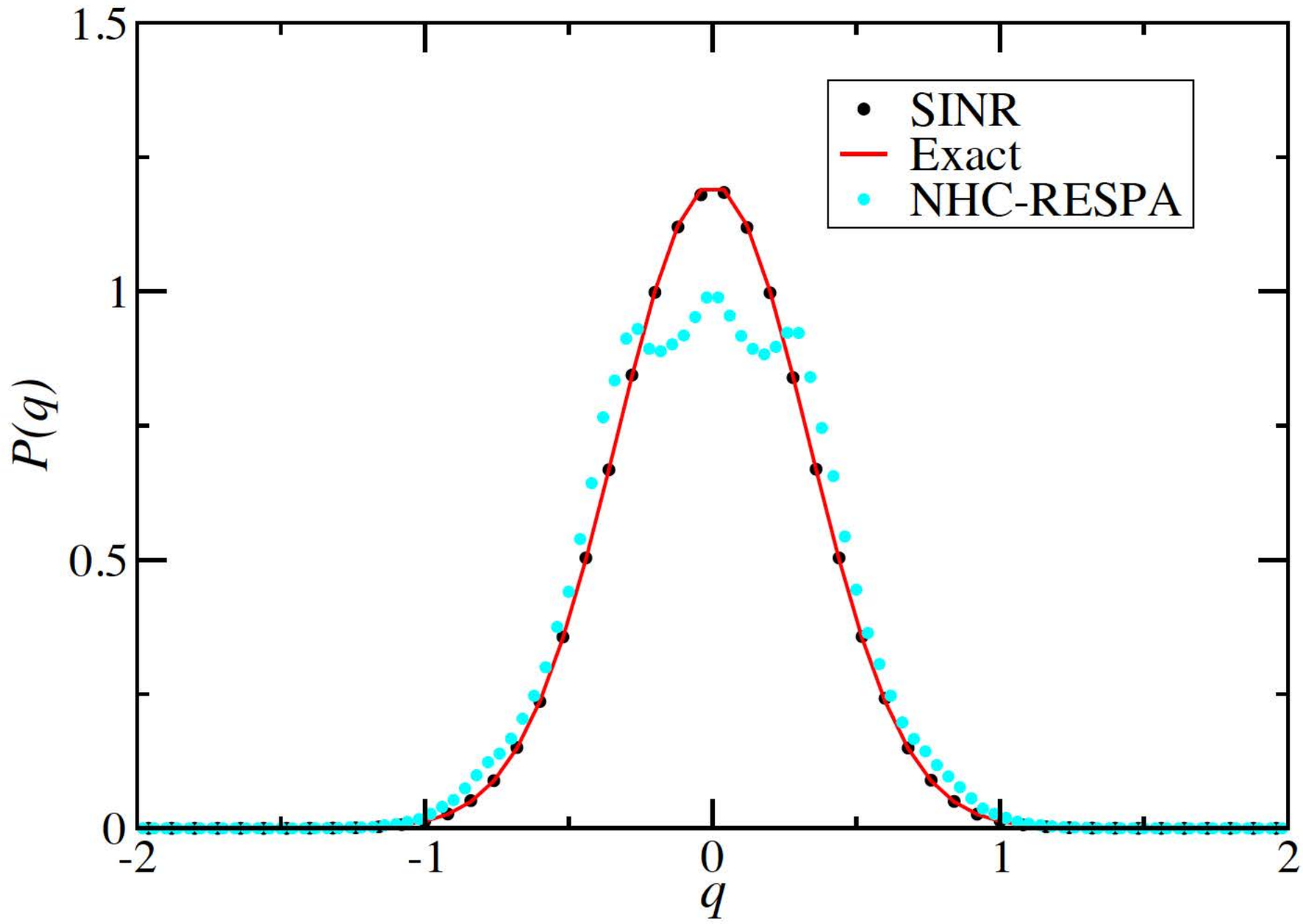
$$\mathrm{Span}\{\boldsymbol{f}, \boldsymbol{g}, [\boldsymbol{f}, \boldsymbol{g}], [\boldsymbol{f}, [\boldsymbol{f}, \boldsymbol{g}]], \boxed{[\boldsymbol{f}, [\boldsymbol{f}, [\boldsymbol{f}, \boldsymbol{g}]]]}\} = T\mathscr{M}$$

The Alexander Davie commutator 😃

Thus conclude (subject to a minorization condition, left as an exercise) that **SIN** is **ergodic** on $\mathscr{M}$ with a unique invariant distribution which, after marginalization reduces to:

$$\rho(x) = e^{-\beta U(x)}$$

sampling $U(q) = \frac{1}{2}\omega^2 q^2 + \frac{1}{4} g q^4$



SIN(R)

Stochastic Isokinetic Nosé-Hoover (RESPA)

$$\mathrm{d}q = p\mathrm{d}t,$$

$$\mathrm{d}p = F\mathrm{d}t - \left(\frac{p(F^{\mathrm{F}} + F^{\mathrm{S}}) - \frac{1}{2}\mu_1\xi_1^2\xi_2}{\Lambda} \right) p\mathrm{d}t,$$

$$\mathrm{d}\xi_1 = -\left(\frac{p(F^{\mathrm{F}} + F^{\mathrm{S}}) - \frac{1}{2}\mu_1\xi_1^2\xi_2}{\Lambda} \right) \xi_1\mathrm{d}t - \xi_2\xi_1\mathrm{d}t,$$

$$\mathrm{d}\xi_2 = \mu_2^{-1}(\mu_1\xi_1^2 - \beta^{-1})\mathrm{d}t - \gamma\xi_2\mathrm{d}t + \sigma\mathrm{d}\mathrm{W},$$

Kinetic part:

$$\mathcal{L}_{\mathrm{K}} = p\frac{\partial}{\partial q},$$

isokinetic Fast force:

$$\mathcal{L}_{\mathrm{F}} = \left[F^{\mathrm{F}} - (p^2/\Lambda)F^{\mathrm{F}}\right]\frac{\partial}{\partial p} - \frac{p\xi_1 F^{\mathrm{F}}}{\Lambda}\frac{\partial}{\partial \xi_1},$$

isokinetic Slow force:

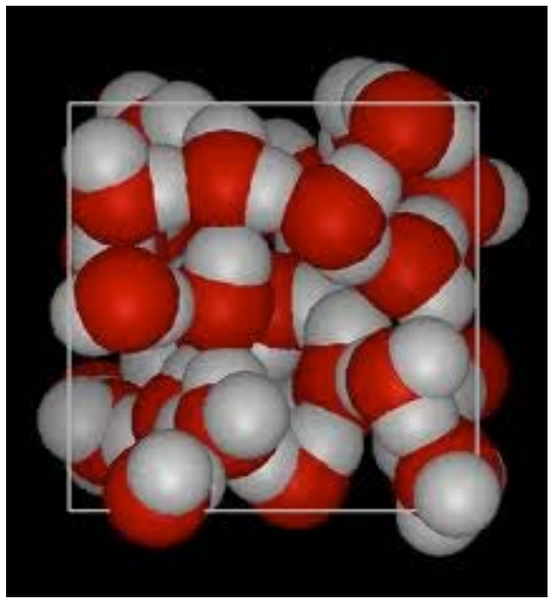
$$\mathcal{L}_{\mathrm{S}} = \left[F^{\mathrm{S}} - (p^2/\Lambda)F^{\mathrm{S}}\right]\frac{\partial}{\partial p} - \frac{p\xi_1 F^{\mathrm{S}}}{\Lambda}\frac{\partial}{\partial \xi_1},$$

isokinetic Nosé terms:

$$\mathcal{L}_{\mathrm{N}} = \frac{1}{2}\frac{\mu_1\xi_1^2\xi_2 p}{\Lambda}\frac{\partial}{\partial p} + \frac{1}{2}\frac{\mu_1\xi_1^3\xi_2}{\Lambda}\frac{\partial}{\partial \xi_1} - \xi_2\xi_1\frac{\partial}{\partial \xi_1} + \mu_2^{-1}[\mu_1\xi_1^2 - \beta^{-1}]\frac{\partial}{\partial \xi_2},$$

Ornstein-Uhlenbeck term:

$$\mathcal{L}_{\mathrm{O}} = -\gamma\xi_2\frac{\partial}{\partial \xi_2} + \frac{\sigma^2}{2}\frac{\partial^2}{\partial \xi_2^2}.$$



periodic box, 25A sides,
512 fully flexible
SPC water molecules (all atom)
periodic boundary conditions
three level splitting

$$U = U_{\mathrm{bond}} + U_{\mathrm{s.r.}} + U_{\mathrm{l.r.}}$$

Coulombic forces: Smooth Particle Mesh Ewald

Long ranged forces includes ‘reciprocal space’ part + screened coulomb interactions within the simulation cell

Short-ranged Regime: e.g. Lennard-Jones cutoffs to 6A, ‘real space’ part of Ewald.

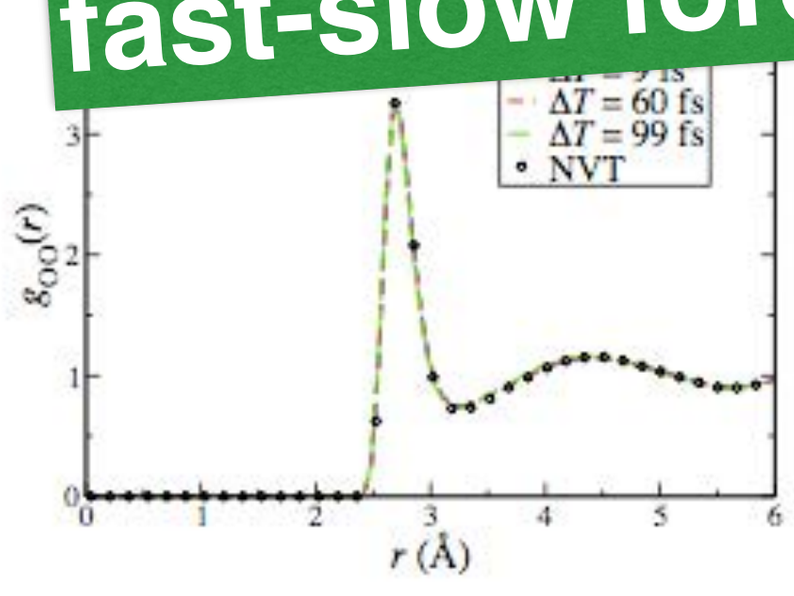
Bond part: Intramolecular interactions, e.g. angle and length bonds.

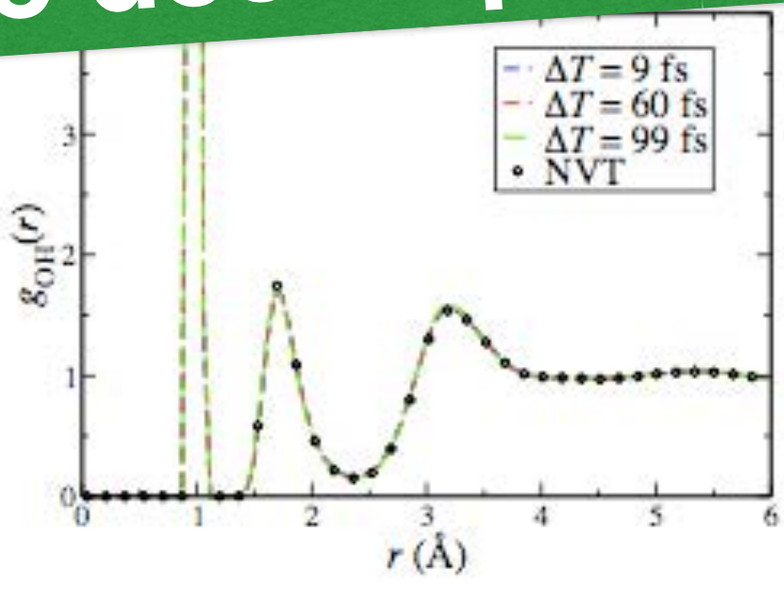
Flexible H2O simulations @ $\delta t = 99\mathrm{fs}$

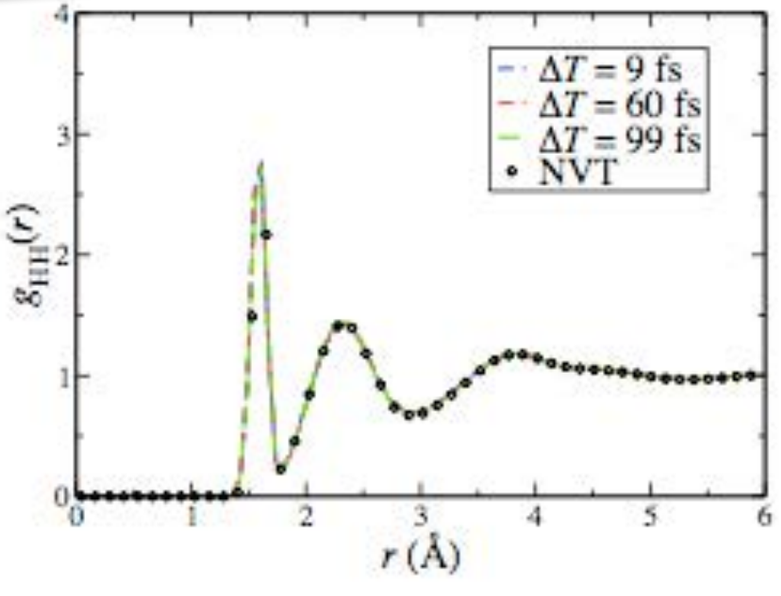
$$\delta t_{\mathrm{inner}} = 0.5\mathrm{fs}, \qquad \delta t_{\mathrm{mid}} = 3\mathrm{fs}$$

Take Home Message #6

Isokinetic constraints can stabilize stochastic multiple timestepping based on fast-slow force decomposition.







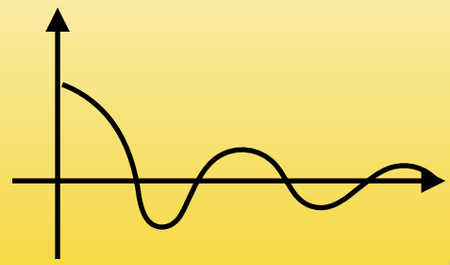
Ensemble Preconditioning

Goal: redesign the dynamics (and integrator) to enhance the rate of convergence for typical observables f

$$\langle f \rangle = \int f(x)\rho(x)\mathrm{d}x = \lim_{N\to\infty} N^{-1} \sum_{t=1}^{N} f(x_t)$$

figure of merit = **Integrated Autocorrelation Time (IAT)**

$$\tau_f = 1 + 2\sum_{t=1}^{\infty} \mathrm{cor}(f(x_t), f(x_0))$$

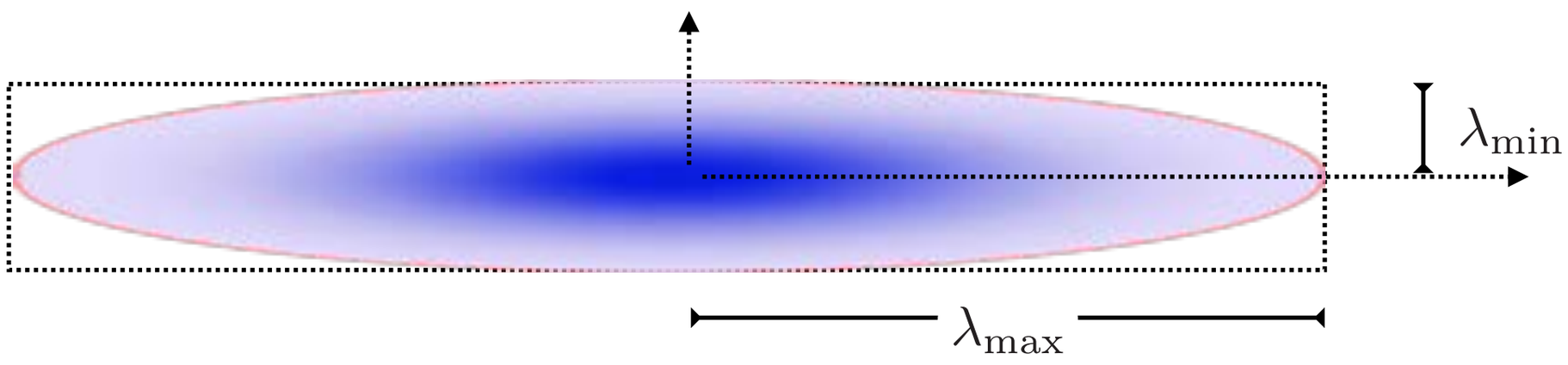


We would like to have τ_f as small as possible

Motivating Example:

$$\pi(x) = \exp\left(-\frac{1}{2}x^T\Sigma^{-1}x\right)$$

Eigenvalues: $0 < \lambda_{\min} =< \ldots < \lambda_{\max} = \rho(\Sigma)$



For MCMC schemes like Euler-Maruyama or Leimkuhler-Matthews, **stability requires** $h = O(\lambda_{\min})$

But for $f(x) = x \cdot \mathbf{e}_{\max}$ $\tau_f = O(\lambda_{\max}/\lambda_{\min})$

Poor Scaling $\Longrightarrow$ **Slow Convergence**

Ensemble Preconditioning

More generally, we wish to sample problems with complicated energy functions, where each basin or local approximation may be very poorly scaled.

Related Concepts

Stochastic Newton schemes

BFGS Method

MC Hammer (Goodman and Weare)

Compare to: Riemannian Manifold HMC (Girolami et al)

Ensemble Preconditioning

Use local information to estimate inverse Hessian matrix; precondition (rescale) dynamics to enhance convergence (reduce IAT)

Wishlist:

- Increase efficiency by reducing the IAT
- Compute the preconditioning based on a local ensemble approximation
- Allow for inertial effects (underdamped Langevin/HMC)

Idea: Use a collection of "walkers" to generate local covariance information and use this to estimate the inverse Hessian adaptively

Procedure

Use an ensemble of L walkers:

$$Q = (q_1, q_2, \ldots, q_L) \in \mathbb{R}^{dL}, \qquad P = (p_1, p_2, \ldots, p_L) \in \mathbb{R}^{dL},$$

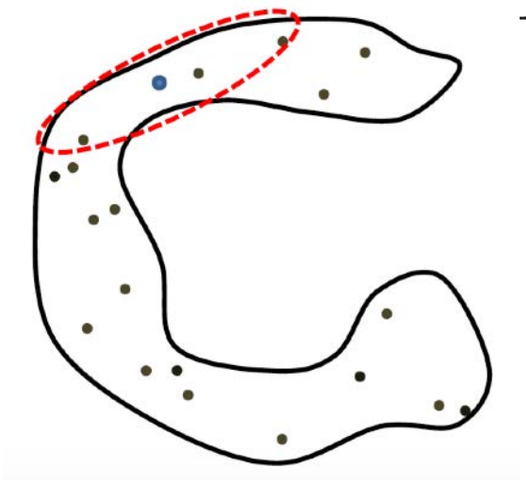
$$\bar{\pi}(Q, P) = \prod_{i=1}^{L} \hat{\pi}(q_i, p_i), \qquad \int \bar{\pi}(Q, P) \mathrm{P} = \prod_{i=1}^{L} \pi(q_i).$$

each walker samples the same target distribution $\pi(q)$

We construct an dynamics in the extended space and compute ensemble averages by marginalisation over the individual walkers.

$$\dot{Q} = B(Q)P,$$
$$\dot{P} = B(Q)^{\mathrm{T}}\nabla \log(\pi(Q)) + \mathrm{div}(B(Q)^{\mathrm{T}}) - \gamma P + \sqrt{2\gamma}\eta(t).$$

$$B(Q) = \mathrm{diag}(B_1(Q), B_2(Q), \ldots, B_L(Q))$$



$$B_i(Q) = \sqrt{I_{\mathrm{d}} + \eta\,\mathrm{wcov}(Q_{[i]}, \omega_{\lambda(Q_{[i]},q_i)})}$$

blend with identity (robustness)

collects covariance info of nearby walkers

$$Q_{[i]} = (q_1, q_2, \ldots, q_{i-1}, q_{i+1}, \ldots, q_L)$$

Basing B_i on other walkers only
eliminates the problems of multiplicative noise

Discretization of SDEs: similar to **BAOAB**

Discretization

BAOABish:

$$p^{(n+1/2)} = p^{(n)} + \frac{\delta t}{2} F(q^{(n)}),$$

$$q^{(n+1/2)} = q^{(n)} + \frac{\delta t}{2} B\left(q^{(n+1/2)}\right) p^{(n+1/2)}$$

$$\hat{p}^{(n+1/2)} = \alpha p^{(n+1/2)} + \frac{(\alpha+1)\delta t}{2} \mathrm{div}\left(B\left(q^{(n+1/2)}\right)^{\mathrm{T}}\right) + \sqrt{1-\alpha^2}\mathrm{R}^{(n)}$$

$$q^{(n+1)} = q^{(n+1/2)} + \frac{\delta t}{2} B\left(q^{(n+1/2)}\right) \hat{p}^{(n+1/2)}$$

$$p^{(n+1)} = \hat{p}^{(n+1/2)} + \frac{\delta t}{2} F(q^{(n+1)})$$

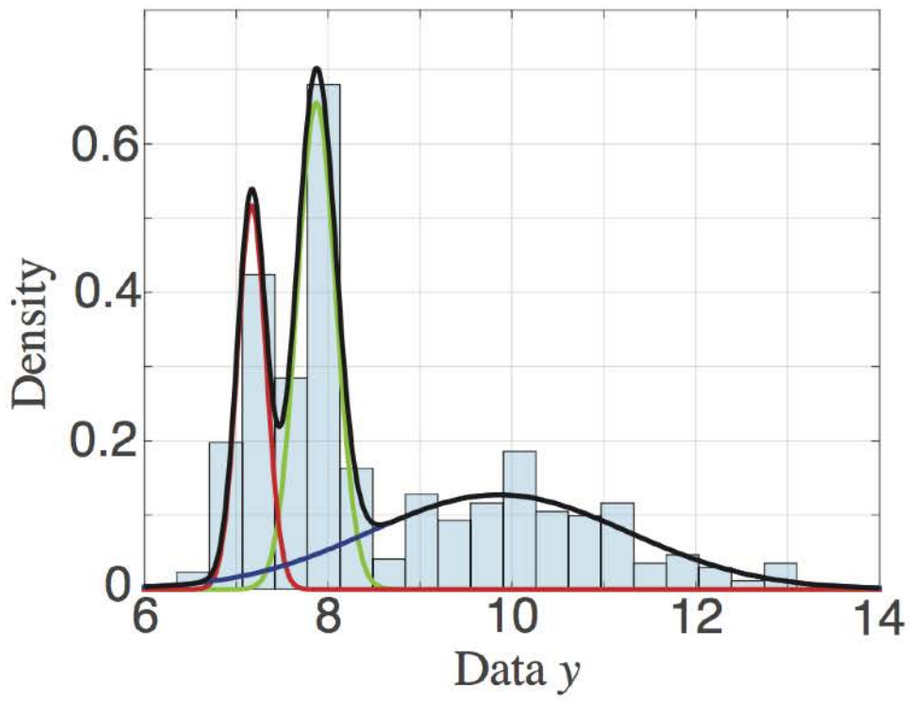
$$\alpha = e^{-\gamma h}$$

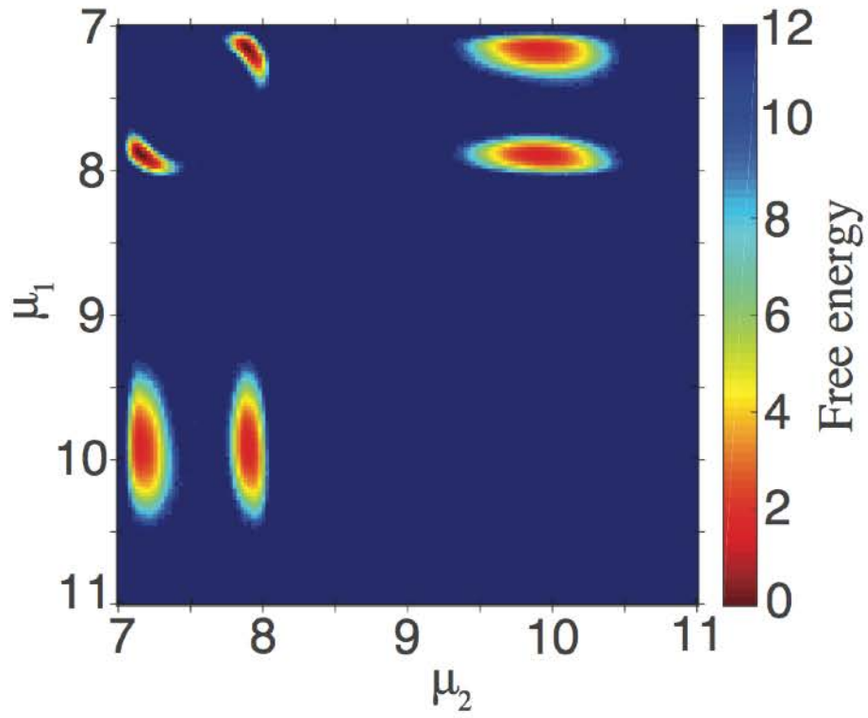
Gaussian Mixture Model: Hidalgo Stamps

"Adventures in Stamp Collecting"

$$\sum_{n=1}^{3} z_n \mathcal{N}(\mu_n, \lambda_n^{-2})$$

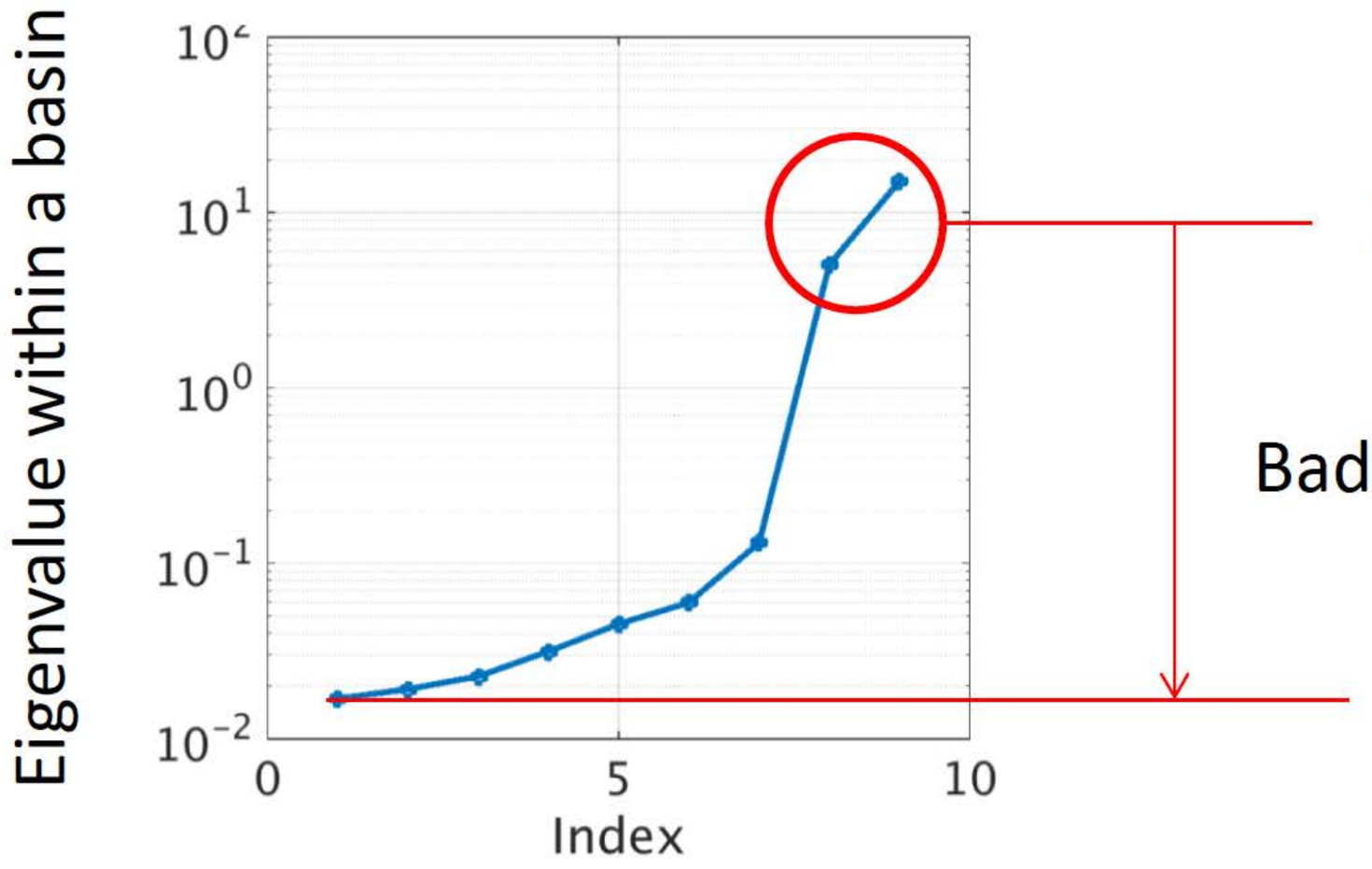
Dataset: Thickness of 485 stamps from Mexico in 1872.





- (moderately) poorly scaled basins
- multimodal due to "label switching" symmetry

Within one basin, we have bad scaling:



Slow modes of the system

Bad scaling :: inefficiency

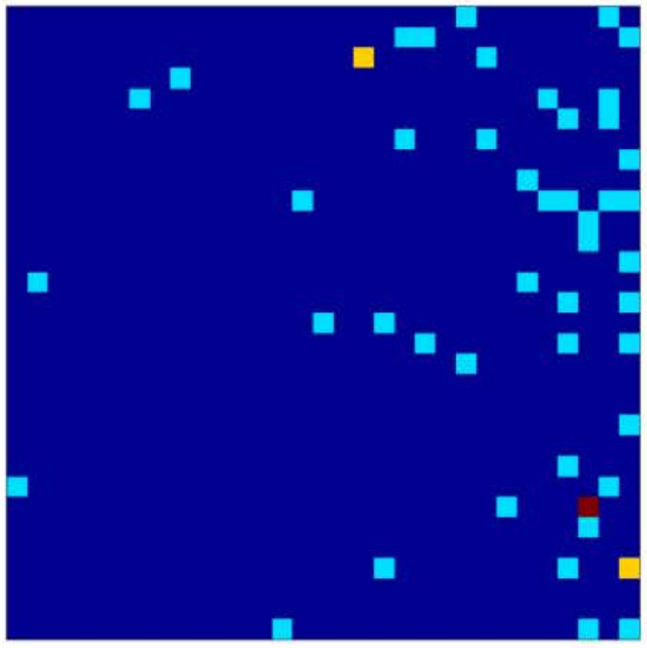
But the eigensystems are **different in different basins**, so the localized covariance is needed…

Gaussian Mixture Model: Hidalgo Stamps

Integrated Autocorrelation Times of Different Schemes

Scheme	$\min(z)$	$\max(\lambda)$	$\min(\mu)$	β
HMC	21495	**42935**	27452	7148
Langevin Dynamics	6825	**13279**	8384	4641
Ensemble Q-N	69	83	98	**115**

Numerical test: Log-Gaussian Cox model



Observations X

Break $[0,1]^2$ into a 32x32 grid.

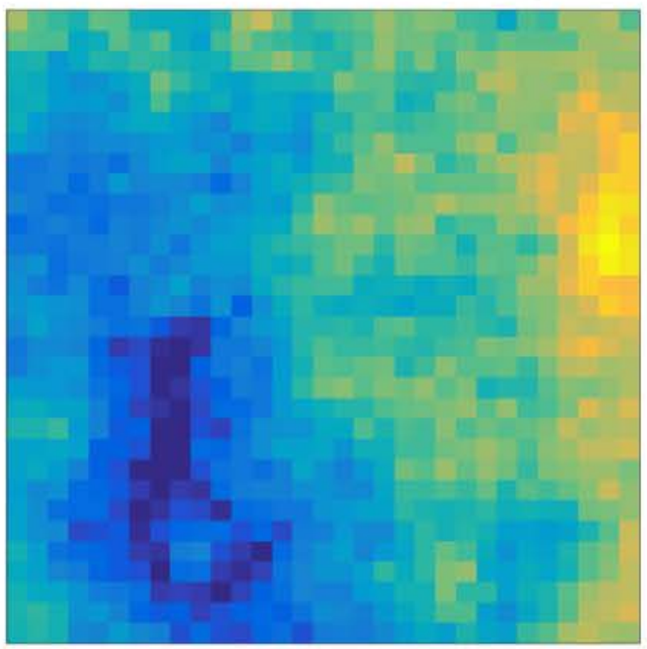
Observed intensity in box (i,j) is $X_{i,j}$,
Poisson distributed with mean

$$\Lambda(i,j)/32^2, \qquad \Lambda(i,j) = \exp(Y_{i,j})$$

where $Y \sim N(\mu, \Sigma)$, with

$$\Sigma_{(i,j),(i',j')} = \sigma^2 \exp[-\sqrt{(i-i')^2 + (j-j')^2}/(32\beta)]$$

Means $\exp(Y_{i,j})$



We generate synthetic data X using

$$\sigma^2 = 1.91, \quad \beta = 1/33, \quad \mu = \log(126) - \sigma^2/2$$

We fix μ and aim to infer likely Y, using hyperparameters σ^2, β, with prior $\mathrm{gamma}(2, \frac{1}{2})$

Log Gaussian Cox Model

Scheme	x	σ^2	β	Efficiency
HMC	800.7	1041.6	**1318.7**	1.0
RMHMC	2158.9	34.0	**1502.0**	0.15
LD				

Take Home Message #7

Using a collection of walker particles can improve conditioning of poorly scaled basins and increase sampling efficiency.

Ensemble Quasi-Newton python package

http://bitbucket.org/c_matthews/ensembleqn

7 Messages of this Lecture

1. The right order of Langevin building blocks provides substantial improvements in accuracy (low sampling bias).

2. Adaptive Langevin dynamics offers a simple method for removing bias in noisy gradient simulations while maintaining high accuracy (low bias).

3. Constraints don't buy much stability in deterministic MD, but they do improve accuracy!

4. Combining constraints and multiple timestepping can improve accuracy and stability

5. Constrained Langevin dynamics, using a multiple timestepping implementation of geodesic flow, can yield big MD sampling improvements with high accuracy.

6. Isokinetic constraints can stabilize stochastic multiple timestepping based on fast-slow force decomposition.

7. Using a collection of walker particles can improve conditioning of poorly scaled basins and increase sampling efficiency.