

# Molecular Dynamics

## Day 4

Ben Leimkuhler

**convergence of averages  
Langevin for biomolecules  
constrained Langevin  
multiple timestepping and “SIN”**

# Fokker-Planck Equation

The stochastic paths of an Itō SDE sample the evolving distribution with probability density satisfying

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

$\mathcal{L}^\dagger$  is the  $L^2$  adjoint of  $\mathcal{L}$  defined by

$$(\mathcal{L}g)(x) = \lim_{\delta t \rightarrow 0} \frac{\mathbb{E}[g(x(\delta t)) | x(0) = x] - g(x)}{\delta t}$$

# SDEs and ergodicity

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

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

**generator**

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

**ergodic  
wrt to  $\pi$**

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

**transfer  
operator**

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

**hypoelliptic**

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

**smooth transition  
density**

# How to prove geometric ergodicity

(1) Check **parabolic Hörmander condition**

(2) Show a **controllability property**

(3) Find a **Lyapunov function**  $\mathcal{L}\mathcal{K} \leq -a\mathcal{K} + b$

In many cases, particularly when we work on compact position spaces, the key step is (1), above.

# Hypoelliptic Property

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

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

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

*A mixing condition for SDEs.*

**Langevin dynamics:**

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

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

# 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” ...

# Splitting Methods

$$\begin{aligned} & \text{[[ABAO]]} \\ \mathbf{q}_{n+1/2} &= \mathbf{q}_n + \frac{(h/2)\mathbf{M}^{-1}}{(h/2)\mathbf{M}^{-1}}\mathbf{p}_n, \\ \mathbf{p}_{n+1/2} &= \mathbf{p}_n - h\nabla U(\mathbf{q}_{n+1/2}), \\ \mathbf{q}_{n+1} &= \mathbf{q}_{n+1/2} + (h/2)\mathbf{M}^{-1}\mathbf{p}_{n+1/2}, \\ \mathbf{p}_{n+1} &= e^{-h\gamma}\mathbf{p}_{n+1/2} + \zeta_2\mathbf{M}^{1/2}\mathbf{R}_n, \end{aligned}$$

$$\begin{aligned} & \text{[[ABOBA]]} \\ \mathbf{q}_{n+1/2} &= \mathbf{q}_n + \frac{(h/2)\mathbf{M}^{-1}}{(h/2)\mathbf{M}^{-1}}\mathbf{p}_n, \\ \mathbf{p}_{n+1/2} &= \mathbf{p}_n - (h/2)\nabla U(\mathbf{q}_{n+1/2}), \\ \hat{\mathbf{p}}_{n+1/2} &= e^{-h\gamma}\mathbf{p}_{n+1/2} + \zeta_2\mathbf{M}^{1/2}\mathbf{R}_n, \\ \mathbf{p}_{n+1} &= \hat{\mathbf{p}}_{n+1/2} - (h/2)\nabla U(\mathbf{q}_{n+1/2}), \\ \mathbf{q}_{n+1} &= \mathbf{q}_{n+1/2} + (h/2)\mathbf{M}^{-1}\mathbf{p}_{n+1}, \end{aligned}$$

$$\begin{aligned} & \text{[[OABAO]]} \\ \mathbf{p}_{n+1/2} &= e^{-h\gamma/2}\mathbf{p}_n + \zeta_1\mathbf{M}^{1/2}\mathbf{R}_n, \\ \mathbf{q}_{n+1/2} &= \mathbf{q}_n + (h/2)\mathbf{M}^{-1}\mathbf{p}_{n+1/2}, \\ \hat{\mathbf{p}}_{n+1/2} &= \mathbf{p}_{n+1/2} - h\nabla U(\mathbf{q}_{n+1/2}), \\ \mathbf{q}_{n+1} &= \mathbf{q}_{n+1/2} + (h/2)\mathbf{M}^{-1}\hat{\mathbf{p}}_{n+1/2}, \\ \mathbf{p}_{n+1} &= e^{-h\gamma/2}\hat{\mathbf{p}}_{n+1/2} + \zeta_1\mathbf{M}^{1/2}\mathbf{R}_{n+1/2}, \end{aligned}$$

$$\begin{aligned} & \text{[[BABO]]} \\ \mathbf{p}_{n+1/2} &= \mathbf{p}_n - \frac{(h/2)\nabla U(\mathbf{q}_n)}{(h/2)\nabla U(\mathbf{q}_n)}, \\ \mathbf{q}_{n+1} &= \mathbf{q}_n + h\mathbf{M}^{-1}\mathbf{p}_{n+1/2}, \\ \hat{\mathbf{p}}_{n+1/2} &= \mathbf{p}_{n+1/2} - (h/2)\nabla U(\mathbf{q}_{n+1}), \\ \mathbf{p}_{n+1} &= e^{-h\gamma}\hat{\mathbf{p}}_{n+1/2} + \zeta_2\mathbf{M}^{1/2}\mathbf{R}_n, \end{aligned}$$

$$\begin{aligned} & \text{[[BAOAB]]} \\ \mathbf{p}_{n+1/2} &= \mathbf{p}_n - \frac{(h/2)\nabla U(\mathbf{q}_n)}{(h/2)\nabla U(\mathbf{q}_n)}, \\ \mathbf{q}_{n+1/2} &= \mathbf{q}_n + (h/2)\mathbf{M}^{-1}\mathbf{p}_{n+1/2}, \\ \hat{\mathbf{p}}_{n+1/2} &= e^{-h\gamma}\mathbf{p}_{n+1/2} + \zeta_2\mathbf{M}^{1/2}\mathbf{R}_n, \\ \mathbf{q}_{n+1} &= \mathbf{q}_{n+1/2} + (h/2)\mathbf{M}^{-1}\hat{\mathbf{p}}_{n+1/2}, \\ \mathbf{p}_{n+1} &= \hat{\mathbf{p}}_{n+1/2} - (h/2)\nabla U(\mathbf{q}_{n+1}), \end{aligned}$$

$$\begin{aligned} & \text{[[OBABO]]} \\ \mathbf{p}_{n+1/2} &= e^{-h\gamma/2}\mathbf{p}_n + \zeta_1\mathbf{M}^{1/2}\mathbf{R}_n, \\ \hat{\mathbf{p}}_{n+1/2} &= \mathbf{p}_{n+1/2} - (h/2)\nabla U(\mathbf{q}_n), \\ \mathbf{q}_{n+1} &= \mathbf{q}_n + h\mathbf{M}^{-1}\hat{\mathbf{p}}_{n+1/2}, \\ \hat{\mathbf{p}}_{n+1} &= \hat{\mathbf{p}}_{n+1/2} - (h/2)\nabla U(\mathbf{q}_{n+1}), \\ \mathbf{p}_{n+1} &= e^{-h\gamma/2}\hat{\mathbf{p}}_{n+1} + \zeta_1\mathbf{M}^{1/2}\mathbf{R}_{n+1/2}, \end{aligned}$$

$$\zeta_j = \left[ k_B T \left( 1 - e^{-j\gamma h} \right) \right]^{1/2}$$

$$\mathbf{R}_n, \mathbf{R}_{n+1/2} \sim \mathcal{N}(0, 1)$$

# Other Popular Methods

## Stochastic Position Verlet (SPV)

$$\mathbf{q}_{n+1/2} = \mathbf{q}_n + (h/2)\mathbf{M}^{-1}\mathbf{p}_n \quad [\text{A(O+B)A}]$$

$$\mathbf{p}_{n+1} = e^{-h\gamma}\mathbf{p}_n - \eta \nabla U(\mathbf{q}_n) + \zeta \mathbf{M}^{1/2} \mathbf{R}_n$$

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

$$\eta = (1 - e^{-h\gamma})/\gamma, \quad \zeta = [k_B T (1 - e^{-2\gamma h})]^{1/2}$$

## Brünger-Brooks-Karplus (BBK)

$$\mathbf{p}_{n+1/2} = (1 - h\gamma/2)\mathbf{p}_n - (h/2)\nabla U(\mathbf{q}_n) + \frac{1}{2}\sqrt{2k_B T h \gamma} \mathbf{M}^{1/2} \mathbf{R}_n$$

$$\mathbf{q}_{n+1} = \mathbf{q}_n + h\mathbf{M}^{-1}\mathbf{p}_{n+1/2}$$

$$\mathbf{p}_{n+1} = [\mathbf{p}_{n+1/2} - (h/2)\nabla U(\mathbf{q}_{n+1}) + \frac{1}{2}\sqrt{2k_B T h \gamma} \mathbf{M}^{1/2} \mathbf{R}'_n]/(1 + h\gamma/2)$$

# Convergence of Averages

Assume we have an ergodic process, so

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

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

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

We would normally like to have  $\tau_f$  as small as possible

# Example: Monte-Carlo

A Monte-Carlo simulation for calculating  $\pi$

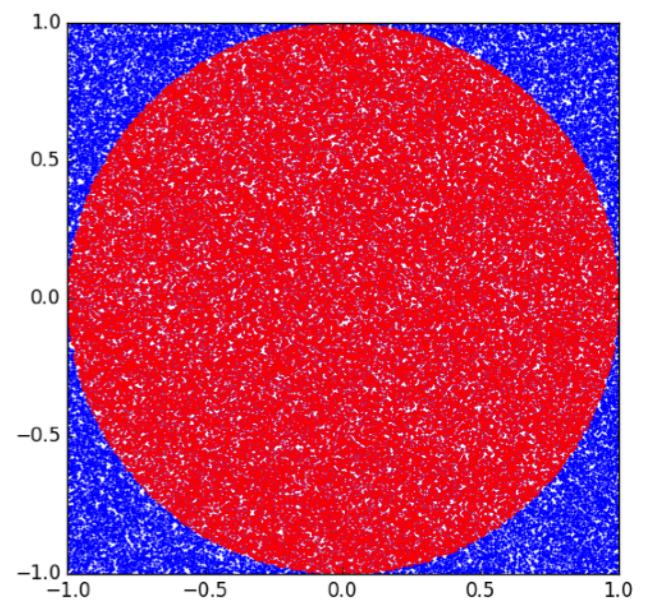
```
import numpy as np
from numpy import random as rr

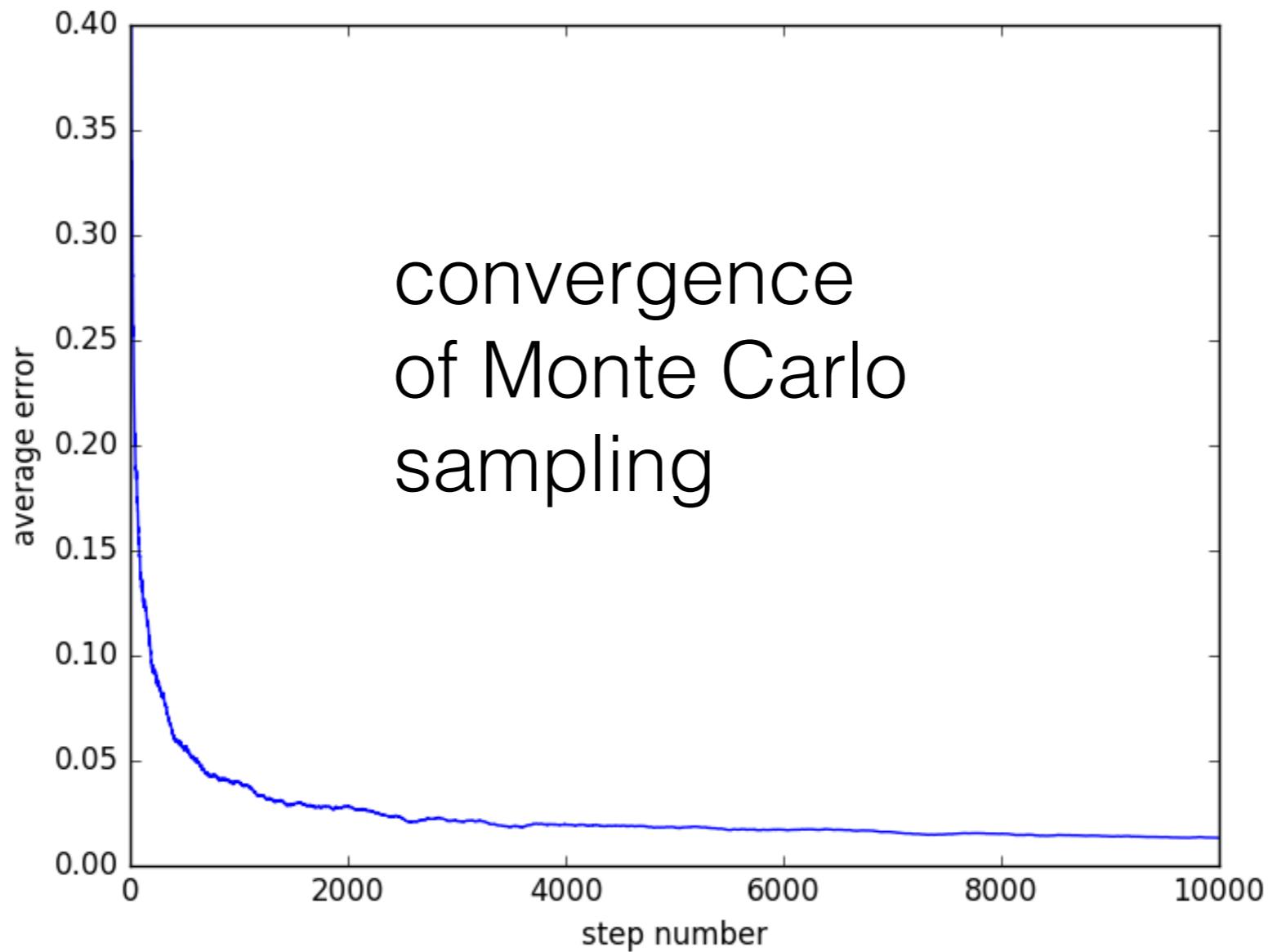
def calculate_pi(N):
    Inside = np.int64(0)
    for step in range(0,N):
        x,y = rr.uniform(-1,1), rr.uniform(-1,1) # draw
        if( x*x+y*y<1):                      # acceptance test
            Inside = Inside + 1
    return (4*np.float64(Inside)/np.float64(N))

print(calculate_pi(100000))
print(calculate_pi(10000000))
```

```
3.14208
3.1416164
```

very easy to program!  
converges rigorously  
slow convergence  $\sim 1/\sqrt{N}$





# Adjusting the Friction in BAOAB LD

$$dq = pdt$$

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

$$\gamma = 0$$

$$\gamma = \gamma^* = ??$$

$$\gamma = \infty$$

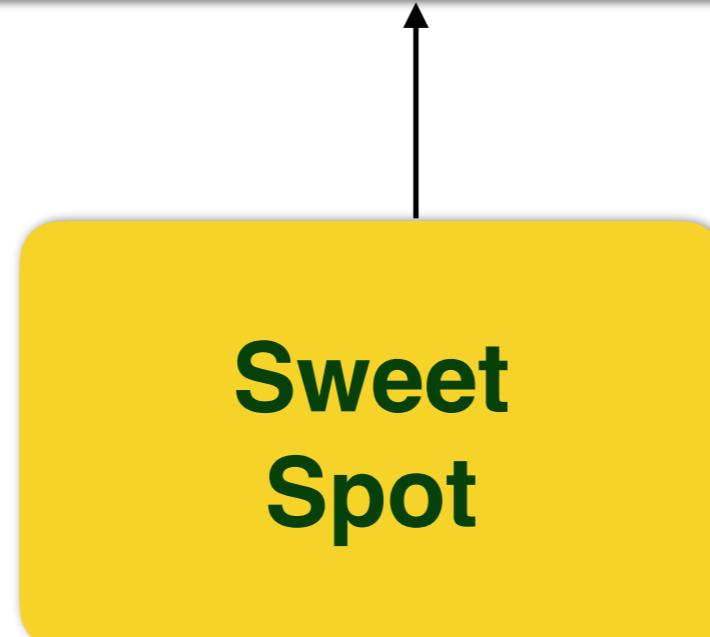
Hamiltonian  
Dynamics

**Poor Sampling**  
(not ergodic!)

Overdamped  
(Brownian Dynamics)

**Slow Convergence**

**Sweet  
Spot**



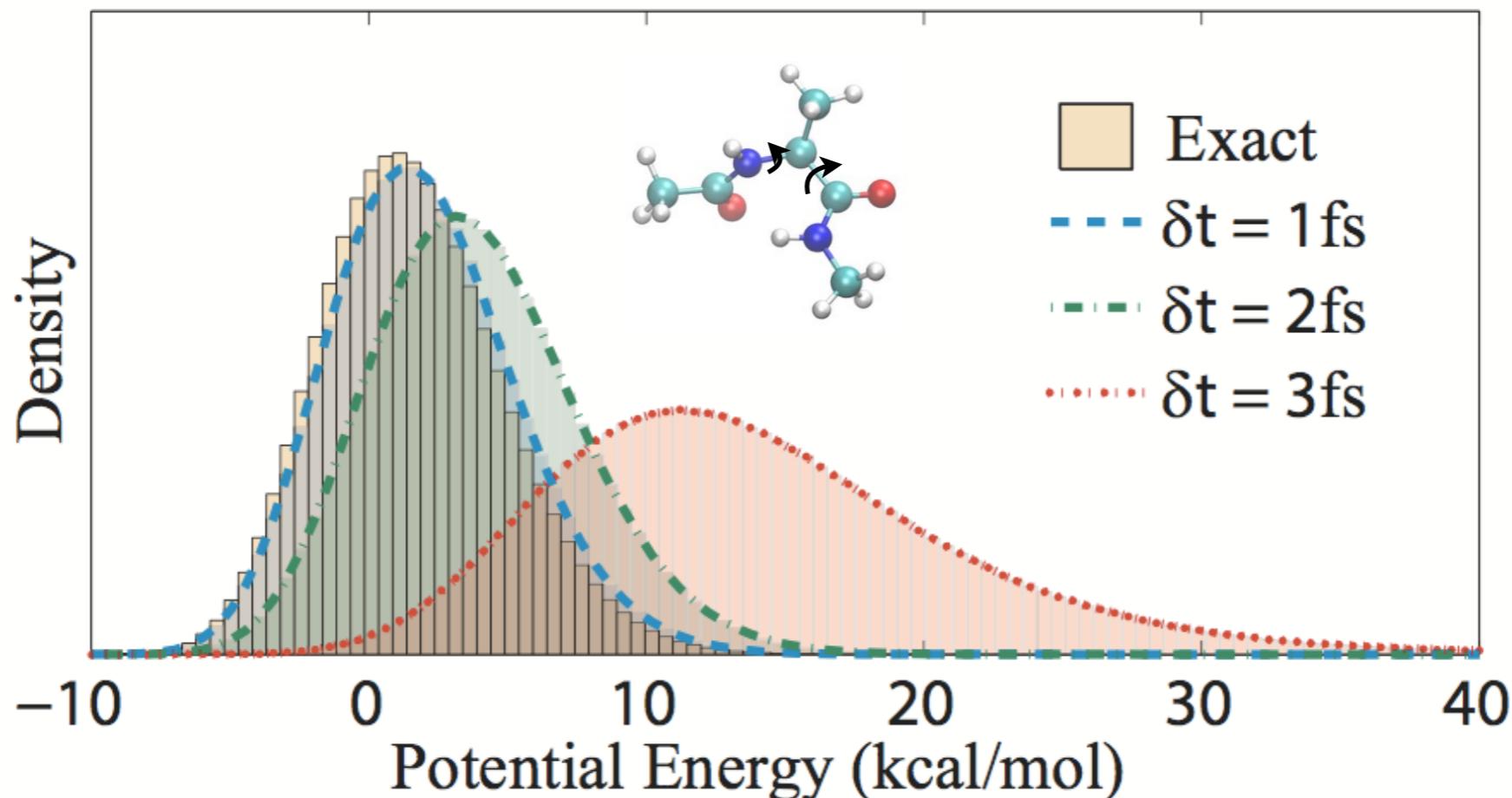
**Uncontrolled Bias**

**Bias  $\approx 0$**   
(Superconvergent!)

# **Application of Langevin Dynamics to biomolecules**

# Numerical Methods for Langevin Dynamics

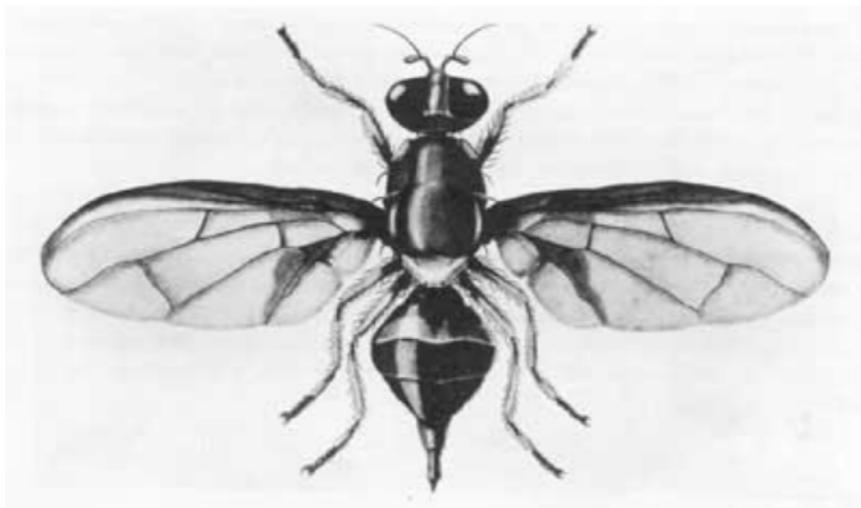
Typical Langevin discretization - **BBK**



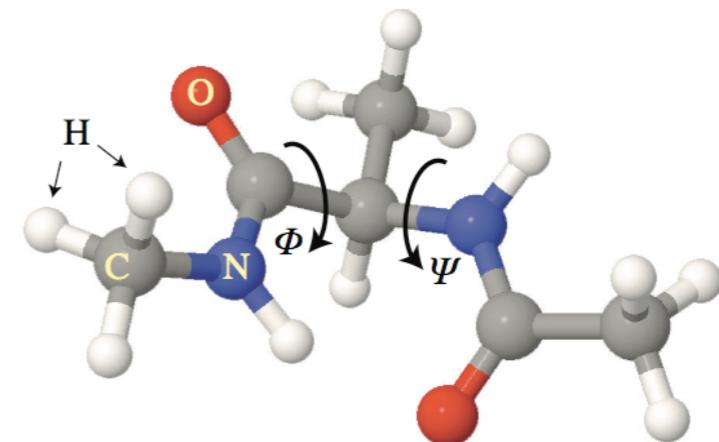
computed potential energy distribution for alanine dipeptide  
using BBK for Langevin dynamics...**but can we do better?**

Focus here on **constant** friction/noise case...

fruit fly



alanine dipeptide  
(the fruit fly of biomolecules)

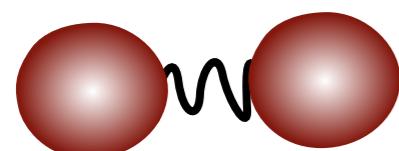
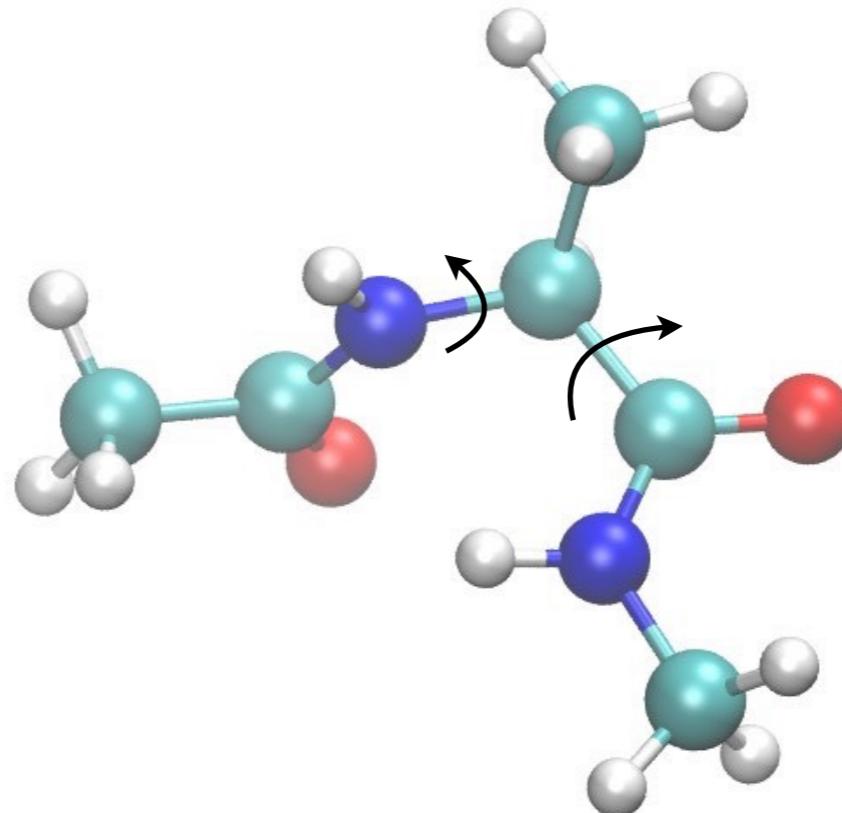


# Oscillatory modes of bio-MD

Vibrational mode	Wavelength of absorption [cm <sup>-1</sup> ]	Period [fs]
O–H stretch	3200–3600	9.8
N–H stretch		
C–H stretch	3000	11.1
O–C–O asymmetric stretch	2400	13.9
C≡C, C≡N stretch	2100	15.9
C=O (carbonyl) stretch	1700	19.6
C=C stretch		
H–O–H bend	1600	20.8
C–N–H bend	1500	22.2
H–N–H bend		
C=C (aromatic) stretch		
C–N stretch (amines)	1250	26.2
Water Libration (rocking)	800	41.7
O–C–O bending	700	47.6
C=C–H bending (alkenes)		
C=C–H bending (aromatic)		

# Biomolecular Sampling

Alanine Dipeptide

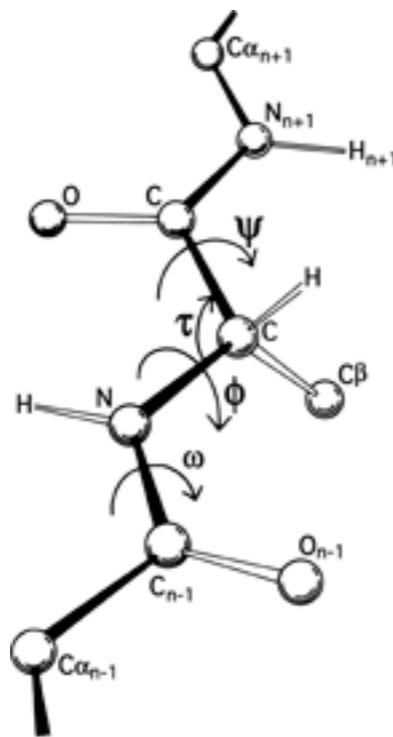


Stability restriction on timestep due to bonds

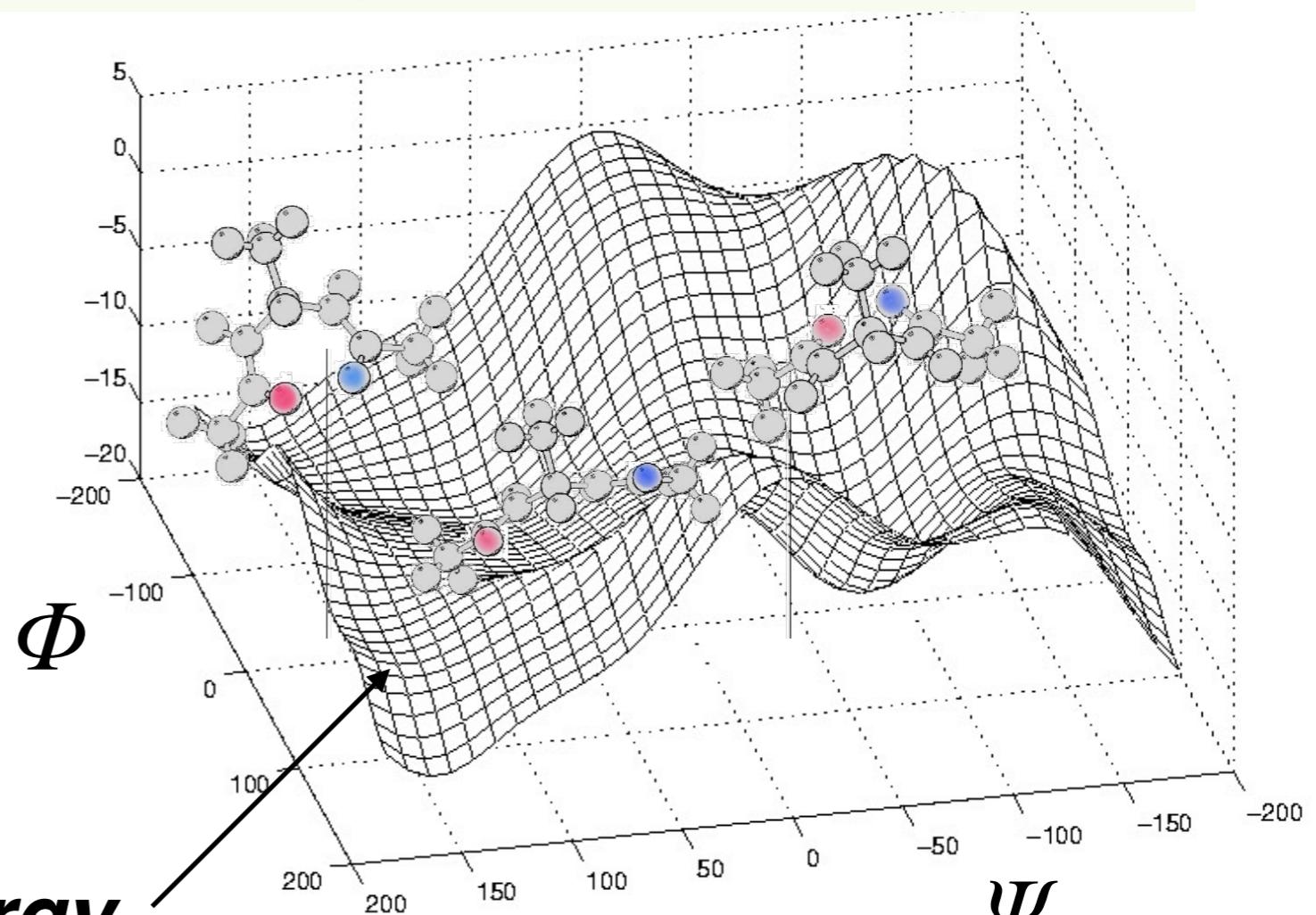
$$\delta t \lesssim \frac{2}{\omega_F} = \text{about } 3\text{fs} = 3 \times 10^{-15}\text{s}$$

# Ramachandran Plot

Reducing the system to just two degrees of freedom - the backbone dihedral angles, yields a greatly simplified **structural** “conformational” landscape.



*Free energy  
landscape*



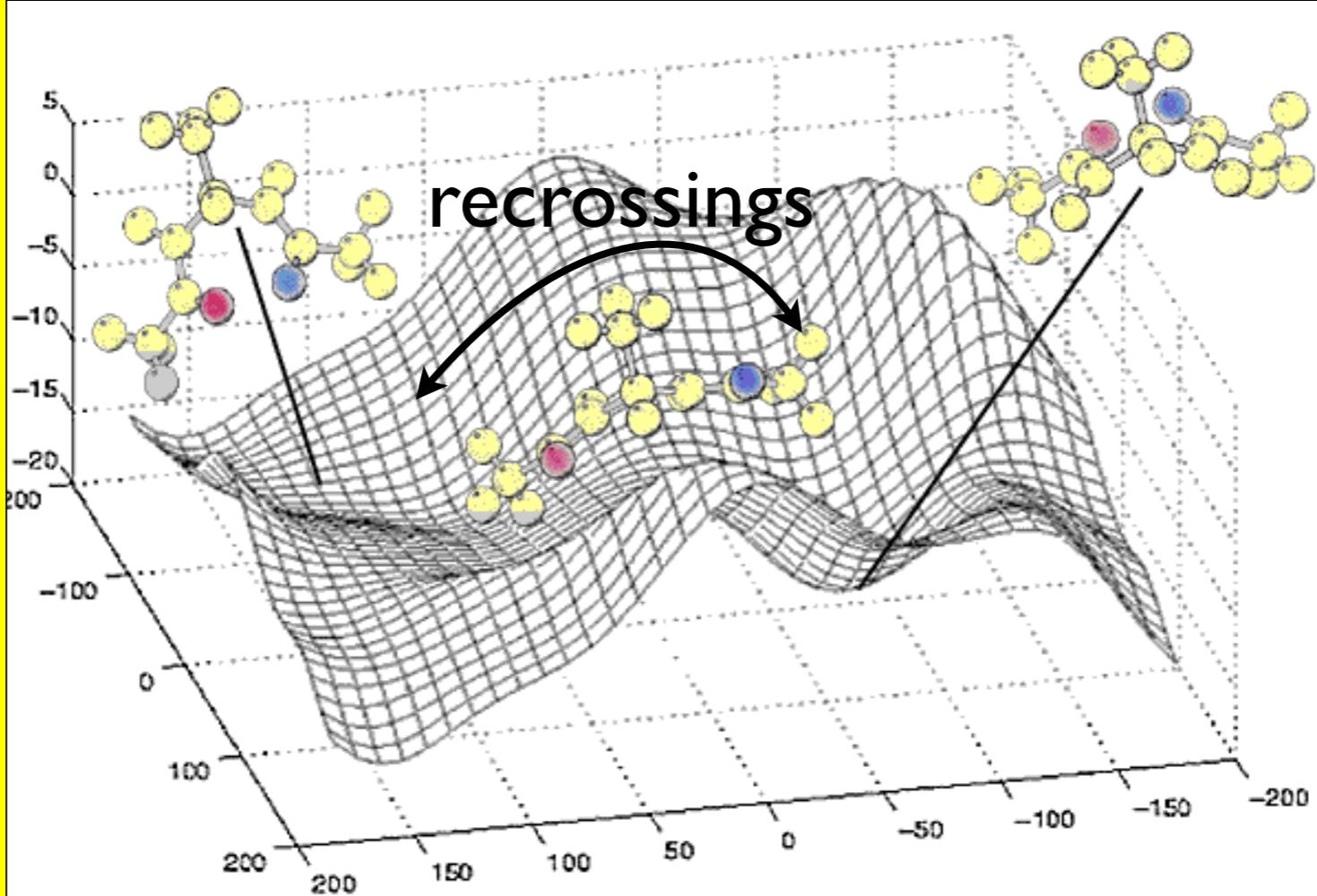
# Testing Methods for Biological Molecules

- Biomolecular models include many force terms  
    ⇒ **complicated landscape**
- Stiff bonds NH, OH.. ⇒ **dominant harmonic terms**
- Presence of H<sub>2</sub>O ⇒ **alternate diffusion mechanism**
- basins + barriers ⇒ long sampling times/**rare events**
- Accuracy requirements not always transparent  
goals of simulation may include dynamics or sampling

Methods need to balance  
**diffusion rate vs accuracy**  
of basin approximation

# Recrossings (rare events, like in the double well)

Alanine Dipeptide Free Energy

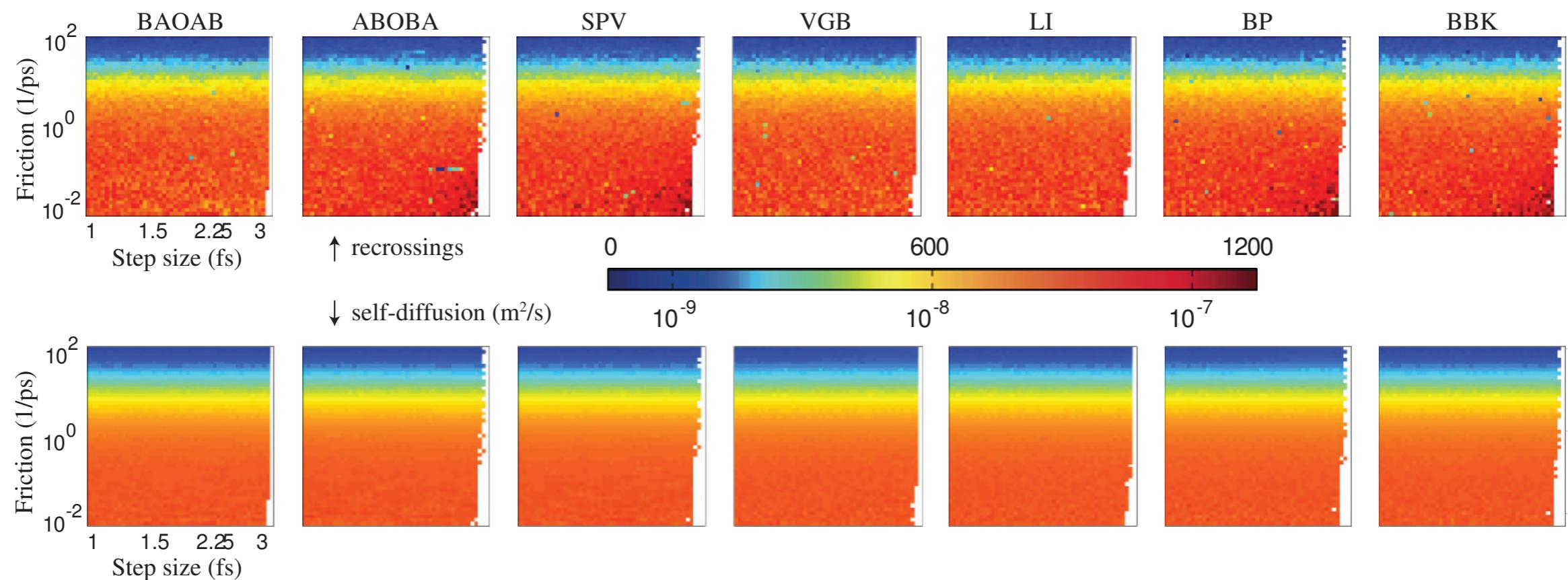


# Diffusion vs Accuracy

With BAOAB, we would imagine it to be better to use large friction coefficient  $\gamma$   
⇒ 4th order

(BAOAB Limit Method, i.e. Leimkuhler-Matthews)

**But...  $\gamma$  strongly affects the diffusion rate**



# Measuring Accuracy

Canonical density for solvated alanine dipeptide is a function in a 3200 dimensional space. Convergence and order of accuracy depend on how we collapse the space.

Examples:

$$\langle p_i^2 \rangle_\beta = \beta^{-1}$$

kinetic temperature

$$\langle U(q) \rangle_\beta$$

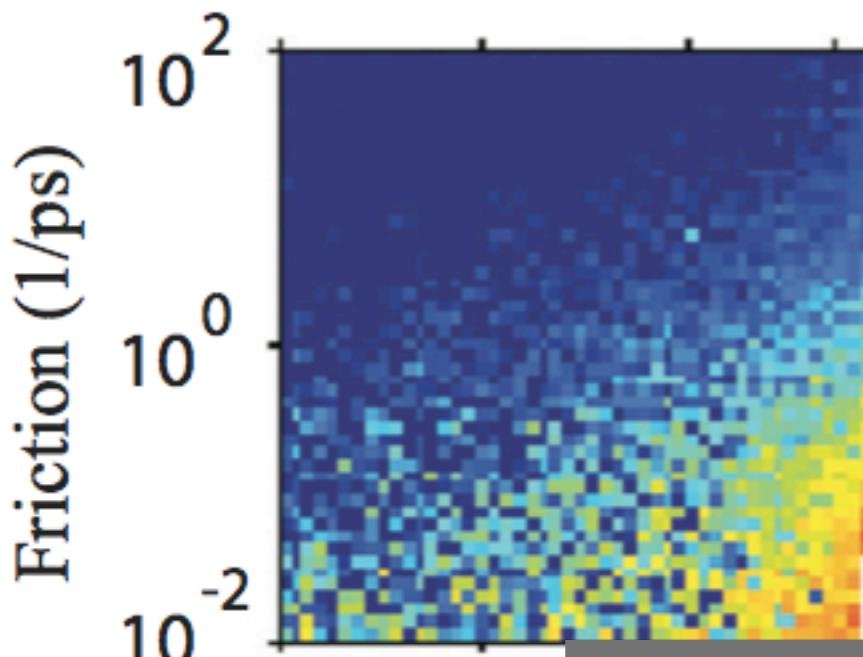
average potential energy

$$\left\langle \sum q_i \cdot \nabla U(q_i) \right\rangle_\beta = \beta^{-1}$$

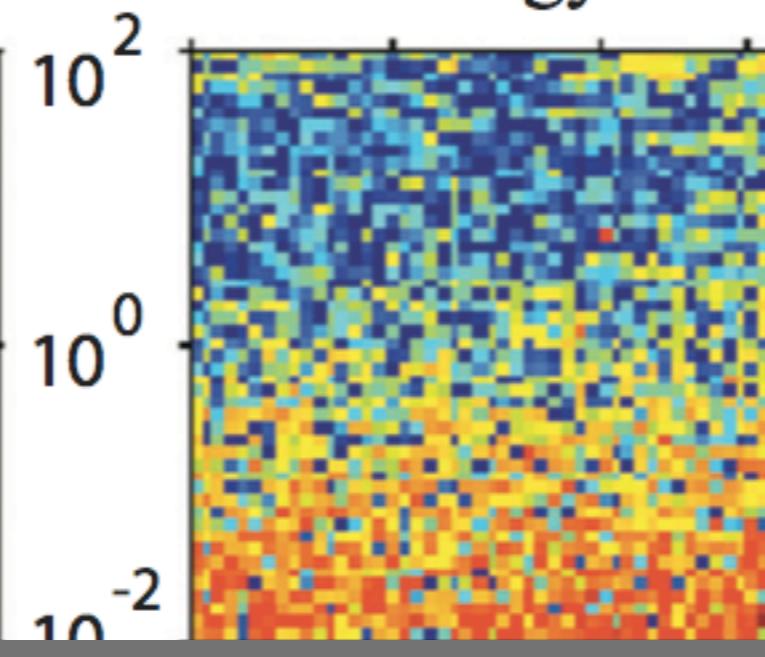
a configurational temp

**BAOAB**

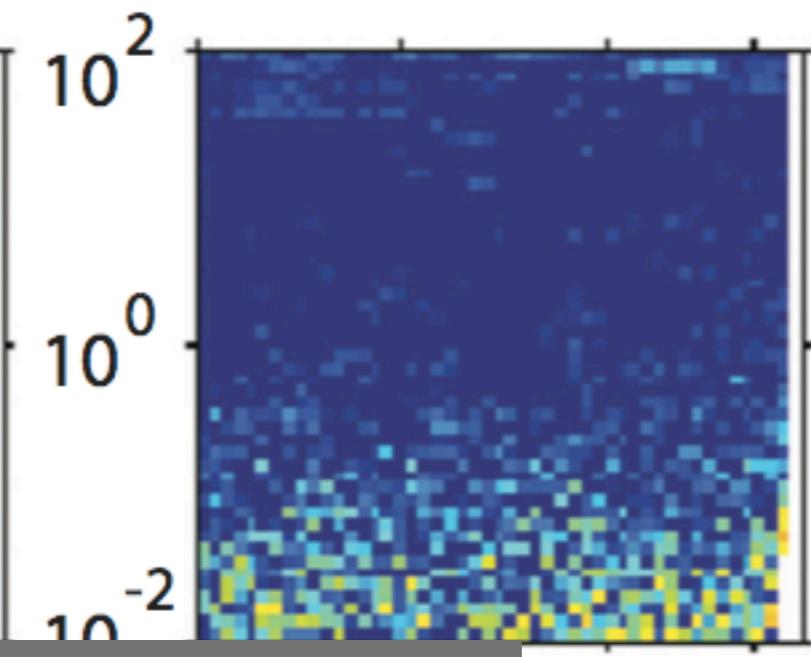
Relative error in  
average  
bond energy



Relative error in  
average  
total potential  
energy

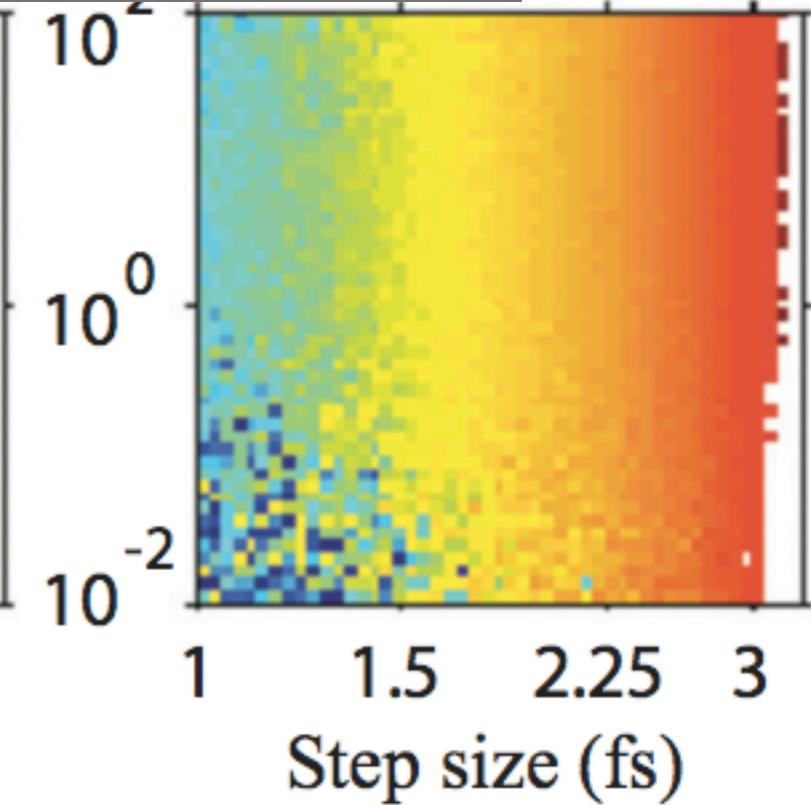
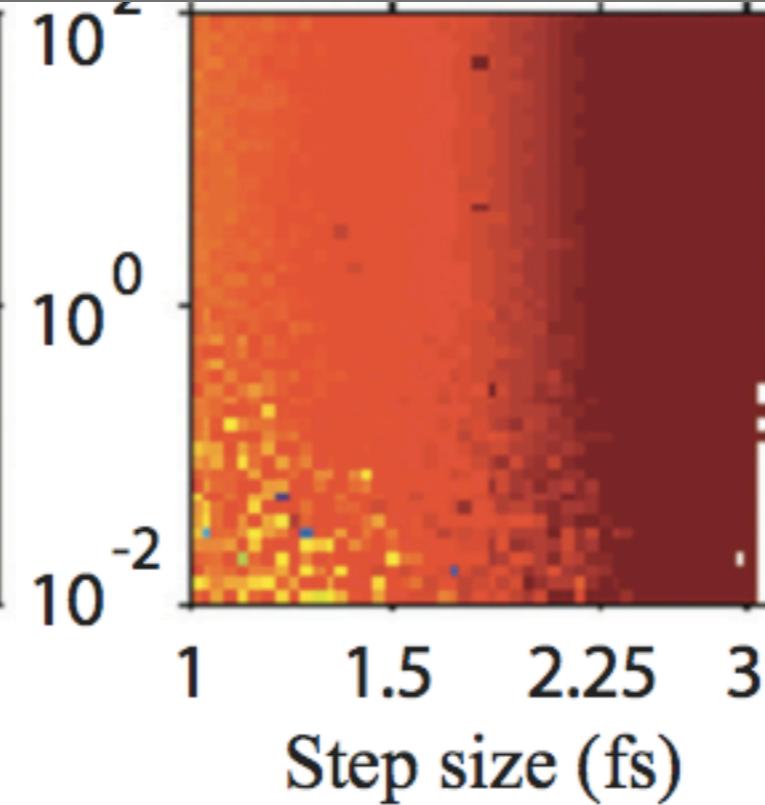
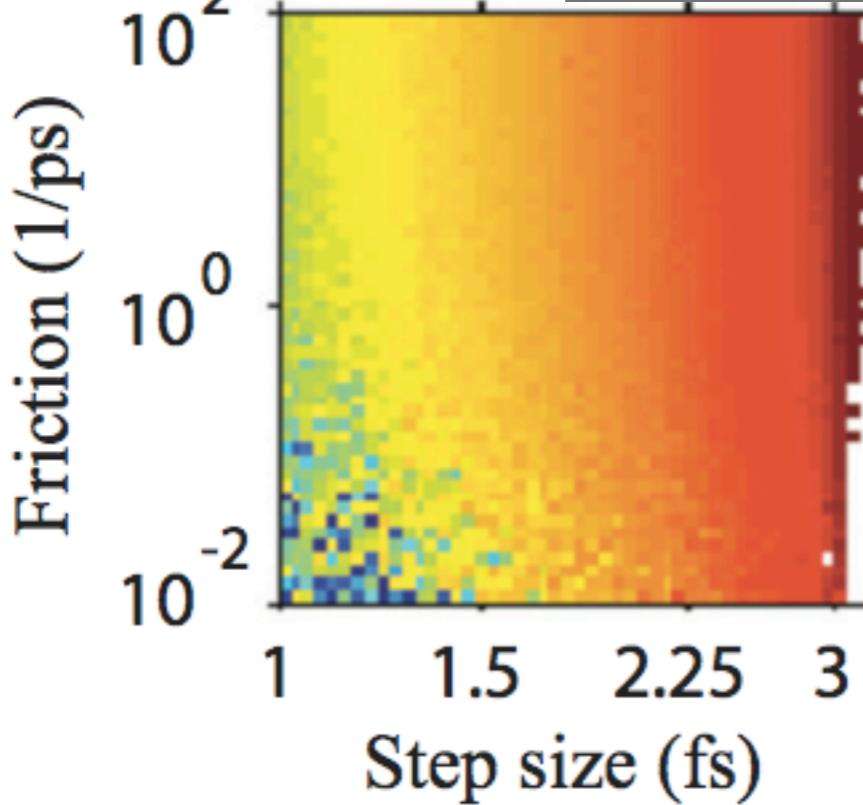


Relative error in  
configurational  
temperature



Red= 100% error; blue = 1% error

**B̄BK**



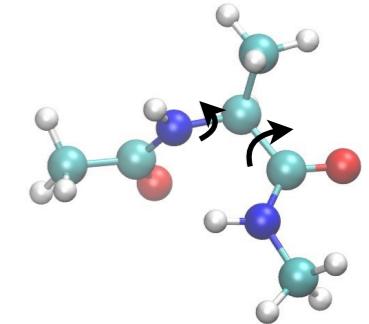
Step size (fs)

Step size (fs)

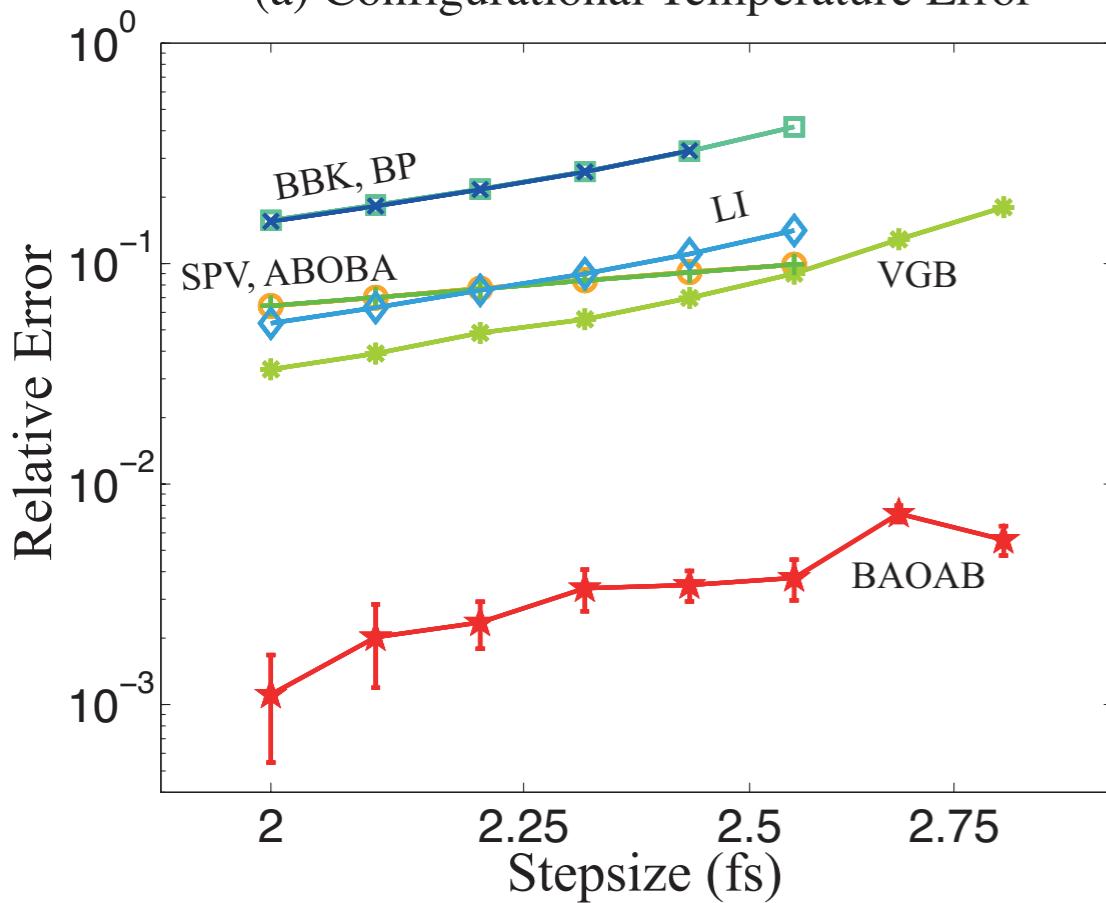
Step size (fs)

# Solvated Systems

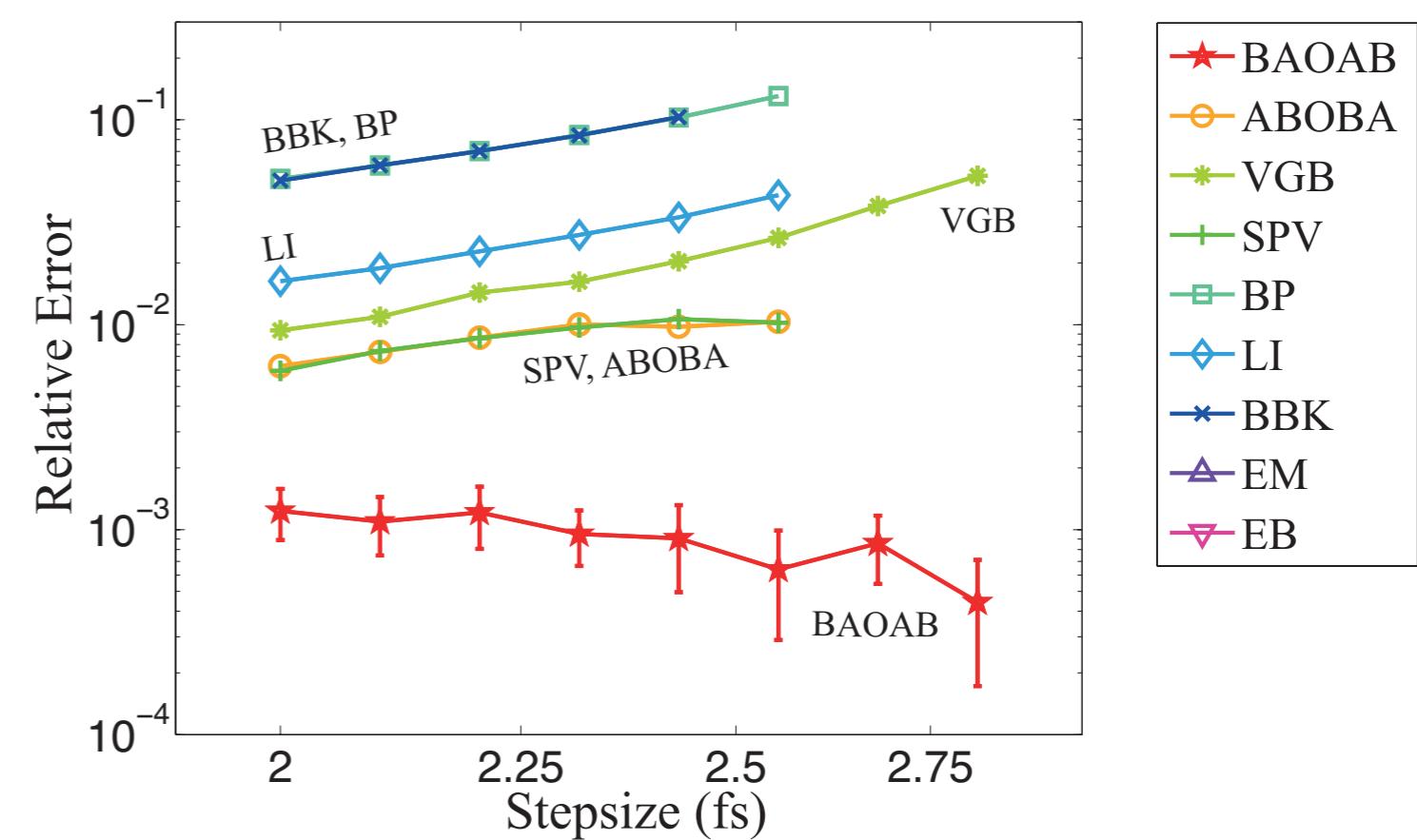
for alanine dipeptide in flexible water



(a) Configurational Temperature Error



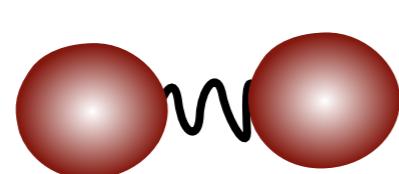
(b) Average Total Potential Energy Error



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

# Why are Bond Energies so accurate?

In solvated simulations, BAOAB shows much more accurate bond energies than other methods--even for modest stepsize



$$H = \frac{p^2}{2} + K \frac{q^2}{2}$$

For Harmonic oscillator model problem

$$\begin{bmatrix} q_{n+1} \\ p_{n+1} \end{bmatrix} \leftarrow \Psi \begin{bmatrix} q_n \\ p_n \end{bmatrix} + \mu_n,$$

solve linear system for canonical averages

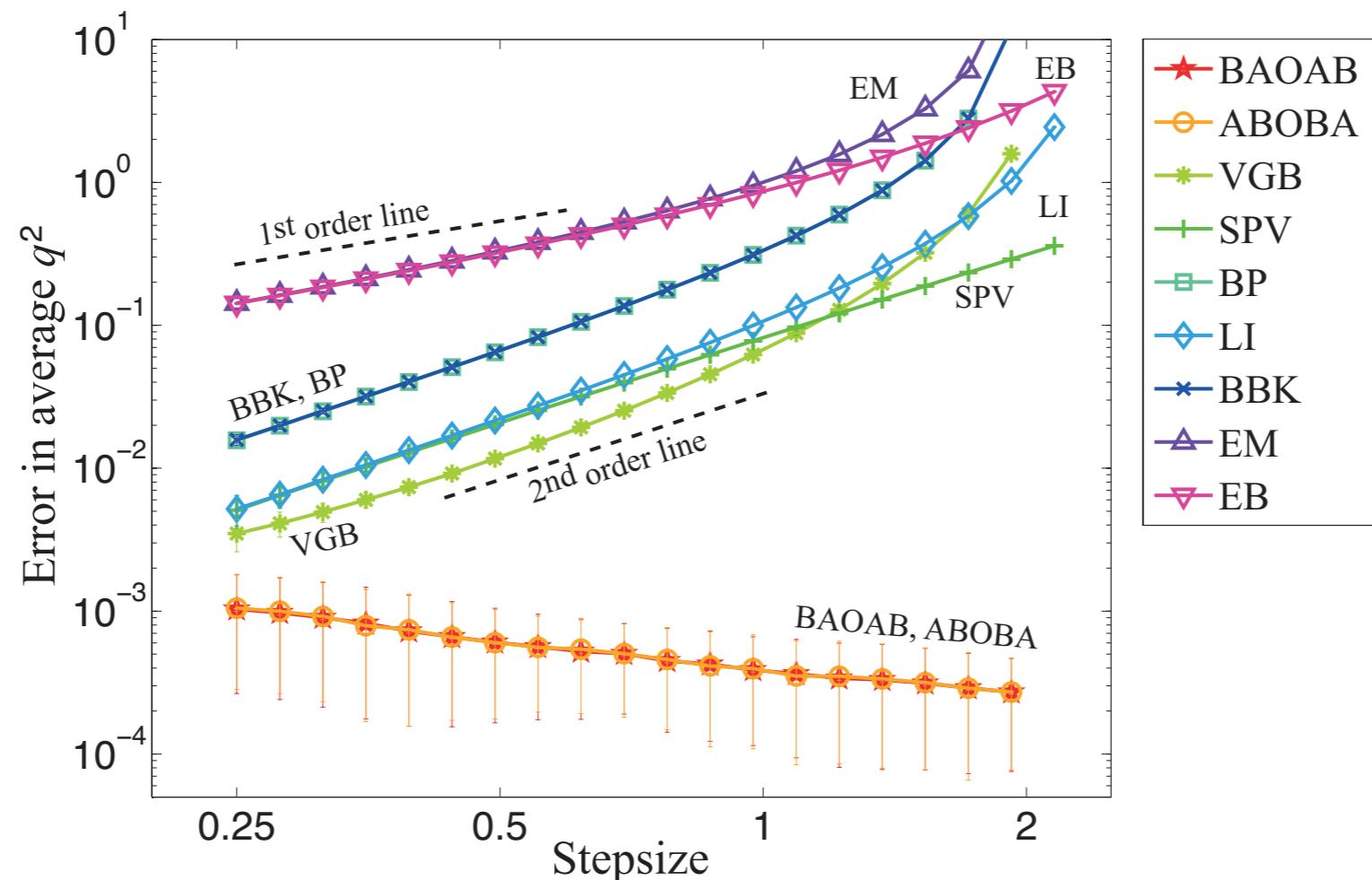
If we are not in the high friction limit, why is **BAOAB** so much better?

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...*

# Anharmonic model problem

$$U(q) = \frac{q^2}{2} + \eta q^4$$

Asymptotic analysis of sampling error of  $\langle q^2 \rangle$

typical, e.g. **OBABO**:  $O(\delta t^2)$

---

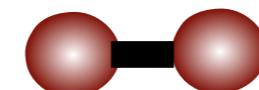
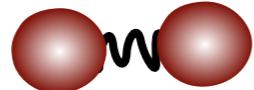
**ABOBA**:  $O(\eta \delta t^2)$

---

**BAOAB**:  $\frac{9\eta^2 \delta t^2}{6\gamma^2 + 8} + O(\eta^2 \delta t^4)$

# **Constrained Langevin Dynamics**

# Constrained Langevin Dynamics

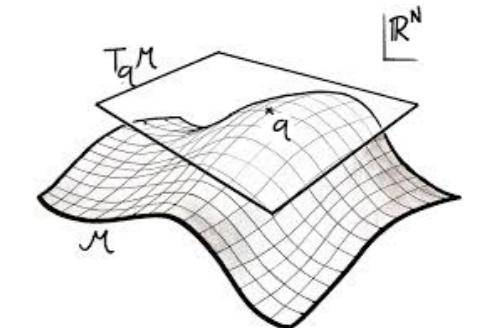


$$\frac{d}{dt}q = \mathbf{M}^{-1}p$$

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

$$0 = g_j(q), \quad j = 1, 2, \dots, m$$

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



Lagrange multipliers

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

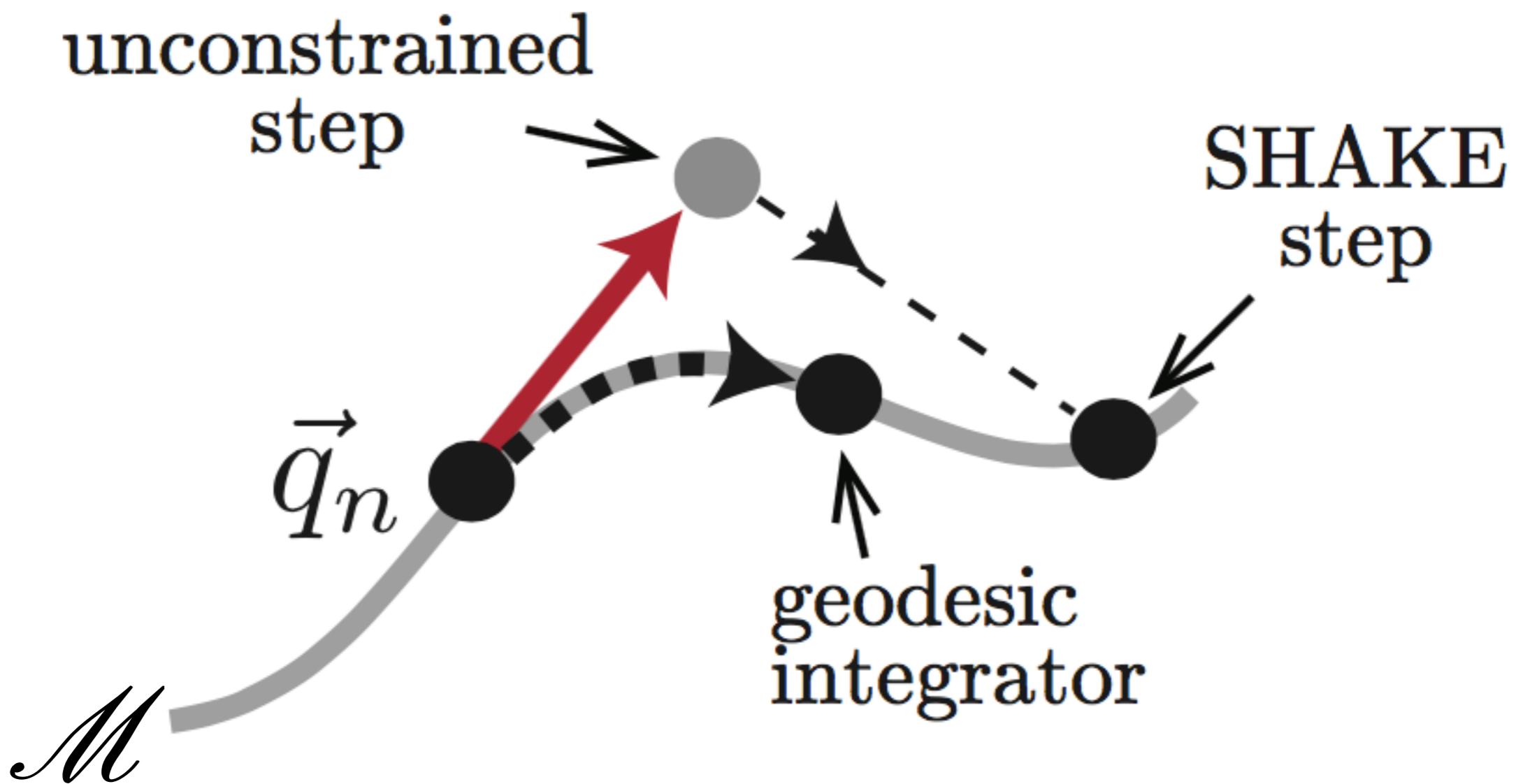
Some technical aspects:

*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.

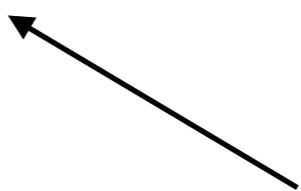
The proper analogue of the **Verlet (BAB)** or **BAOAB**

$$g\text{-Verlet: } \Phi_{h,H_{T^*\mathcal{M}}} \approx \Phi_{h/2,U_{T^*\mathcal{M}}} \circ \Phi_{h,T_{T^*\mathcal{M}}} \circ \Phi_{h/2,U_{T^*\mathcal{M}}}$$

Combines: **geodesic flow**  
**projected “kicks”**

# Constrained OU

$$dp = -\gamma \Pi p dt + \sqrt{2\gamma k_B T} \Pi M^{1/2} dW$$



Projector onto the co-tangent space

Easily solved using Rodrigues' formula:

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

assuming  $\Pi p(0) = p(0)$

# 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.

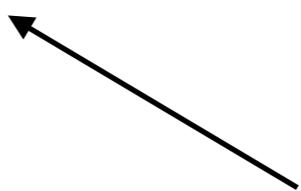
The natural constrained analogue of **Verlet (BAB)** or **BAOAB**

$$g\text{-Verlet: } \Phi_{h,H_{T^*\mathcal{M}}} \approx \Phi_{h/2,U_{T^*\mathcal{M}}} \circ \Phi_{h,T_{T^*\mathcal{M}}} \circ \Phi_{h/2,U_{T^*\mathcal{M}}}$$

Combines: **geodesic flow   projected “kicks”  
projective OU**

# Constrained OU

$$dp = -\gamma \Pi p dt + \sqrt{2\gamma k_B T} \Pi M^{1/2} dW$$



Projector onto the co-tangent space

Easily solved using Rodrigues' formula:

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

assuming  $\Pi p(0) = p(0)$

# Geodesic Integrator

## Numerical Analysis

Constrained Hamiltonian system

in coordinates     $\hat{H}(\theta, p_\theta) = \frac{1}{2} p_\theta^T \mathbf{J}(\theta) p_\theta + \hat{U}(\theta)$

↓

Splittings



BCH (nastier commutators)

“BAOAB” 2nd order

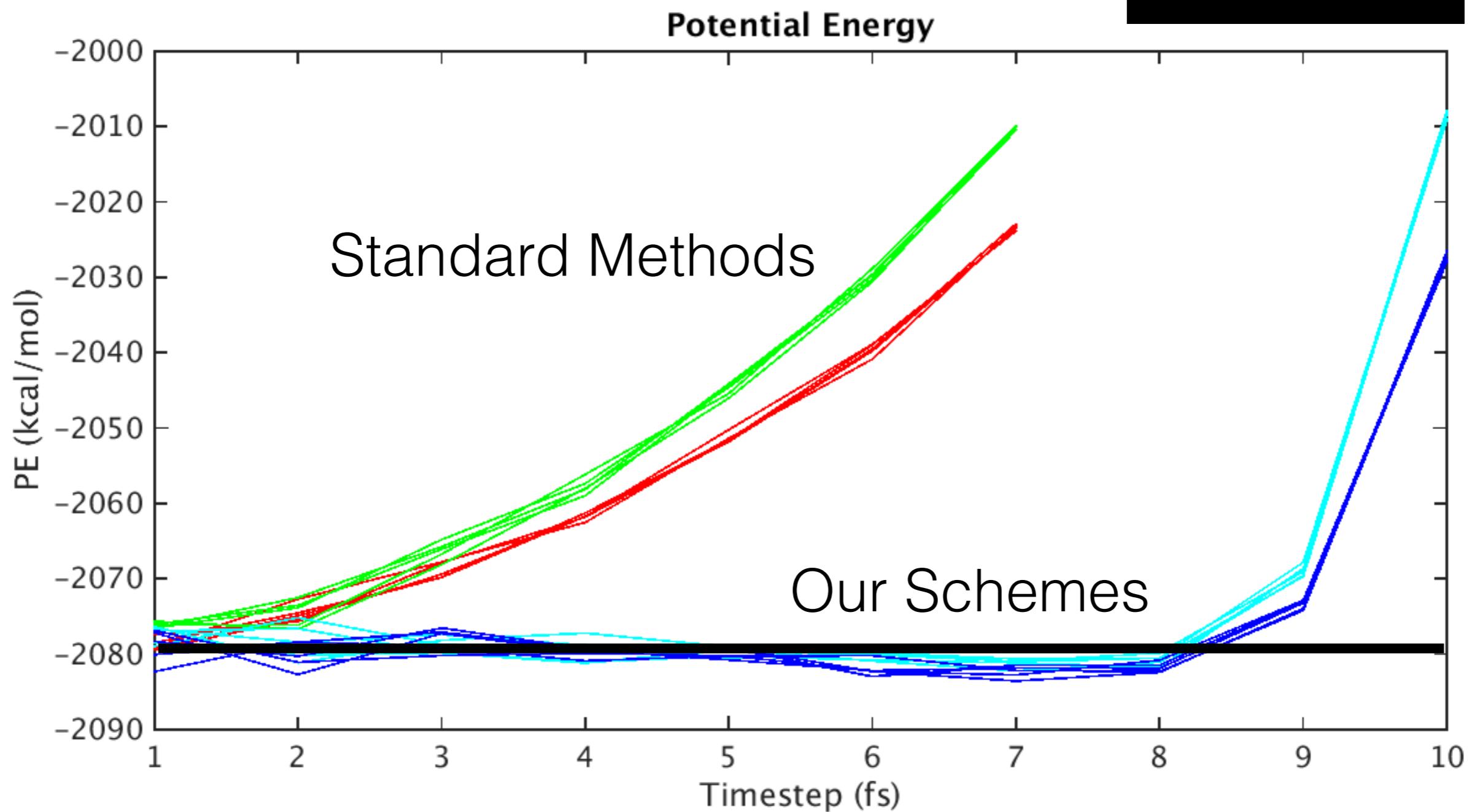
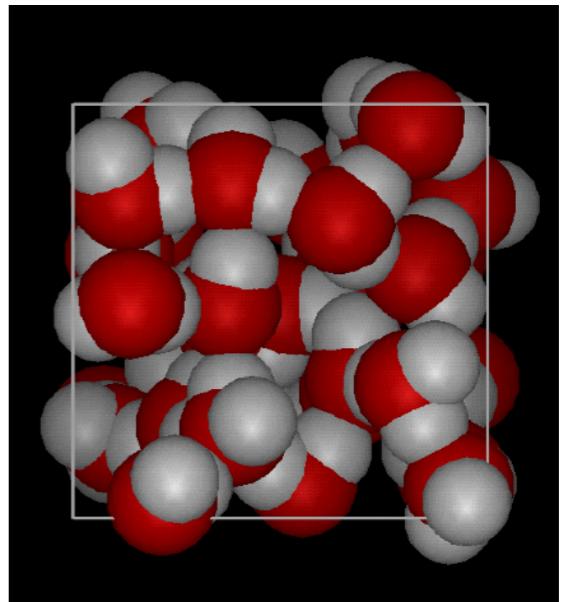


superconvergence



Box of H<sub>2</sub>O

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



# Self-Diffusion Coefficient of TIP3P Water

Scheme	3fs	5fs	7fs	9fs
Tinker default	$4.59 \pm 0.13$	$5.08 \pm 0.37$	$5.81 \pm 0.37$	fail
MEVME	$4.73 \pm 0.30$	$4.85 \pm 0.23$	$5.69 \pm 0.34$	fail
g-OBABO (1 RATTLE)	$4.65 \pm 0.18$	$5.33 \pm 0.42$	$5.69 \pm 0.34$	fail
g-OBABO (5 RATTLES)	$4.79 \pm 0.31$	$5.14 \pm 0.38$	$5.40 \pm 0.40$	$6.39 \pm 0.15$
g-BAOAB (1 RATTLE)	$4.48 \pm 0.32$	$4.41 \pm 0.19$	$4.54 \pm 0.26$	$4.61 \pm 0.31$
g-BAOAB (5 RATTLES)	$4.52 \pm 0.34$	$4.51 \pm 0.20$	$4.67 \pm 0.45$	$4.65 \pm 0.30$

Table 1: Computed diffusion coefficient of water ( $\times 10^{-5} \text{cm}^2/\text{s}$ )

Self-diffusion coeff of TIP3P water in water  $\sim \mathbf{4.51} \times 10^{-5}$

Appx 2x the experimental value!  $\mathbf{2.34} \pm 0.05 \text{cm}^2/\text{s}$

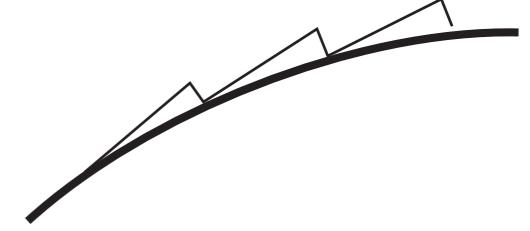
# Multiple-Timestepping Implementation

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

$$p \leftarrow p + \frac{\delta t}{2} F(q) + \sum_j \mu_j G_j(q) \in T_q^* \mathcal{M},$$

**Cotangent  
space projection**

For  $k$  from 1 to  $K_r$  do:

$$(q, p) \leftarrow A \left( q, p, \frac{\delta t}{2K_r} \right)$$

**Several Rattle  
(geodesic flow) Steps**

end do

$$p \leftarrow a_2 p + b_2 \mathbf{M}^{1/2} \mathbf{R} + \sum_j \mu_j G_j(q) \in T_q^* \mathcal{M},$$

**O Step**

For  $k$  from 1 to  $K_r$  do:

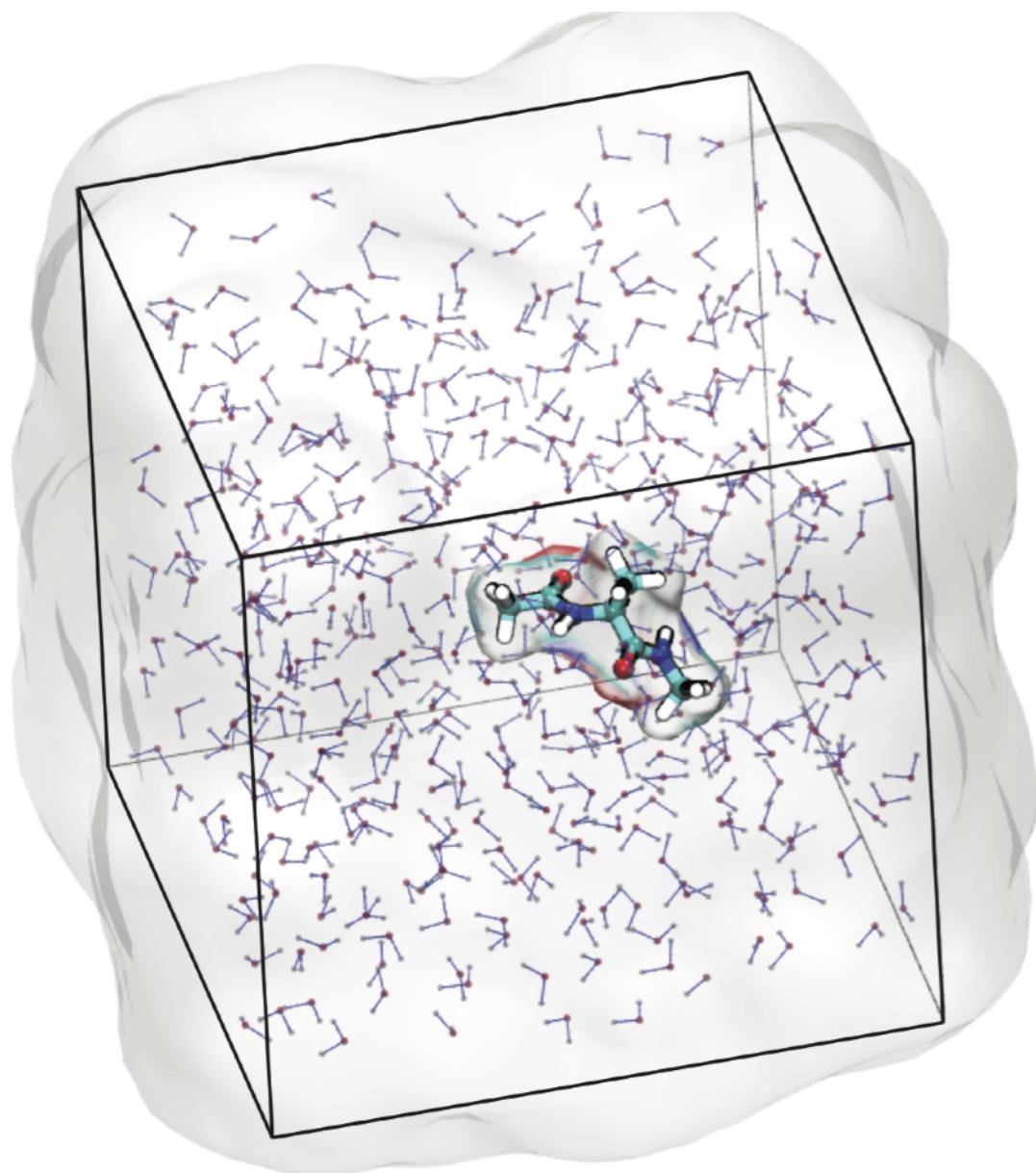
$$(q, p) \leftarrow A \left( q, p, \frac{\delta t}{2K_r} \right)$$

end do

$$p \leftarrow p + \frac{\delta t}{2} F(q) + \sum_j \mu_j \nabla g_j(q) \in T_q^* \mathcal{M},$$

# 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)

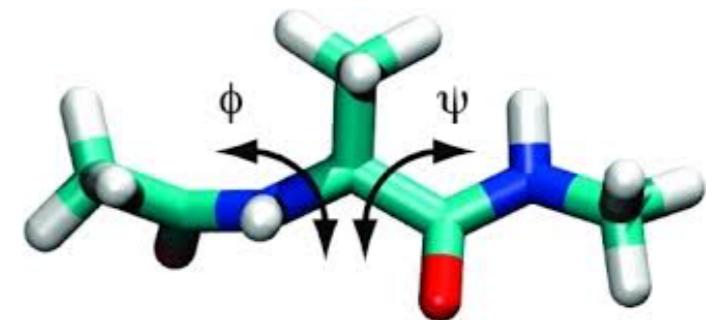
**SS** dominates the computational cost. **PP**, **PS** determine the stepsize. Therefore, consider  
***stochastic two-level multiple timestepping:***

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

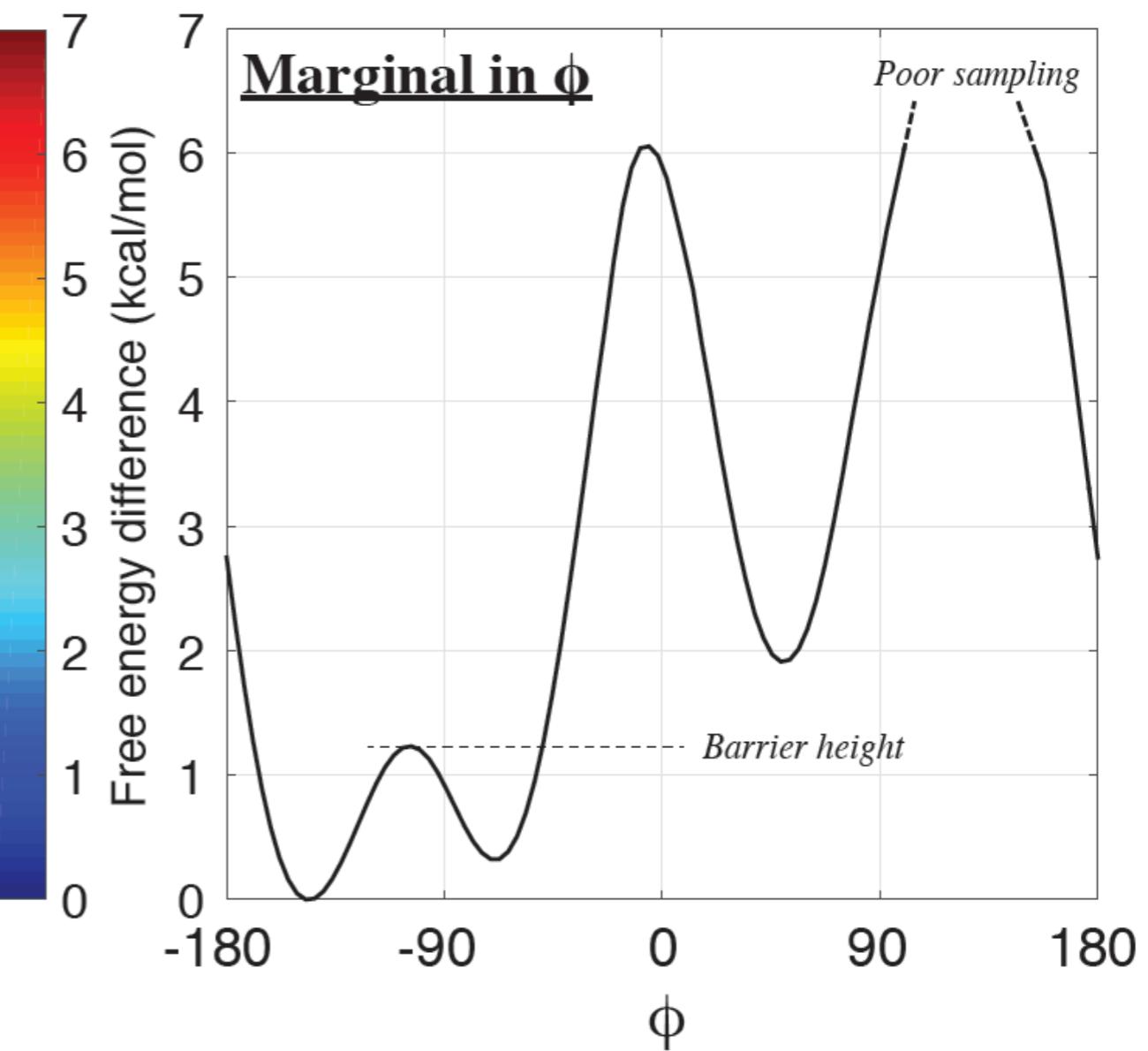
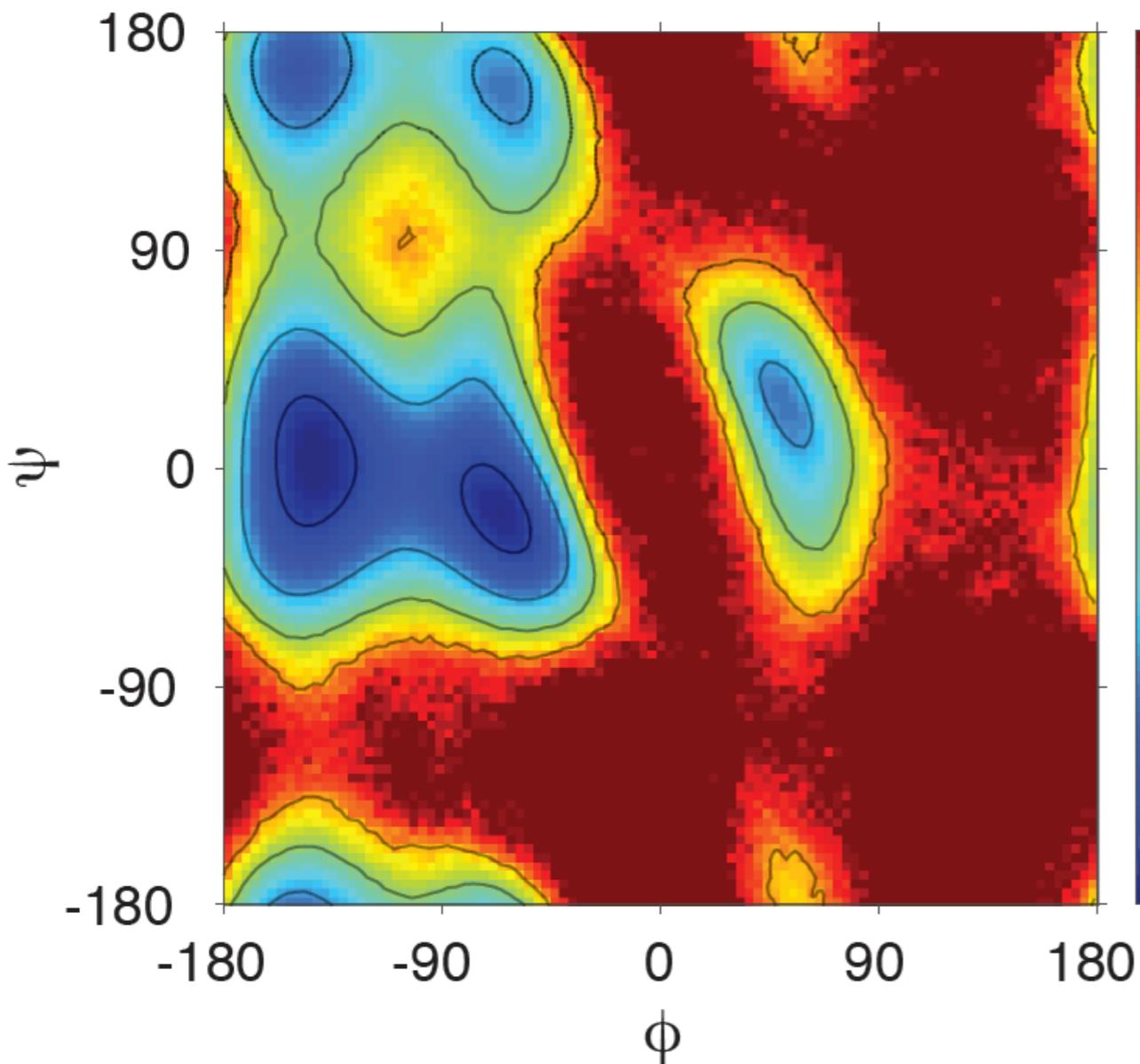
**m=2 or m=3 iterations  
of a geodesic Langevin integrator  
using several RATTLE substeps**

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

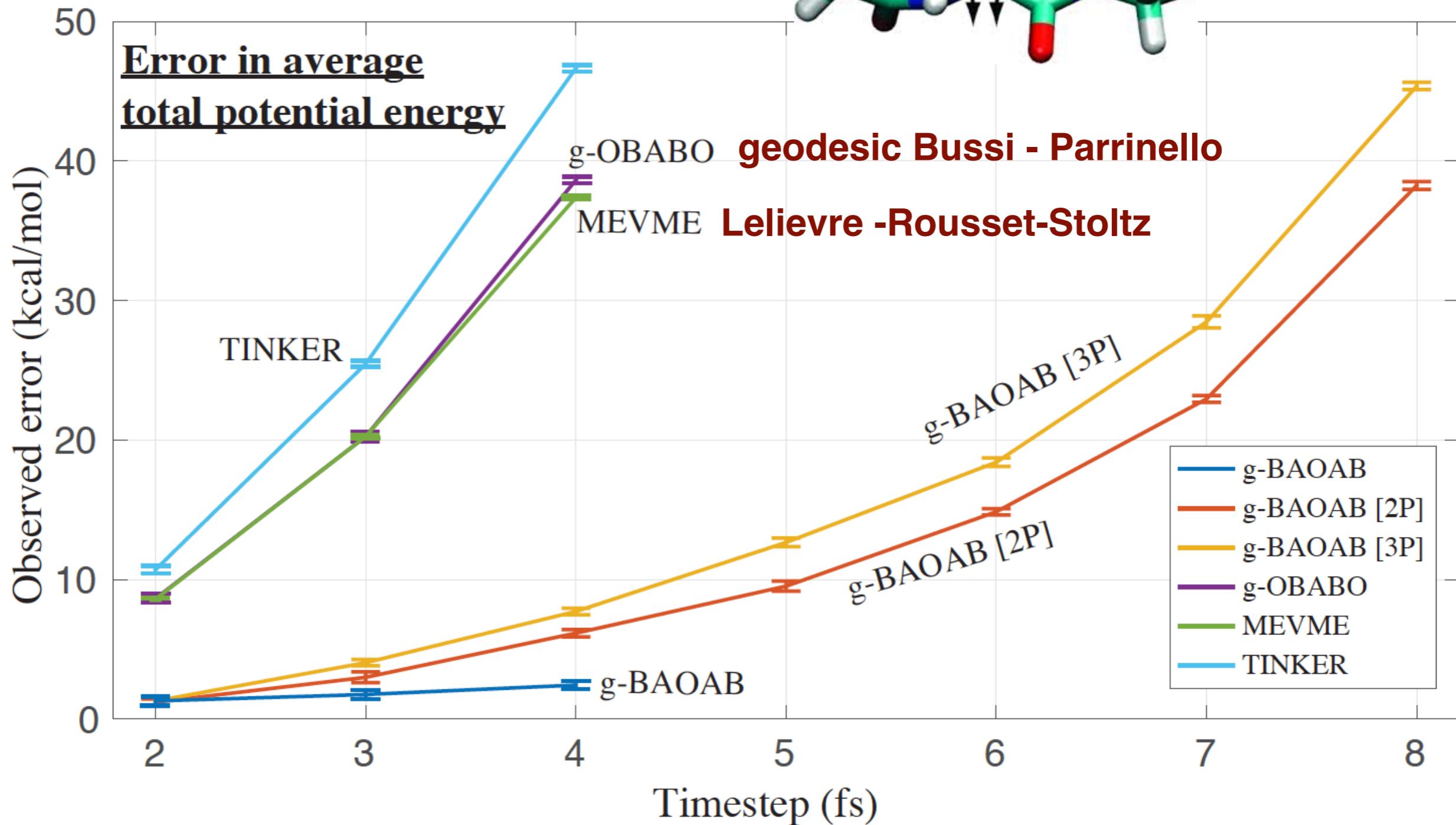
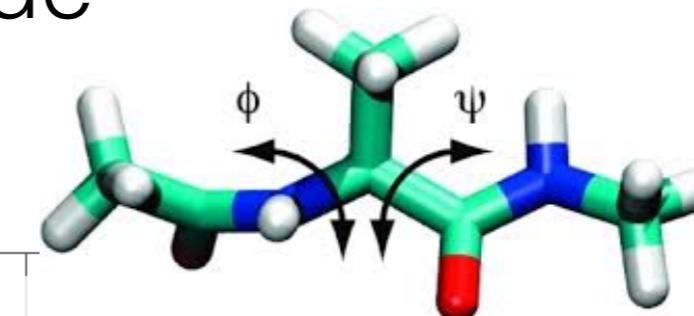
**TINKER**: Tinker's internal scheme

**OBABO** with constraint projections

**MEVME**: Scheme of Lelievre, Rousset, Stoltz 2010

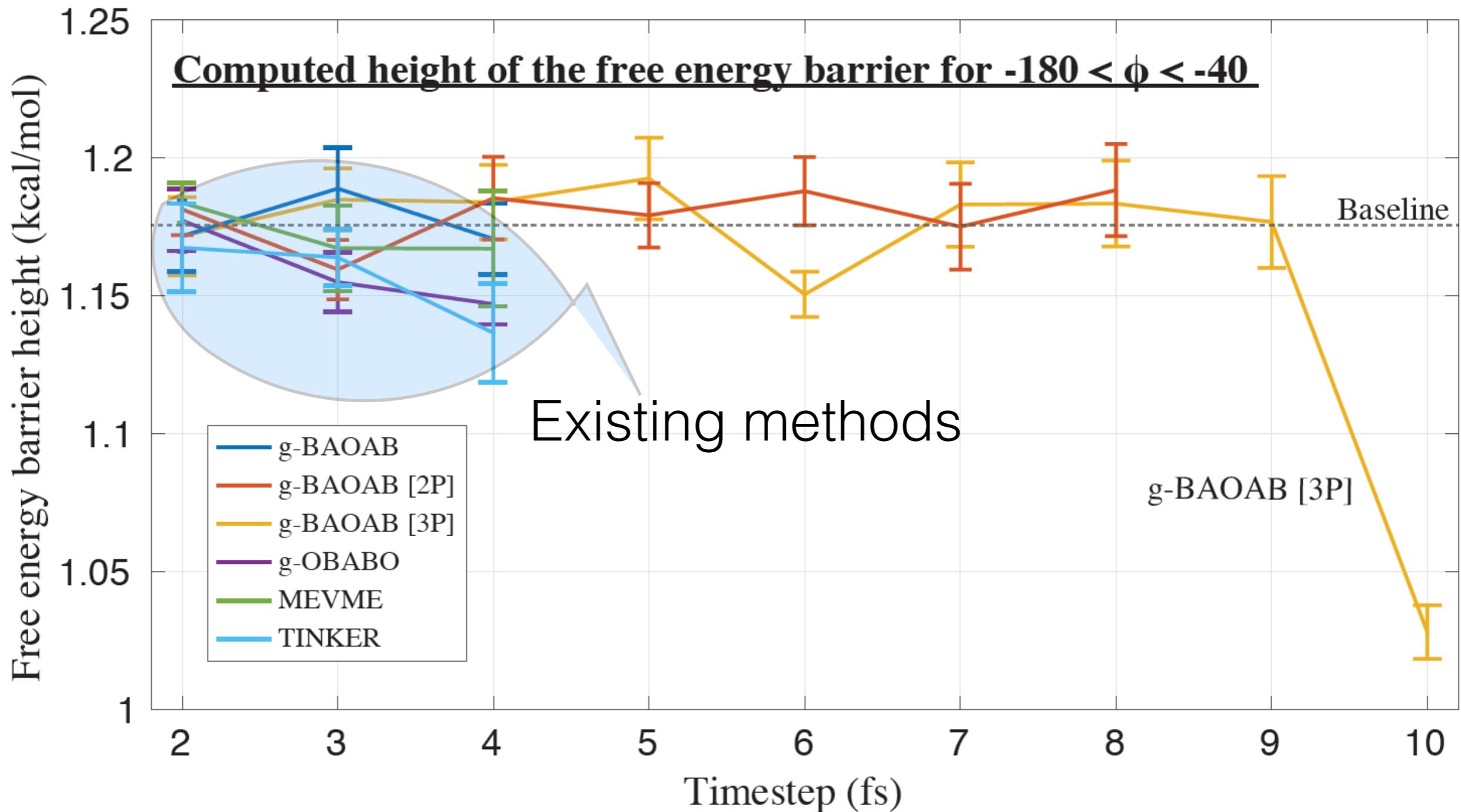
**MTS-BAOAB**: geodesic-MTS method, using 20 steps to compute the geodesics and m=3 (PP+PS) steps per step

# Alanine Dipeptide



WHY do we get this extreme accuracy?  
not explained by current BAOAB analysis

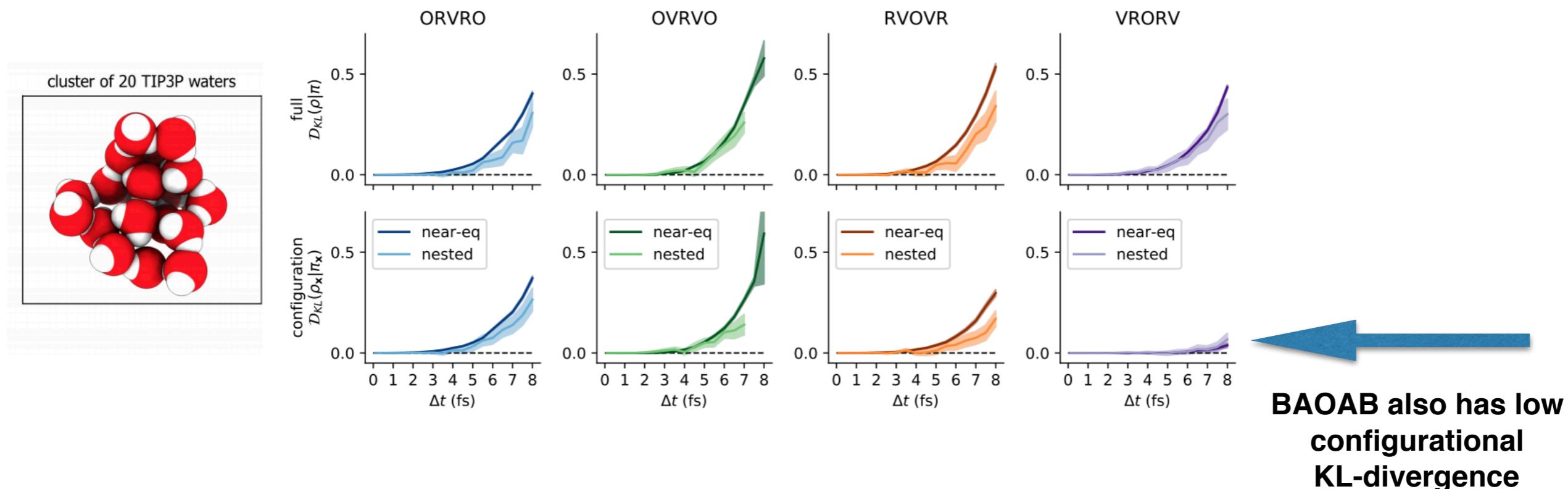
# Alanine Dipeptide Effective Free Energy Barrier Height *using geodesic solvent-solute splitting*



# 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

## Configurational Kullback-Leibler (KL)–divergence



# Stochastic Backpropagation and Approximate Inference in Deep Generative Models

Danilo J. Rezende, Shakir Mohamed, Daan Wierstra

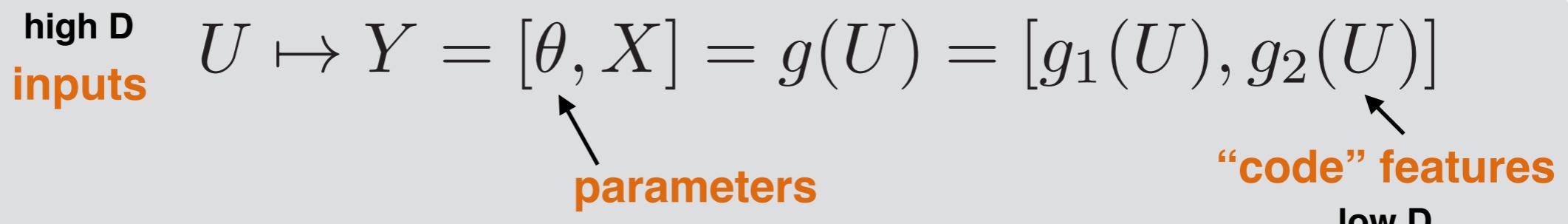
{danilor, shakir, daanw}@google.com

Google DeepMind, London

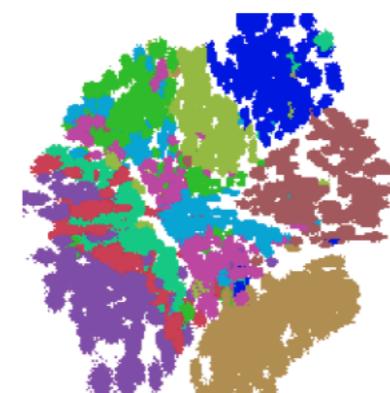
## Abstract

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.



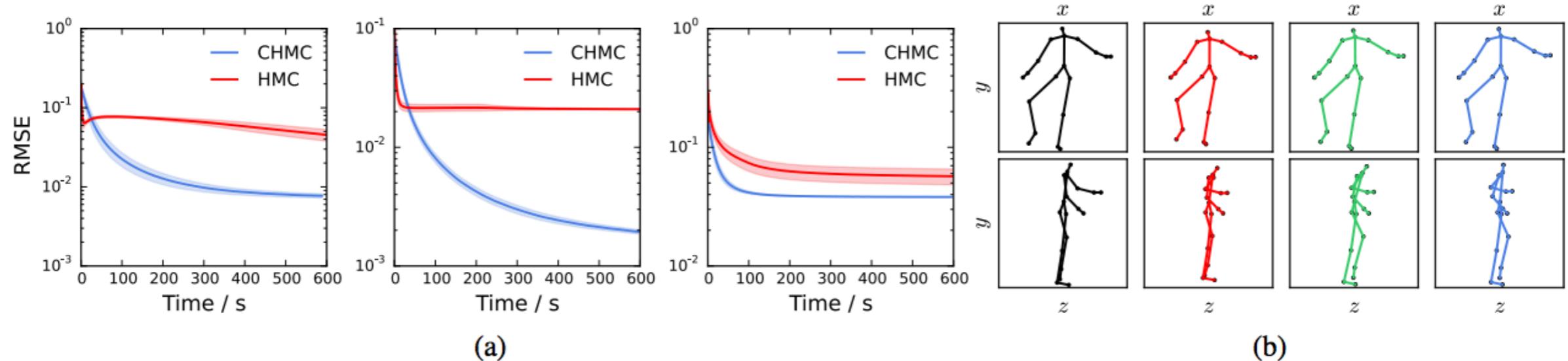
0 2 2 3 8 6 7 3 8 8	0 6 7 0 5 7 8 9 0 9	6 6 9 4 5 7 9 8 0 9
9 0 5 5 0 9 7 6 4 8	3 5 7 6 6 9 5 1 3 0	3 5 7 6 6 9 5 1 3 0
4 6 3 2 4 1 7 1 7 7	4 1 4 5 5 4 0 6 4 9	4 1 4 5 5 4 0 6 4 9
5 1 8 4 8 6 6 5 4 9	8 9 7 7 2 9 0 7 4 8	8 9 7 7 2 9 0 7 4 8
3 3 0 6 1 3 2 6 2 3	6 5 4 0 0 9 4 2 2 8	6 5 4 0 0 9 4 2 2 8
6 4 5 0 1 1 4 5 8 1	8 9 5 6 1 5 0 7 7 6	8 9 5 6 1 5 0 7 7 6
7 8 3 7 9 7 1 6 7 9	5 6 2 9 7 6 9 4 0 9	5 6 2 9 7 6 9 4 0 9
0 0 1 7 3 3 1 3 2 1	2 3 1 3 4 1 5 6 4 0	2 3 1 3 4 1 5 6 4 0
3 3 9 3 6 9 8 7 8 6	1 2 5 7 6 9 9 5 3 7	1 2 5 7 6 9 9 5 3 7
2 4 8 4 9 5 1 6 8 8	6 2 3 8 7 4 0 9 4 3	6 2 3 8 7 4 0 9 4 3



# 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,

$$\begin{bmatrix} dq \\ dp \end{bmatrix} = \underbrace{\begin{bmatrix} p dt \\ -\Omega^2 q dt - \gamma p dt + \sqrt{2\gamma/\beta} dW \end{bmatrix}}_{\text{Fast}} + \underbrace{\begin{bmatrix} 0 \\ -q dt \end{bmatrix}}_{\text{Slow}}$$

# 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:

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

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

**Combine with kicks by the slow force**

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

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

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

**Stability if**

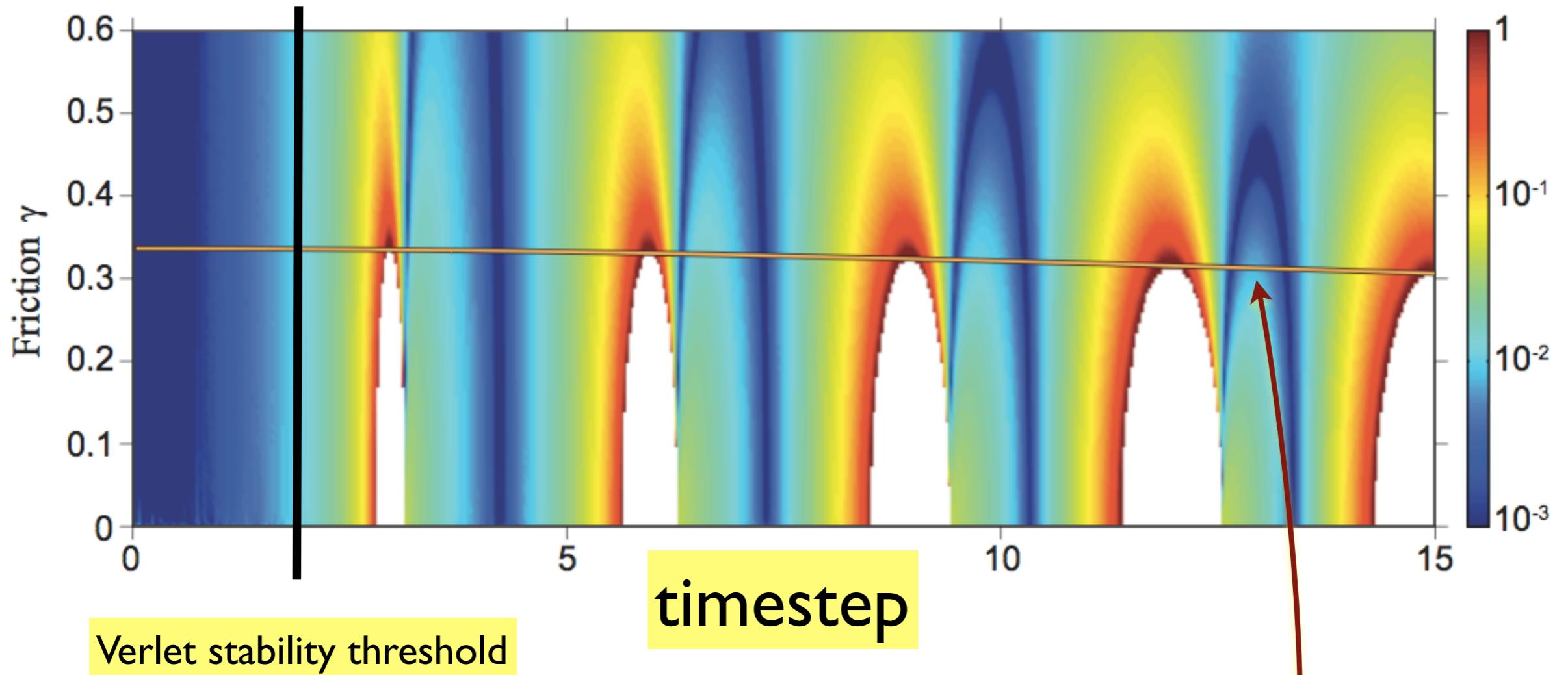
$$\rho(X) \leq 1$$

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

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

# sampling error

noise strength



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

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

$$\dot{q} = p$$

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

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

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

$$\lambda = \frac{2p \cdot F - \xi_1^2\xi_2}{2K(p, \xi_1)}$$
 to make  $K(p, \xi_1) := p \cdot p + \frac{\xi_1^2}{2} = kT$   
**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:

$$C_j(x) = c_j, j = 1, 2, \dots, k$$

phase space compressibility:

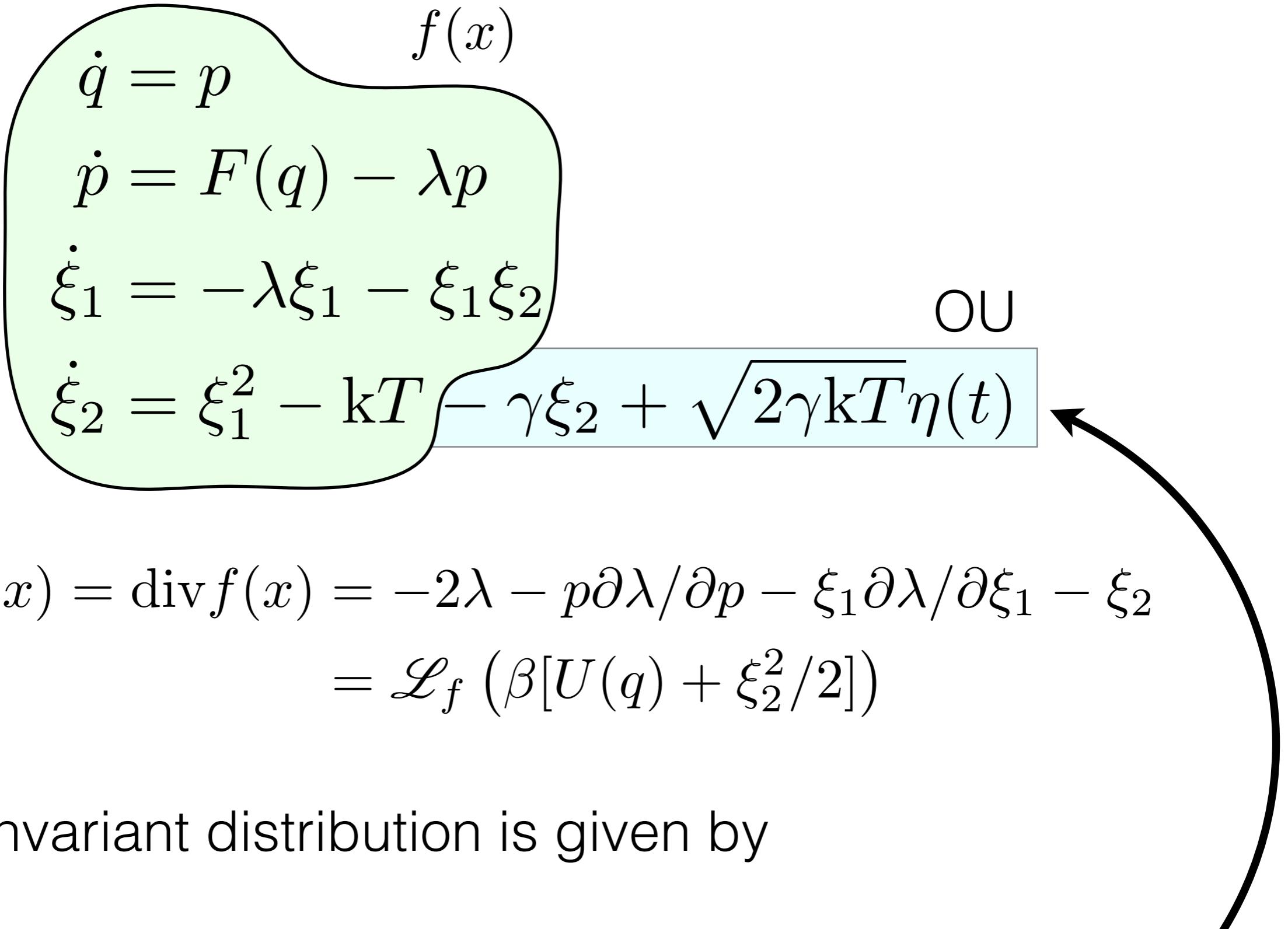
$$\kappa(x) = \operatorname{div} f(x)$$

and find:

$$\kappa = \mathcal{L}_f w$$

Then, if ergodic, the partition function is:

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



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

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

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

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

**Lyapunov function:** 😊

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

# Hörmander Condition

SDE

$$\dot{x} = b_0(x) + \sum_{i=1}^m b_i(x) \eta_i(t)$$

$$b_j \in T\mathcal{M}$$

Lars Hörmander b. 1931



Fields medal: 1962

Existence of a unique invariant measure on  $\mathcal{M}$  is assured if the collection of iterated commutators of the vector fields spans the tangent space at every point.

$$\text{span}\{b_{i=0}^m, [b_i, b_j]_{i,j=0}^m, \dots\} = T\mathcal{M}$$

*1 constraint, 4 variables, thus 3-dimensional tangent space*

# Ergodicity Property of SIN

$$f = p\partial_q - (q + \lambda p)\partial_p - (\lambda\xi_1 + \xi_2\xi_1)\partial_{\xi_1} + [(\xi_1^2 - k_B T) - \gamma\xi_2]\partial_{\xi_2},$$

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

$$\text{Span}\{\mathbf{f}, \mathbf{g}, [\mathbf{f}, \mathbf{g}], [\mathbf{f}, [\mathbf{f}, \mathbf{g}]], [\mathbf{f}, [\mathbf{f}, [\mathbf{f}, \mathbf{g}]]]\} = T\mathcal{M}$$

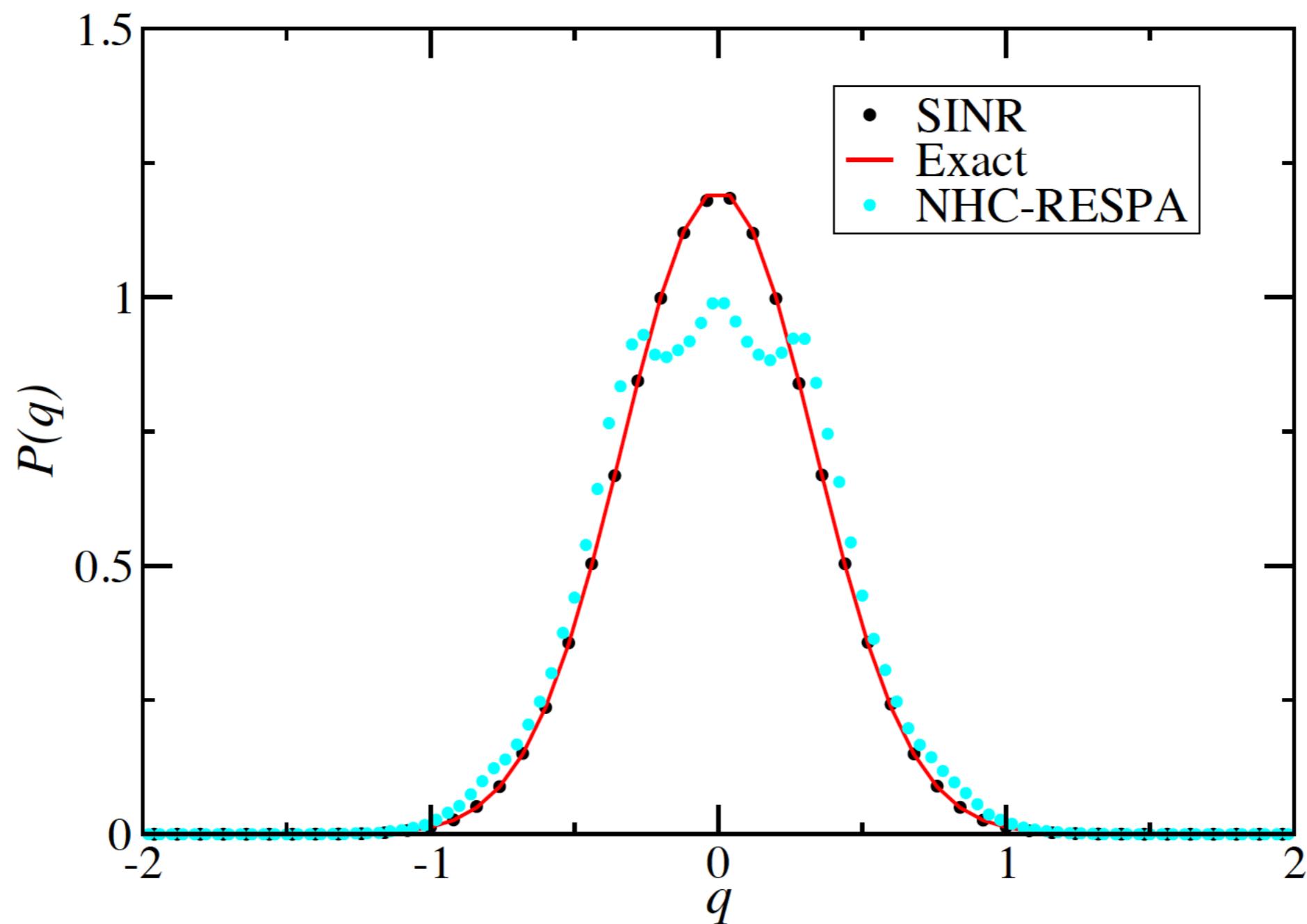
*The Alexander Davie commutator*



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

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

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



# SIN(R)

## Stochastic Isokinetic Nosé-Hoover (RESPA)

$$dq = pdt,$$

$$dp = Fdt - \left( \frac{p(F^F + F^S) - \frac{1}{2}\mu_1\xi_1^2\xi_2}{\Lambda} \right) pdt,$$

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

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

*Kinetic part:*

$$\mathcal{L}_K = p \frac{\partial}{\partial q},$$

*isokinetic Fast force:*

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

*isokinetic Slow force:*

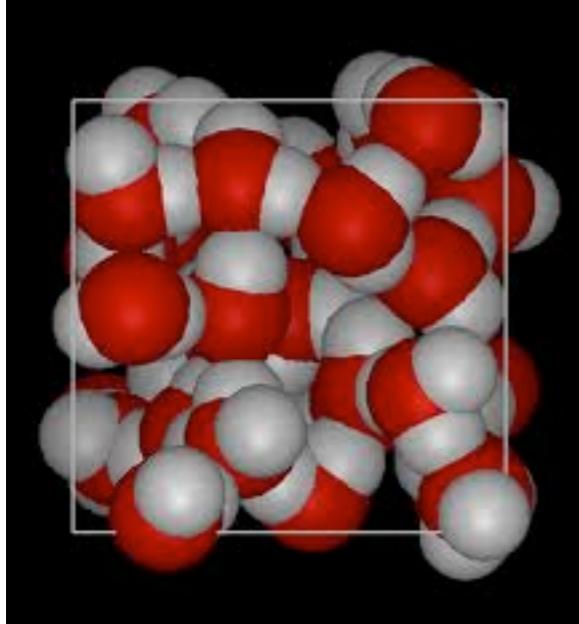
$$\mathcal{L}_S = [F^S - (p^2/\Lambda)F^S] \frac{\partial}{\partial p} - \frac{p\xi_1 F^S}{\Lambda} \frac{\partial}{\partial \xi_1},$$

*isokinetic Nosé terms:*

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

*Ornstein-Uhlenbeck term:*

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



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

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

**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 6Å, ‘real space’ part of Ewald.

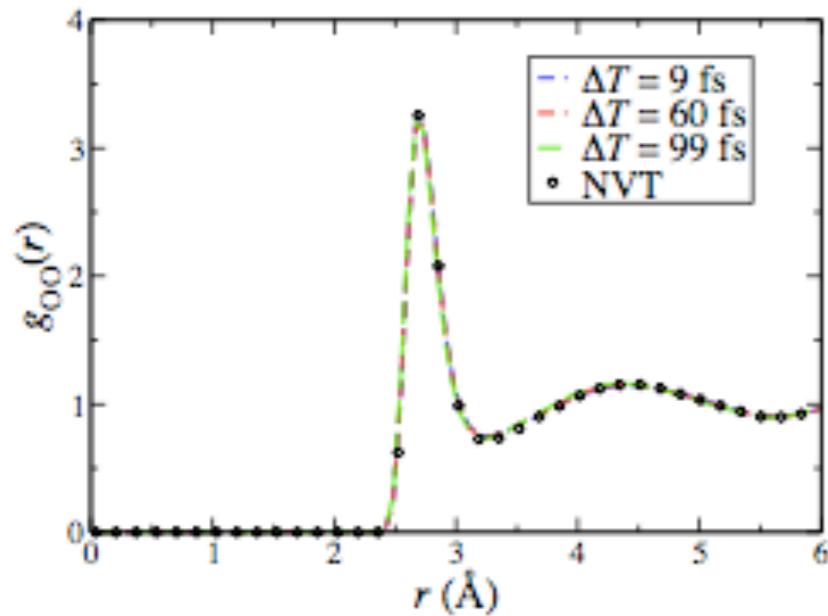
**Bond part:** Intramolecular interactions, e.g. angle and length bonds.

# Flexible H<sub>2</sub>O simulations @ $\delta t = 99\text{fs}$

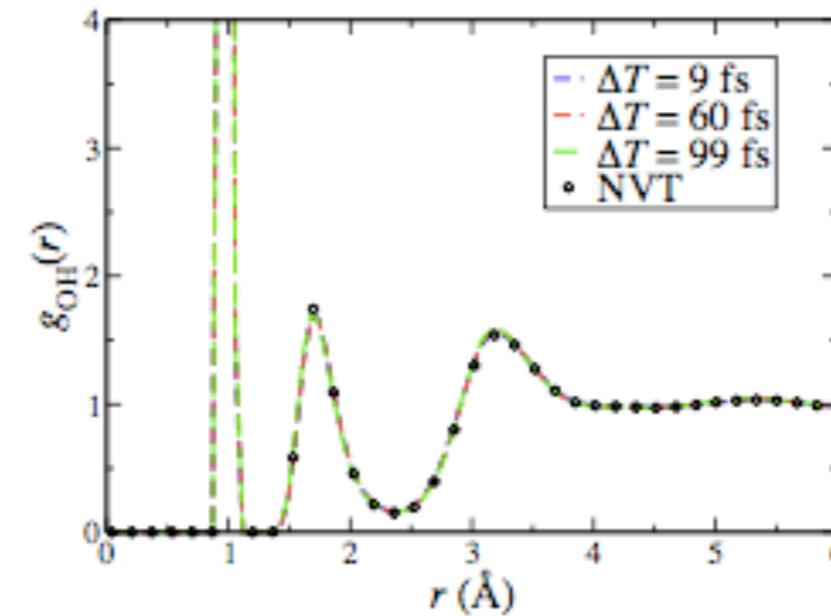
$$\delta t_{\text{inner}} = 0.5\text{fs}, \quad \delta t_{\text{mid}} = 3\text{fs}$$

## radial distributions

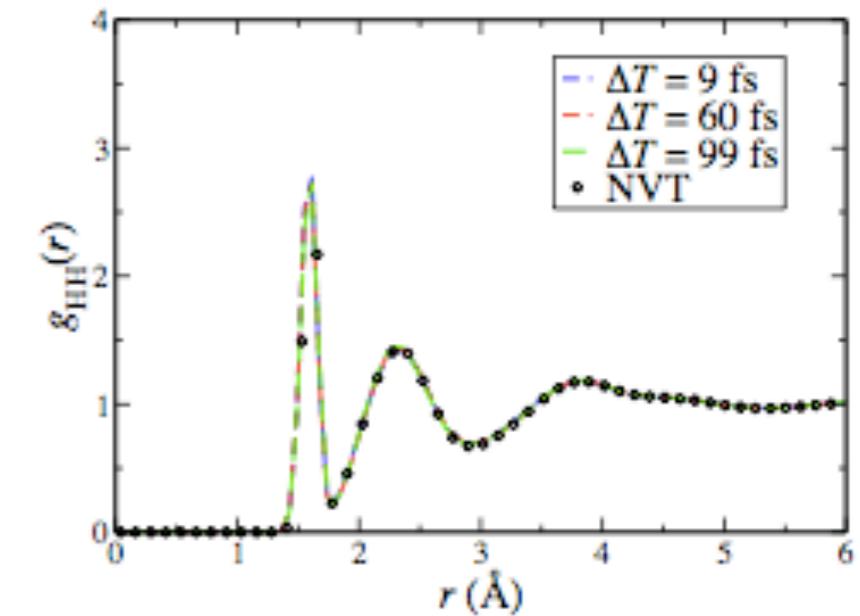
O-O



H-O



H-H



# Comparisons for biomolecules/detailed H<sub>2</sub>O

	timestep	Speedup	Notes
typical MD	1fs	1	fully flexible
typical Langevin MD	2fs	2	fully flexible
BAOAB Langevin	2.7fs	2.7	optimized for config sampling
MTS (RESPA)	3.5fs	~3	resonances limit stability
SHAKE MD	4fs	3.5	symmetric Newton
SHAKE Langevin	4fs	3.5	MEVME, g-OBABO, etc.
Equilibrium-MOLLY Izaguirre, Reich & Skeel (JCP '99)	6fs	4+	deterministic MTS-based force decomposition
g-BAOAB-MTS	8-9fs	6-7	Langevin, geodesics (RATTLE), PP+PS+SS decomposition
Colored noise thermostats to control resonance in MTS Ceriotti et al	10-12fs	??	pre-analysis of fundamental frequencies ( <b>Hessians</b> /normal modes) and tuning of parameters
LN, Langevin normal mode analysis, Barth&Schlick	48fs	10	requires <b>Hessians</b> + force decomposition
Stochastic-Isokinetic Nose (L., Margul, Tuckerman)	~100fs	>10	force decomposition (Ewald) q-sampling *only*

# Next Up

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