# Part V

# Generalized shadow hybrid Monte Carlo (GSHMC): An efficient Monte Carlo sampling method

# Outline

1. Introduction to Monte Carlo and Markov Chain Monte Carlo methods
2. Summary of the generalized hybrid Monte Carlo (HMC) method.
3. Analysis of acceptance rates.
4. Using modified energies to increase the acceptance rate in HMC methods – shadow hybrid Monte Carlo methods.
5. Applications

References: Jun S Liu "Monte Carlo Strategies in Scientific Computing", Jari Kaipio & Erkii Somersalo "Statistical and Computational Inverse Problems"

Work on GSHMC in collaboration with Elena Akhmatskaya (Fujitsu Laboratories Europe, London), and Chze Ling Wee, Mark Sansom (Biochemistry, Oxford).

# 1. Monte Carlo Methods

Given a probability density function (PDF) $\pi$ over a configuration space $\mathbb{R}^n$, we are often interested in expectation values

$$\mathbb{E}[f] = \int_{\mathbb{R}^n} f(\mathbf{x})\, \pi(\mathbf{x})\, d\mathbf{x}$$

for a given function (observable) $f$. If $n \gg 1$, numerical quadrature can become prohibitively expensive and Monte Carlo methods are often the only alternative.

The basic Monte Carlo approach assumes that we can draw independent and identically distributed (i.i.d) random samples $\mathbf{x}_i$ from the PDF $\pi$. Then an approximation to $\mathbb{E}[f]$ can be obtained as

$$\bar{f}_N = \frac{1}{N} \sum_{i=1}^{N} f(\mathbf{x}_i).$$

The law of large numbers states that the average of i.i.d. random samples tends towards their common mean. The convergence rate is assessed by the central limit theorem, which states that

$$\sqrt{N}(\bar{I}_N - \mathbb{E}[f]) \to \mathsf{N}(0, \sigma^2) \quad \text{(in distribution)},$$

where

$$\sigma^2 = \mathsf{var}[f] = \mathbb{E}[(f - \mathbb{E}[f])^2], \quad \text{(variance)}.$$

Hence the error of the Monte Carlo method is approximately $\mathcal{O}(N^{-1/2})$ regardless of the dimension $n$ of the configuration space.

Note that the convergence rate is slower than that for any standard quadrature rule in $\mathbb{R}$ (i.e. $n = 1$).

How to generate i.i.d. samples from a given PDF $\pi$? Almost all methods rely on the assumption that we can draw samples from the uniform distribution U[0, 1] and/or the normal distribution N(0, 1). Note that often $\pi$ is known only up to a normalizing constant, i.e., a $l(\mathbf{x}) = c\,\pi(\mathbf{x})$ is computable.

The following rejection method goes back to von Neumann (1951):

Assume that we can draw i.i.d. samples from a PDF $\rho$ and that there exists a constant $M$ such that $M\,\rho(\mathbf{x}) \geq l(\mathbf{x})$. Then we may apply the following procedure to produce an i.i.d. sample from $\pi$:

(a) Draw a sample $\mathbf{x}$ from the PDF $\rho$ and compute the ratio

$$r = \frac{l(\mathbf{x})}{M\,\rho(\mathbf{x})} = \frac{c\,\pi(\mathbf{x})}{M\,\rho(\mathbf{x})} \leq 1$$

(b) Draw a sample $u$ from U[0, 1] and accept $\mathbf{x}$ if $u \leq r$ otherwise return to (a) till an acceptable sample has been generated.

This works since the PDF $\pi'$ for the accepted samples $\mathbf{x}$ is equal to the proposal PDF times the acceptance rate times a normalization constant, i.e.

$$\pi'(\mathbf{x}) = \rho(\mathbf{x}) \times \frac{l(\mathbf{x})}{M\,\rho(\mathbf{x})} \times C = \pi(\mathbf{x})\frac{c\,C}{M}$$

and we pick $C = M/c$ to achieve $\pi' = \pi$.

## 2. Markov Chain Monte Carlo (MCMC) Methods

The following strategy allows one to produce statistically dependent samples from a given PDF in an elegant manner (Metropolis et al, 1953). It is based on the idea that one can construct a Markov process with the desired $\pi$ as the (only) invariant PDF.

Let $A(\mathbf{y}|\mathbf{x})$ denote the transition rule of a Markov process, i.e., given a PDF $\rho_i$ the next iterate is provided by

$$\rho_{i+1}(\mathbf{y}) = \int_{\mathbb{R}^n} \rho_i(\mathbf{x})\, A(\mathbf{y}|\mathbf{x})\, d\mathbf{x}.$$

and invariance of $\pi$ amounts to

$$\pi(\mathbf{y}) = \int_{\mathbb{R}^n} \pi(\mathbf{x})\, A(\mathbf{y}|\mathbf{x})\, d\mathbf{x}.$$

A stronger assumption is that of detailed balance

$$\pi(\mathbf{x})\, A(\mathbf{y}|\mathbf{x}) = \pi(\mathbf{y})\, A(\mathbf{x}|\mathbf{y}).$$

Metropolis-Hastings algorithm. Given the current state $\mathbf{x}_i$:

- Draw $\mathbf{y}$ from the proposal distribution $T(\mathbf{y}|\mathbf{x}_i)$.

- Draw $u \sim U[0,1]$ and update

$$\mathbf{x}_{i+1} = \begin{cases} \mathbf{y}, & \text{if } u \leq r(\mathbf{y}, \mathbf{x}_i) \\ \mathbf{x}_i & \text{otherwise.} \end{cases}$$

Here

$$r(\mathbf{y}, \mathbf{x}) = \frac{\delta(\mathbf{y}, \mathbf{x})}{\pi(\mathbf{x})\, T(\mathbf{y}|\mathbf{x})}$$

and $\delta(\mathbf{y}, \mathbf{x})$ is any symmetric function in $\mathbf{x}$ and $\mathbf{y}$ that makes $r(\mathbf{x}, \mathbf{y}) \leq 1$ for all $\mathbf{x}$, $\mathbf{y}$.

The actual transition probability from $\mathbf{x}$ to $\mathbf{y} \neq \mathbf{x}$ is given by

$$A(\mathbf{y}|\mathbf{x}) = T(\mathbf{y}|\mathbf{x})\, r(\mathbf{y},\mathbf{x}) = \pi(\mathbf{x})^{-1}\, \delta(\mathbf{y},\mathbf{x}).$$

Similarly,

$$A(\mathbf{x}|\mathbf{y}) = T(\mathbf{x}|\mathbf{y})\, r(\mathbf{x},\mathbf{y}) = \pi^{-1}(\mathbf{y})\, \delta(\mathbf{x},\mathbf{y}).$$

Since $\delta(\mathbf{x},\mathbf{y}) = \delta(\mathbf{y},\mathbf{x})$ the detailed balance condition

$$\pi(\mathbf{x})\, A(\mathbf{y}|\mathbf{x}) = \pi(\mathbf{y})\, A(\mathbf{x}|\mathbf{y})$$

follows.

The most popular choice for $\delta(\mathbf{y},\mathbf{x})$ is

$$\delta(\mathbf{y},\mathbf{x}) = \min(\pi(\mathbf{x})\, T(\mathbf{y}|\mathbf{x}), \pi(\mathbf{y})\, T(\mathbf{x}|\mathbf{y}))$$

leading to

$$r(\mathbf{y},\mathbf{x}) = \min\left(1, \frac{\pi(\mathbf{y})\, T(\mathbf{x}|\mathbf{y})}{\pi(\mathbf{x})\, T(\mathbf{y}|\mathbf{x})}\right).$$

A simple random walk MCMC algorithm is provided by the proposal step

$$\mathbf{y} = \mathbf{x}_i + \varepsilon_i$$

where $\varepsilon_i$ are i.i.d. according to some given PDF $\rho(\mathbf{x})$. Provided that $\rho(\mathbf{x}) = \rho(-\mathbf{x})$, we obtain $T(\mathbf{y}|\mathbf{x}) = T(\mathbf{x}|\mathbf{y})$ and the acceptance function reduces to

$$r(\mathbf{y}, \mathbf{x}) = \min\left(1, \frac{\pi(\mathbf{y})}{\pi(\mathbf{x})}\right).$$

An efficient MCMC method should lead to a rapid decorrelation of the accepted samples $\{\mathbf{x}_i\}_{i=1}^N$. This is measured by the integrated auto-correlation function $\tau_{int}(h)$ for a given observable $h(\mathbf{x})$.

Let $\sigma^2 = \mathrm{var}[h]$ and

$$\rho_j = \mathrm{corr}\{h(\mathbf{x}_i), h(\mathbf{x}_{i+j})\},$$

which becomes independent of $i$ once the MCMC has equilibrated, i.e., all $\mathbf{x}_i$ are assumed to follow the law $\pi$. Then

$$\begin{aligned} N\mathrm{var}\left\{\frac{h(\mathbf{x}_1)+\cdots+h(\mathbf{x}_N)}{N}\right\} &= \sigma^2\left[1+2\sum_{j=1}^{N-1}\left(1-\frac{j}{m}\right)\rho_j\right] \\ &\approx \sigma^2\left[1+\sum_{j=1}^{N-1}\rho_j\right] =: 2\sigma^2\,\tau_{int}(h). \end{aligned}$$

Hence the variance in the MCMC estimator is equal to that of

$$\frac{N}{\tau_{int}(h)}$$

independent samples (e.g. from a rejection method).

## 4. Generalized hybrid Monte Carlo method

A Markov chain will converge to some distribution of configurations if it is constructed out of Markov chain Monte Carlo (MCMC) updates each of which has the desired distribution as a fixed point, and which taken together are ergodic.

The generalized hybrid Monte Carlo (GHMC) algorithm of Horowitz (1991) and Kennedy & Pendleton (2001) for sampling from the canonical ensemble with density function

$$\rho_T(\Gamma) = \pi(\mathbf{q}, \mathbf{p}) \propto \exp(-\beta \mathcal{H}(\mathbf{q}, \mathbf{p})),$$

$\beta = 1/k_B T$, on the contrary, is defined as the concatenation of a molecular dynamics Monte Carlo (MDMC) and a partial momentum refreshment Monte Carlo (PMMC) step.

Molecular dynamics Monte Carlo (MDMC):

(i) *Molecular dynamics* (MD): an approximate integration of Hamilton's equations of motion with a time-reversible and symplectic method $\Psi_h$ over $L$ steps and step-size $h$.

The resulting time-reversible and sympletic (and hence volume conserving) map from the initial to the final state is denoted by $U_\tau : (\mathbf{q}, \mathbf{p}) \to (\mathbf{q}', \mathbf{p}')$, $\tau = L\,h$.

(ii) A *momentum flip* $\mathcal{F} : (\mathbf{q}, \mathbf{p}) \to (\mathbf{q}, -\mathbf{p})$.

(iii) *Monte Carlo* (MC): a Metropolis accept/reject test

$$(\mathbf{q}', \mathbf{p}') = \begin{cases} \mathcal{F}U_\tau(\mathbf{q}, \mathbf{p}) & \text{with probability } \min(1, \exp(-\beta\,\delta H)) \\ (\mathbf{q}, \mathbf{p}) & \text{otherwise} \end{cases},$$

with

$$\delta H := \mathcal{H}(U_\tau(\mathbf{q}, \mathbf{p})) - \mathcal{H}(\mathbf{q}, \mathbf{p}) = \mathcal{H}(\mathcal{F}U_\tau(\mathbf{q}, \mathbf{p})) - \mathcal{H}(\mathbf{q}, \mathbf{p}).$$

## Partial momentum refreshment Monte Carlo (PMMC)

We first apply an extra momentum flip $\mathcal{F}$ so that the trajectory is reversed upon an MDMC rejection (instead of upon an acceptance). The momenta $\mathbf{p}$ are now mixed with a normal (Gaussian) i.i.d. distributed noise vector $\mathbf{u} \in \mathbb{R}^m$ and the complete partial momentum refreshment step is given by

$$\begin{pmatrix} \mathbf{u}' \\ \mathbf{p}' \end{pmatrix} = \begin{pmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ -\mathbf{p} \end{pmatrix} \tag{1}$$

where

$$\mathbf{u} = \beta^{-1/2} \mathcal{M}(\mathbf{q})^{1/2} \xi, \quad \xi = (\xi_1, \ldots, \xi_m)^T, \quad \xi_i \sim \mathrm{N}(0,1),$$

and $0 \leq \phi \leq \pi/2$. Here $\mathrm{N}(0,1)$ denotes the normal distribution with zero mean and unit variance.

Remark. If $\mathbf{p}$ and $\mathbf{u}$ are both distributed according to the same normal (Gaussian) distribution, then so are $\mathbf{p}'$ and $\mathbf{u}'$. This special property of Gaussian random variables under an orthogonal transformation (1) makes it possible to conduct the partial momentum refreshment step without a Metropolis accept/reject test.

Special cases of GHMC.

- The standard hybrid Monte Carlo (HMC) algorithm is the special case where $\phi = \pi/2$. The momentum flips may be ignored in this case since $\mathbf{p}' = \mathbf{u}$ in (1) and the previous value of $\mathbf{p}$ is entirely discarded.

- The choice $\phi = 0$ corresponds to constant energy molecular dynamics (provided $\delta H = 0$).

- The Langevin Monte Carlo algorithm of Horowitz (1991) corresponds to $L = 1$; i.e., a single MD time-step with $\tau = h$, and $\phi$ arbitrary. Langevin Monte Carlo recovers stochastic Langevin dynamics

$$\dot{\mathbf{q}} = \mathcal{M}^{-1}\,\mathbf{p}, \qquad \dot{\mathbf{p}} = -\nabla_{\mathbf{q}} V(\mathbf{q}) - \gamma \mathbf{p} + \sigma \dot{\mathbf{W}},$$

provided $\phi = \sqrt{2\gamma h} \ll 1$. Here $\gamma > 0$ is a constant, $\mathbf{W}(t)$ is an $m$-dimensional Wiener process, and $\sigma$ is determined by the standard fluctuation-dissipation relation. Indeed, we find that (1) without the momentum flip $\mathcal{F}$ and with $\phi = \sqrt{\gamma h} \ll 1$ reduces to

$$\mathbf{p}' \approx (1 - \gamma h)\,\mathbf{p} + (2\gamma h)^{1/2}\mathbf{u}.$$

The overdamped first-order Langevin dynamics limit follows for $\phi = \pi/2$.

Detailed balance of MDMC step:

Given two subsets $A$ and $B$ of phase space $\mathbb{R}^{2m}$, let $\pi_{AB}$ denote the probability to go from $A$ to $B$. Then

$$\begin{aligned}
\pi_{AB} &= \int_B \int_A A(\Gamma'|\Gamma)\, \rho_T(\Gamma)\, d\Gamma\, d\Gamma' \\
&= \int_{\mathbb{R}^{2m}} \chi_A(\Gamma)\, \chi_B(\mathcal{F}U_\tau(\Gamma))\, \min\left\{1, \frac{\rho_T(\mathcal{F}U_\tau(\Gamma))}{\rho_T(\Gamma)}\right\} \rho_T(\Gamma)\, d\Gamma \\
&= \int_{\mathbb{R}^{2m}} \chi_A(\Gamma)\, \chi_B(\mathcal{F}U_\tau(\Gamma))\, \min\left\{\rho_T(\Gamma), \rho(\mathcal{F}U_\tau(\Gamma))\right\} d\Gamma \\
&= \int_{\mathbb{R}^{2m}} \chi_A(\mathcal{F}U_\tau(\hat{\Gamma}))\, \chi_B(\hat{\Gamma})\, \min\left\{\rho_T(\mathcal{F}U_\tau(\hat{\Gamma})), \rho_T(\hat{\Gamma})\right\} d\hat{\Gamma} \\
&= \pi_{BA}
\end{aligned}$$

where we used the substitution $\hat{\Gamma} = \mathcal{F}U_\tau(\Gamma)$ (volume conserving!) and $(\mathcal{F}U_\tau)^2 = \mathrm{id}$!

# 5. Analysis of acceptance rates

We now analyze the acceptance rate of the molecular dynamics Monte Carlo step ($\beta = 1$ for simplicity).

Step 1. We obtain

$$\begin{aligned}\langle \exp(-\delta H)\rangle &= \frac{1}{C}\int_{-\infty}^{\infty} \exp(-\delta H)\, \exp(-\mathcal{H})\, d\mathbf{z} \\ &= \frac{1}{C}\int_{-\infty}^{\infty} \exp(-H \circ U_\tau)\, d\mathbf{z} = \frac{1}{C}\int_{-\infty}^{\infty} \exp(-H)\, d\mathbf{z}' = 1\end{aligned}$$

Step 2. Cumulants:

$$\begin{aligned}0 &= \log\left(\langle \exp(-\delta H)\rangle\right) \\ &= \sum_{n=1}^{\infty} \frac{\kappa_n}{n!} = -\langle \delta H\rangle + \frac{1}{2}\langle (\delta H - \langle \delta H\rangle)^2\rangle + \ldots\end{aligned}$$

Hence

$$\langle \delta H\rangle \approx \frac{1}{2}\langle (\delta H - \langle \delta H\rangle)^2\rangle = \mathcal{O}(m\, h^{2p})$$

$p \geq 1$ the order of the method and $m$ the number of DOFs.

Step 3. Hence we may assume that $\delta H$ is N$(\sigma_0^2/2, \sigma_0^2)$ distributed with $\sigma_0^2 = \langle(\delta H - \langle\delta H\rangle)^2\rangle$. The average acceptance rate is determined by

$$\begin{aligned}\langle P_{acc}\rangle &= \frac{1}{\sqrt{2\pi}\sigma_0}\int_{\infty}^{\infty} \text{min }(1, \text{exp}(-x))\ \text{exp}\left(-\frac{(x-\langle\delta H\rangle)^2}{2\sigma_0^2}\right) dx \\ &= \text{erfc}\left(\frac{1}{2}\langle\delta H\rangle\right).\end{aligned}$$

Step 4. To keep the average acceptance rate constant as the system size changes we have to keep the variance $\sigma_0^2$ constant, i.e.,

$$h^{2p}\, m = \text{constant}.$$

However, the MDMC step becomes increasingly computationally demanding as we either decrease h (i.e., increase the number of MD steps $L$) or increase the order $p$ of the method.

## 6. Shadow hybrid Monte Carlo methods

The performance of hybrid Monte Carlo methods degrades as the system size increases.

Izaguirre & Hampton (2004) suggested to implement the standard HMC method with respect to a modified canonical density

$$\widetilde{\rho}_T \propto \exp(-\beta\widetilde{\mathcal{H}}),$$

where

$$\widetilde{\mathcal{H}} = \max\{\mathcal{H}, \mathcal{H}_h - c\} \tag{2}$$

and $c$ is an arbitrary constant that limits the amount by which $\widetilde{\mathcal{H}}$ is allowed to depart from $\mathcal{H}_h$. Here $\mathcal{H}_h$ is a modified energy in the sense of backward error analysis.

Note that any modified target density is suitable as along as

$$\widetilde{\rho}_T(\mathbf{q},\mathbf{p}) = \widetilde{\rho}_T(\mathbf{q},-\mathbf{p}), \qquad \text{i.e.} \qquad \widetilde{\mathcal{H}}(\mathbf{q},\mathbf{p}) = \widetilde{\mathcal{H}}(\mathbf{q},-\mathbf{p}).$$

Ideally one would like to choose $\widetilde{\mathcal{H}} = \mathcal{H}_h$ to get an optimal acceptance rate in the MDMC part of HMC.

However, since the modified energy $\mathcal{H}_h$ is no longer separable and the momenta $p$ are no longer Gaussian distributed, the generation of a new set of momenta via a simple acceptance-reject method becomes very costly. This has motivated the specific choice of $\widetilde{\mathcal{H}}$ in (2).

The shadow hybrid Monte Carlo (SHMC) method of Izaguirre & Hampton requires fine tuning the parameter $c$ to optimize the performance of the method.

## 7. Generalized shadow hybrid Monte Carlo methods

Our work has been motivated by the desire to use

$$\widetilde{\rho}_T \propto \exp(-\beta \mathcal{H}_h),$$

where $\mathcal{H}_h$ is a modified/shadow energy computed by the methods described earlier (Part IV). We have mostly used 4th and 6th order modified energies.

This choice achieves an optimal acceptance rate $P_{\text{acc}}$ in the MDMC part (arbitrarily close to one even as the system size $m$ increases).

Since $P_{\text{acc}} \approx 1$ in the MDMC part, the negative impact of the undesirable momentum flip $\mathcal{F}$ upon rejection can be minimized in the GHMC method at essentially no additional cost.

This makes GHMC a viable alternative to the standard HMC method. (A low acceptance rate leads to an undesirable *Zitter-bewegung* due to an accumulation of momentum flips.)

What about the momentum refreshment step?

The GHMC methodology allows one to implement the partial momentum refreshment (PMMC) step as a Markov chain Monte Carlo method with invariant density $\widetilde{\rho}_T$.

## 8. Partial momentum update hybrid Monte Carlo

In the partial momentum refreshment step, we consider $\mathcal{H}_h$ for fixed $\mathbf{q}$ as a function of $\mathbf{p}$ and introduce $\mathcal{H}_{h,\mathbf{q}}(\mathbf{p})$ as a short hand the emphasize this point.

We now construct a standard hybrid Monte Carlo method in the "position" vector $\mathbf{p}$ and a set of "conjugate momenta" $\xi \in \mathbb{R}^m$. The required "Hamiltonian" is provided by

$$H(\mathbf{p}, \xi) = \mathcal{H}_{h,\mathbf{q}}(\mathbf{p}) + \frac{1}{2}\xi \cdot \xi$$

and defines the reference density

$$\widehat{\rho} \propto \exp(-\beta H)$$

for the Metropolis acceptance criterion.

We also need a symplectic and time-reversible (in $\mathbf{p}$ and $\xi$) propagator $U_\phi$, $\phi$ the "step-size". The propagator could, for example, be provided by the exact solution to the linear Hamiltonian

$$\bar{H}(\mathbf{p}, \xi) = \frac{1}{2}\mathbf{p} \cdot \left[\mathcal{M}(\mathbf{q})^{-1}\mathbf{p}\right] + \frac{1}{2}\xi \cdot \xi$$

(This has essentially been used in the original GHMC method!) or by a symplectic time-stepping method.

Following the previous analysis and the fact that $U_\phi - \mathrm{id} = \mathcal{O}(\phi)$, we obtain

$$\langle \delta H \rangle = \mathcal{O}(m\,\phi^2)$$

for $\phi \ll 1$.

We keep the acceptance rate in the PMMC step constant by picking $\phi$ such that

$$m\,\phi^2 = \text{constant}. \tag{3}$$

Remarks.

- Since the NVE ensemble approaches the NVT ensemble as $m \to \infty$, it makes a lot of sense to have $\phi \to 0$ as $m \to \infty$. It essentially means that the momentum vector gets less and less stochastically perturbed as the number of DOFs increases.
- The hybrid Monte Carlo method in $\mathbf{p}$ and $\xi$ is very inexpensive since no force evaluations are required for the propagator $U_\phi$!

## 9. Generalized shadow hybrid Monte Carlo (GSHMC)

A GHMC method with respect to a modified energy has been introduced by Akhmatskaya & Reich (2005,2008) and is called the generalized shadow HMC (GSHMC) method.

The GSHMC method achieves an $m$-indepedent acceptance rate in both the MDMC and PMMC steps for given fixed step-size $h$ in the MD propagator $\Psi_h$.

GSHMC can be viewed as an importance sampling method using the modified ensemble

$$\widetilde{\rho}_T \propto \exp(-\beta \mathcal{H}_h);$$

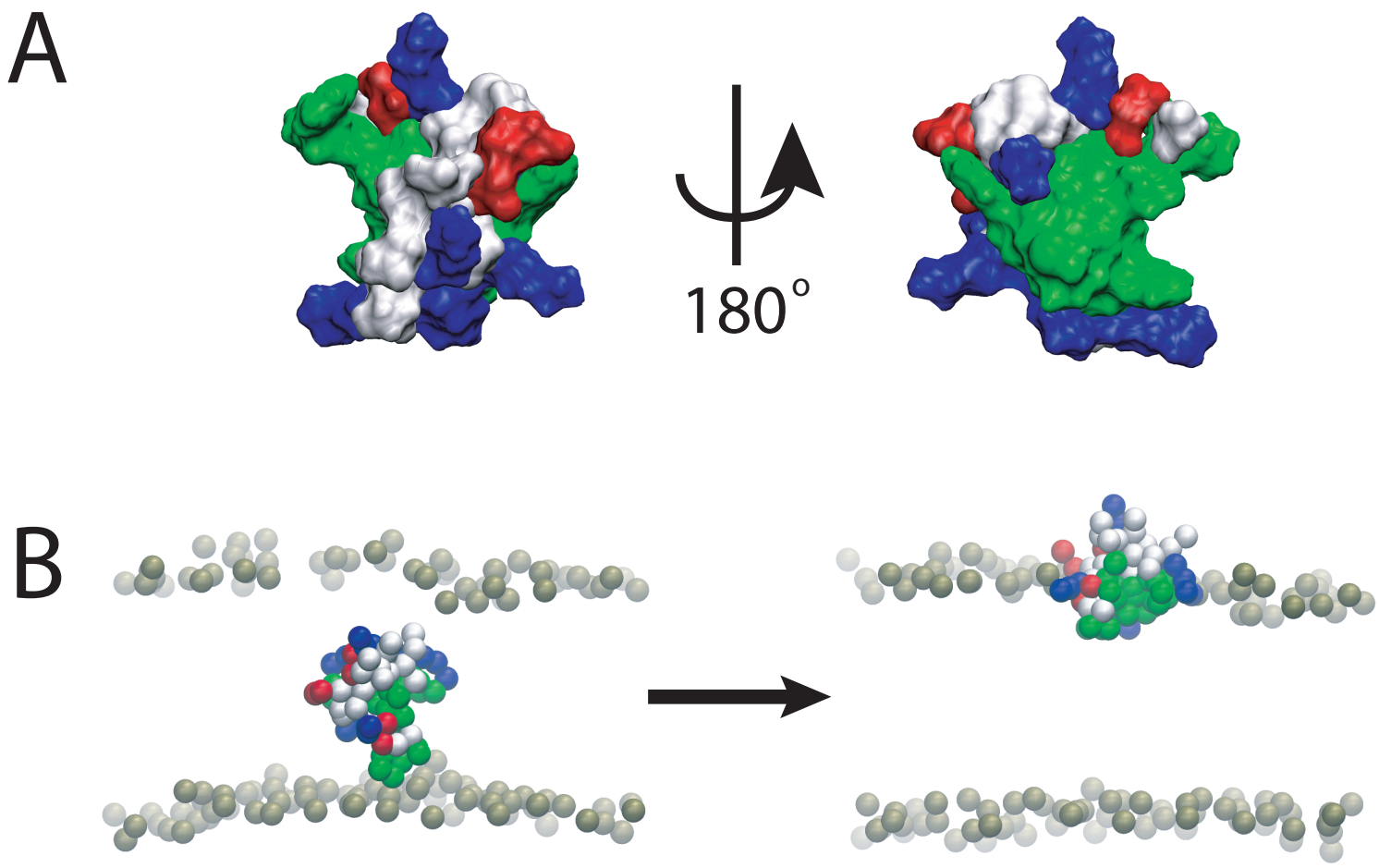
hence a re-weighting of the Monte Carlo sampling is necessary.

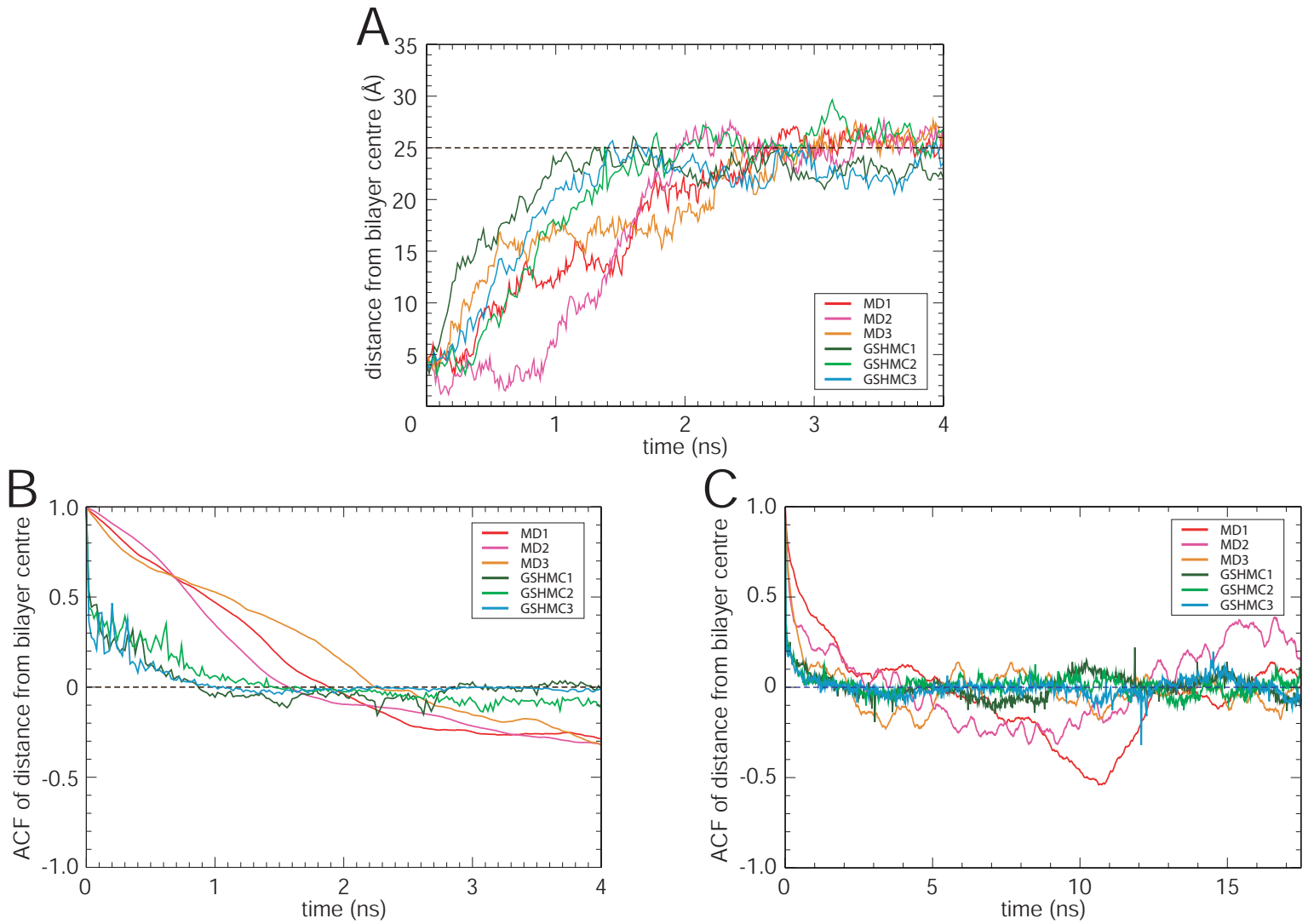
The computational overhead compared to GHMC consists in two evaluations of the modified energy $\mathcal{H}_h$ per time step (and hence in a small number of additional force evaluations).

## 9. Application of GSHMC to MD

We provide a numerical example for an interfacial membrane protein (Akhmatskaya, Reich, Sansom, 2008). We compare the sampling efficiency of GSHMC to standard molecular dynamics with a Berendsen thermostat.

We studied a coarse grained model of a toxin/bilayer system initially located at the "wrong" position in the membrane. We monitor the drift (distance) to the preferred location over 50 ns simulations.


A
180°
B


A
distance from bilayer centre (Å)
time (ns)
MD1
MD2
MD3
GSHMC1
GSHMC2
GSHMC3
B
ACF of distance from bilayer centre
time (ns)
C
ACF of distance from bilayer centre
time (ns)

## 9. Bayesian parameter estimation

Another natural application of GSHMC is in Bayesian parameter estimation (and other statistical inference problems) (Neal, 1996).

We are currently investigating the application of GSHMC to data assimilation and initialization of numerical weather prediction (3D and 4D Var).

In Bayesian estimation, we consider the unknown parameters $\mathbf{x}$ as random variables with some *a priori* distribution $\rho$. The *a priori* information is combined with the measurement $\mathbf{y}$ through a conditional density function $\rho(\mathbf{y}|\mathbf{x})$.

Recall that Bayes' rule is given by

$$\rho(\mathbf{x}|\mathbf{y}) = \frac{\rho(\mathbf{y}|\mathbf{x})\,\rho(\mathbf{x})}{\rho(\mathbf{y})},$$

where

$$\rho(\mathbf{y}) = \int_{-\infty}^{+\infty} \rho(\mathbf{y}|\mathbf{x})\,\rho(\mathbf{x})\,d\mathbf{x}$$

is just a normalization constant since $\mathbf{y}$ is a set of known quantities.

Typically, Monte Carlo methods are used to generate samples $\{\mathbf{x}_i\}$ from the distribution $\rho(\mathbf{x}|\mathbf{y})$.

To put this into the context of hybrid Monte Carlo methods, we introduce the potential $U(\mathbf{x})$ through

$$C \exp(-U(\mathbf{x})) = \rho(\mathbf{x}|\mathbf{y})$$

and the "guide Hamiltonian"

$$\mathcal{H} = \frac{1}{2}\pi \cdot \left[M^{-1}\pi\right] + U(\mathbf{x}).$$

The "mass matrix" $M$ should be used as a "preconditioner" to enhance sampling.

One could think, of course, of more complex kinetic energy functions $K(\pi)$.

In the context of data assimilation (3D and 4D Var) the *prior* is provided by the last forecast and the additional information comes from observations.