Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Computational Physics 200 (2004) 581–604
JOURNAL OF COMPUTATIONAL PHYSICS
www.elsevier.com/locate/jcp

Shadow hybrid Monte Carlo: an efficient propagator in phase space of macromolecules

Jesús A. Izaguirre *, Scott S. Hampton

Department of Computer Science and Engineering, University of Notre Dame, 326C Cushing Hall, Notre Dame, IN 46556, USA

Received 29 December 2003; received in revised form 19 April 2004; accepted 20 April 2004
Available online 7 June 2004

Abstract

Shadow hybrid Monte Carlo (SHMC) is a new method for sampling the phase space of large molecules, particularly biological molecules. It improves sampling of hybrid Monte Carlo (HMC) by allowing larger time steps and system sizes in the molecular dynamics (MD) step. The acceptance rate of HMC decreases exponentially with increasing system size N or time step δt. This is due to discretization errors introduced by the numerical integrator. SHMC achieves an asymptotic $O(N^{1/4})$ speedup over HMC by sampling from all of phase space using high order approximations to a shadow or modified Hamiltonian exactly integrated by a symplectic MD integrator. SHMC satisfies microscopic reversibility and is a rigorous sampling method. SHMC requires extra storage, modest computational overhead, and a reweighting step to obtain averages from the canonical ensemble. This is validated by numerical experiments that compute observables for different molecules, ranging from a small n-alkane butane with four united atoms to a larger solvated protein with 14,281 atoms. In these experiments, SHMC achieves an order magnitude speedup in sampling efficiency for medium sized proteins. Sampling efficiency is measured by monitoring the rate at which different conformations of the molecules' dihedral angles are visited, and by computing ergodic measures of some observables.

Keywords: Sampling methods; Hybrid Monte Carlo; Symplectic integrator; Modified Hamiltonian; Conformational sampling

1. Introduction

The sampling of the configuration space of complex biological molecules is an important and formidable problem. One major difficulty is the high dimensionality of this space, roughly $3N$, with the number of atoms N typically in the thousands. Other difficulties include the presence of multiple time and length scales, and the rugged energy hyper-surfaces that make trapping in local minima common, cf. [1].

* Corresponding author.
E-mail addresses: izaguirr@cse.nd.edu (J.A. Izaguirre), shampton@cse.nd.edu (S.S. Hampton).

0021-9991/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jcp.2004.04.016

This paper [1] introduces shadow hybrid Monte Carlo (SHMC), a propagator through phase space that enhances the scaling of hybrid Monte Carlo (HMC) with space dimensionality.

Sampling of configuration space can be done with Markov chain Monte Carlo (MC) methods or using molecular dynamics (MD). MC methods are rigorous sampling techniques. However, their application for sampling large biological molecules is limited because of the difficulty of specifying good moves for dense systems [2] and the large cost of computing the long range electrostatic energy, cf. [3, p. 380]. MD, on the other hand, can be readily applied, and it enables relatively large steps in phase space as well as allowing global updates of all the positions and momenta in the system. Nevertheless, the numerical implementation of MD introduces a bias due to finite step size in the numerical integrator of the equations of motion.

The HMC, introduced in [4], uses MD to generate a global MC move and then uses the Metropolis criterion to accept or reject the move. HMC rigorously samples the canonical distribution and eliminates the bias of MD due to finite step size. It is sufficient that the numerical integrator for MD be reversible and preserve volume in phase space to ensure detailed balance.

Unfortunately, the acceptance rate of HMC decreases exponentially with increasing system size N or time step δt. This is due to discretization errors introduced by the numerical integrator, cf. [5,6]. These errors can be reduced by using higher order integrators for the MD step. This was attempted for lattice-gauge simulations in [7], but with disappointing results. Furthermore, higher order integrators are not an efficient alternative for MD because the evaluation of the force is very expensive, and these integrators typically require more than one force evaluation per step.

SHMC is a biased variation on HMC. It uses a highly accurate approximation to the shadow Hamiltonian to sample more efficiently through phase space. The shadow Hamiltonian is exactly conserved by the numerical integrator, and a cheap and arbitrarily accurate approximation has been proposed in [8]. SHMC samples a noncanonical distribution defined by high order approximations to the shadow Hamiltonian, which greatly increases the acceptance rate of the method. A reweighting of the observables is performed in order to obtain proper canonical averages, thus eliminating the bias introduced by the shadow Hamiltonian. The overhead introduced by the method is modest in terms of time, involving only dot products of the history of positions and momenta generated by the integrator. There is moderate extra storage to keep this history.

In this paper, SHMC is derived, along with: (i) a proof that it preserves microscopic reversibility, which makes it a rigorous sampling method; (ii) an analysis of the asymptotic speedup of SHMC over HMC, which is shown to be $\Theta(N^{1/4})$ when using Verlet/Leapfrog or r-RESPA/Impulse as the integrator; and (iii) results of evaluating correctness and efficiency of sampling in a number of molecular systems: butane, decalanine, and BPTI, ranging in size from 4 to 14,281 atoms. Correctness is evaluated by computing: (i) the average torsion energy for butane and (ii) the average potential energy for decalanine and BPTI. Efficiency of sampling for the molecular systems is measured by the computer time per new conformation visited. This depends on both the acceptance rate of HMC and the number of new conformations discovered per simulation. Ergodic measures for the potential energy are also computed.

1.1. Sampling problem

The problem of sampling can be thought of as estimating expectation values for a function $A(\Gamma)$ with respect to a probability distribution function (p.d.f.) $\rho(\Gamma)$, where $\Gamma = \begin{pmatrix} X \\ P \end{pmatrix}$, and X and P are the vectors of collective positions and momenta. For the case of continuous components of Γ,

$$\langle A(\Gamma)\rangle_\rho = \int A(\Gamma)\rho(\Gamma)\mathrm{d}\Gamma. \tag{1}$$

[1] Work similar in scope to this paper appears in the proceedings of the International Conference on Computational Science (ICCS'04).

Examples of observables A are potential energy, pressure, free energy, and distribution of solvent molecules in vacancies [3,9]. For the sampling of configuration space of biological molecules, ρ typically corresponds to a constant temperature T and volume V ensemble (canonical ensemble),

$$\rho_{\mathrm{NVT}}(\Gamma) = \frac{\exp(-\mathscr{H}(\Gamma)/(k_{\mathrm{B}}T))}{\int \exp(-\mathscr{H}(\Gamma)/(k_{\mathrm{B}}T))\mathrm{d}\Gamma}, \tag{2}$$

such that $\mathscr{H}$ is the Hamiltonian or total energy of the system, k_{B} is Boltzmann's constant, and T is temperature.

2. Derivation of shadow hybrid Monte Carlo

The derivation of SHMC follows a discussion of MD, Markov chains, and HMC. The structure of the proofs for microscopic reversibility of SHMC are modeled after [10].

2.1. Molecular dynamics as a sampling method

MD is an important sampling method for biomolecules. It can be readily applied as long as one has a "force field" description of all the atoms and interactions among atoms in a molecule. MD finds changes over time in conformations of the molecule. Conformations are semi-stable geometric configurations, and are defined more precisely in Section 4.3.2.

MD typically solves Newton's equations of motion, a Hamiltonian system of equations,

$$\dot{\gamma}(t) = J\mathscr{H}_{\gamma}(\gamma(t)), \quad J = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}, \quad \gamma(t) = \begin{bmatrix} x(t) \\ p(t) \end{bmatrix}, \tag{3}$$

with a Hamiltonian

$$\mathscr{H}(x,p) = \frac{1}{2}p^{\mathrm{T}}M^{-1}p + U(x), \tag{4}$$

where M is a diagonal matrix of masses, $U(x)$ is the potential energy of the system, and $p = M\dot{x}$ are the momenta. Here, we use a lower case γ to denote a variable, whereas later on we specify a particular value of γ by using an upper case, Γ. This notation is also used for the algorithmic descriptions in the following sections.

Eq. (3) can be rewritten as

$$\dot{x}(t) = M^{-1}p(t), \quad \dot{p}(t) = F(x(t)), \tag{5}$$

where the conservative forces $F(x(t)) = -\nabla U(x(t))$.

Numerical integrators for MD generate a solution $\Gamma^n \approx \Gamma(n\,\delta t)$, where the step size or time step used in the discretization is δt. Typical integrators can be expressed as

$$\Gamma^{n+1} = \Psi(\Gamma^n). \tag{6}$$

The implementation of SHMC in this paper uses the Verlet or Leapfrog discretization of Eq. (5). This can be written as

$$x^{n+1} = x^n + \delta t M^{-1} p^n - \frac{1}{2}\delta t^2 M^{-1} F(x^n),$$
$$p^{n+1} = p^n - \frac{1}{2}\delta t\big(F(x^n) + F(x^{n+1})\big). \tag{7}$$

If the system is ergodic (cf. Section 2.2), modifications to the equations of motion for MD can give the correct averages for sampling from ρ_{NVT}, as for example Nosé thermostat methods [11]. In any case, the use of a finite time step δt introduces a bias in the estimate of averages using MD.

2.2. Markov chain Monte Carlo

In contrast to MD, Markov chain Monte Carlo methods are rigorous sampling methods. Let $\rho_{\mathrm{x}}(x)$ be the desired probability density function (p.d.f.) for configuration space. Sampling from $\rho_{\mathrm{x}}(x)$ may be performed by simulating a Markov chain Monte Carlo. Moves are suggested according to $\rho_{\mathrm{s}}(x'|x)$, the conditional p.d.f. for the new configuration x' given that the previous configuration is x.

Algorithm 1. Markov chain Monte Carlo

Given X:

(1) Generate X'

(2) Accept X' with probability

$$\min\left\{1, \frac{\rho_{\mathrm{x}}(X')\rho_{\mathrm{s}}(X|X')}{\rho_{\mathrm{x}}(X)\rho_{\mathrm{s}}(X'|X)}\right\}.$$

(3) If rejected, choose X.

Algorithm 1 is a generalization of Metropolis Monte Carlo to nonsymmetric proposal distributions, cf. [12,13, p. 129].

Definition 1. Microscopic reversibility means that

$$\rho_{\mathrm{s}}(x|x')\rho_{\mathrm{x}}(x') = \rho_{\mathrm{s}}(x'|x)\rho_{\mathrm{x}}(x),$$

cf. [14, p. 116].

2.3. Hybrid Monte Carlo

HMC combines an MD trajectory with an MC rejection step. It takes advantage of the long steps in phase space that can be achieved through MD. It also eliminates the inaccuracies due to a finite time step and other numerical artifacts through the MC step.

Let $\gamma = \begin{pmatrix} x \\ p \end{pmatrix}$, and $\Psi = \begin{pmatrix} \Psi_1 \\ \Psi_2 \end{pmatrix}$ be an MD integrator that propagates γ through phase space, that is, $\gamma^{n+1} = \Psi(\gamma^n)$. Assume that the integrator preserves phase space volume, $\det \partial_\gamma \Psi(\gamma) = 1$, and is reversible, $\Psi^{-1}(\gamma) = R\Psi(R\gamma)$, where $R = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}$. Then HMC, given by Algorithm 2, satisfies microscopic reversibility in configuration space [10].

Algorithm 2. Hybrid Monte Carlo (HMC)

(1) **MC step**: Generate P from $\rho_{\mathrm{p}}(p)$

(2) **MD step**: Given (X,P):
(a) $(X',P') = \Psi(X,P)$
(b) Accept X' with probability

$$\min\left\{1, \frac{\rho(\Psi(X,P))}{\rho(X,P)}\right\},$$

where $\rho(x,p) = \rho_x(x)\rho_p(p)$
(c) If rejected, choose X.

2.4. Shadow hybrid Monte Carlo

HMC's performance degrades when δt and N grow. This is due to errors in the energy, which increase with δt and N and cause an extremely high rejection rate, cf. Section 2.5. SHMC is a generalization of HMC, sampling is in all of phase space rather than configuration space alone.

Let $\tilde{\rho}(x,p)$ be the target density of SHMC, where

$$\tilde{\rho}(x,p) \propto \exp\Big(-\beta\tilde{\mathscr{H}}(x,p)\Big), \tag{8}$$

$$\tilde{\mathscr{H}}(x,p) = \max\left\{\mathscr{H}(x,p), \mathscr{H}_{[2k]}(x,p) - c\right\}. \tag{9}$$

Here, $\mathscr{H}_{[2k]}(x,p)$ is the highly accurate shadow Hamiltonian, defined in Section 3.1, and c is an arbitrary constant that limits the amount by which $\mathscr{H}_{[2k]}$ is allowed to depart from $\mathscr{H}$. It is assumed that $\tilde{\mathscr{H}}(x,Rp) = \tilde{\mathscr{H}}(x,p)$.

Algorithm 3 lists the steps required in calculating SHMC. First, a new set of momenta P are generated, typically from a Gaussian distribution. The new momenta are accepted based on a Metropolis acceptance step relative to the difference between the total and shadow energies. This is repeated until a set of momenta are accepted. The number of attempts required for accepting the momenta can be controlled by the parameter c, which is discussed further in the following paragraph and in Section 4.3.6. Next, the system is integrated using MD and accepted with probability proportional to Eq. (8). Finally, in order to calculate unbiased values, the observables must be reweighted. Even though the reweighting is listed as part of the algorithm, it is in fact a postprocessing step.

Algorithm 3. Shadow hybrid Monte Carlo (SHMC)
(1) **MC step**: Given $X = x$, generate P with p.d.f. $\tilde{\rho}(X,p)$, using the acceptance–rejection method:
(a) Generate P having p.d.f. $\rho_p(p)$
(b) Accept with probability

$$\min\left\{1, \frac{\exp(-\beta(\mathscr{H}_{[2k]}(X,P) - c))}{\exp(-\beta\mathscr{H}(X,P))}\right\}. \tag{10}$$

(c) Repeat (1a)–(1b) until P is accepted.
(2) **MD step**: Given $\Gamma = \begin{pmatrix} X \\ P \end{pmatrix}$,
(a) $\Gamma' = R\Psi(\Gamma)$ (where Ψ nearly conserves $\mathscr{H}_{[2k]}$)
(b) Accept Γ' with probability

$$\min\left\{1, \frac{\tilde{\rho}(\Gamma')}{\tilde{\rho}(\Gamma)}\right\}.$$

(c) If rejected, choose Γ.

Reweighting Post-processing: Given sequence of $\{A, \Gamma\}$, reweight observable A using $\rho(\Gamma)/\tilde{\rho}(\Gamma)$ before computing averages.

For example, to obtain proper canonical distributions:

$$\langle A \rangle_{\rho\,\mathrm{NVT}} = \frac{\sum_{i=1}^{m} w_i A_i}{\sum_{i=1}^{m} w_i},$$

where

$$w_i = \frac{\exp(-\beta \mathscr{H}(\Gamma_i))}{\exp\left(-\beta \tilde{\mathscr{H}}(\Gamma_i)\right)}.$$

SHMC's parameter c appears in the momenta and position acceptance equations and is also used during the reweighting step. The effect of c on the simulation is readily apparent. For large positive c, SHMC is equivalent to HMC with a different set of momenta. In this HMC-like regime, the acceptance rate of the MD step is low and that of the MC step is high. Also, the deviation of the reweighted values is small. Conversely, for large negative c, the acceptance rate of the MD step is high while the acceptance of the MC step is low. In this regime, there is large variation in the reweighted values. Currently, the value of c is chosen proportional to the average difference between the Hamiltonian and the shadow Hamiltonian. A small number of simulation steps gives a good approximation to this parameter, see Section 4.3.6.

Let $\Delta\mathscr{H} = \mathscr{H}_{[2k]} - \mathscr{H}$. Experiments suggest that $\Delta\mathscr{H}$ is predominantly positive in MD simulations. Fig. 1 shows a typical plot of the total energy of a 66-atom decalanine molecule simulated using Leapfrog. In addition, the 4th and 8th order shadow Hamiltonians are also shown. The accuracy of the shadow Hamiltonians is apparent.

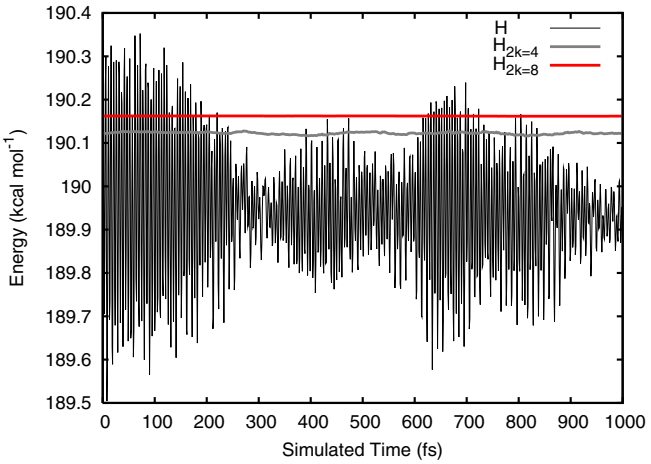

Fig. 1. Total energy, 4th and 8th order shadow Hamiltonians for a simulated decalanine.

Proposition 1. *SHMC satisfies microscopic reversibility.*

Proof. Both the MD and MC steps separately satisfy microscopic reversibility, as is shown in the next two propositions. Thus, SHMC, which is the combination of the MD and the MC steps, also satisfies it. □

Proposition 2. *The MD step of SHMC satisfies microscopic reversibility.*

Proof. Sufficient to consider $\gamma' \neq \gamma$. Let π_{AB} be the probability of $\Gamma \in A$ and $\Gamma' \in B$, assuming Γ has p.d.f. $\tilde{\rho}(\gamma)$. Then

$$\pi_{AB} = \int_A \int_B \tilde{\rho}(\gamma'|\gamma)\tilde{\rho}(\gamma)\mathrm{d}\gamma'\,\mathrm{d}\gamma.$$

It suffices to show that $\pi_{AB} = \pi_{BA}$ for arbitrary $A \cap B = \phi$. Now,

$$\pi_{AB} = \int\int \chi_A(\gamma)\chi_B(R\Psi(\gamma)) \min\left\{1, \frac{\tilde{\rho}(\Psi(\gamma))}{\tilde{\rho}(\gamma)}\right\}\tilde{\rho}(\gamma)\mathrm{d}\gamma,$$

where χ_A and χ_B are indicator, or characteristic, functions of sets A and B, and then

$$\pi_{AB} = \int\int \chi_A(\gamma)\chi_B(R\Psi(\gamma)) \min\left\{\tilde{\rho}(\gamma), \tilde{\rho}(\Psi(\gamma))\right\}\mathrm{d}\gamma.$$

Replacing γ by $\Psi^{-1}(\gamma) = R\Psi(R\gamma)$:

$$\pi_{AB} = \int\int \chi_A(R\Psi(R\gamma))\chi_B(R\gamma) \min\left\{\tilde{\rho}(R\Psi(R\gamma)), \tilde{\rho}(\gamma)\right\}\mathrm{d}\gamma.$$

Finally, replacing γ by $R\gamma$:

$$\pi_{AB} = \int\int \chi_A(R\Psi(\gamma))\chi_B(\gamma) \min\left\{\tilde{\rho}(\gamma), \tilde{\rho}(\Psi(\gamma))\right\}\mathrm{d}\gamma = \pi_{BA.} \quad \square$$

Proposition 3. *The MC step generates* $\tilde{\rho}(x,p)$.

Proof. The main result is due to von Neumann ([15]; as quoted in [14, p. 349]; cf. [16, pp. 171 ff]). The acceptance–rejection method generates a random number with a complicated p.d.f. $f(z)$. The distribution function is split as $f(z) = Cg(z)h(z)$, where $h(z)$ is a simple p.d.f., $C \geqslant 1$, and $0 \leqslant g(z) \leqslant 1$. A random variable Z with p.d.f. $h(z)$ is generated; then a uniform random number U from (0, 1) is generated. If $U \leqslant g(Z)$, then Z has the p.d.f $f(z)$. Otherwise, repeat the process.

Let $z = P$, then

$$f(z) = \tilde{\rho}(x,p) = \rho_{\mathrm{x}}(x)\rho_{\mathrm{p}}(p) \min\left\{1, \exp\left(-\beta(\mathscr{H}_{[2k]}(x,p) - c - \mathscr{H}(x,p))\right)\right\}.$$

Here, $C = \rho_{\mathrm{x}}(x)$, $h(z) = \rho_{\mathrm{p}}(p)$, e.g., the Gaussian distribution of velocities in HMC, and $g(z) = \min\{1, \exp(-\beta(\mathscr{H}_{[2k]}(x,p) - c - \mathscr{H}(x,p)))\}$. □

Proposition 4. *The MC step of SHMC satisfies microscopic reversibility.*

Proof. Since previous P' and P are independent, the probabilities of going from one to the other are clearly symmetric. □

2.5. Theoretical performance of SHMC

The cost of HMC as a function of system size N and time step δt has been investigated in [5,10,17]. HMC drives (x,p) towards an equilibrium with a coupled probability

$$\rho(x,p) = \rho_x(x)\rho_p(p) \propto \exp(-\beta \mathscr{H}(x,p)).$$

Using the property of volume preservation, the expected value over this distribution is

$$\langle \exp(-\beta \delta \mathscr{H}(x,p)) \rangle_\rho = 1, \tag{11}$$

where the discretization error $\delta\mathscr{H} = \mathscr{H}(x',p') - \mathscr{H}(x,p)$. Taking the log of both sides of Eq. (11):

$$\log \langle \exp(-\beta \delta \mathscr{H}(x,p)) \rangle_\rho = 0.$$

Henceforth the parameters of the Hamiltonian $\mathscr{H}$ and the density ρ are omitted. Since exp is a convex function, $\langle \delta\mathscr{H} \rangle \geqslant 0$, with equality possible only if the MD integrator Ψ exactly conserves energy. For small $\delta\mathscr{H}$, the expansion of Eq. (11) up to a third order term is

$$\langle \exp(-\beta \delta \mathscr{H}) \rangle = 1 - \beta \langle \delta \mathscr{H} \rangle + \frac{\beta^2}{2} \langle \delta \mathscr{H}^2 \rangle + \mathrm{O}(\delta \mathscr{H}^3) = 1.$$

Thus,

$$\langle \delta \mathscr{H} \rangle = \frac{\beta}{2} \langle \delta \mathscr{H}^2 \rangle + \mathrm{O}(\delta \mathscr{H}^3). \tag{12}$$

What is the expected value of the probability of acceptance in HMC? In order to obtain a nonzero average acceptance probability in the limit $N \to \infty$, $\delta t \to 0$. In this limit, the distribution of the average discretization error becomes Gaussian, since the higher order terms vanish. Substituting this result in the definition of the acceptance probability, one gets that

$$P_A = \mathrm{erfc}\left(\tfrac{1}{2}\sqrt{\beta \langle \delta \mathscr{H} \rangle}\right).$$

For an MD integrator Ψ that is $\mathrm{O}(\delta t^m)$ accurate, Eq. (12) means that

$$\langle \delta \mathscr{H} \rangle = \mathrm{O}(N \delta t^{2m}),$$

concretely, $\delta t \propto N^{-1/(2m)}$.

Let L be the MD trajectory length needed to produce an uncorrelated sample. Assuming L is fixed, the cost of producing uncorrelated samples increases as

$$C_{\mathrm{MD}} N^{1/(2m)} + C_{\mathrm{MC}}, \tag{13}$$

where $N^{1/(2m)}$ is the number of MD steps to achieve a trajectory of length L. C_{MD} is the cost of each MD step, which depends on the cost of the force evaluation, and will be anywhere from $\Theta(N)$ for cutoff computation to $\Theta(N^2)$ for all pairs evaluation, or more typically $\Theta(N \log N)$ for tree methods or FFT-based methods. C_{MC} is the cost of generating P, basically the generation of $\Theta(N)$ random numbers from a Gaussian distribution.

A similar argument can be made for the acceptance rate of SHMC, but now with respect to the equilibrium p.d.f. $\tilde{\rho}$. Thus, $\langle \delta \tilde{\mathscr{H}} \rangle = \mathrm{O}(\delta t^{2p})$, where $p = 2k$ is the order of the shadow Hamiltonian $\mathscr{H}_{[2k]}$, and the cost of producing uncorrelated samples increases as

$$S_{\mathrm{MD}} N^{1/(2p)} + S_{\mathrm{MC}}. \tag{14}$$

The asymptotic speedup of SHMC over HMC is given by the quotient of Eqs. (13) and (14). Note that $S_{\mathrm{MD}} \gg S_{\mathrm{MC}}$ and $C_{\mathrm{MD}} \gg C_{\mathrm{MC}}$. However, $S_{\mathrm{MC}} \geqslant C_{\mathrm{MC}}$ due to the additional rejection step for the momenta in SHMC. The asymptotic speedup is given by

$$\eta = \frac{N^{1/(2m)}}{N^{1/(2p)}} = N^{\frac{1}{2}\frac{p-m}{mp}}, \tag{15}$$

and becomes $\mathrm{O}(N^{1/4})$ for high order shadow Hamiltonians.

3. Implementation of SHMC

Although SHMC may seem complicated, the complications are more in the formulas than in the logic, so it is not difficult to program. Pseudocode for SHMC is in Appendix A, and an explanation of this code follows.

3.1. Approximation to the shadow Hamiltonian

In the case of a Hamiltonian system, Eq. (3), the modified equation for an integrator is Hamiltonian if and only if the integrator is symplectic, cf. [18, p. 129–136]. The integrator is symplectic if $\partial_\gamma \Psi(\gamma)^{\mathrm{T}} J \partial_\gamma \Psi(\gamma) \equiv J$. There is also evidence that the numerical solution of symplectic integrators stays close to the solution of a modified Hamiltonian $\mathscr{H}^{\delta t}(x,p)$ for very long times [19]. Work by Skeel and Hardy [8] shows how to compute an arbitrarily accurate approximation to the modified Hamiltonian integrated by symplectic integrators based on splitting. The goal is to compute

$$\mathscr{H}_{[2k]}(x,p) = \mathscr{H}^{\delta t}(x,p) + \mathrm{O}(\delta t^{2k}), \tag{16}$$

where $\mathscr{H}_{[2k]}$ is a shadow Hamiltonian of order $2k$.

The construction adds a new position variable and a conjugate momentum variable β to get an extended Hamiltonian $\bar{\mathscr{H}}(y)$ which is homogeneous of order 2. For this Hamiltonian, $\bar{\mathscr{H}}(y) = \frac{1}{2}\dot{y}(t)^{\mathrm{T}} \bar{J} y(t)$. An approximation to $y(t)$ is formed using the numerical solution of the extended Hamiltonian system $\bar{\mathscr{H}}$. The resulting shadow Hamiltonian approximation satisfies Eq. (16).

The shadow Hamiltonian is a combination of trajectory information, that is, k copies of available positions and momenta generated by the MD integration, and an extra degree of freedom β that is propagated along with the momenta. By construction, $\mathscr{H}_{[2k]}$ is exact for quadratic Hamiltonians, which is useful, because harmonic motion dominates applications such as MD. Details can be found in the original reference.

Formulae for the 4th and 8th shadow Hamiltonians, $k = 2$ and $k = 4$ respectively, follow:

$$\mathscr{H}_{[4]} = \frac{1}{2\delta t}\left(\mathrm{A}_{10} - \frac{1}{6}\mathrm{A}_{12}\right), \tag{17}$$

$$\mathscr{H}_{[8]} = \frac{1}{2\delta t}\left(\mathrm{A}_{10} - \frac{2}{7}\mathrm{A}_{12} - \frac{19}{210}\mathrm{A}_{14} + \frac{5}{42}\mathrm{A}_{30} + \frac{13}{105}\mathrm{A}_{32} - \frac{1}{140}\mathrm{A}_{34}\right). \tag{18}$$

Define the ith centered difference formula to be $\delta^{[i]}\omega^n$. So, for example, $\delta^{[2]}x^n$ would represent the 2nd centered difference of the positions:

$$\delta^2 x^n = x^{n+1} - 2x^n + x^{n-1}. \tag{19}$$

The averaging operator is similarly defined by $\mu\omega^n = \frac{1}{2}w^{n+1/2} + \frac{1}{2}w^{n-1/2}$.

Now, define A_{ij}, where i is odd and j is even, as follows:

$$A_{ij} = \begin{cases} \mu\delta^{[i]}x^n \cdot \delta^{[j]}p^nM - \delta^{[j]}x^n \cdot \mu\delta^{[i]}p^nM - \mu\delta^{[i]}\beta^n & : \quad j = 0 \\ \mu\delta^{[i]}x^n \cdot \delta^{[j]}p^nM - \delta^{[j]}x^n \cdot \mu\delta^{[i]}p^nM & : \quad j \neq 0 \end{cases} \tag{20}$$

Finally, the β term propagated by Leapfrog is

$$\beta^{n+1} := \beta^n - \delta t(x^n \cdot F^n + 2U(x^n)). \tag{21}$$

3.2. SHMC implementation details

SHMC is implemented inside of an existing framework for MD simulations, ProtoMol [20,21]. However, SHMC is unlike any of the existing integrators due to the fact that it requires values both 'forward' and 'backward' in time to properly calculate it at the current time step. In order to simulate calculating forward elements, the simulation is run and values are saved as needed. When enough steps have been saved, the shadow is computed and stored along with the data corresponding to the step at which it was to be calculated. The shadow Hamiltonian only needs to be computed at the beginning and end of the MD step, which is typically hundreds of time steps long. The history of values needed by SHMC are stored in a data structure known as a *deque*. A deque is a double ended queue, which means elements can be accessed from both the beginning and the end of the queue. A deque was chosen since values must occasionally be inserted or removed from both ends, depending on whether the simulation is going forwards or backwards in time.

After saving the current state of the system, momenta are randomly generated with a Gaussian distribution proportional to the temperature of the system. The total energy, $\mathscr{H}(\Gamma)$, is calculated and stored. In order to calculate $\mathscr{H}_{[2k]}(\Gamma)$, $k/2$ "past" and $k/2$ "future" values of the state of the system are required. At the beginning of the MD step, a backwards simulation is run for $k/2$ steps and the history needed by SHMC is stored at each step. It is now necessary to restore the system to the state in which it originally existed with the new momenta. The simulation is run forward for $k/2$ steps and again the data needed by SHMC is stored. It is now possible to compute $\mathscr{H}_{[2k]}(\Gamma')$ and determine whether or not this new set of momenta are accepted. If the momenta are accepted, the simulation resumes from time step $k/2$. If the momenta are not accepted then restore the complete system to the state in which it existed before the new momenta were generated. This process is repeated until a new set of momenta are accepted. The number of required repetitions can be reduced by increasing c.

At the end of the MD step, the state of the system is saved and then run for an additional $k/2$ steps. Once the data needed to calculate the shadow is available, the system is restored back to its state at time n_{md}. If the new positions are accepted then this cycle is complete. If the new positions are rejected, then the original positions must be restored. In addition, the forces and energies associated with these positions must also be returned to their original values.

4. Numerical tests

Testing is used to determine the validity of the following hypotheses:

(H1) SHMC samples with reasonable efficiency as the system size N and the time step δt are increased. Slight performance degradation is expected as N increases due to greater truncation and round off error. Similarly, as δt approaches its upper bound, set both by instabilities in the MD integrator

and the fastest motions of the system, performance is affected. However, these effects are negligible for reasonable values. These results are in Section 4.3.5.

(H2) SHMC has an asymptotic speedup over HMC of $\Theta(N^{\frac{1}{4}})$. Experimental results in Section 4.3.5 agree with the analysis of Section 2.5.

(H3) The overhead of computing the shadow Hamiltonian is small for values of L needed to produce uncorrelated samples, where $L = n_{\mathrm{md}} * \delta t$.

(H4) Any bias introduced by sampling from the shadow Hamiltonian is removed by a re-weighting of sampled values, cf. Algorithm 3. Reweighted averages for potential energies of molecular systems tested are indistinguishable from correct estimates obtained using HMC, cf. Section 4.3.3.

4.1. Test systems

SHMC was tested on molecular systems ranging from the simple alkane *n*-butane with only four (united) atoms to a more complex solvated protein, BPTI, with 14,281 atoms. While *n*-butane is a simple molecule, it is a common test case because results can be verified analytically [22]. Table 1 lists the test molecules, the corresponding number of atoms, and the length of L, the length of each MC step.

Testing was done on a 44 node Linux cluster administered by the Department of Computer Science and Engineering at the University of Notre Dame. Each node contains 2, 2.4 GHz Xeon processors and 1 GB RDRAM. There is nearly a terabyte of storage as well as fast Ethernet and Myrinet networking.

4.2. Method parameters

HMC and SHMC have several parameters that affect their performance, including the random number generator, the integrator chosen for the MD, and integrator parameters: time step δt and trajectory length L. SHMC also needs a tuning parameter c to indicate allowed divergence between the shadow and the total energy. HMC and SHMC currently use the random number generator `drand48()`, although methods with longer periods may be preferable, cf. [3, p. 355].

Any time reversible and volume preserving integrator can be used for HMC. SHMC requires in addition that the integrator be symplectic in order to compute the shadow Hamiltonian. In this work, implementations of HMC and SHMC use the Verlet/Leapfrog discretization of Eq. (7), which satisfies the constraints for both propagators.

The choices of L and δt have dramatic effects on performance of HMC and SHMC. L should be long enough so that the longest correlation times of interest are sampled during an MD step, thus avoiding the random walk behavior of MC. One way of approximating the correlation times in a molecular system is to compute the normal modes through a linearization of the interaction forces of interest, and take the maximum period $\tau_{\max}$ as a desirable value for L. Ideally, δt should be close to $\tau_{\min}$, and thus $n_{\mathrm{md}} \equiv \tau_{\max}/\tau_{\min}$. In practice, however, numerical artifacts like instability and resonance force $\delta t \ll \tau_{\min}$. Based on suggestions in [23], a value of n_{md} is chosen from a distribution $[0.7L/\delta t, 1.3L/\delta t]$. The maximum time step δt is fixed by stability limits of Verlet/Leapfrog. The efficiency of the methods are compared for fixed expected

Table 1
Listing of molecules used for testing SHMC

Molecule	Number of atoms	L (fs)
n-butane	4	630
Decalanine	66	300
Solvated BPTI	1101	42
Solvated BPTI	14,281	15

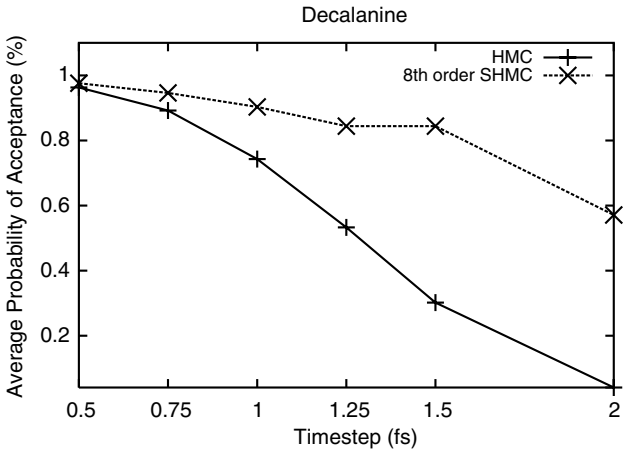

Fig. 2. The average probability of accepting the MD step for a 66-atom decalanine.

value of the trajectory L, such that $E[L] = L$. This way, direct comparisons can be made between simulations with different δt.

4.3. Test metrics

Several metrics have been used to test SHMC against HMC. The efficiency of sampling is measured by computing the cost to generate a new geometric conformation. The statistical error is measured by computing several observables using both methods. Additionally, the acceptance rate and the conformation count of the methods is computed to understand the method's behavior.

4.3.1. Acceptance rate

The acceptance rate (AR) is the ratio of accepted moves to total moves attempted in Algorithm 2. The best AR is the one that samples configuration space more quickly, cf. [3, p. 376].

Fig. 2 shows the average probability of accepting a step based on the time step for HMC and SHMC. Both methods decay with increasing time step, but HMC does so much faster than SHMC.

4.3.2. Conformations

The number of molecular conformations visited by HMC and SHMC is determined using a method suggested in [24,25]. This method monitors the molecules' dihedral angles. In a preprocessing step, each dihedral angle is divided into wells. Divisions between wells are placed at the maxima of the dihedral potential energy. At each HMC or SHMC step, every angle is assigned an integer corresponding to the well currently occupied by the angle. This results in a string that identifies the conformation. A conformation is considered a unique structure if an equivalent string has not yet been recorded. Angles that involve hydrogens are not considered in the analysis. One feature of this method is that it can be generalized to all dihedral angles. More importantly, it is independent of the propagator used to actually generate new conformations. This method is justified by the fact that once a dihedral angle is within the well formed between two local maxima, the natural behavior of the angle is to approach the local minimum.

As the number of atoms increases, the possible states of these strings increases exponentially. Even a small protein such as BPTI has more than 2000 dihedral angles. The majority of these contain hydrogen atoms and are immediately discounted from analysis due to the fast motion of the hydrogen. Even so, there

are still hundreds of dihedrals each with multiple wells. With roughly 200 dihedrals from BPTI, say with three wells each, there would be a state space of 3^{200} different strings. In practice, there are many non-physical states (e.g., overlapping configurations) and inaccessible states due to high local energy barriers. The rugged nature of the energy hyper-surface can not be eliminated solely by efficient propagators, but requires methods such as potential smoothing or multi-canonical ensembles, cf. [1].

A dihedral (torsion) angle occurs when, within a molecule, there are four atoms connected linearly end-to-end. The first three atoms form a plane, as do the last three atoms. The dihedral angle is the angle between the two planes. [2] The potential energy for a single dihedral angle is defined as follows:

$$U^{\mathrm{dih}}(\phi) = \sum_{i=1}^{m} \frac{1}{2} f_i (1 + \cos(n_i \phi - \delta_i)). \tag{22}$$

Associated with each term in the potential energy for a dihedral angle ϕ is a force constant f, a periodicity n, a phase-shift δ and a multiplicity m. Many dihedral angles are defined with $m = 1$, but there are some that can only be constructed as a combination of multiple terms.

Algorithm 4 is used to count the number of conformations discovered by the propagators. The pre-processing step enumerates the possible states or wells for the dihedrals. It finds all the local minima and maxima of the dihedral function using Brent's method. Thus, it is able to define the domain of each conformation. During simulation, the matching step generates a string identifying the conformation. If a new string is generated, it is stored in a set container, along with the time at which it was generated. The preprocessing step, if done on all the elements of a force field and stored, need only be done once.

Algorithm 4. Method for counting molecular conformations.

(1) **Preprocessing**:
   (a) Remove dihedrals containing H
   (b) Find maxima of Eq. (22) using Brent's method
   (c) If the phase shift is nonzero, shift the critical points accordingly
   (d) Enumerate wells for dihedrals

(2) **Matching**:
   (a) Determine which well each dihedral occupies
   (b) Form *conformation* string based on wells of dihedrals
   (c) Update counter and time step for conformation string.

4.3.3. Average dihedral energy

To test correctness of SHMC and the ability to use it as an unbiased estimator, the average dihedral energy for butane, and the average potential energy for decalanine and BPTI are computed. The average dihedral energy for butane can be obtained analytically at any temperature T, where $\beta = 1/k_{\mathrm{B}}T$, by

$$\left\langle U^{\mathrm{dih}}(\phi) \right\rangle_{\beta} = \frac{\int_0^{2\pi} U^{\mathrm{dih}}(\phi) \exp(-\beta U^{\mathrm{dih}}(\phi)) \mathrm{d}\phi}{\int_0^{2\pi} \exp(-\beta U^{\mathrm{dih}}(\phi)) \mathrm{d}\phi}. \tag{23}$$

The analytical result at $T = 300$ K is compared against averages obtained using HMC and SHMC. The parameters for the butane dihedral are those in CHARMM 22 [26]:

$$U^{\mathrm{dih}}(\phi) = 1.6(1 + \cos(3\phi - \pi)) + 0.6(1 + \cos(\phi - \pi)), \tag{24}$$

[2] An alternate definition has the dihedral defined to be the angle between the normals to the two planes.

where $U^{\mathrm{dih}}(\phi)$ has units of kcal mol^{-1} (1 kcal mol^{-1} = 4.1868 kJ mol^{-1}). Substituting Eq. (24) into Eq. (23), and evaluating this integral numerically for $\beta = 1/(k_{\mathrm{B}}300\ \mathrm{K})$, where $k_{\mathrm{B}} = 0.00198719$ kcal mol^{-1} K^{-1} gives

$$\left\langle U^{\mathrm{dih}}(\phi)\right\rangle_{\beta=1/(k_{\mathrm{B}}\,300\ \mathrm{K})} = 0.62848\ \mathrm{kcal\ mol^{-1}}. \tag{25}$$

This expectation value is correct for butane because the dihedral angle coordinate decouples from the rest of the internal coordinates (bonds and angles), and the functional determinant from Cartesian to internal motion is 1, cf. [22].

Simulations of four united-atom CH_3–CH_2–CH_2–CH_3 butane were performed. The mass of CH_3 is 15.035 u and CH_2 is 14.027 u. A total of 16 simulations of total length 114 ns at $T = 300$ K were run. Values for the expected MD trajectory length are $L = \{630, 450, 234, 72\ \mathrm{fs}\}$, where 630 fs $\approx \tau_{\max}$ for butane. Time steps $\delta t = \{8, 6, 3, 1\ \mathrm{fs}\}$, where 8 fs is close to the stability limit of Leapfrog for butane. The values computed from the simulations for $L = 630$ fs are shown in Table 2.

The error bar is estimated as the standard deviation computed using the block averaging method of Flyvbjerg and Petersen [27,p. 530]. It can be seen that all the values agree with the analytical result of Eq. (25).

4.3.4. Average potential energy

The average potential energy for BPTI is computed using CHARMM 22 parameters for proteins [28,29]. Table 3 shows the average potential energy (PE) for decalanine. There is no statistically significant difference among the values. All of the reweighted values are within at least one standard deviation of the unweighted HMC values. Additionally, the reweighted standard deviation is acceptable in all cases.

The values for solvated BPTI with 1101 atoms are listed in Table 4. Here, there is still good agreement between the reweighted PE and that of HMC. However, the standard deviation is not as close for the smaller time steps. A larger value of c would alleviate this but it would also decrease the probability of acceptance. It should be noted that the time steps considered here are rather large for the size of this input. HMC was unable to accept even a single move for $\delta t = 1$, which is why it has no value listed.

Finally, consider the PE of the solvated BPTI, with 14,281 atoms, in Table 5. These results are not quite as good when compared to the previous two molecules. This behavior is most likely due to a bad choice of c. Unfortunately, the time required to execute simulations at this size molecule is on the order of days even

Table 2
Expected value of the torsional energy U^{dih} for n-butane

δt (fs)	HMC ($L = 630$ fs)		SHMC ($L = 630$ fs)	
	$\langle U^{\mathrm{dih}}(\phi)\rangle$	AR	$\langle U^{\mathrm{dih}}(\phi)\rangle$	AR
1	0.62 ± 0.01	(100%)	0.64 ± 0.01	(100%)
3	0.63 ± 0.01	(96%)	0.63 ± 0.01	(100%)
6	0.63 ± 0.01	(79%)	0.64 ± 0.01	(100%)
8	0.65 ± 0.02	(51%)	0.67 ± 0.01	(99%)

Table 3
Average potential energy and standard deviation for decalanine for HMC and SHMC using the 8th order shadow Hamiltonian

Method	Time step					
	0.5	0.75	1.0	1.25	1.5	2
HMC	97.5 ± 6.5	97.4 ± 6.9	100 ± 6.6	99.8 ± 6.7	98.1 ± 7.1	97.4 ± 9.1
SHMC	103 ± 6.7	102 ± 7	96.8 ± 7.2	98.9 ± 6.8	97.3 ± 8	99.7 ± 8.4

The MD step trajectory length $L = 100$ fs.

Table 4
Potential energy and standard deviation for BPTI with 1101 atoms for HMC and SHMC using the 8th order shadow Hamiltonian

Method	Time step		
	0.25	0.5	1.0
HMC	-1590 ± 33	-1566 ± 21	–[a]
SHMC	-1583 ± 122	-1581 ± 30	-1562 ± 35

The MD step trajectory length $L = 42$ fs.
[a] HMC was unable to accept an adequate number of steps for this time step.

Table 5
Potential energy and standard deviation for BPTI with 14,281 atoms

Method	Time step		
	0.1	0.25	0.5
HMC	-40293 ± 227	-40080 ± 88	–[a]
8th	-40798 ± 2217	-40739 ± 290	-40520 ± 506

The MD step trajectory length $L = 15$ fs.
[a] HMC was unable to accept an adequate number of conformations for this time step.

for short runs. Therefore, this limits the amount of experimentation to determine an optimal value for c. Better ways to estimate c are being investigated.

4.3.5. Sampling efficiency

The sampling efficiency of HMC and SHMC is defined as the computational cost per new conformation. This value is calculated by dividing the running time of the simulation by the number of conformations discovered.

$$\text{Cost of sampling (CS)} = \frac{\text{Execution time (ET)}}{\text{Unique conformations (C)}}. \tag{26}$$

Comparisons of this value are only valid for simulations run using the same platform and the same input. However, this is a fair metric when comparing different sampling methods, since it takes care of the overhead of more sophisticated trial moves, and any other effects on the quality (or lack thereof, e.g., correlation) of samples produced by different sampling techniques.

In Fig. 3 we show the number of conformations per second as a function of the time step for a 66-atom decalanine. Lower values denote a better sampling efficiency. At its best, HMC is only as good as SHMC for one time step, $\delta t = 1$. In terms of efficiency, SHMC shows great speedup over HMC when the optimal values for both methods are used.

$$\text{Speedup} = \frac{\text{Best efficiency HMC}}{\text{Best efficiency SHMC}}. \tag{27}$$

Since the cost per conformation for SHMC still appears to be decreasing as the time step increases, it is possible that the speedup is larger. Figs. 4 and 5 show even more dramatic results for BPTI with 1101 and 14,281 atoms, respectively. This is expected, since the speedup increases asymptotically as $\mathcal{O}(N^{1/4})$.

Table 6 lists the results recorded as well as the theoretical asymptotic speedup for all three molecules. As can be seen from the table, there is good agreement for the smaller molecule and excellent agreement for the larger molecule. The BPTI containing 1101 atoms does not agree very well with the asymptotic speedup.

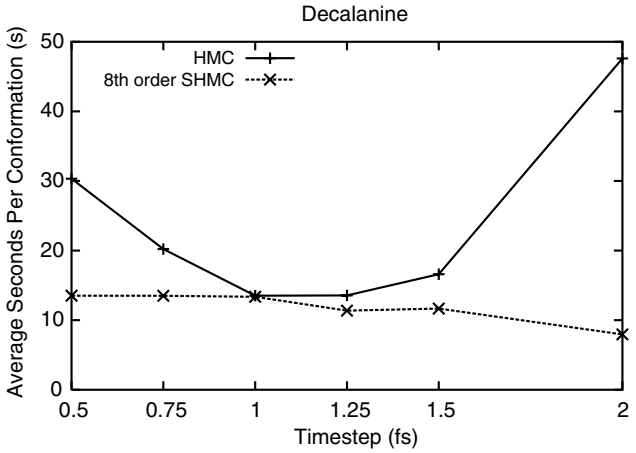

Fig. 3. Average computer time per discovered conformation for 66-atom decalanine.

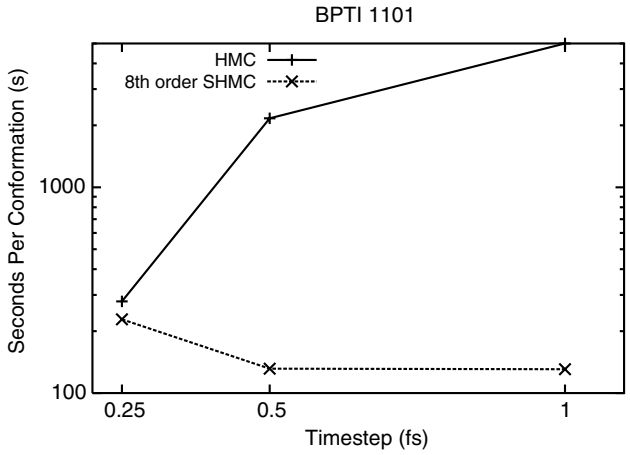

Fig. 4. Average computer time per discovered conformation for 1101-atom BPTI.

This could be due to several factors. Most likely, a poor choice of c was made when doing the initial runs. Additionally, these values were chosen such that the variance of the observed potential energy was minimized among all given runs. Even so, there is still a 2-fold increase in efficiency over standard HMC. BPTI with 14,281 atoms achieves a speedup of 10.

Zhou et al. [30,31] use a method that can approximately determine whether a sampling algorithm is ergodic. By studying how long it takes two independent trajectories to achieve self-averaging of a property, we can gauge the efficiency of the method. The ergodic measure for the potential energy of a system is defined by

$$\mathrm{d}u(t) = \left|\bar{u}^{\alpha}(t) - \bar{u}^{\beta}(t)\right|^2, \tag{28}$$

where $\bar{u} = \frac{1}{t}\int_0^t u(s)\mathrm{d}s$ is the average potential energy from time 0 to t, and α and β are independent trajectories. Fig. 6 shows the results for a pair of simulations of decalanine. In this graph we compare the results of HMC and SHMC for a time step of 0.5 fs. The value for the constant c is 0.6, a relatively large value for this size molecule. The y-axis is the value of $\mathrm{d}u(t)$ normalized by $\mathrm{d}u(0)$ and plotted with a log scale.

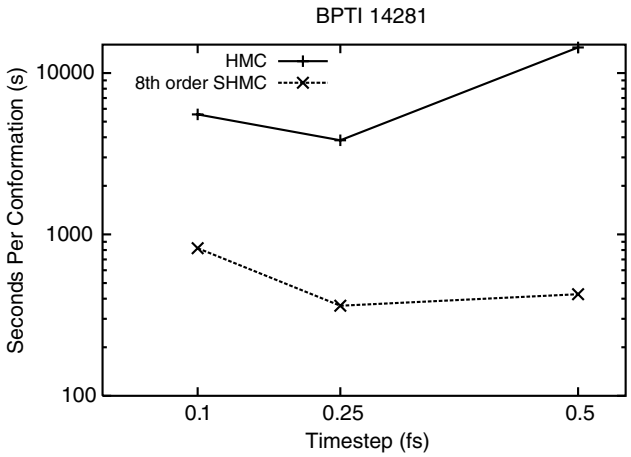

Fig. 5. Average computer time per discovered conformation for 14,281-atom BPTI.

Table 6
Speedup of SHMC over HMC based on conformational efficiency and moderate variance

Number of atoms	4th order SHMC	8th order SHMC	Theoretical
66	1.6	1.7	2.4
1101	2.2	2.1	5.6
14281	9.7	10.1	10.0

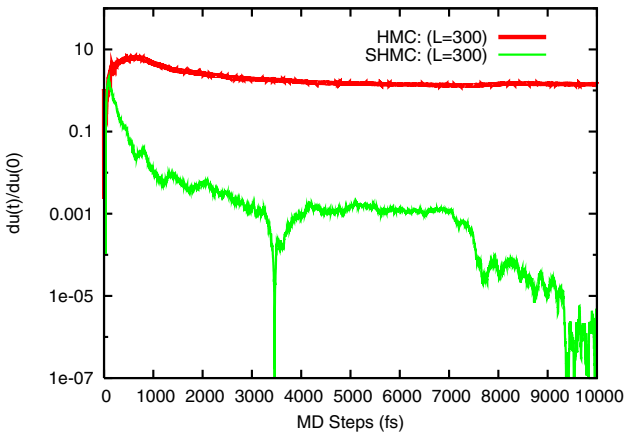

Fig. 6. Ergodic measure of the potential energy for a 66-atom decalanine with time step 0.5 fs.

Table 7
Speedup of SHMC over HMC based on conformational efficiency only

Number of atoms	4th order SHMC	8th order SHMC	Theoretical
66	1.6	2.5	2.4
1101	2.6	3.5	5.6
14281	12	16.5	10.0

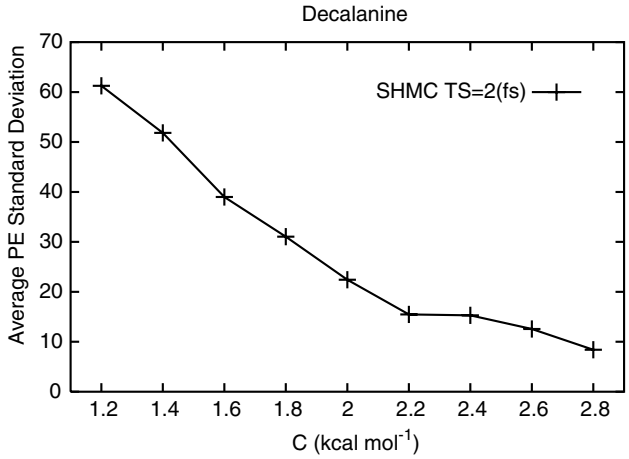

Fig. 7. The effect of c on the standard deviation of the potential energy.

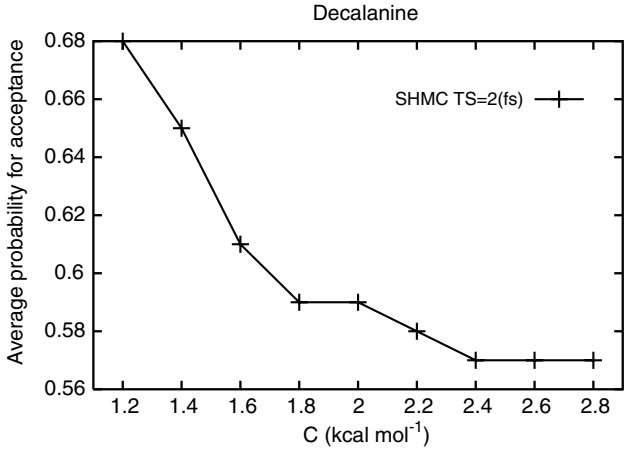

Fig. 8. The effect of c on the average probability of accepting the MD step.

The x-axis represents the number of simulated steps for each algorithm. If two trajectories are self-averaging, then the plot of $\mathrm{d}u(t)$ should approach 0. The graph shows that the ergodic measure for HMC is decreasing, albeit much more slowly than the ergodic measure for SHMC. This is further evidence that SHMC is a more efficient sampling method than HMC.

The emphasis for all data collected in this paper was to minimize the variance. However, larger speedups are possible if one disregards the variance, cf. Table 7. Here, similar results for the smaller molecules are seen, but the speedup for the larger BPTI is quite impressive. In order to achieve these speedups, methods that minimize the variance, such as control variates, are being considered.

4.3.6. The parameter c

Choosing a good value for c is very important to the performance of SHMC. Currently, the best method for verifying a particular choice is through experimentation. In practice, the following steps offer a good initial starting point:

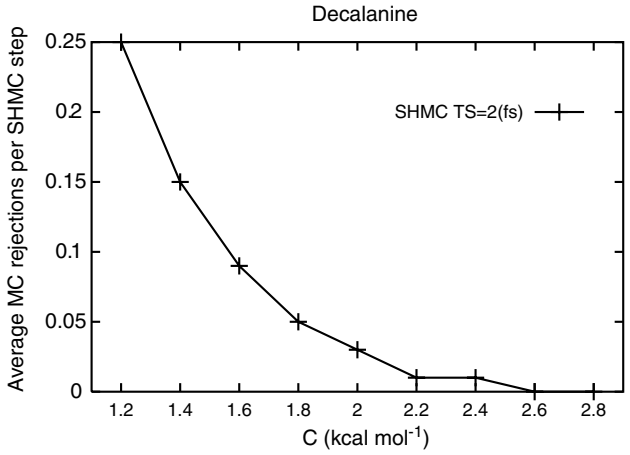

Fig. 9. The effect of c on the number of MC rejections of the momenta per SHMC step.

- Run SHMC on the system of interest and record $\Delta\mathscr{H} = \mathscr{H}_{[2k]} - \mathscr{H}$ at each step. The more data points recorded, the more accurate c will be.
- Determine the expected value and standard deviation of $\Delta\mathscr{H}$.
- Choose c to be approximately $\langle\Delta\mathscr{H}\rangle - 1.5 * \sigma$.

Through numerous experiments, this policy has shown to be a good heuristic, although it is sometimes necessary to further adjust the value of c. Additional experiments around the starting value of c are usually beneficial.

In the following graphs, the effect of c on the simulation is shown. Fig. 7 shows a plot of the standard deviation of the potential energy as a function of the value chosen for c. The system is decalanine, with a time step of 2 fs. This is a pattern echoed throughout the numerical experiments.

As can be seen in Fig. 8, the average probability of accepting the MD move also decreases as c increases. This shows the importance of making c as small as possible.

In Fig. 9, the number of MC rejections per SHMC step are plotted. As c increases, the number decreases with the value going to 0 as c gets large. This puts a lower bound on the value of c. If too small of a value is chosen for c, then there will be more rejections per step and the efficiency will drop because of the extra overhead.

As can be seen from the previous figures, the choice of c must be considered before any serious application of SHMC. This is especially true as the size of the system and the time step increase.

5. Discussion

SHMC is a generalization of HMC that samples from a p.d.f. in all of phase space. This p.d.f. is induced by a modified Hamiltonian. Since the modified Hamiltonian is by construction close to the true Hamiltonian, the reweighting does not damage the variance. The additional parameter, c, of SHMC, measures the amount by which the modified and the true Hamiltonian can depart. Different regions of phase space may need different optimal parameters. Here, c is chosen to satisfy both bounds on the statistical error of sampling and an acceptable performance. A rule of thumb is that it should be close to the difference between the true and the modified Hamiltonian. Other criteria are possible, and it would be desirable to provide "optimal" choices.

For SHMC and methods that sample from noncanonical distributions, it is important to check the efficiency of sampling for the canonical ensemble after reweighting. The appropriate metric for efficiency depends on the application. The metric used here seems adequate for sampling of biomolecules. One could also only include the Φ [3, p. 83] or Ψ dihedrals of biological molecules. Even more general are the various ergodic measures that measure the self-averaging properties of algorithms, cf. [31].

The efficiency of Monte Carlo methods can be improved using other variance reduction techniques. For example, [32] improves the acceptance rate of HMC by using "reject" and "accept" windows. It accepts whether to move to the accept window or to remain in the reject window based on the ratio of the sum of the probabilities of the states in the accept and the reject windows. We are also actively exploring to see if the method of control variates could be used in SHMC, cf. [16, pp. 277 ff].

Other generalizations of HMC enhance barrier crossing in rugged energy hyper-surfaces. The adaptive temperature HMC method [22] has an additional parameter that is adapted to enhance sampling at low temperatures. It samples from a mixed canonical ensemble. HMC is combined with tempering by Neal [33]. Other multicanonical protocols are the original protocol of Berg and Neuhaus [34], J-walking [35], multicanonical J-walking [36], $1/k$ sampling [37], and simulated tempering [38]. Some methods sample from non-Boltzmann ensembles, such as Tsallis' [39]. Other methods make the potential energy surface smooth, e.g., [40,41].

HMC has been combined with multiple time stepping (MTS) algorithms such as r-RESPA [42,43], which allows longer time steps in the MD step. Extension of SHMC to MTS would require construction of shadow Hamiltonians for MTS integrators. It might also be advantageous to combine it with more stable symplectic MTS algorithms. One such method is MOLLY [44–47]. This combination with SHMC should provide even larger speedups, since it would make it possible to get closer to large MD time steps.

SHMC can be combined with other MC moves. This flexibility is one of the advantages of MC over MD for sampling. For example, it might be useful to perform moves on the dihedral angles directly [48,49]. Another HMC method combines MD with a knowledge base potential for the acceptance criterion in the MC step, cf. [50]. The motivation is to discriminate folded or native from metastable structures, which is hard for current MD force fields [51,52].

Conformational dynamics [53,54] is an application that might benefit from SHMC. It performs many short HMC simulations in order to compute the stochastic matrix of a Markov chain. Then it identifies almost invariant sets of configurations, thereby allowing a reduction of the number of degrees of freedom in the system. Another possible application of SHMC is as a component of the *transition path sampling* approach to sample rare events such as barrier crossings [55,56]. The method provides an importance sampling from which transition states can be characterized using statistical mechanics.

Nonbiological applications of SHMC are possible. Typically, the momenta will not have any meaningful interpretation. For example, HMC has been used for Bayesian inference in multi-layer perceptrons, cf. [57].

Acknowledgements

This work was partially supported by an NSF Career Award ACI-0135195. Computations were performed in part through a Beowulf cluster supported by NSF grant DMR-0079647 and also by an Equipment Renovation grant through the University of Notre Dame. Scott Hampton was supported through an Arthur J. Schmitt fellowship. The Theoretical and Computational Biophysics group at the University of Illinois hosted Jesus Izaguirre as a visiting research assistant professor during the summer 2003, and provided access to its computational resources. Robert Skeel from Illinois gave valuable advice on making SHMC a rigorous sampling method. The following have also contributed useful suggestions: Hong Hu, David Hardy, Edward Maginn, and Gary Huber.

Appendix A. Pseudocode for SHMC

The following is a high level pseudocode based on our actual implementation of computing SHMC with a 4th-order shadow.

```
RUN_SHMC(number_of_steps)
01.   k := 2          //shadow order is 2 * k.
02.   shadow_c := #          //Define constant c.
03.   save_positions_forces();          //Save current state of the system.
04.    for i:= 1 to number_of_steps
05.      get_new_momenta();
06.      run_md(cyclelength);
GET_NEW_MOMENTA()
01.   while(! accept_momenta)
02.      calc_rand_velocity(initial_temperature)
03.      init_tot_energy := compute_total_energy()
04.      run_time_reverse_leapfrog(k/2)
05.      restore_state_0()
06.      run_leapfrog(k/2)
07.      init_shadow_energy := calc_shadow()
08.      momenta_ener_diff := init_shadow_energy – shadow_c – init_tot_energy
08.      accept_momenta := accept_new_momenta(momenta_ener_diff)
09.      if (not accept_momenta)
10.         restore_positions()
RUN_MD()
01.   md_init_ener := max(init_shadow_energy – shadow_c, init_tot_energy)
02.   for i := k/2+1 to cyclelength
03.      run_leapfrog()
04.   fin_total_ener := compute_total_energy()
05.   for i := 1 to k/2
06.      run_leapfrog()
07.   restore_prev_state()
08.   fin_shadow_ener := calc_shadow()
09.   md_fin_ener := max(fin_shadow_ener – shadow_c, fin_total_ener)
10.   accept_md := accept_new_positions(md_fin_ener – md_init_ener)
11.   if(not accept_md)
12.      restore_positions_forces()
RUN_LEAPFROG()
01.   do_half_kick()      //Update momenta
02.   beta := beta + calc_beta()      //Update beta using Eq. (21)
03.   do_drift()      //Update positions
04.   update_forces()      //Update forces
05.   do_half_kick()
06.   beta := beta + calc_beta()
07.   save_system_state(positions, velocities, beta)
CALC_SHADOW()
          //This is 4th order shadow Hamiltanion
01.   Real pos_1 vel_0, ..., a10, a12
02.   for i := 1 to number_atoms
```

//Compute the dot products, see Eqs. (19) and (20)

03. $\text{pos}_1\text{vel}_0 = \text{pos}_1\text{vel}_0 + \text{mass[i]} * 0.5 * (\text{positions}^{n+1}[\text{i}] - \text{positions}^{n-1}[\text{i}]) \cdot \text{velocities}^{n}[\text{i}]$

04. $\text{vel}_1\text{pos}_0 = \text{vel}_1\text{pos}_0 + \text{mass[i]} * 0.5 * (\text{velocities}^{n+1}[\text{i}] - \text{velocities}^{n-1}[\text{i}]) \cdot \text{positions}^{n}[\text{i}]$

05. $\text{pos}_1\text{vel}_2 = \text{pos}_1\text{vel}_2 + \text{mass [i]} * 0.5 * (\text{positions}^{n+1}[\text{i}] - \text{positions}^{n-1}[\text{i}]) \cdot (\text{velocities}^{n+1}[\text{i}] - 2 * \text{velocities}^{n}[\text{i}] + \text{velocities}^{n-1}[\text{i}])$

06. $\text{vel}_1\text{pos}_2 = \text{vel}_1\text{pos}_2 + \text{mass[i]} * 0.5 * (\text{velocities}^{n+1}[\text{i}] - \text{velocities}^{n-1}[\text{i}]) \cdot (\text{positions}^{n+1}[\text{i}] - 2 * \text{positions}^{n}[\text{i}] + \text{positions}^{n-1}[\text{i}])$

//Compute the a_i_j.

07. $\text{a10} := \text{pos}_1\text{vel}_0 - \text{vel}_1\text{Pos}_0 - 0.5 * (\text{beta}^{n+1} - \text{beta}^{n-1})$

08. $\text{a12} := \text{pos}_1\text{vel}_2 - \text{vel}_1\text{pos}_2$

//Calculate the 4th order shadow.

09. shadow4: = c*(a10 – a12/6.0)

References

[1] B.J. Berne, J.E. Straub, Novel methods of sampling phase space in the simulation of biological systems, Curr. Topics Struct. Biol. 7 (1997) 181–189.

[2] A. Brass, B.J. Pendleton, Y. Chen, B. Robson, Hybrid Monte Carlo simulations theory and initial comparison with molecular dynamics, Biopolymers 33 (1993) 1307–1315.

[3] T. Schlick, Molecular Modeling and Simulation – An Interdisciplinary Guide, Springer-Verlag, New York, NY, 2002.

[4] S. Duane, A.D. Kennedy, B.J. Pendleton, D. Roweth, Hybrid Monte Carlo, Phys. Lett. B 195 (1987) 216–222.

[5] M. Creutz, Global Monte Carlo algorithms for many-fermion systems, Phys. Rev. D 38 (4) (1988) 1228–1238.

[6] A.D. Kennedy, B. Pendleton, Acceptances and autocorrelations in hybrid Monte Carlo, Nucl. Phys. B (Proc. Suppl.) 20 (1991) 118–121.

[7] M. Creutz, A. Gocksch, Higher-order hybrid Monte Carlo algorithms, Phys. Rev. Lett. 63 (1) (1989) 9–12.

[8] R.D. Skeel, D.J. Hardy, Practical construction of modified Hamiltonians, SIAM J. Sci. Comput. 23 (4) (2001) 1172–1188.

[9] A.R. Leach, Molecular Modelling: Principles and Applications, Addison-Wesley, Reading, MA, 1996.

[10] B. Mehlig, D.W. Heermann, B.M. Forrest, Hybrid Monte Carlo method for condensed-matter systems, Phys. Rev. B 45 (2) (1992) 679–685.

[11] S.D. Bond, B.J. Leimkuhler, B.B. Laird, The Nosé–Poincaré method for constant temperature molecular dynamics, J. Comput. Phys. 151 (1) (1999) 114–134.

[12] W.K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, Biometrika 57 (1970) 97–109.

[13] R.M. Neal, Probabilistic inference using Markov chain Monte Carlo methods, Tech. Rep. CRG-TR-93-1, University of Toronto, 1993.

[14] M.P. Allen, D.J. Tildesley, Computer Simulation of Liquids, Clarendon Press, Oxford, 1987, reprinted in paperback in 1989 with corrections.

[15] J. von Neumann, Various techniques used in connection with random digits, in: Monte Carlo Method, Applied Mathematics Series, 12, National Bureau of Standards, Washington, DC, 1951.

[16] G.S. Fishman, Monte Carlo: Concepts, Algorithms, and Applications, Springer Series in Operations Research, Springer-Verlag, New York, 2000.

[17] S. Gupta, A. Irbäck, F. Karsch, B. Petersson, The acceptance probability in the hybrid Monte-Carlo method, Phys. Lett. B 242 (1990) 437–443.

[18] J.M. Sanz-Serna, M.P. Calvo, Numerical Hamiltonian Problems, Chapman and Hall, London, 1994.

[19] S. Reich, Backward error analysis for numerical integrators, SIAM J. Numer. Anal. 36 (1999) 1549–1570.

[20] T. Matthey, Framework Design, Parallelization and Force Computation in Molecular Dynamics, Ph.D. thesis, University of Bergen, Bergen, Norway, 2002.

[21] T. Matthey, T. Cickovski, S. Hampton, A. Ko, Q. Ma, T. Slabach, J.A. Izaguirre, PROTOMOL: an object-oriented framework for prototyping novel algorithms for molecular dynamics, ACM Trans. Math. Software, in press.

[22] A. Fischer, F. Cordes, C. Schütte, Hybrid Monte Carlo with adaptive temperature in mixed-canonical ensemble: efficient conformational analysis of RNA, J. Comp. Chem. 19 (15) (1998) 1689–1697.
[23] P.B. Mackenzie, An Improved Hybrid Monte Carlo method, Tech. Rep. FERMILAB-Pub-89/100-T, Fermi National Accelerator Laboratory, 1989.
[24] P.D. Kirchhoff, M.B. Bass, B.A. Hanks, J. Briggs, A. Collet, J.A. McCammon, Structural fluctuations of a cryptophane host: a molecular dynamics simulation, J. Am. Chem. Soc. 118 (1996) 3237–3246.
[25] M.J. Potter, P.D. Kirchhoff, H.A. Carlson, J.A. McCammon, Molecular dynamics of cryptophane and its complexes with tetramethylammonium and neopentane using a continuum solvent model, J. Comp. Chem. 20 (1999) 956–970.
[26] S.E. Feller, D. Yin, R.W. Pastor, A.D. MacKerell Jr., Molecular dynamics simulation of unsaturated lipids at low hydration: parametrization and comparison with diffraction studies, Biophys. J. 73 (5) (1997) 2269–2279.
[27] D. Frenkel, B. Smit, Understanding Molecular Simulation, second ed., Academic Press, San Diego, 2002.
[28] A.D. MacKerell Jr., D. Bashford, M. Bellott, R.L. Dunbrack Jr., J. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, I.W.E. Reiher, B. Roux, M. Schlenkrich, J. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, All-hydrogen empirical potential for molecular modeling and dynamics studies of proteins using the CHARMM22 force field, J. Phys. Chem. B 102 (1998) 3586–3616.
[29] A.D. MacKerell Jr., D. Bashford, M. Bellott, R.L. Dunbrack Jr., J. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, B. Roux, M. Schlenkrich, J. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, Self-consistent parameterization of biomolecules for molecular modeling and condensed phase simulations, FASEB J. A 143 (1992) 6.
[30] D. Thirumalai, R.D. Mountain, T.R. Kirkpatrick, Ergodic behavior in supercooled liquids and in glasses, Phys. Rev. A 39 (1989) 3563.
[31] R.H. Zhou, B.J. Berne, Smart walking – a new method for Boltzmann sampling of protein conformations, J. Chem. Phys. 107 (21) (1997) 9185–9196.
[32] R.M. Neal, An improved acceptance procedure for the hybrid Monte Carlo algorithm, J. Comput. Phys. 111 (1994) 194–203.
[33] R.M. Neal, Sampling from multimodal distributions using tempered transitions, Stat. Comput. 6 (1996) 353–366.
[34] B.A. Berg, T. Neuhaus, Multicanonical ensemble – a new approach to simulate 1st-order phase-transitions, Phys. Rev. Lett. 68 (1992) 9–12.
[35] D.D. Frantz, D.L. Freeman, J.D. Doll, Reducing quasi-ergodic behavior in Monte Carlo simulation by J-walking: Applications to atomic clusters, J. Chem. Phys. 93 (1990) 2769.
[36] H. Xu, B.J. Berne, Multicanonical jump walking: a method for efficiently sampling rough energy landscapes, J. Chem. Phys. 110 (21) (1999) 10299–10306.
[37] B. Hesselbo, R.B. Stinchcombe, Monte Carlo simulation and global optimization without parameters, Phys. Rev. Lett. 74 (1995) 2151–2155.
[38] E. Marinari, G. Parisi, Simulated tempering: a new Monte Carlo scheme, Europhys. Lett. 19 (1992) 451–458.
[39] I. Andricioaei, J.E. Straub, Generalized simulated annealing algorithms using Tsallis statistics: Application to conformational optimization of a tetrapeptide, Phys. Rev. E 53 (1996) R3055–R3058.
[40] Z. Liu, B.J. Berne, Method for accelerating chain folding and mixing, J. Chem. Phys. 99 (8) (1993) 6071–6077.
[41] L. Piela, J. Kostrowicki, H.A. Scheraga, The multiple minima problem in the conformational analysis of molecules: deformation of the protein energy hypersurface by the diffusion equation method, J. Phys. Chem. 93 (1989) 3339–3346.
[42] M. Tuckerman, B.J. Berne, G.J. Martyna, Reversible multiple time scale molecular dynamics, J. Chem. Phys. 97 (3) (1992) 1990–2001.
[43] D.G. Gromov, J.J. de Pablo, Structure of binary polymer blends – multiple time step hybrid Monte Carlo simulations and self-consistent integral-equation theory, J. Chem. Phys. 103 (18) (1995) 8247–8256.
[44] B. Garc'ıa-Archilla, J.M. Sanz-Serna, R.D. Skeel, Long-time-step methods for oscillatory differential equations, SIAM J. Sci. Comput. 20 (3) (1998) 930–963.
[45] J.A. Izaguirre, S. Reich, R.D. Skeel, Longer time steps for molecular dynamics, J. Chem. Phys. 110 (19) (1999) 9853–9864.
[46] J.A. Izaguirre, D.P. Catarello, J.M. Wozniak, R.D. Skeel, Langevin stabilization of molecular dynamics, J. Chem. Phys. 114 (5) (2001) 2090–2098.
[47] Q. Ma, J.A. Izaguirre, Targeted mollified impulse – a multiscale stochastic integrator for long molecular dynamics simulations, Multiscale Model. Simul. 2 (1) (2003) 1–21.
[48] F. Guarnieri, W.C. Still, A rapidly convergent simulation method – mixed Monte Carlo stochastic dynamics, J. Comp. Chem. 15 (11) (1994) 1302–1310.
[49] L.J. LaBerge, J.C. Tully, A rigorous procedure for combining molecular dynamics and Monte Carlo simulation algorithms, Chem. Phys. 260 (1–2) (2000) 183–191.
[50] H. Zhang, A new hybrid Monte Carlo algorithm for protein potential function test and structure refinement, Proteins 34 (1999) 464–471.

[51] J. Novotny, R. Bruccoleri, M. Karplus, An analysis of incorrectly folded proteins – implications for structure prediction, J. Mol. Biol. 177 (4) (1984) 787–818.
[52] Y. Wang, H. Zhang, L. Wei, R. Scott, Discriminating compact nonative structures from the native structure of globular proteins, Proc. Natl. Acad. Sci. USA 92 (1995) 709–713.
[53] C. Schütte, A. Fischer, W. Huisinga, P. Deuflhard, A direct approach to conformational dynamics based on hybrid Monte Carlo, J. Comput. Phys 151 (1) (1999) 146–168.
[54] C. Schütte, Conformational dynamics: modelling, theory, algorithm, and application to biomolecules, Tech. rep., Konrad–Zuse–Zentrum für Informationstechnik Berlin, July 1999, SC 99-18.
[55] C. Dellago, P.G. Bolhuis, S.C. Csajka, D. Chandler, Transition path sampling and the calculation of rate constants, Phys. Rev. E 108 (5) (1998) 1964–1977.
[56] G.E. Crooks, D. Chandler, Efficient transition path sampling for nonequilibrium stochastic dynamics, Phys. Rev. E 64 (2000) 026109.
[57] R.M. Neal, Bayesian learning for neural networks, in: Lecture Notes in Statistics, vol. 118, Springer-Verlag, New York, 1996.