TECHNICAL MATERIAL

New Hybrid Monte Carlo Methods for Efficient Sampling: from Physics to Biology and Statistics

Elena AKHMATSKAYA [1-3*] and Sebastian REICH [4]

[1] *Fujitsu Laboratories of Europe Ltd, Hayes Park Central, Hayes End Road, Hayes, Middlesex, UB4 8FE, UK*
[2] *Basque Center for Applied Mathematics, Bizkaia Technology Park, Building 500, E48160, Derio, Spain*
[3] *Ikerbasque, Basque Foundation for Science, E-48011 Bilbao, Spain*
[4] *Universitat Potsdam, Institut fur Mathematik, Am Neuen Palais 10, D-14469, Potsdam, Germany*

We introduce a class of novel hybrid methods for detailed simulations of large complex systems in physics, biology, materials science and statistics. These generalized shadow Hybrid Monte Carlo (GSHMC) methods combine the advantages of stochastic and deterministic simulation techniques. They utilize a partial momentum update to retain some of the dynamical information, employ modified Hamiltonians [1-3] to overcome exponential performance degradation with the system's size and make use of multi-scale nature of complex systems. Variants of GSHMCs were developed for atomistic simulation, particle simulation and statistics: GSHMC (thermodynamically consistent implementation of constant-temperature molecular dynamics), MTS-GSHMC (multiple-time-stepping GSHMC), meso-GSHMC (Metropolis corrected dissipative particle dynamics (DPD) method), and a generalized shadow Hamiltonian Monte Carlo, GSHmMC, (a GSHMC for statistical simulations). All of these are compatible with other enhanced sampling techniques and suitable for massively parallel computing allowing for a range of multi-level parallel strategies.

A brief description of the GSHMC approach, examples of its application on high performance computers and comparison with other existing techniques are given. Our approach is shown to resolve such problems as resonance instabilities of the MTS methods and non-preservation of thermodynamic equilibrium properties in DPD, and to outperform known methods in sampling efficiency by an order of magnitude[4].

KEYWORDS: Hybrid, Monte Carlo, Hamiltonian, sampling

I. Introduction

The statistical sampling method (later known as the Monte Carlo (MC) method) has a very long history[5]. Lord Kelvin used the method in 1901 for testing the equipartition theorem[6] whereas E. Fermi employed the statistical sampling to study neutron moderation in 1934[7]. However only after the advent of the first electronic computers, the method has received an increasing interest as a powerful tool for solving important physical problems. The first published report on the method appeared in 1953. In their paper Metropolis et al.[8] introduced what is today known as Monte Carlo importance sampling, also referred to as the Metropolis-Hasting algorithm for the Boltzmann distribution or Metropolis Monte Carlo. The algorithm generates a random walk using a proposal density and includes a method for rejecting / accepting the proposed moves. Nowadays the algorithm is used in nearly every aspect of scientific inquiry and in the year 2000 it was ranked as one of the "10 algorithms with the greatest influence on the development and practice of science and engineering in the 20th century"[9].

Another method for molecular simulation was reported soon after the introduction of Metropolis Monte Carlo. The first molecular dynamics simulation was presented at a conference in Brussels in 1956 by Alder and Wainwright and then published in 1957[10]. Molecular dynamics (MD) is a computer simulation technique where the time evolution of a set of interacting atoms is studied by numerically integrating their Newton's equations of motion.

In the following decades molecular dynamics and Monte Carlo were considered as two basic techniques for molecular simulations. Both methods have their advantages and disadvantages. Interestingly, the methods appear to be surprisingly complementary: where one method fails another succeeds.

Molecular dynamics is a deterministic procedure yielding smooth trajectories where time has a clear interpretation. The method generates kinetic and thermodynamic data. On the contrary, Monte Carlo is a stochastic (random) method yielding discontinuous trajectories. The time has no clear meaning in MC and the method does not provide kinetic information.

MD allows only for slow exploration of configurational

*Corresponding Author, E-mail:akhmatskaya@bcamath.org

space through a sequence of many small steps in the discretized version of the equations of motion which in turn leads to the systematic discretization error. By contrast, MC does not have an equivalent of a time step error, its sampling is exact with only statistical errors involved. The method is flexible in the choice of sequences of steps which potentially can lead to rapid exploration of configurational space. However for complex dense systems (e.g. biomolecules) it is becoming very hard to choose a good trial move[11] that could help to achieve sampling performance better than that of the MD.

The MD method utilizes an expensive force computation and requires additional modifications to sample from a constant temperature ensemble. The MC method does not require force calculation; it maintains the temperature by construction through the Metropolis test function.

From computational point of view, the MD algorithm is not, by its nature, embarrassingly parallel although a significant progress has been achieved in developing efficient parallel MD software codes[12-21]. The most obvious approach to parallelize MC is simply to run multiple chains in parallel[22-23]. This parallel chains approach does not require any parallel programming although more sophisticated schemes combined with this approach for further performance improvement can be effective [24].

Efforts have been made to take advantage of both simulation techniques by uniting them in yet another simulation method. The Hybrid Monte Carlo (HMC) method introduced by Duane, Kennedy, Pendleton and Roweth in 1987[25] combines short constant energy MD trajectories with an MC rejection step. Each new trajectory is accepted with Metropolis probability after which momenta are randomly reset. HMC can be viewed either as an efficient MC with a "smart" collective move or an MD with corrupted dynamics which rigorously samples from the target temperature. The method can also be adapted to the simulation at constant pressure [26]. The HMC approach has not become as popular a sampling method as MD or MC outside the lattice filed community (to whom it was originally addressed) because of its two essential drawbacks. First, the HMC cannot provide detailed dynamic information. Second, it suffers from exponential performance degradation (low acceptance rate) with respect to the system's size and the time step. This is due to discretization errors introduced by the numerical integrator [27, 28].

To keep more of the dynamic information, the HMC has been generalized to allow for partial momentum updates between Monte Carlo steps, instead of complete replacements of momenta by random samples from the appropriate Boltzmann distribution [29,30]. However, to satisfy the detailed balance condition [31], the generalized hybrid Monte Carlo (GHMC) method requires a momentum reversal (flip) in case of rejection of the molecular dynamics proposal. This means that the GHMC method essentially reverses its direction upon rejection condition [32] which, in combination with the poor acceptance rate discussed above, makes the approach unsuitable for effective simulation of large complex systems. The acceptance rate of the HMC method can be increased by using importance sampling with respect to a shadow Hamiltonian as first proposed in [1] and further developed in [33]. Efficient implementation of shadow Hamiltonians in the context of the GHMC method was achieved in the generalized shadow hybrid Monte Carlo (GSHMC) method [2,3]. The method offers a rigorous, flexible and efficient approach to conformational sampling in atomistic simulations. In the GSHMC, a modified Metropolis criterion and a flexible momentum update significantly improve the acceptance rate compared with standard hybrid Monte Carlo simulation and at the same time allow more dynamical information to be retained, making GSHMC a powerful tool for simulation of large molecules. It is also well suited to massively parallel computation. Parallelism in the momentum update step and in the molecular dynamics trajectories can be combined with any inherent parallelism in the application of interest and with other enhanced sampling techniques to achieve high levels of efficiency.

The GSHMC was originally designed for atomistic simulations. In this paper we adopt the method to three other simulation methodologies: macromolecular simulation, particle simulation and statistical simulation. Accordingly, this yields three new simulation methods: multiple-time-stepping generalized shadow hybrid Monte Carlo (MTS-GSHMC), meso-GSHMC and generalized shadow Hamiltonian Monte Carlo (GSHmMC).

The MTS-GSHMC method belongs to the class of multiple-time-stepping (MTS) methods [34-39] which make use of the multi-scale nature of the macromolecular systems. In MTS methods, savings of computational time can be realized if the slowly varying forces are held constant over longer intervals than the more rapidly varying forces. Standard integration procedures in MD can then be modified by splitting the forces into fast (short-range) and slow (long-range) components and evaluating the former more frequently than the latter. The ratio between frequencies of evaluation of the long-range forces and short-range forces measures the gain in simulation time and will be further referred as "the step-size ratio".

Though in theory MTS methods can dramatically speed up MD simulations by reducing the number of expensive slow force evaluations, in practice such impulse MTS methods, e.g., the popular Verlet-I/r-RESPA [34,35] suffer from severe resonance instabilities that limit their practical performance gain [36,37]. Performance of impulse MTS methods was recently improved in the Langevin stabilized MTS methods [37] by reducing resonance induced instabilities through the introduction of mollified MTS methods (see Izaguirre et al. [38]) and by introducing of a weak coupling to a stochastic heat bath (Langevin dynamics) [37]. However, accurate simulations still restrict the step-size ratio [37,38].

In addition to the limitations which are specific to each of the above methods, all of them share a common drawback – they do not exactly sample from the target temperature even if the simulations are stable and are subject to a thermostat

[40,41]. This error can be controlled only with a loss of computational efficiency.

In this paper we introduce a novel MTS method, MTS-GSHMC, which takes advantage of the recent advances and removes the bottlenecks of existing multi-scale methods.

MTS-GSHMC utilizes the GSHMC to provide the desired weak stochastic stabilization for the MD multi-scale integrator to enhance computational performance [37]. It also employs a mollified impulse MTS method [38] in the Molecular Dynamics step of the GSHMC to eliminate resonance induced instabilities. We also introduce a new technique for computing shadow Hamiltonians for multiple-time-stepping (MTS) methods, which was successfully applied in MTS-GSHMC.

The second novel method discussed in this paper inherits the sampling abilities of its predecessor, GSHMC, and also respects the Galilean invariance of the underlying conservative dynamics so that it can be applied for efficient sampling in advanced simulation at mesoscopic scales. This method, called meso-GSHMC, puts dissipative particle dynamics, DPD [42-51], within the framework of Markov chain Monte Carlo (MCMC) methods which implies rigorous sampling from the canonical distribution regardless of the chosen time step. Meso-GSHMC employs a novel (local) momentum refreshment Monte Carlo step, which conserves the Boltzmann velocity distribution as well as total linear and angular momentum. The proposed method can be viewed as Metropolis corrected time-stepping method similar to the MALA [52] scheme for first-order Brownian dynamics.

Finally, the third new method discussed here, GSHmMC, is a reformulation of GSHMC designed for solving statistical inference problems. The idea to adopt HMC for statistical simulation belongs to R. Neal [53-54]. It was further developed in [31,55-60]. Recently the name "Hamiltonian Monte Carlo" mainly replaced "Hybrid Monte Carlo" term in statistical computation despite the fact that the dynamics considered in that case is fictitious, in contrast to molecular simulation applications. Hamiltonian Monte Carlo makes efficient use of gradient information to reduce random walk behavior. The gradient indicates in which direction one should go to find states with high probability. The end result of HMC is that proposals move across the sample space in larger steps and are therefore less correlated and converge to the target distribution more rapidly.

It was shown that for indirect observation models, the nonlinear state-space model, the non-linear random effects model, neural network models and the generalized linear models, HMC can be more efficient than the standard Markov Chain Monte Carlo methods [54-57]. However for high dimensional systems the HMC method becomes less efficient due to increased rejections. With a large number of hidden units, or when the data set becomes large, the success of this method may require a very large number of evaluations of the posterior distribution and its partial derivatives. In situations where the posterior distribution is computationally costly to evaluate, this may lead to an unacceptable computational load for HMC. In this case reduction of the number of required evaluations is the obvious solution. GSHmMC offers all the advantages of HMC while greatly reducing the number of such evaluations. In general, the method largely eliminates the dependence of the acceptance rate on the system size [3].

The outline of the paper is as follows. In Section II we summarize the GSHMC algorithm, discuss its implementation on massively parallel computers and its application to simulation of real physical systems. In Section III and IV we present the MTS-GSHMC and meso-GSHMC methods respectively and show, on simple test models, their advantages as compared to the commonly used similar techniques. We reformulate the GSHMC algorithm for statistical inference problems in Section V and discuss the possibilities for further improvement of the GSHmMC sampling abilities and its applicability to a wider range of distributions. Concluding remarks can be found in Section VI.

II. GSHMC

1. Ideas behind GSHMC

The objective for developing the generalized shadow hybrid Monte Carlo was to improve the acceptance rate in GHMC[29,30] for large system sizes by reducing a discretization error in the numerical solution. This would lead to more efficient sampling, fewer momentum reversals and thus to more dynamical information being retained.

Let us assume a system of N atoms and introduce the collective atomic position vector X, the diagonal mass matrix of atomic masses M, the collective atomic momentum vector $P=M\dot{X}$, the potential energy U and the Hamiltonian (energy)

$$H(X,P)=\frac{1}{2}P^{T}M^{-1}P+U(X). \qquad (1)$$

Throughout the paper, we will use the following conventions. We assume that X and P are column vectors of length $3N$. We also introduce the collective atomic velocity vector $V=M^{-1}P=\dot{X}$. Y^{T} in (1) denotes the transpose of a vector Y, i.e., Y^{T} is a row vector if Y is a column vector and vice versa.

The standard HMC method consists of two sub-steps: molecular dynamics and momentum refreshment.

Molecular dynamics requires an approximate integration of Newton's equations of motion for a classical unconstrained simulation,

$$M\ddot{X}=-\nabla U(X) \qquad (2)$$

using a time-reversible and volume-preserving method, e.g., that of Verlet [62], over N_{MD} steps and step-size Δt.

The associate canonical Hamiltonian equations of motion are given by

$$\dot{X} = \nabla_P H(X,P) = M^{-1}P, \quad \dot{P} = -\nabla_X H(X,P) = -\nabla U(X) \ . \qquad (3)$$

MD simulations within HMC methods are performed at a constant number of particles *N*, a volume V and an energy *E* (*N*V*E* ensembles) whereas HMC samples either in the canonical ensembles under constant number of particles *N*, volume V and temperature T (*N*VT ensembles) or in the isobaric-isothermal *N*PT ensembles i.e. under constant number of particles *N*, pressure P and temperature T. In this paper we concentrate on *N*VT ensembles simulations.

Expectation values from a *N*VT ensemble are characterized by the canonical distribution

$$\rho_{canon} \propto \exp(-\beta H) , \qquad (4)$$

where $\beta = 1/k_B \mathrm{T}$, and k_B the Boltzmann constant.

On completion of a molecular dynamics step, the new trajectory is accepted with the probability

$$\min(1, \exp(-\beta \Delta H)) \qquad (5)$$

where $\Delta H = H^n - H^0$ and H^0 and H^n are values of the Hamiltonian at the beginning and at the end of the molecular dynamics step, respectively.

HMC only retains the accepted position vector *X*. The accepted momentum vector is replaced by a new momentum vector drawn from the Boltzmann distribution at the given temperature T. This completes a single HMC step.

It is a well known observation that the energy fluctuations along numerical trajectories are scaled with system sizes and time-steps [27, 28]:

$$\left\langle H^n - H^0 \right\rangle = O(N\Delta t^{2p}) \qquad (6)$$

assuming that a time-stepping method of order *p* was applied.

This means that the acceptance rate of the molecular dynamics proposal step governed by (5) deteriorates with increasing time-steps and increasing system sizes. One option to counteract this effect is to apply higher-order symplectic time-stepping methods. However this also increases the computational cost of HMC. An alternative, less expensive approach is based on the concept of modified/shadow Hamiltonians $\hat{H}_{\Delta t}$ for symplectic time-stepping methods. The key observation is that symplectic time-stepping methods conserve a modified/shadow Hamiltonians/energy to much higher accuracy than the accuracy of the method itself [63-64]:

$$\langle \Delta \hat{H}_{\Delta t} \rangle = \left\langle \hat{H}^n_{\Delta t} - \hat{H}^0_{\Delta t} \right\rangle = O(N\Delta t^{2m}) \ , \qquad (7)$$

where $m \geq 4$.

This suggests to implement HMC in the framework of importance sampling with respect to such shadow Hamiltonians.

The GSHMC method is based on the simple idea that the canonical ensemble (4) may be replaced by the shadow ensemble

$$\tilde{\rho} \propto \exp(-\beta \hat{H}_{\Delta t}) \ . \qquad (8)$$

However, the shadow ensemble (8) makes the momentum refreshment part of HMC more complex, since the momenta are no longer Boltzmann distributed under $\tilde{\rho}$. The GSHMC method adopted the momentum refreshment step introduced for a generalized HMC methodology [29,30]. The key idea of the GSHMC momentum refreshment step is to use the GHMC partial momentum update combined with the modified Metropolis acceptance criterion to recover the Boltzmann distribution.

Another consequence of using the shadow ensemble in GSHMC is the need for re-weighting of the observables of the GSHMC simulation in order to obtain proper canonical averages, thus eliminating the bias introduced by the shadow Hamiltonian.

In the next section we formulate the GSHMC algorithm and discuss some practical issues.

2. GSHMC: Algorithmic Summary

The GSHMC simulation algorithm can be summarized as follows:

Step 1. Given the positions *X* and momenta *P* generate momenta P' and a noise vector Ξ':

$$\begin{aligned} P' &= \cos(\varphi) P + \sin(\varphi) \Xi \\ \Xi' &= -\sin(\varphi) P + \cos(\varphi) \Xi \end{aligned} \qquad (9)$$

where $0 < \phi \leq \pi/2$ is a given angle and $\Xi \sim \mathcal{N}[0, \beta M^{-1}]$. Here $\mathcal{N}[0, \beta M^{-1}]$ denotes the (3*N*)-dimensional normal distribution with zero mean and covariance matrix βM^{-1}. The angle φ in equation (9) is a tuneable parameter of the GSHMC method. Special cases are $\varphi = \pi/2$ for performing sampling only and $\varphi = \sqrt{2\gamma \Delta t}$ to provide statistically rigorous implementation of stochastic Langevin dynamics. The friction coefficient γ is a positive constant and N_{MD} has to be set to 1 [61].

Step 2. Evaluate shadow Hamiltonians $\hat{H}_{\Delta t}$ at (*X*, *P*) and at (*X*, P') using either the methodology described in [3] or the method for evaluating shadow Hamiltonians as proposed by Skeel and Hardy [6].

The new momenta P' are accepted with probability

$$\min\left\{1, \frac{\exp(-\beta[\hat{H}_{\Delta t}(X,P') + \frac{1}{2}(\Xi')^T M^{-1} \Xi'])}{\exp(-\beta[\hat{H}_{\Delta t}(X,P) + \frac{1}{2}(\Xi)^T M^{-1} \Xi])}\right\} \ . \qquad (10)$$

If P' is accepted, choose $P_{new} = P'$ as the new momentum, otherwise set $P_{new} = P$.

Step 3. Given (*X*, P_{new}), run MD simulation for a

fixed number of steps N_{MD} with time step Δt to generate new positions X_{new}.

Step 4. Evaluate modified Hamiltonians $\hat{H}_{\Delta t}$ at (X_{new}, P_{new}). Accept the new state (X_{new}, P_{new}) with probability

$$\min\left\{1, \exp(-\beta[\hat{H}_{\Delta t}(X_{new}, P_{new}) - \hat{H}_{\Delta t}(X, P_{new})])\right\}. \qquad (11)$$

If accepted, choose (X_{new}, P_{new}) as a starting state for the next GSHMC step, otherwise take X and negate momenta P_{new}.

Step 5. Return to Step 1.

We note that one can repeat the refreshment steps 1-2 before continuing with the molecular dynamics part of GSHMC. Hence the complete GSHMC cycle in this case will consist of a molecular dynamics Monte Carlo step and a Monte Carlo momentum update step, followed by one or more additional Monte Carlo momentum update steps. In other words, GSHMC becomes the concatenation of three or more Markov processes with identical invariant distribution functions (here the canonical distribution with respect to a modified Hamiltonian). This is important in achieving a high acceptance rate for the momentum update step and high overall efficiency.

Let assume that K cycles of five steps forming the GSHMC algorithm are required to accomplish the simulation and $\Omega(X, P) = \{ \Omega_1, \Omega_2, \ldots \Omega_K\}$ are observables along a sequence of states (X_i, P_i), $i = 1, 2, \ldots K$. Then due to the use of a modified ensemble the averages are computed using re-weighted Ω_i values [1,33]:

$$\langle \Omega \rangle = \frac{\sum_{i=1}^{K} w_i \Omega_i}{\sum_{i=1}^{K} w_i} \quad , \quad w_i = \exp\left(-\beta[H(X_i,P_i) - \hat{H}_{\Delta t}(X_i,P_i)]\right). \qquad (12)$$

We have to stress that by construction the modified Hamiltonians $\hat{H}_{\Delta t}$ are approximations of the true Hamiltonians and a value of the time-step Δt has to be chosen in GSHMC such that the backward error analysis is valid. This implies that the true and the modified Hamiltonians should stay close to each other during the simulation. Under this condition the variance in the weights in (12) remains small. Though non-constant weights reduce the efficiency of the implied estimators, this reduction in efficiency is minor in our case. Also, in any case, such reduction in efficiency is much less dramatic than rejecting many proposal steps which induces a high correlation between samples and thus degrades the performance of the estimator. In other words, one has to balance the efficiency of the estimator as determined by the weight factors versus the higher correlation of samples due to a higher rejection rate under a non-modified Metropolis test. Our experience is that the former is to be preferred over the latter.

The tunable parameters of the GSHMC method are the time-step Δt, the number of MD steps N_{MD} and the angle φ. The acceptance rate generally decreases for φ approaching the value $\pi/2$. On the other hand, if φ is too small, the sampling of the GSHMC method can become less efficient. An optimal value of φ is typically found by requiring that the rejection rate in the velocity refreshment step should be about or less than 10 %. To keep the rejection rate in the velocity update at this level for a given angle φ it is advised to repeat this computationally inexpensive step a sufficient number of times, as it was already mentioned above. This strategy can be efficiently implemented in parallel (see Section 3). Further details and other options to increase the acceptance rate of the partial momentum update step of GSHMC can be found in Ref.[3].

Finally, it should be noted that a rejection of an MD proposal step leads to a simple negation of momenta (a momentum flip) and the next Monte Carlo cycle would start with a time reversed MD trajectory. This is needed to guarantee the accuracy of the method through obeying the detailed balance condition and to increase its sampling efficiency. However momentum flips interfere strongly with the dynamics of the underlying Langevin equations. This is a well-known deficiency of the generalized hybrid Monte Carlo methods [30,61]. In order to reduce this undesired effect, the time-step of integration in the MD stage of GSHMC, the parameter φ and the order of shadow Hamiltonian should be chosen such that the probability of having both the position as well as momentum refreshment steps being simultaneously rejected is sufficiently small (e.g. less than 1 %).

Alternatively, one can use the GSHMC method without momentum flip as suggested in [61]. The method is a viable option for thermostatted dynamics since it meddles less with the natural autocorrelation functions of the underlying (stochastic) dynamics model provided the rejection rate of MD proposals is kept below 10% and φ is sufficiently small.

We now move to practical use of the GSHMC algorithm and discuss the implementation of the GSHMC method on modern computers.

3. Parallelization of GSHMC

Computer simulation of complex physical systems (e.g., in biology or nanotechnology) remains one of the current challenges. Namely, the challenge lies in being able to cope with nearly ten orders of magnitude of time scales, with the conformational flexibility and multiple interactions which lead to systems of enormous complexity. One possible way to fill the gap between the timescales currently attainable by simulation techniques and important physical phenomena is to develop simulation methodologies with better sampling characteristics like the ones we discuss in this paper. Another possibility to extend the reachable timescales is to better utilize powerful compute resources via massive parallelization. These approaches are complementary.

There are two places in the GSHMC algorithm where parallelism can always be employed. As pointed out earlier,

the momentum update step can be run an arbitrary number of times (with different random seeds) to try to ensure acceptance of the momentum refreshment step. These independent momentum update steps can be run in parallel in order to reduce the real computation time. One needs, however, to be careful to ensure that this parallelism does not result in excessive communication. For example, if in a cluster system distributed nodes are used for running the parallel momentum update steps, then communications (in our implementation, copying the data across NFS) are required after completion of each momentum update step and, thus, become expensive for a large number of processors. If possible, preference should be given to using shared memory processors for the momentum update step. It might be beneficial to keep the number of processors required for the momentum update step low and use an alternative strategy for increasing the acceptance rate of the momentum update, e.g. applying GS2HMC approach [3]. Alternatively, one can perform multiple momentum update steps on a single processor to avoid expensive communication and to get a better control over acceptance rates in momentum update step.

The second place where parallelism can be employed is MD trajectory step. This simply uses a parallel version of the MD software – for example, GROMACS [15], NAMD [14], etc. In addition to the parallelism in the momentum update and the MD trajectory, the particular application of interest might contain its own inherent parallelism. For example, if one is interested in the sampling of phase space only, an ensemble of independent GSHMC simulations can be run to achieve an efficient sampling of the phase space. This approach is embarrassingly parallel and can be employed successfully for sampling rare events such as finding transition states in simulations of large systems when the kinetic details are not of interest. Similarly, for applications requiring the collection of statistics at different values of parameters, e.g. free energy calculation using non-equilibrium work values, each GSHMC simulation can be run in parallel for different values of those parameters. No communication is needed at this stage and the choice of the number of processors will be regulated by the requirements on parameters.

Finally, we note that the flexibility of GSHMC makes it a natural sampling device to be combined with other enhanced sampling strategies such as multicanonical replica exchange [66], multiplexed-replica exchange [67] and ConfJump [68]. The latter was specially designed for sampling with the hybrid Monte Carlo method.

The calculation of absolute free energies of binding using the Bennett acceptance ratio (BAR) method [69] for computing free-energy differences between intermediate λ states is a good example of how one might go about parallelizing GSHMC simulations. In this case, independent simulations are required for *N_states* values of λ for each solvated ligand and solvated protein–ligand system. The procedure used to parallelize such calculations is shown in **Fig. 1**. In Fig. 1, the parallel computer consists of a set of multi-processor / core nodes (as found for example in a cluster system). Each independent λ value is treated by exactly one node, and this implies that there is no need for inter-node communication. Within a node both the momentum update step and the MD trajectory are parallelized over *N_proc* SMP processors / cores (*N_proc* =

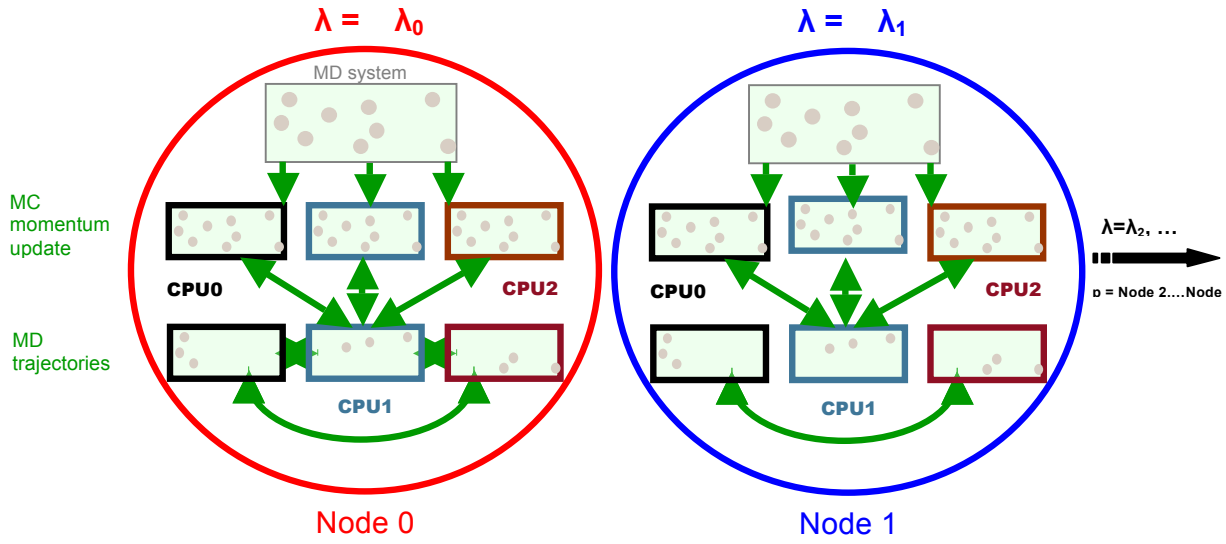

Fig. 1 Scheme used to parallelize GSHMC calculation of absolute free energies of binding.

3 in Fig. 1). This implies that the value of φ must be chosen in such a way that *N_proc* momentum update steps yield a sufficiently high acceptance rate. Alternatively this approach can be combined with the multiple momentum update steps performed serially on each processor. The choice of number of nodes should be made on the basis of achieving good load balance for *N_states* values of λ (λ =33 in our example): optimal values are the factors of *N_states*, as any other choice will lead to some degree of load imbalance. With this parallel decomposition we can achieve high parallel efficiencies for these calculations.

In the next section we demonstrate the effectiveness of the GSHMC approach for simulation of complex systems.

4. Simulations Using GSHMC

The first example of the usefulness of GSHMC comes from its comparison with MD for a coarse-grained representation (CG-MD) of a small peptide toxin interacting with phospholipid bilayer [4]. Specifically, we demonstrate that GSHMC allows for a quicker localization of the toxin to the head-group/water interface of the bilayer.

The simulation system comprises the gating-modifier peptide toxin, Voltage-Sensor Toxin 1 (VSTx1, taken from spider venom), in the palmitoyl-oleoyl-phosphatidyl choline (POPC) membrane environment. The coarse-grained protein model introduced by Bond and Sansom [70] is used. The toxin is initially "buried" in the hydrophobic core of the POPC bilayer (see **Fig. 2**). The objective then is to measure the rate of drift of the toxin along the bilayer normal to the preferred interfacial location. The GSHMC simulations use N_{MD} = 1000 and φ = 0.32 with the order of the modified Hamiltonian being 6. This choice leads to acceptance rates for MD step and momentum update steps of ~99% and ~90%, respectively. All MD simulations in this study are performed using the GROMACS code [15].

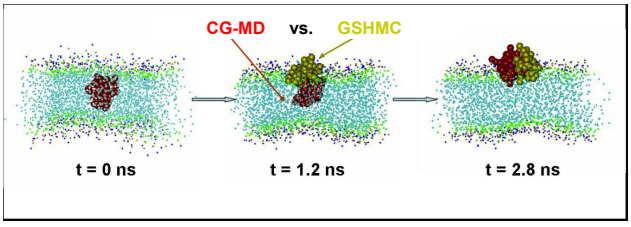

Fig. 2 The evolution of the toxin from the initial position within the membrane to its preferred position at the surface of the bilayer observed during CG-MD (red) and GSHMC (golden) simulations. GSHMC finds a preferred position of toxin faster than CG-MD does.

Three CG-MD and three GSHMC simulations each of duration 50 ns are performed with different initial velocities but identical initial coordinates. In all simulations the toxin drifts towards the head-group/water interface of the bilayer, to an average distance d of ~23–25 Å as measured from the bilayer centre. However the GSHMC method leads to a much more rapid repositioning of the toxin compared to a CG-MD simulation (Fig. 2).

The sampling superiority of GSHMC over CG-MD is also confirmed by the analysis of autocorrelation functions (ACF). As can be seen from **Fig. 3**, GSHMC converges to the equilibrium values more rapidly (up to 12 times more rapidly according to the examination of integrated ACFs) than the CG-MD simulations. Also, fluctuations about the mean in GSHMC (not shown) are of shorter duration compared with the CG-MD simulations which means that GSHMC locates its equilibrium value more frequently (up to 8 times) than CG-MD simulations [4].

One more application which can obviously benefit from using the GSHMC approach is calculation of free energies of binding.

Our final test of the efficiency and accuracy of GSHMC comes from calculations of absolute free energies of binding of three ligands to FKBP protein. Comparison is made with the study of Fujitani et al.[71] which uses methods based on standard MD simulation.

FKBP is a 107-residue protein, best known as the target of the widely used immunosuppressive drug FK506. We have examined the binding of three ligands (L12, LG2 and

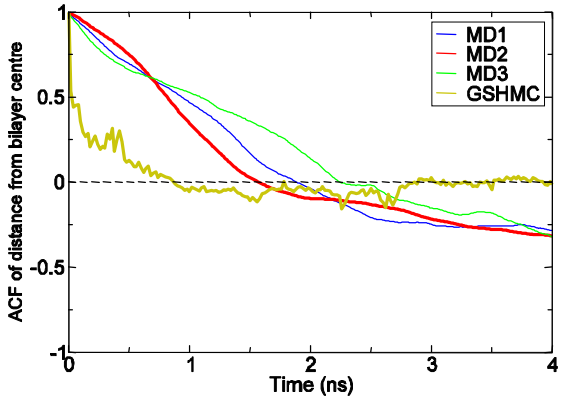

Fig. 3 Autocorrelation function of *d* computed over data from 0 to 8 ns.

LG3) to FKBP using methods very similar to those described in ref.[71], namely the AMBER99 force field for the protein, GAFF force field for the ligands, TIP3P water and the BAR method [69] for computing free-energy differences. Binding energies are evaluated using the double-annihilation method. This requires free-energy calculations for the solvated ligand system and the solvated FKBP–ligand system. We started our simulations using the equilibrated structures from ref.[71], and use the same number of λ points (33) as in the earlier study. The GSHMC simulations are carried out using N_{MD} = 500, φ = 0.35 for the solvated ligand systems, φ = 0.2 for the solvated FKBP–ligand systems, and with modified Hamiltonians of order 6. The chosen parameters provide acceptance rates for MD and momentum update steps of ~99% and 95% respectively.

In **Table 1** we compare the computed absolute binding energies with the values reported in Fujitani et al.

In all three cases the binding energies agree well. GSHMC achieves the same accuracy as MD in one seventh of the computer time.

Table 1 Absolute free energies of binding computed using MD (Fujitani et al.) and GSHMC

Ligand	Fujitani et al.	GSHMC
L12	–7.1	–7.52 ± 0.42
LG2	–4.3	–4.57 ± 0.46
LG3	–4.9	–4.96 ± 0.35

Finally, we note that in the study of Fujitani et al. long (20 ns) equilibration simulations were required prior to gathering free-energy data. Because of its efficient sampling (see the previous toxin test), GSHMC could be also very valuable for this initial equilibration phase.

The basic GSHMC algorithm, originally designed for atomistic simulation of complex systems, can be modified in many ways to make it useful for a wider range of applications. The methods we will discuss further in this paper take all advantages of GSHMC which makes them competitive alternatives to the existing approaches currently dominating in the specific applications areas.

III. MTS-GSHMC

1. Motivation

The potential energy U in Newton's equations of motion (2) is typically given by a sum of bonded and nonbonded interactions:

$$U = U^{bonded} + U^{nonbonded}, \tag{13}$$

$$U^{bonded} = U^{bond} + U^{angle} + U^{dihedral} + U^{improper}, \tag{14}$$

$$U^{nonbonded} = U^{Lennard-Jones} + U^{electrostatics}, \tag{15}$$

and the gradient vector $-\nabla U(X)$ is the force. The computational complexity of bonded interactions is proportional to N while it scales with N^2 for the non-bonded interactions. Simple cut-off schemes have been devised to reduce the computational cost of non-bonded case. But it has also been found that cut-off schemes lead to poor approximations for highly charged systems such as biomolecular ones [72].

The approximate computation of ensemble averages requires performing MD simulations over as large as possible number of time steps. The length of a MD simulation is, on the other hand, limited by the length of the time-step that can be used. Making use of the multi-scale structure of the molecular force fields, MD simulations have been greatly accelerated by the use of multiple-time-stepping (MTS) methods, such as the Verlet-I/r-RESPA method [34,35] which is based on approximating "slow" forces as widely separated impulses.

One derives impulse MTS methods by first rewriting (2) as

$$M\ddot{X} = F^{fast}(X) + F^{slow}(X), \tag{16}$$

where $F^{fast} = -\nabla U^{fast}$ and $F^{slow} = -\nabla U^{slow}$ subject to

$$U = U^{fast} + U^{slow} \quad . \tag{17}$$

Partitioning the potential U into a "fast" part U^{fast} and a "slow" part U^{slow} is done in such a way that an appropriate outer time-step Δt for the slow part is significantly larger than an inner time-step δt for the fast part, and evaluations of the fast force field F^{fast} are computationally much less expensive than evaluations of F^{slow}.

Given an integer $p > 1$ such that the outer time-step Δt and the inner time-step δt satisfy $\Delta t = p\delta t$, an impluse MTS method can now be stated as

$$M\ddot{X}(t) = \sum_{m=0}^{L} c_m \Delta t \delta_X(t - m\Delta t) F^{slow}(X(t)) + \sum_{n=0}^{pL} d_n \delta t \delta_X(t - n\delta t) F^{fast}(X(t)) \tag{18}$$

for $t \in [0, t' = N_{MD}\Delta t]$, δ_x is the Dirac delta function, and $c_m = d_n = 1$ except when $m = n = 0$ or $m = N_{MD}$, $n = pN_{MD}$, respectively, in which case $c_m = d_n = 1/2$.

MTS methods, such as (18) can dramatically speed up MD simulations since the expensive force field evaluations F^{slow} need to be performed only at the larger outer time-step Δt. In this context, note that, for biomolecular applications, the computational complexity of the fast force field evaluations scales linearly in the number of atoms N while it scales quadratically in N for the slow force field evaluations. Furthermore, while the short-ranged (bonded) fast forces are easy to compute in parallel, long-ranged (non-bonded) slow forces require global data communication and hence are more difficult to parallelize efficiently.

However, the impulse MTS method (18) suffers from resonance instabilities. For example, for solvated biomolecular systems, one has to restrict the step-size ratio $p = \Delta t/\delta t$ to $p \le 4$ with an inner time-step of $\delta t = 1$ fs [38].

Recent research has partially resolved some of these issues. On the one hand, resonance induced instabilities of impulse MTS methods have been eliminated through the introduction of mollified MTS methods by Izaguirre et al.[38]. Further improvements have been achieved by weak coupling to a stochastic heat bath (Langevin dynamics) [37]. However, accurate simulations are still limited by a step-size ratio $p = \Delta t/\delta t$ in the range of $p = 6 - 12$ for solvated biomolecular systems [37,38].

Another problem is that common time-stepping methods do not exactly sample from the target temperature T even if the simulations are stable and are subject to a thermostat [40, 41]. This error can be controlled by making the step-size Δt sufficiently small and, in particular, much smaller than

required by stability considerations alone. However, reducing Δt also reduces computational efficiency.

In constrast, as discussed in the previous sections, the GSHMC method rigorously samples from the canonical ensemble while only weakly interfering with the underlying molecular dynamics model. Here we propose to extend the GSHMC method such that a thermodynamically consistent weak stochastic stabilization of mollified MTS methods can be achieved. We will demonstrate that the combined MTS-GSHMC method results in an improved stability of mollified MTS methods.

The following changes in GSHMC are required to achieve those goals. First, a mollified impulse MTS method has to be implemented in the Molecular Dynamics Monte Carlo step of the GSHMC method to eliminate resonance induced instabilities. Second, the derivation of shadow Hamiltonians for the mollified MTS methods is necessary since the shadow Hamiltonians used in GSHMC are specific to the single time-step Störmer-Verlet method. To overcome this limitation we employed a highly efficient method for evaluating shadow Hamiltonians as proposed by Skeel and Hardy [65] for symplectic splitting methods [63,64]. However, due to the multi-scale nature of MTS methods, a non-trivial modification of the approach of Skeel and Hardy is required to make it applicable to the MTS-GSHMC method [73].

We summarize the new results and the GSHMC-MTS algorithm in the next section. More details can be found elsewhere [73].

2. MTS-GSHMC: Summary

We now demonstrate how to combine GSHMC with a mollified MTS method. To do so, we assume that a splitting of the potential energy function U into a fast contribution U^{fast} and a slow contribution U^{slow} (17) is given. We also assume that an averaging operator

$$\overline{X} = A(X), \tag{19}$$

which assigns a filtered, averaged position $\overline{X}$ [38] to an instantaneous collective atomic position vector X has been defined. The averaging operator is then applied to the slow potential to yield a mollified slow potential

$$U^{slow}_{molly}(X) = U^{slow}(A(X)). \tag{20}$$

Also, we assume that an extended mollified MTS method (21) as well as a shadow Hamiltonian (22) of order $q \geq 4$ have been implemented.

As we showed in [73] the extended mollified MTS methods (with introduction of an additional variable $b \in \mathbb{R}$ required for computing shadow Hamiltonians [73]) can be expressed compactly as follows:

$$\begin{aligned} M\ddot{X}(t) &= \sum_{m=0}^{L} c_m \Delta t \delta_X(t - m\Delta t) F^{slow}_{molly}(X(t)) + \sum_{n=0}^{pL} d_n \delta t \delta_X(t - n\delta t) F^{fast}(X(t)), \quad (21) \\ \dot{b}(t) &= -\sum_{m=0}^{L} c_m \Delta t \delta_X(t - m\Delta t) \left[X(t)^T F^{slow}_{molly}(X(t)) + 2U^{slow}_{molly}(X(t)) \right] \\ &\quad - \sum_{n=0}^{pL} d_n \delta t \delta_X(t - n\delta t) \left[X(t)^T F^{fast}(X(t)) + 2U^{fast}(X(t)) \right]. \end{aligned}$$

Besides, in [73] we derived the q-th order shadow Hamiltonians to be used in MTS-GSHMC.

Let $H^{[q],MTS}_{\Delta t}$ denote the q-th order shadow Hamiltonian for an MTS method according to the construction of Skeel and Hardy [65]. Let $H^{[q],MTS,fast}_{\Delta t}$ denote the q-th order shadow Hamiltonian for the same MTS method with F^{slow}_{molly} set equal to zero. Finally, let $H^{[q],SV,fast}_{\delta t}$ denote the q-th order shadow Hamiltonian for the Störmer-Verlet method applied to the fast system with step-size δt. Then the shadow Hamiltonian for use in the new MTS-GSHMC method is given by:

$$H^{[q]}_{\Delta t} = H^{[q],MTS}_{\Delta t} - H^{[q],MTS,fast}_{\Delta t} + H^{[q],SV,fast}_{\delta t}. \tag{22}$$

We now summarize the proposed MTS-GSHMC method. As GSHMC, the MTS-GSHMC is defined as the concatenation of two MCMC steps: a molecular dynamics Monte Carlo (MDMC) and a partial momentum refreshment Monte Carlo (PMMC) steps. The algorithmic scheme introduced in section II.2 for GSHMC remains the same for MTS-GSHMC. The difference is that instead of performing conventional MD in step 3 one should apply the extended mollified MTS method (21). Also, in steps 2 and 4 the shadow Hamiltonians have to be computed according to (22).

We wish to point out that in step 3 (MDMC) it is possible to skip negating momenta in case of rejection of the MDMC proposal following the suggestions in [61].

Reweighting of averages is also required in MTS-GSHMC and has to be performed according to (12).

The free parameters of the MTS-GSHMC scheme include the angle ϕ in (9), the inner and outer step-sizes δt and Δt, respectively, as well as the number of outer time-steps N_{MD} and the order of the shadow Hamiltonian. We will always assume that

$$\phi = \sqrt{2\gamma\tau} \quad << 1. \tag{23}$$

Here $\gamma > 0$ is the collision frequency of an underlying Langevin model and $\tau = \Delta t N_{MD}$.

Note that only differences in b appear in the formulas for shadow Hamiltonians and, hence, one can set $b = 0$ at the beginning of each MDMC and PMMC step. See [73] for more details and explanations.

3. Numerical Test

We consider a one-dimensional chain of diatomic molecules interacting through Lennard-Jones potentials. One-dimensional test problems such as the one considered here have been widely used to test MTS algorithms (see, for example, Hairer et al. [63]).

We assume periodic boundary conditions over a domain of length $l = 20$ and set the number of atoms to $N = 20$. The potential energy of the system is given by

$$U(X)=\frac{K}{2}\sum_{i=1}^{N/2}(|x_{2i-1}-x_{2i}|-1)^2+2\sum_{i=1}^{N}\sum_{j\in n(i)}\left[\left(\frac{\sigma}{|x_i-x_j|}\right)^{12}-\left(\frac{\sigma}{|x_i-x_j|}\right)^{6}\right], \quad (24)$$

where $n(i)$ contains the indices of the $(N-3)$ nearest non-bonded neighbors of particle i, position $x_i \in [0,l]$, $\sigma^6 = 1/2$, and $K = 154210$.

We perform constant *N*VT simulations at $\beta = 1/k_B T = 1$ using the proposed MTS-GSHMC method and compare the results to Langevin stabilized and mollified MTS simulations (MTS-LD) [37]. Following [37], the collision frequency is $\gamma = 0.2$ ps^{-1} in all cases. The slow and fast potential energy contributions are defined by the stiff bonded interactions and the Lennard-Jones interactions, respectively. The inner time-step δt was set to 1 fs for both MTS methods.

The MTS-GSHMC method was implemented with fourth-order shadow Hamiltonian (22), $N_{MD} = 1$ and $\phi = \sqrt{2\gamma\Delta t}$. The number of Monte Carlo samples I is given by the closest to $10000/\Delta t$ so that each simulation covers approximately the same ``MD time-span" of 10 microseconds.

We found that the largest achievable outer time-step for MTS-LD is $\Delta t = 10$ fs. This might be surprising at the first glance since constant energy simulations with Lennard-Jones interactions alone allowed for a larger time-step of $\Delta t = 15$ fs. However, one has to keep in mind that thermostatted MD simulations lead to relatively large fluctuations in instantaneous values of total energy which allow for rare high energy Lennard-Jones collisions. These collisions can destabilize MTS-LD. Hence the achievable outer step-size of MTS-LD is determined by stability and not by accuracy of the method. MTS-GSHMC, on the contrary, was found to be stable for all implemented outer step-sizes ranging from $\Delta t = 10$ fs to $\Delta t = 22$ fs. Given that errors in mean total energy and in the computational temperature should not exceed 5% of the reference values (obtained with Langevin dynamics) and that the acceptance rate should stay above 70% [61], we found that the largest acceptable outer time-step of MTS-GSHMC is $\Delta t = 20$ fs. To measure the efficiency of MTS-GSHMC we introduced the effective outer step-size which takes into account acceptance rates achieved in the method:

$$\Delta t_{eff} = \Delta t \times (acceptance\ rate(AR)\ of\ MDMC/100). \quad (25)$$

The efficiency gain is then defined as the ratio of Δt_{eff} to the largest achievable outer step-size for MTS-LD ($\Delta t = 10$). Using this measure, it was found that the efficiency gain of MTS-GSHMC over MTS-LD is about 1.9.

We refer the reader to **Table 2** for the computed values of mean total energy $\langle H \rangle$, temperatures and acceptance rates of MDMC (AR MDMC) and PMMC (AR PMMC) proposal steps. To get a better impression on the accuracy of MTS-GSHMC as a function of the outer time-step we also computed the center of mass velocity ACFs (**Fig. 4**). Given the fact that we cover a relative large range of time-steps beyond the stability limit of MTS-LD, the computed ACFs agree remarkably well. The corresponding ACF from the small time-step reference simulation LD are also provided for comparison.

Table 2 Mean total energies, temperatures and acceptance rates computed using MTS-GSHMC and MTS-LD methods for a range of outer step-sizes values

Method	Δt fs	Δt_{eff} fs	numerical $\langle H \rangle$	numerical temperature	AR PMMC	AR MDMC
MTS-GSHMC	10	10.0	28.0218	0.9948	98.5%	99.9%
MTS-GSHMC	12	12.0	28.5266	1.0086	98.4%	99.7%
MTS-GSHMC	14	13.9	29.2522	1.0259	98.2%	99.1%
MTS-GSHMC	16	15.3	28.0826	0.9974	98.0%	98.3%
MTS-GSHMC	18	17.5	28.2329	0.9965	97.8%	97.3%
MTS-GSHMC	20	19.2	28.5648	1.0178	97.6%	96.0%
MTS-LD	10	N/A	29.3044	1.0041	N/A	N/A

There are limitations of our model system compared to biomolecular simulations that need to be pointed out clearly. In particular, numerical evidence suggests that the achievable outer step size of mollified MTS methods is limited by resonance instabilities at about 8 fs for explicit water simulations [38]. Additional resonance instabilities are found for even larger outer step-sizes [37]. These instabilities can be masked by using Langevin dynamics with increasingly large values of the collision frequencies γ [37]. On the other hand, the achievable outer step-size for our model system is limited by the stability of Störmer-Verlet method with respect to Lennard-Jones interactions. This instability cannot be overcome by Langevin dynamics even with increasingly large values of γ. This (non-resonance)

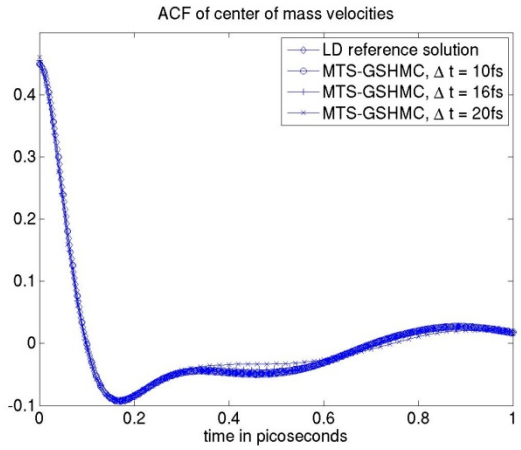

Fig. 4 Autocorrelation functions (ACFs) of diatomic center of mass velocities for MTS-GSHMC as a function of the outer time-step *Δt*.

stability barrier also limits the possible efficiency gains for MTS-GSHMC and we expect larger possible increases of outer-step sizes for simulations of biomolecular systems. Finally, note that short range contributions of the Lennard-Jones interactions have been treated as part of the fast forces in [37,38] which again points towards larger achievable outer time-steps for MTS-GSHMC in biomolecular applications.

IV. Meso-GSHMC

1. Motivation

Mesoscopic scales (or meso-scales) are transitional regions between macroscopic and microscopic regimes. Mesoscopic phenomena embrace a diverse range of systems including liquid crystals, colloids, biomembranes, micro-/nano-channels, polymer solutions and melts. However, simulation at mesoscopic level is beyond the practical limits of atomistic (microscopic) simulation methods. Prediction of mesoscopic phenomena is not within the reach of continuum (macroscopic) simulations either as the latter neglect all microstructure.

Dissipative particle dynamics (DPD) [42-51] has become a powerful and popular method to perform meso-scale simulations. Its computational cost scales linearly with the number of particles if the DPD algorithm is properly implemented, and hence very large systems can be modelled. The method can be used in complex-geometry domains. Mathematically, DPD predicts the behavior of systems consisting of particles which are interacting through a combination of conservative, dissipative and fluctuation forces.

Following [74] the DPD method can be formulated via a stochastic differential equation (SDE):

$$dX = M^{-1}Pdt,$$
$$dP = -\nabla_X U(X)dt - \sum_{k=1}^{K} \nabla_X h_k(X)\left[\gamma \dot{h}_k(X)dt - \sigma dW_k\right] \qquad (26)$$

where

$$\dot{h}_k(X) = \nabla_X h_k(X)\cdot V = \nabla_X h_k(X)\cdot M^{-1}P, \qquad (27)$$

and the functions $h_k(X), k = 1,...K,$ can be chosen quite arbitrarily. To reproduce a constant temperature ensemble, the friction coefficients $\gamma > 0$ and the noise amplitudes $\sigma > 0$ have to satisfy the fluctuation dissipation relation

$$\sigma = \sqrt{2k_B T\gamma}\ . \qquad (28)$$

Finally, W_k are independent Wiener processes.

An intriguing aspect of the DPD equations (26) is that they can be made to satisfy Newton's third law which implies conservation of total linear momentum as well as total angular momentum.

Despite its advantages, DPD has certain practical problems. For example, none of the existing numerical implementations of DPD can reproduce correctly the simulation temperature under the full DPD dynamics. Since the fluctuation-dissipation terms in DPD can be comparable to the conservative contributions, the non-preservation of thermodynamic equilibrium properties poses a serious obstacle for practical simulations.

A similar problem arises in classical molecular simulations when they are performed under constant temperature using a thermostat.

In this section we demonstrate how to implement a DPD –type momentum refreshment step within the GSHMC methodology such that the resulting method, meso-GSHMC, samples exactly from the canonical distributions and can be applicable to meso-scale particle simulations.

2. Momentum Update in Meso-GSHMC

To derive a time-stepping method for stochastic thermostats with position-dependent fluctuation-dissipation terms such as DPD, the composition approach can be applied [62,64]. It consists of a composition of two time-stepping methods: one for the conservative part and one for the fluctuation-dissipation part. Note that none of these methods is necessarily the exact propagator to the corresponding stochastic or deterministic equations of motion. As before (see section II), we suggest to discretize the conservative dynamics part by the standard Störmer-Verlet method. For the fluctuation-dissipation part, we proposed a methodology [75] that exactly samples from the underlying Boltzmann distribution and can be applied to arbitrary position dependent fluctuation-dissipation terms. The key idea is to obtain a numerical momentum update step in (26) as the solution of a generating differential equation for the fluctuation-dissipation part of the general DPD formulation (26):

$$\frac{dP}{ds} = \sum_{k=1}^{K} \nabla_X h_k(X) R_k, \qquad (29)$$
$$\frac{dR_k}{ds} = -\nabla_X h_k(X)\cdot M^{-1}P, \quad k = 1,.....K$$

at

$$s = \sqrt{2\gamma\Delta t}\ , \ \Delta t \text{ is the step-size} \qquad (30)$$

for given initial conditions

$$P(0) = P^n, \quad R_k(0) = R_k{}^n \sim N[0, k_B T], \quad k = 1,.....K. \qquad (31)$$

The solution conserves the extended canonical distribution

$$\rho_{ext\_canon} \propto \exp(-\beta H_{ext}), \qquad (32)$$

where

$$H_{ext}(P,R) = H + \frac{1}{2}R^T R. \qquad (33)$$

Using the implicit midpoint rule

$$P' = P + \frac{\Delta s}{2}\sum_{k=1}^{K}\nabla_X h_k(X)(R'_k + R_k),$$
$$R'_k = R_k - \frac{\Delta s}{2}\nabla_X h_k(X)\cdot M^{-1}(P'+P), \quad k = 1,.....K \tag{34}$$

the solution can be found by a simple fixed point iteration or some other iterative solver for Δs sufficiently small. The suggested procedure can be viewed as a momentum update step compatible to DPD. This completes the time-stepping method for the fluctuation-dissipation part.

The proposed composed approximation of two methods does not exactly sample from the *N*VT ensemble since the conservative time-stepping method does not simultaneously conserve energy and volume. The Metropolis acceptance criterion can be applied for correction of numerically induced errors in the conservation of total energy in the way as it was done in GSHMC (see section II).

Also, as in the two previously described GSHMC-based methods, in meso-GSHMC we want to increase the acceptance rate in the conservative dynamics part through the use of a shadow Hamiltonian. Thus we replace the canonical density (4) with (8) in the conservative time-stepping method. Similarly, at the DPD momentum update step the extended canonical density (32) has to be substituted by

$$\tilde{\rho}_{ext\_canon} \propto \exp(-\beta\hat{H}_{\Delta t} - \beta/2\sum_{k=1}^{K}(R_k)^2)\,. \tag{35}$$

Thus, we have obtained an extension of the GSHMC method to the generalized DPD equations (26), which we call meso-GSHMC. See [75] and the algorithmic summary below for details.

3. Algorithmic Summary of Meso-GSHMC

The meso-GSHMC method is defined through an energy/Hamiltonian (1), a shadow energy $\hat{H}_{\Delta t}$, inverse temperature $\beta = 1/k_B T$, a set of position-dependent functions, $h_k(X), k = 1,..K$, friction constant $\gamma > 0$, time-step Δt, and number of time-steps N_{MD}. We now summarize a single step of the meso-GSHMC method.

Step 1. Given (X_0, P_0), numerically integrate the Hamiltonian equations of motion (3) for a fixed number of steps N_{MD} with time step Δt with the Störmer-Verlet method to generate a new pair (X_{new}, P_{new}).

Step 2. Evaluate modified Hamiltonians $\hat{H}_{\Delta t}$ at (X_0, P_0) and at (X_{new}, P_{new}) using either the methodology described in [3] or the method for evaluating shadow Hamiltonians as proposed by Skeel and Hardy [65].
Accept the new state (X_{new}, P_{new}) with probability

$$\min\left\{1, \exp(-\beta[\hat{H}_{\Delta t}(X_{new}, P_{new}) - \hat{H}_{\Delta t}(X_0, P_0)])\right\}. \tag{36}$$

If accepted, choose (X_{new}, P_{new}) as new positions X and momenta P, otherwise take X_0 and negate momenta P_0.

Step 3. Generate a sequence of i.i.d. random numbers $R_k \sim N[0, k_B T], \quad k = 1,.....K$. Choose initial momentum and fixed position to be equal to P and X respectively. Using the implicit midpoint rule implementation of the momentum refreshment step, solve the system (34) for (P', R_k') by fixed point iteration.

Step 4. Evaluate modified Hamiltonians $\hat{H}_{\Delta t}$ at (X, P'). The new momenta P'' are accepted with probability

$$\min\left\{1, \frac{\exp(-\beta[\hat{H}_{\Delta t}(X, P') + \frac{1}{2}\sum_{k=1}^{K}(R'_k)^2])}{\exp(-\beta[\hat{H}_{\Delta t}(X, P) + \frac{1}{2}\sum_{k=1}^{K}(R_k)^2])}\right\}. \tag{37}$$

The newly accepted pair of position and momentum vectors is provided by X (from the conservative dynamics part) and P'' (from the momentum refreshment step), respectively if P'' is accepted or by (X, P') otherwise.

Step 5. Return to Step 1.

Since the meso-GSHMC method samples with respect to a modified canonical ensemble, it is necessary to re-weight the computed sample of the observables according to (12).

4. Numerical Test

The numerical experiment is conducted for Model A as described in [47] and compared to the MD-VV implementation of DPD [47].

We consider a total of $N = 4000$ particles with mass $m = 1$ in a cubic domain of size 10×10×10 with periodic boundary conditions. The conservative forces are set equal to zero, i.e., the Hamiltonian (1) reduces to

$$H(X,P) = \frac{1}{2}P^T M^{-1} P$$

and the resulting equations of motion can be solved exactly. Since proposal steps are always accepted, meso-GSHMC is also equivalent to a stochastic splitting method without Metropolis test. We set $N_{MD} = 1$ in the conservative dynamics part of meso-GSHMC, i.e., $\tau = \Delta t$ and perform experiments for different values of the step-size Δt. For the chosen units, we have $k_B T = 1$ in dimensionless variables. The functions $h_k(r) = \varphi(r), k = 1,..K$ for the momentum refreshment step are defined via

$$\varphi'(r) = \begin{cases} 1 - r/r_c, & r \le r_c \\ 0, & r > r_c \end{cases}. \tag{38}$$

The cutoff distance is set equal to $r_c = 1$. We also use $\Delta s = \sqrt{2\gamma\Delta t}$ in (34) with $\gamma = 4.5$.

The reference experiments with the DPD method [47] use the same parameter settings. The numerical results from the meso-GSHMC and traditional DPD simulations can be found in **Fig. 5-6**. Following the argument of [50], Fig. 5 demonstrates that the deviations of $g(r/r_c)$ the numerically computed radial distribution function, from its exact value $g = 1$ is of purely statistical origin (finite sample size) for the meso-GSHMC method. The same does not hold for the DPD method unless $\Delta t \le 0.05$. We also note that the meso-GSHMC method exactly reproduces the target inverse

temperature β = 1 for all values of Δt, while the DPD method leads to a nearly linear increase in the numerically observed temperature with respect to the step-size Δt (see Fig. 6).

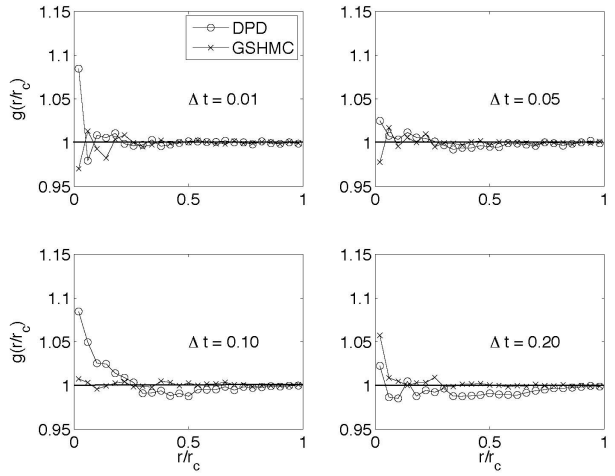

Fig. 5 Radial distribution function $g(r/r_c)$ for different values of the step–size Δt in Model A.

Further results for other DPD time-stepping methods can be found in the study [48] which recommended the methods of

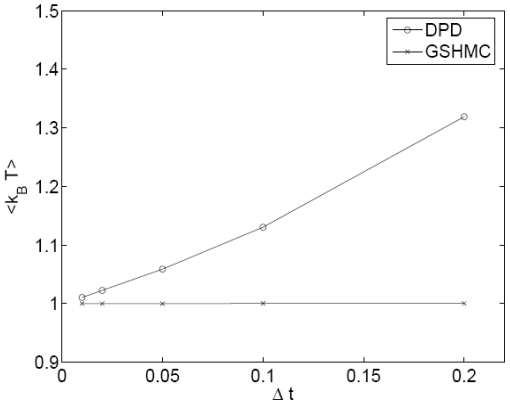

Fig. 6 Numerically observed temperature $\langle k_B T \rangle$ vs. the step-size Δt in Model A.

Lowe [45] and Shardlow [46] as promising candidates for DPD simulations. While the Lowe scheme leads to exact temperature preservation similar to the meso-GSHMC method, the Shardlow scheme leads to a small systematic shift in temperature under force free motion. However we have to stress that the meso-GSHMC is the only method to sample correctly under the full DPD setting.

V. GSHmMC

GSHMC can be applied to solve statistical inference problems in the same manner as the hybrid Monte Carlo (HMC) method [57].

The target posterior distribution can always be written out explicitly, up to a normalization constant, as

$$\pi(X) \propto f(Y \mid X)\pi_0(X) \equiv \exp(-U(X)), \tag{39}$$

where f is the probabilistic model that connects data Y with unknown parameters X, π_0 is the prior distribution in X (which is often assumed to be Gaussian), and

$$U(X) = -\log f(Y \mid X) - \log \pi_0(X) \cdot \tag{40}$$

In order to use GSHMC to sample the posterior distribution (39), we introduce an auxiliary 'momentum' variable P, a (constant) symmetric positive-definite 'mass matrix' M, and the guide Hamiltonian

$$H(X,P) = \frac{1}{2} P^T M^{-1} P + U(X) \tag{41}$$

with associated Newtonian equations of motion

$$\dot{X} = \nabla_P H(X,P) = M^{-1}P, \qquad \dot{P} = -\nabla_X H(X,P) = -\nabla U(X). \tag{42}$$

The matrix M takes the role of a preconditioner, i.e. one should select M such that the dynamics of the guide Hamiltonian system is as uniform as possible to enhance sampling. The 'mass matrix' is typically diagonal and is often a scalar multiple of the identity matrix. The equations (42) look exactly like (3) but 'position', 'momentum' and 'mass' in (42) are entirely artificial and do not have physical meaning. These equations can be integrated in time by a symplectic and time-reversible method such as Störmer-Verlet.

In order to sample (39) with GSHMC the partial derivatives of the log of the density function have to exist except perhaps on a set of points with zero probability, for which some fictitious value could be returned.

There are two principal differences between GSHmMC and Hamiltonian Monte Carlo which, in fact, make GSHmMC more efficient for sampling than Hamiltonian MC, particularly for high dimensionality systems.

First, in the GSHMC method guide Hamiltonians in the Metropolis criterion are replaced with modified guide Hamiltonians which are asymptotic expansions of guide Hamiltonians in powers of the step size and are more sensitive indicators than guide Hamiltonians of drift in the 'energy' caused by instability. As a result, as already was discussed in this paper and in [3], one can achieve nearly 100% acceptance rate independent of the system size provided that high-order modified guide Hamiltonians are used. The modified guide Hamiltonians are computed diagnostically at the beginning and the end of a simulation interval only and produce little computational overhead. We have to stress here that GSHmMC is designed as a purely sampling device. Momentum flips described in section II.2 and claimed to be a drawback of GSHMC do not have any negative effect on the performance of GSHmMC. Moreover as it was discussed in [61] momentum flips improve sampling efficiency of generalized hybrid Monte Carlo methods.

Second, instead of completely resetting 'momentum' as done in Hamiltonian MC, GSHmMC introduces a partial

momentum refreshment procedure accompanied by a modified Metropolis acceptance criterion. Such a procedure improves the overall sampling rate [4].

The partial momentum refreshment part and the re-weighting procedure for expectation values in GSHmMC do not differ from the ones introduced in GSHMC.

Therefore the GSHmMC algorithm has precisely the same formulation as GSHMC as given in section II.2 even though it uses different definitions.

There is a potential for further increase in efficiency of GSHmMC. Thus, improving the accuracy of the integration phase by replacing the 'standard' Störmer-Verlet integrator with a more appropriate one, e.g. linearly implicit methods, will allow keeping a comparatively large step size even for highly oscillatory guide Hamiltonians. Combining partial momentum refreshment with the "window" method by Neal [76] may lead to a substantial improvement which would increase with dimensionality [57].

The well known bottleneck of the Hamiltonian MC method, also inherited by GSHmMC, is the amount of tuning required to obtain reasonable rates of acceptance. The number of integration steps, the size of each step, parameter φ in (9) and the 'mass' matrix are the parameters to tune. Whilst the choice of the first three parameters can be tuned based on the overall acceptance rate of the GSHmMC sampler using different adaptive schemes, (see for example [77-79]), it is unclear how to select the values of the weight matrix M in any automated manner that does not require some knowledge of the target density.

One of the possible ways to overcome this problem was suggested in [59]. The authors picked up the idea of Zlochin and Baram [80] to define the Hamiltonian in general form on a Riemann manifold. Then the deterministic proposal is guided not only by the derivative information of the target density but also by the local geometric structure of the manifold as determined by the metric tensor. As a result, matrix M, defining a globally constant metric, is now replaced with the position specific metric. This removes the need for tuning the 'mass' matrix that dramatically affects the performance of Hamiltonian MC. They also suggested using the time reversible volume preserving Generalized Leapfrog algorithm [64] for solving the non-separable Hamiltonian equations to ensure a correct MCMC scheme satisfying detailed balance and convergence to the desired target density.

Unfortunately essential overheads are associated with this approach. One is the developing of analytical or numerical procedures for calculating the metric tensor and the associated derivatives. The second is the $O(N^3)$ scaling of solving the linear systems when inverting the metric tensor.

This approach, called RM-HMC, can be implemented in GSHmMC. However, we recommend using it mainly for high dimensional problems with strong correlations such as inferring the highly dimensional latent Gaussian field [81] where HMC fails due to the high levels of spatial correlation in the latent field.

Combining GSHmMC with "tempering" methods such as parallel tempering [82], simulated tempering [83], tempered transitions [84], annealed importance sampling [85] to enhance sampling efficiency of the method is also feasible. The further ideas for performance improvement, theoretical analysis of limiting rates of convergence, ergodicity, optimal step sizes and acceptance rates for HMC-based methods can be found in [86].

As to the computational aspects the automatic or adjoint differentiation methods [87] may be of use in GSHmMC.

One of the important applications where GSHmMC can be effective is data assimilation. Data assimilation is a technique for combining mathematical models of physical systems with measurements in order to estimate either the state of the system or the parameters of the model. Mathematically, data assimilation leads to a deformation of the underlying probability density functions which is often achieved by Monte Carlo type methods (e.g. particle and ensemble Kalman filters) [88]. As already pointed out, the necessary sampling of posterior probability density functions can be achieved efficiently by GSHmMC provided the density functions are sufficiently smooth.

VI. Conclusion

We introduced a class of hybrid methods which were designed to make feasible efficient and detailed simulation of large and complex systems.

The methods are based on Hamiltonian Dynamics and Monte Carlo. They provide an exact temperature control during the molecular simulation and are able to extract dynamical information in a rigorous manner. They proved to be efficient sampling tools and overcome certain problems associated with MD, MC, conventional hybrid MC methods, MTS approaches and the simulation techniques commonly used for mesoscopic simulation.

Due to their flexibility the introduced methods are easy to combine with other enhanced sampling methods and to efficiently implement on modern supercomputers.

References

1) J. A. Izaguirre and S. S. Hampton, "Shadow Hybrid Monte Carlo: An efficient propagator in phase space of macromolecules," *J. Comput. Phys.*, **200,** 581–604 (2004).
2) E. Akhmatskaya, S. Reich, "The targeted shadowing hybrid Monte Carlo (TSHMC) method," in B. Leimkuhler et al, editor, *New Algorithms for Macromolecular Simulations*, volume 49 of Lecture Notes in Computational Science and Engineering, Berlin, Springer-Verlag, 145–158 (2006).
3) E. Akhmatskaya, S. Reich, "GSHMC: An efficient method for molecular simulations," *J. Comput. Phys*., **227**, 4934–4954 (2008).
4) C. L. Wee, M. S. P. Sansom, S. Reich, E. Akhmatskaya, "Improved sampling for simulations of interfacial membrane proteins: Application of generalized shadow hybrid Monte Carlo to a peptide toxin/bilayer system," *J. Phys. Chem. B*, **112**, 5710–5717 (2008).
5) J. M. Hammersley, D. C. Handscomb, *Monte Carlo Methods*, Chapman and Hall, London, 1 (1964).
6) L. Kelvin, "Nineteenth century clouds over the dynamical

theory of heat and light*," Philos. Mag*., **2**,1 (1901).
7) E. Serge, *From X-Rays to quarks*, W.H. Freeman & Co., San Francisco (1980).
8) N. Metropolis, A. W. Rosenbluth, M. N. Rosenblut, A. H. Teller, E. Teller, "Equation of state calculations by fast computing machines," *J. Chem. Phys*., **21**, 1087–1092 (1953).
9) J. Dongarra, F. Sullivan, "Guest Editors' Introduction: The Top 10 Algorithms," *Comput. Sci. Eng*., **2**[1], 22–23 (2000).
10) B. J. Alder, T. E. Wainwright, "Phase transition for hard sphere system," *J. Chem. Phys*., **27**, 1208 (1957).
11) A. Brass, B. J. Pendleton, Y. Chen, B. Robson, "Hybrid Monte Carlo simulations theory and initial comparison with molecular dynamics," *Biopolymers*, **33**, 1307–1315 (1993).
12) S. Plimpton, B. Hendrickson, "A new parallel method for molecular dynamics simulation of macromolecular systems," *J. Comp. Chem*., **17**, 326-337 (1996).
13) J. C. Phillips, G. Zheng, S. Kumar, L. V. Kalé, "NAMD: biomolecular simulation on thousands of processors," in *Proceedings of SC* 2002, 1-18 (2002).
14) S. Kumar, Chao Huang, G. Almasi, L. V. Kale, "Achieving strong scaling with NAMD on blue gene/l," in *Parallel and Distributed Processing Symposium*. IPDPS 2006, 20th International, 10 (2006).
15) B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, "Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation," *Journal of Chemical Theory and* Computation, **4**, 435-447 (2008).
16) M. Snir, "A note on n-body computations with cutoffs," *Theory of Computing Systems*, **37**, 295-318 (2004).
17) Y. Shan, J. L. Klepeis, M. P. Eastwood, R. O. Dror, D. E. Shaw, "Gaussian split ewald: A fast ewald mesh method for molecular simulation," *The Journal of Chemical Physics*, **122**, 054101 (2005).
18) K. J. Bowers, E. Chow, H. Xu, R. O. Dror, M. P. Eastwood, B. A. Gregersen, J. L. Klepeis, I. Kolossvary, M. A. Moraes, F. D. Sacerdoti, J. K. Salmon, Y. Shan, D. E. Shaw, "Scalable algorithms for molecular dynamics simulations on commodity clusters", Tampa, Florida, USA (2006).
19) K. J. Bowers, R. O. Dror, D. E Shaw, "Zonal methods for the parallel execution of range-limited n-body problems," *J. Comput. Phys*., **221**, 303-329 (2007).
20) D. E. Shaw, "A fast, scalable method for the parallel evaluation of distance-limited pairwise particle interactions," *J. Comput. Chem*., **26**, 1318-1328 (2005).
21) K. J. Bowers, R. O. Dror, D. E. Shaw, "The midpoint method for parallelization of particle simulations," *The Journal of Chemical Physics*, **124**, 184109 (2006).
22) J. S. Rosenthal, "Parallel computing and Monte Carlo algorithms," *Far East Journal of Theoretical Statistics*, **4**, 207–236 (2000).
23) I. Azzini, R. Girardi, M. Ratto, "Paralellization of matlab codes under windows platform for Bayesian estimation: A dynare application," *Working Paper 1*, Euro-area Economy Modelling Centre (2007).
24) I. Strid, "Efficient parallelisation of Metropolis–Hastings algorithms using a prefetching approach," *Computational Statistics & Data Analysis*, **54**, 2814-2835 (2010).
25) S. Duane, A. Kennedy, B. Pendleton, D. Roweth, "Hybrid Monte-Carlo," *Phys. Lett. B*, **195**, 216-222 (1987).
26) R. Faller, J. J. de Pablo, "Constant pressure hybrid Molecular Dynamics–Monte Carlo simulations," *J. Chem. Phys*., **116**, 55-59 (2002).
27) M. Creutz, "Global Monte Carlo algorithms for many-fermion systems," *Phys. Rev. D*, **38** [4], 1228–1238 (1988).
28) A. D. Kennedy, B. Pendleton, "Acceptances and autocorrelations in hybrid Monte Carlo," *Nucl. Phys. B (Proc. Suppl.),* **20**, 118–121 (1991).
29) A. M. Horowitz, "A generalized guided Monte-Carlo algorithm," *Phys. Lett. B*, **268**, 247–252 (1991).
30) A. D. Kennedy, B. Pendleton, "Cost of the generalized hybrid Monte Carlo algorithm for free field theory," *Nucl. Phys. B*, **607**, 456–510 (2001).
31) J. S. Liu, *Monte Carlo strategies in scientific computing*, Springer-Verlag, New York (2001).
32) C.W. Gardiner, *Handbook on Stochastic Methods*, third ed., Springer-Verlag (2004).
33) C. R. Sweet, S. S. Hampton, R. D. Skeel, J. A. Izaguirre, "A separable shadow Hamiltonian hybrid Monte Carlo method," *J Chem Phys*., **131**[17], 174106 (2009).
34) H. Grubmüller, H. Heller, A. Windemuth, K. Schulten, "Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions," *Mol. Sim*., **6**, 121-142 (1991).
35) D. Humphreys, R. Friesner, B. Berne, "A multiple-time-step molecular dynamics algorithm for macromolecules," *J. Phys. Chem*., **98[**27], 6685-6892 (1994).
36) Q. Ma, J. Izaguirre, R. Skeel, "Verlet-I/R-Respa/Impulse is limited by nonlinear instabilities," *SIAM J. Sci. Comput*., **24** [6]. 1951-1973 (2003).
37) J. Izaguirre, D. Catarello, J. Wozniak, R. Skeel, "Langevin stabilization of molecular dynamics," *J. Chem. Phys*., **114**, 2090-2098 (2001).
38) J. Izaguirre, S. Reich, R. Skeel, "Longer time steps for molecular dynamics," *J. Chem. Phys*., **110**, 9853-9864 (1999).
39) G. Zhang, T. Schlick, "LIN: A new algorithm to simulate the dynamics of biomolecules by combining implicit-integration and normal mode techniques," *J. Comp. Chem*., **14**, 1212 (1993).
40) R. Pastor, B. Brooks, A. Szabo, "An analysis of the accuracy of Langevin and molecular dynamics algorithms," *Mol. Phys*., **65**, 1409-1419 (1988).
41) S. Bond, B. Leimkuhler, "Molecular dynamics and the accuracy of numerically computed averages," *Acta Numerica*, **16**, 1–65 (2007).
42) P. J. Hoogerbrugge, J. M. V. A. Koelman, "Simulating microscopic hydrodynamic phenomena with dissipative particle dynamics," *Europhys. Lett*, **19**, 155–160 (1992).
43) P. Español, P. B. Warren, "Statistical mechanics of dissipative particle dynamics," *Europhys.Lett*., **30**, 191–196 (1995).
44) I. Pagonabarraga, M. H. J. Hagen, D. Frenkel, "Self-consistent dissipative particle dynamics," *Europhys. Lett*., **42**, 377–382 (1998).
45) C. P. Lowe, "An alternative approach to dissipative particle dynamics," *Europhys. Lett*., **47**, 145–151 (1999).
46) T. Shardlow, "Splitting for dissipative particle dynamics," *SIAM J. Sci. Comput*., **24**, 1267–1282 (2003).
47) I. Vattulainen, M. Karttunen, B. Besold, J. M. Polson, "Integration schemes for dissipative particle dynamics simulations: From softly interacting systems towards hybrid models," *J. Chem. Phys*., **116**, 3967–3979 (2002).
48) P. Nikunen, M. Karttunen, I. Vattulainen, "How would you integrate the equations of motion in dissipative particle dynamics," *Computer Physics Communications*, **153**, 407–423 (2003).

49) E. A. Koopman, C. P. Lowe, "Advantages of a Lowe-Andersen thermostat in molecular dynamics simulations," *J. Chem. Phys*., **124**, 204103 (2006).
50) E. A. J. F. Peters, "Elimination of time step effects in DPD," *Europhys. Lett*., **66**, 311–317 (2004).
51) M. Serrano, G. De Fabritiis, P. Español, P. V. Coveney, "A stochastic Trotter integration scheme for dissipative particle dynamics," *Mathematics and Computers in Simulation*, **72**, 190–194 (2006).
52) G. O. Roberts, R. L. Tweedie, "Exponential convergence of langevin diffusions and their discrete approximations," *Bernoulli*, **2**[4], 341–363 (1995).
53) R. M. Neal, "Probabilistic inference using Markov Chain Monte Carlo methods," *Technical report CRG-TR-93-1*, Dept. of Computer Science, University of Toronto (1993).
54) R. M. Neal, *Bayesian Learning for Neural Networks*, Springer-Verlag, New York (1996).
55) P. Gustafson, "Large hierarchical Bayesian analysis of multivariate survival data," *Biometrics*., **53**, 230 – 242 (1997).
56) H. Ishwaran, "Applications of hybrid Monte Carlo to generalized linear models: quasicomplete separation and neural networks", *Journal of Computational and Graphical Statistics*, **8**, 779–799 (1999).
57) R. N. Neal, "MCMC using Hamiltonian dynamics," *Handbook of Markov Chain Monte Carlo*, Chapman & Hall / CRC Press (2010).
58) M. N. Schmidt, "Function factorization using warped Gaussian processes," *Proceedings,* Twenty-Six International Conference on Machine Learning (2009).
59) M. Girolami, B. Calderhead, "Riemannian Manifold Langevin and Hamiltonian Monte Carlo methods, *J.R.Statist.Soc. B*, **73**[2], 1-37 (2011).
60) S. H. Cheung, J. L. Beck, "Bayesian model updating using hybrid Monte Carlo simulation with application to structural dynamic models with many uncertain parameters," *Journal of Engineering Mechanics*, **135**[4], 243-255 (2009).
61) E. Akhmatskaya, N. Bou-Rabee, S. Reich, "Generalized hybrid Monte Carlo methods without momentum flip," *J. Comput. Phys*., **228**, 2256–2265 (2009).
62) M. Allen, D. Tildesley, *Computer Simulation of Liquids*. Clarendon Press, Oxford (1987).
63) E. Hairer, C. Lubich, G. Wanner, *Geometric Numerical Integration*, Springer-Verlag, Berlin Heidelberg (2002).
64) B. Leimkuhler, S. Reich, *Simulating Hamiltonian Dynamics*, Cambridge University Press, Cambridge (2005).
65) R. Skeel, D. Hardy, "Practical construction of modified Hamiltonians," *SIAM J. Sci. Comput*., **23**, 1172-1188 (2001).
66) Y. Sugita, Y. Okamoto, "Replica exchange multicanonical algorithm and multicanonical replica-exchange method for simulating systems with rough energy landscape," *Chem. Phys. Lett*., **329**, 261 (2000).
67) Y. M. Rhee, V. S. Pande, "Multiplexed-replica exchange molecular dynamics method for protein folding simulation", *Biophys. J.,* **84**, 775 (2003).
68) L. Walter, M. Weber, "ConfJump: a fast biomolecular sampling method which drills tunnels through high mountains," *Technical report*, ZIB (2006).
69) C. H. Bennett, "Efficient Estimation of Free Energy Differences from Monte Carlo Data," *J. Comput. Phys*., **22**, 245 (1976).
70) P. J. Bond, M. S. P. Sansom, "Insertion and assembly of membrane proteins via simulation," *J. Amer. Chem. Soc*., **128**[8], 2697 (2006).
71) H. Fujitani, Y. Tanida, M. Ito, G. Jayachandran, C. D. Snow, M. R. Shirts, E. J. Sorin, V. S. Pande, "Direct calculation of the binding free energies of FKBP ligands," *J. Chem. Phys*., **123**, 084108 (2005).
72) T. Schlick, *Molecular Modeling and Simulation*, Springer-Verlag, New York (2002).
73) S. Reich, E. Akhmatskaya, "A multiple-time-stepping shadow generalized hybrid Monte Carlo (MTS-GSHMC) method.," *Technical Report*, Fujitsu Laboratories of Europe (2009).
74) C. J. Cotter, S. Reich, "An extended dissipative particle dynamics model," *Europhys. Lett*., **64**, 723–729 (2003).
75) E. Akhmatskaya, S. Reich, "Meso-GSHMC: a stochastic algorithm for meso-scale constant temperature simulations," *Technical Report*, submitted (2010).
76) R. M. Neal, "An improved acceptance procedure for the hybrid Monte Carlo algorithm," *Journal of Computational Physics*, **111**, 194-203 (1994).
77) C. Andrieu, J. Thoms, "A tutorial on adaptive MCMC," *Statistics and Computing*, **18**, 343–373 (2008).
78) R. M. Neal, "The short-cut Metropolis method," *Technical Report*, N 0506, Department of Statistics, University of Toronto (2005).
79) R. M. Neal, "Short-cut MCMC: An alternative to adaptation," *Third Workshop on Monte Carlo Methods*, Harvard (2007).
80) M. Zlochin, Y. Baram, "Manifold stochastic dynamics for Bayesian learning," *Neural Computation*, **13**, 2549–2572 (2001).
81) O. F. Christensen, G. O. Roberts, J. S. Rosenthal, "Scaling limits for the transient phase of local Metropolis-Hastings algorithms," *Journal of the Royal Statistical Society: Series B*., **67**[2], 253–268 (2005).
82) C. J. Geyer, "Markov chain Monte Carlo maximum likelihood," in E. M. Keramidas, editor, *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, 156-163 (1991).
83) E. Marinari, G. Parisi, "Simulated tempering: A new Monte Carlo scheme," *Europhysics Letters*, **19**, 451-458 (1992).
84) R. M. Neal, "Sampling from multimodal distributions using tempered transitions," *Statistics and Computing*, **6** (1996).
85) R. M. Neal, "Annealed importance sampling. Statistics and Computing," **11**, 125-139 (2001).
86) A. Beskos, N. Pillai, G. Roberts, S. Serna, A. Stuart, "Optimal tuning of the Hybrid Monte-Carlo algorithm," *Technical Report*, Department of Statistical Science, UCL (2010).
87) K. M. Hanson, "Use of probability gradients in Hybrid MCMC and a new convergence test," *Los Alamos Report LA-UR-02-4105*, summary of talk presented at 7th Valencia International Meeting on Bayesian Statistics (2002).
88) G. Evensen, *Data Assimilation: The Ensemble Kalman Filter*, 2nd edition, Springer-Verlag, New York (2009).