

Optimal implementation of the Shadow Hybrid Monte Carlo method.

Christopher R. Sweet, Scott S. Hampton, Jesús A. Izaguirre

Department of Computer Science

and Engineering

University of Notre Dame

Notre Dame, IN 46556

USA

July 12, 2006

Abstract

Abstract.

keywords: Sampling methods; Hybrid Monte Carlo; Symplectic integrator; Shadow Hamiltonian; Conformational sampling.

1 Introduction

This paper provides an analytical framework for the implementation of the Shadow Hybrid Monte Carlo (SHMC) method [5] allowing the additional parameter required by SHMC, on which the efficiency of the method and quality of observables is dependent, to be determined explicitly. The SHMC method [5] is a generalization of the Hybrid Monte Carlo method (HMC) which samples from the extended phase-space of the shadow Hamiltonian. This approach overcomes the exponentially decreasing acceptance rate of the HMC method with increasing system size N or time step Δt . The formulation of the SHMC method requires the introduction of a parameter c since the shadow Hamiltonian $\hat{\mathcal{H}}$ can have a significant separation from the original Hamiltonian \mathcal{H} as seen in Figure 1, which is generally positive for molecular dynamics simulations. Both the re-weighting and momenta generation steps are determined by exponential functions of the Hamiltonian difference $\Delta\mathcal{H} = \hat{\mathcal{H}} - \mathcal{H}$, the parameter c is used to reduce this difference to an acceptable value by choosing the Hamiltonian used for sampling as

$$\tilde{\mathcal{H}} = \max\{\mathcal{H}, \hat{\mathcal{H}} - c\}. \quad (1)$$

In the original work [5] the c parameter was determined empirically, usually by an iterative procedure based on some initial guess for the value.

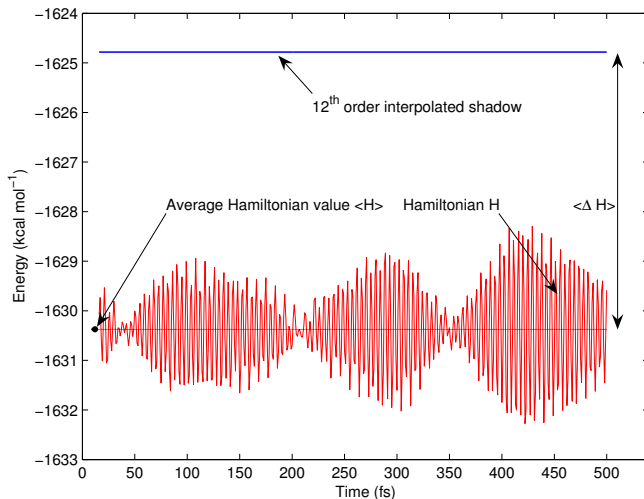


Figure 1: Hamiltonian \mathcal{H} , average Hamiltonian $\langle \mathcal{H} \rangle$ and shadow Hamiltonian $\hat{\mathcal{H}}$ for a box of 216 flexible water molecules with a time-step of 1.0 fs. Here $\Delta \mathcal{H} = \hat{\mathcal{H}} - \mathcal{H}$.

It has been shown [6] that using a symplectic numerical integrator when propagating the equations of motion derived from a Hamiltonian that the numerical results are the exact solution to a shadow Hamiltonian, which can be determined either by an asymptotic expansion in the time-step or numerically using a cheap and arbitrarily accurate approximation [2]. Since the shadow Hamiltonian is exactly conserved the molecular dynamics (MD) step in the SHMC method will have a very high probability of acceptance but the resulting observables then need to be re-weighted to eliminate the bias introduced by the shadow Hamiltonian and obtain the correct canonical averages. It was observed by Izaguirre and Hampton [5] that the variance of observables generated by SHMC was greater than that when using HMC and is dependent on the additional SHMC parameter c . In addition the asymptotic expansion for the shadow Hamiltonian shows that the probability density function (pdf) is not separable in momenta \mathbf{p} and position \mathbf{x} requiring that a novel method be introduced to generate momenta with the correct pdf. Again it was shown [5] that and that the momenta generation (MG) stage can have a poor acceptance rate and is dependent on c .

The Authors provide proofs that both the increase in the variance of the observables and the acceptance ratios for both the MG and MD stages are dependent only on the parameter c and the mean $\mu_{\Delta \mathcal{H}}$ and variance $\sigma_{\Delta \mathcal{H}}^2$ of $\Delta \mathcal{H}$, the difference between the shadow Hamiltonian and the Hamiltonian when calculated for sets of momenta drawn from the Boltzmann distribution. The fundamental assumptions made in the analysis are given in Section 2.

The MD stage acceptance will be dependent on the values of $\Delta \mathcal{H}$ which are accepted at the MG stage, denoted $\Delta \mathcal{G}$ with corresponding variance $\sigma_{\Delta \mathcal{G}}^2$ and mean

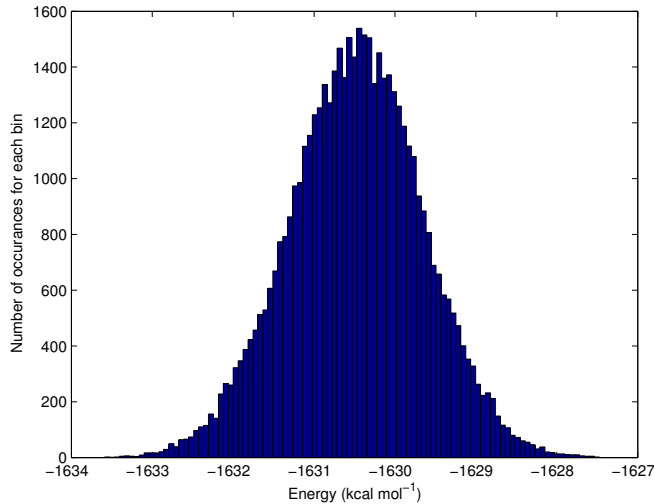


Figure 2: Distribution of the Hamiltonian for the 216 water molecule model with time-step 1.0 fs.

$\mu_{\Delta\mathcal{G}}$, and the values of $\Delta\mathcal{H}$ during the MD stage, denoted $\Delta\mathcal{D}$ with variance $\sigma_{\Delta\mathcal{D}}^2$ and mean $\mu_{\Delta\mathcal{D}}$. In Section 3 we derive the relationship between these parameters and $\mu_{\Delta\mathcal{H}}$ and $\sigma_{\Delta\mathcal{H}}$. In Section 4 the relationship between the efficiency of the method and c is determined leading to a comparison of the efficiencies of the HMC and SHMC in Section 5. In Section 6 the issue of the increase in the variance of the observables is addressed.

2 Assumptions.

The main assumption made in the proofs is that the values for the Hamiltonian are normally distributed, which is observed in practice for many systems as seen in Figure 2 for an example of a box of 216 water molecules with a time-step of 1 fs. We would also expect this from a theoretical analysis; if the solution for the phase-space variables were exact then we would have \mathcal{H} constant but since the numerical solution is exact only for the shadow Hamiltonian $\hat{\mathcal{H}}$ the variation in \mathcal{H} comes from the additional terms in the asymptotic expansion [6]

$$\hat{\mathcal{H}} = \mathcal{H} + \frac{\Delta t^2}{24} (-(\mathbf{U}')^T M^{-1} \mathbf{U}' + 2p^T M^{-1} \mathbf{U}'' M^{-1} p) + \mathcal{O}(\Delta t^4), \quad (2)$$

where M^{-1} is the matrix of inverse masses, p the momenta, \mathbf{U}' the derivative of the potential energy w.r.t. the positions and \mathbf{U}'' its Hessian. If we ignore terms in Δt^4 and higher then for sufficiently large systems, where the HMC method performs poorly, we could reasonably expect that different parts of the system are sufficiently decoupled to be regarded as independent variables and we would then expect the term in Δt^2 to be normally distributed from the Central Limit theorem.

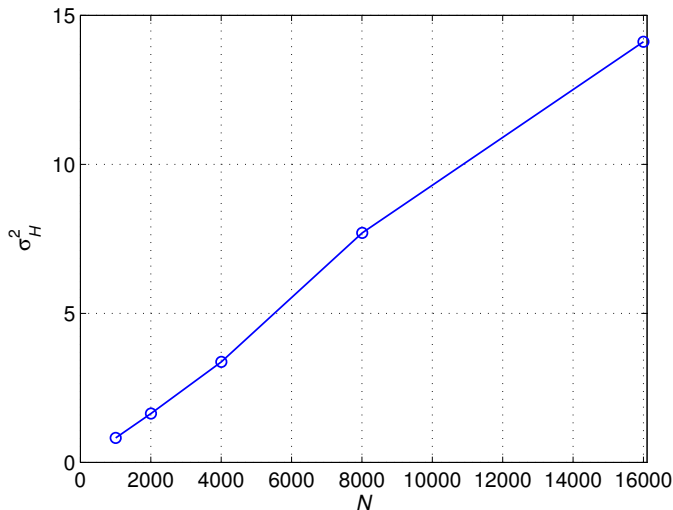


Figure 3: Change in variance of \mathcal{H} with N for boxes of water molecules when simulated with a time-step of 1 fs.

From (2) we have that $\sigma_H \propto \Delta t^2$ and hence the variance σ_H^2 scales with Δt^4 . For *independent* variables A and B we have $\sigma_{A+B}^2 = \sigma_A^2 + \sigma_B^2$ and for many molecular systems, including our boxes of water and many solvated proteins, different parts of the system are sufficiently decoupled such that if the number of molecules is increased then the variance of the Hamiltonian will increase in direct proportion. For our calculations we will assume that the variance of a system is proportional to N , the number of atoms, which coincides with the results for various size boxes of water in Figure 3.

3 The mean and variance of $\Delta\mathcal{H}$ for successful MG steps and the MD stage.

Where the derived results are dependent on the mean and variance of $\Delta\mathcal{H}$ for choices of momenta which were accepted by the MG stage we need to know the relationship to $\mu_{\Delta\mathcal{H}}$ and $\sigma_{\Delta\mathcal{H}}^2$ which is the mean and variance for all of the momenta drawn from the Boltzmann distribution. We will denote this subset of $\Delta\mathcal{H}$ as $\Delta\mathcal{G}$ with mean and variance $\mu_{\Delta\mathcal{G}}$ and $\sigma_{\Delta\mathcal{G}}^2$. We then have

Proposition 3.1 For $\Delta\mathcal{G}(\mathbf{x}, \mathbf{p}) \in \Delta\mathcal{H}(\mathbf{x}, \mathbf{p})$ where $\Delta\mathcal{G}(\mathbf{x}, \mathbf{p})$ has been selected s.t. momenta \mathbf{p} with p.d.f. $\rho(\mathbf{x}, \mathbf{p}) \propto \exp(-\beta\mathcal{H})$ has been accepted with probability $\min\{1, \exp(-\beta(\Delta\mathcal{H} - c))\}$ we have

$$\mu_{\Delta\mathcal{G}} = \mu_{\Delta\mathcal{H}} - \frac{\beta\sigma_{\Delta\mathcal{H}}^2}{1 + f_1(\mu_{\Delta\mathcal{H}}, \sigma_{\Delta\mathcal{H}}, c, \beta)} \quad (3)$$

$$\sigma_{\Delta\mathcal{G}}^2 = \sigma_{\Delta\mathcal{H}}^2 + \frac{\beta^2 \sigma_{\Delta\mathcal{H}}^4 f_1(\mu_{\Delta\mathcal{H}}, \sigma_{\Delta\mathcal{H}}, c, \beta)}{(1 + f_1(\mu_{\Delta\mathcal{H}}, \sigma_{\Delta\mathcal{H}}, c, \beta))^2} + f_3(\mu_{\Delta\mathcal{H}}, \sigma_{\Delta\mathcal{H}}, c, \beta), \quad (4)$$

where

$$f_1(\mu_{\Delta\mathcal{H}}, \sigma_{\Delta\mathcal{H}}, c, \beta) = \frac{1 - \operatorname{erf}\left(\frac{\mu_{\Delta\mathcal{H}} - c}{\sqrt{2}\sigma_{\Delta\mathcal{H}}}\right)}{f_2(\mu_{\Delta\mathcal{H}}, \sigma_{\Delta\mathcal{H}}, c, \beta)}, \quad (5)$$

$$f_2(\mu_{\Delta\mathcal{H}}, \sigma_{\Delta\mathcal{H}}, c, \beta) = \left(1 - \operatorname{erf}\left(\frac{c + \beta\sigma_{\Delta\mathcal{H}}^2 - \mu_{\Delta\mathcal{H}}}{\sqrt{2}\sigma_{\Delta\mathcal{H}}}\right)\right) \quad (6)$$

$$\times \exp\left(\frac{1}{2}\beta(2c + \beta\sigma_{\Delta\mathcal{H}}^2 - 2\mu_{\Delta\mathcal{H}})\right), \quad (7)$$

$$f_3(\mu_{\Delta\mathcal{H}}, \sigma_{\Delta\mathcal{H}}, c, \beta) = \frac{\sqrt{2}\beta\sigma_{\Delta\mathcal{H}}^3 \exp\left(-\frac{(\mu_{\Delta\mathcal{H}} - c)^2}{2\sigma_{\Delta\mathcal{H}}^2}\right)}{\sqrt{\pi} \left(\operatorname{erf}\left(\frac{(\mu_{\Delta\mathcal{H}} - c)}{\sqrt{2}\sigma_{\Delta\mathcal{H}}}\right) - 1 - f_2(\mu_{\Delta\mathcal{H}}, \sigma_{\Delta\mathcal{H}}, c, \beta)\right)}. \quad (8)$$

Proof: The MG rejection ratio will depend on $\mu_{\Delta\mathcal{H}}$ and $\sigma_{\Delta\mathcal{H}}^2$ and we can relate these values to $\mu_{\Delta\mathcal{G}}$ and $\sigma_{\Delta\mathcal{G}}^2$ since the latter terms are derived from the set of generated momenta which have been accepted according to the above probability. As described in Section 1, \mathcal{H} and hence $\Delta\mathcal{H}$ (assuming \mathcal{H} is constant) will be normally distributed which will lead to the following expression for the expected value of $\Delta\mathcal{G}$

$$\begin{aligned} \mu_{\Delta\mathcal{G}} &= \frac{\int_{-\infty}^{\infty} x \min\{\exp(-\beta(x-c)), 1\} \exp\left(\frac{-(x-\mu_{\Delta\mathcal{H}})^2}{2\sigma_{\Delta\mathcal{H}}^2}\right) dx}{\int_{-\infty}^{\infty} \min\{\exp(-\beta(x-c)), 1\} \exp\left(\frac{-(x-\mu_{\Delta\mathcal{H}})^2}{2\sigma_{\Delta\mathcal{H}}^2}\right) dx} \\ &= \frac{\int_c^{\infty} x \exp(-\beta(x-c)) \exp\left(\frac{-(x-\mu_{\Delta\mathcal{H}})^2}{2\sigma_{\Delta\mathcal{H}}^2}\right) dx + \int_{-\infty}^c x \exp\left(\frac{-(x-\mu_{\Delta\mathcal{H}})^2}{2\sigma_{\Delta\mathcal{H}}^2}\right) dx}{\int_c^{\infty} \exp(-\beta(x-c)) \exp\left(\frac{-(x-\mu_{\Delta\mathcal{H}})^2}{2\sigma_{\Delta\mathcal{H}}^2}\right) dx + \int_{-\infty}^c \exp\left(\frac{-(x-\mu_{\Delta\mathcal{H}})^2}{2\sigma_{\Delta\mathcal{H}}^2}\right) dx} \\ &= \mu_{\Delta\mathcal{H}} - \frac{\beta\sigma_{\Delta\mathcal{H}}^2}{1 + f_1(\mu_{\Delta\mathcal{H}}, \sigma_{\Delta\mathcal{H}}, c, \beta)}. \end{aligned}$$

similarly we can derive $\langle\Delta\mathcal{G}^2\rangle$ to find $\sigma_{\Delta\mathcal{G}}^2 = \langle\Delta\mathcal{G}^2\rangle - \mu_{\Delta\mathcal{G}}^2$. \square

In addition the variance and mean of the subsequent MD stage, denoted $\mu_{\Delta\mathcal{D}}$ and $\sigma_{\Delta\mathcal{D}}^2$, are required to calculate the average acceptance probability for the MD stage. The relationship of $\mu_{\Delta\mathcal{D}}$ and $\sigma_{\Delta\mathcal{D}}^2$ to $\mu_{\Delta\mathcal{G}}$ and $\sigma_{\Delta\mathcal{G}}^2$ is shown in the following proposition.

Proposition 3.2 *For values of $\Delta\mathcal{H}$ generated during the MD stage we have*

$$\mu_{\Delta\mathcal{D}} = \mu_{\Delta\mathcal{H}}, \quad (9)$$

$$\sigma_{\Delta\mathcal{D}}^2 = \sigma_{\Delta\mathcal{H}}^2 + \beta\sigma_{\Delta\mathcal{H}}^2. \quad (10)$$

Proof: The values of $\Delta\mathcal{H}$ generated during the MG stage have to be accepted according to $\min\{1, \exp(-\beta(\Delta\mathcal{H} - c))\}$ since it is clear that phase-space points

which have a higher probability in $\tilde{\mathcal{H}}$ than in \mathcal{H} cannot be generated using the proposed scheme. However the resulting $\Delta\mathcal{H}$ steps during the MD stage have no such restriction and the expected value is given by

$$\begin{aligned}\mu_{\Delta\mathcal{D}} &= \frac{\int_{-\infty}^{\infty} x \exp(-\beta x) \exp\left(\frac{-(x-\mu_{\Delta\mathcal{H}})^2}{2\sigma_{\Delta\mathcal{H}}^2}\right) dx}{\int_{-\infty}^{\infty} \exp(-\beta x) \exp\left(\frac{-(x-\mu_{\Delta\mathcal{H}})^2}{2\sigma_{\Delta\mathcal{H}}^2}\right) dx} \\ &= \mu_{\Delta\mathcal{H}} - \beta\sigma_{\Delta\mathcal{H}}^2.\end{aligned}$$

similarly we can derive $\langle\Delta\mathcal{H}^2\rangle$ for the MD stage to find $\sigma_{\Delta\mathcal{D}}^2 = \sigma_{\Delta\mathcal{H}}^2$. \square

4 The dependence of the method efficiency on c

Since the SHMC method uses a Monte Carlo method to select the initial momenta, and several MD steps are required to calculate the Shadow Hamiltonian, a large rejection ratio at this stage can have a significant effect on the efficiency of the method. For example the box of 216 water molecules simulated with a step size of 1.0 fs and $c = 3.2$ accepts an average of 1 in 122 momenta generation steps and, for $H_{[8]}$, we require 5 MD steps giving a total average of 610 steps. This compares poorly with the 300 steps which were being used in the MD stage (for this test set) giving scope for almost a factor of 3 improvement in efficiency. Since the MG acceptance probability is $P_A = \exp(-\beta(\Delta\mathcal{H} - c))$ then the value of c could be selected to reduce the number of MG steps, however as c approaches $\langle\Delta\mathcal{H}\rangle$ then half of the MD steps will be accepted based on the original Hamiltonian at both the start and end of the MD stage and hence, for large systems with large time-step, will generally be rejected. Clearly there will be some value of c that will provide optimum efficiency and we can estimate this by considering three calculations; (i) The dependence of the MG acceptance ratio on c , (ii) The dependence on c of the number of MD stages where \mathcal{H} dominates the acceptance and finally (iii) The amount of work apportioned to the MG and MD stages.

4.1 The dependence of the MG acceptance ratio on c

During the MG stage the shadow Hamiltonian $\hat{\mathcal{H}}$ and Hamiltonian \mathcal{H} values will be calculated for each set of momenta generated to sample from the Boltzmann distribution, leading to a difference $\Delta\mathcal{H}$ with mean $\mu_{\Delta\mathcal{H}}$ and variance $\sigma_{\Delta\mathcal{H}}^2$. The acceptance ratio for the MG step will be the integral over all possible values of $\Delta\mathcal{H}$ of the product of the probabilities of acceptance of the momenta P_A and the probability for that value of $\Delta\mathcal{H}$, which is assumed to be normally distributed as shown in Section 1.

Proposition 4.1 *The acceptance ratio of the MG step is*

$$R_{MG} \approx \exp\left(\frac{1}{2}\beta(2c + \beta\sigma_{\Delta\mathcal{H}}^2 - 2\mu_{\Delta\mathcal{H}})\right) \quad \text{for } 0 \leq c \leq \mu_{\Delta\mathcal{H}} - \beta\sigma_{\Delta\mathcal{H}}^2. \quad (11)$$

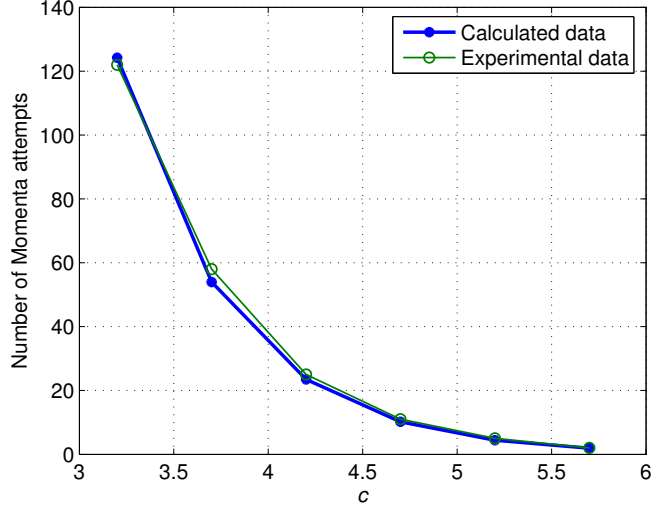


Figure 4: Predicted and experimental MG rejection ratio data for a 216 molecule box of water model.

Proof: The MG acceptance ratio R_{MG} is just

$$\begin{aligned}
R_{MG} &= \frac{1}{\sqrt{2\pi}\sigma_{\Delta\mathcal{H}}} \int_{-\infty}^{\infty} \min\{\exp(-\beta(x-c)), 1\} \exp\left(\frac{-(x-\mu_{\Delta\mathcal{H}})^2}{2\sigma_{\Delta\mathcal{H}}^2}\right) dx \\
&= \frac{1}{2} \left(1 - \operatorname{erf}\left(\frac{(\mu_{\Delta\mathcal{H}}-c)}{\sqrt{2}\sigma_{\Delta\mathcal{H}}}\right) + f_2(\mu_{\Delta\mathcal{H}}, \sigma_{\Delta\mathcal{H}}, c, \beta) \right) \\
&\approx \exp\left(\frac{1}{2}\beta(2c + \beta\sigma_{\Delta\mathcal{H}}^2 - 2\mu_{\Delta\mathcal{H}})\right) \quad \text{for } 0 \leq c \leq \mu_{\Delta\mathcal{H}} - \beta\sigma_{\Delta\mathcal{H}}^2.
\end{aligned}$$

□

We can compare data calculated from (11) to the experimentally obtained results, where $\mu_{\Delta\mathcal{H}} = 6.6120$, $\sigma_{\Delta\mathcal{H}}^2 = 0.6892$ and $\beta = 1.6667$, as seen in Figure 4 where the number of momenta attempts is $1/R_{MG}$.

4.2 The dependence of the number of MD stages where H dominates the acceptance on c

The value of c will determine the values for which the Hamiltonian \mathcal{H} is greater than $\hat{\mathcal{H}} - c$ (hereafter called ‘ \mathcal{H} -points’) and hence the MD stage acceptance is dependent on \mathcal{H} (as in the original HMC scheme). First we need to determine the ratio of the number of \mathcal{H} -points to the total number of samples, S_H .

Proposition 4.2 *The ratio of the number of \mathcal{H} -points to the total number of sam-*

ples, where $\Delta\mathcal{H}$ has expected value and variance of μ and σ^2 respectively, is

$$S = \frac{1}{2} - \frac{1}{2} \operatorname{erf} \left(\frac{\mu - c}{\sqrt{2}\sigma} \right). \quad (12)$$

Proof: The ratio of the number of \mathcal{H} -points to the total number of samples can be found by integrating the probability function for all values above the point $\hat{\mathcal{H}} - c$ (the range of the \mathcal{H} -points), then assuming again that the Hamiltonian is normally distributed

$$\begin{aligned} S &= \frac{1}{\sqrt{2\pi}\sigma} \int_{2\mu-c}^{\infty} \exp \left(\frac{-(x-\mu)^2}{2\sigma^2} \right) dx \\ &= \frac{1}{2} - \frac{1}{2} \operatorname{erf} \left(\frac{\mu - c}{\sqrt{2}\sigma} \right). \end{aligned}$$

□

Since the probability of acceptance of the non- \mathcal{H} -points is close to 1, as the shadow Hamiltonian is almost exactly conserved, a lower bound on the acceptance ratio will just be $(1 - S_{MD})(1 - S_{MG})$ where S_{MD} is the number of \mathcal{H} -points during the MD stage and S_{MG} the number after successful MG sections.

A more accurate bound can be determined by considering the points which were \mathcal{H} -points after the MG section but where not at the end of the MD stage. This condition will occur with probability $(1 - S_{MD})S_{MG}$ and are guaranteed to be accepted as the total energy is less at the end of the stage when considering $\tilde{\mathcal{H}}$. The remaining contributions will be small for large N as the variance of $\Delta\mathcal{H}$ will be large, giving

$$P_H = (1 - S_{MD})(1 - S_{MG}) + (1 - S_{MD})S_{MG}. \quad (13)$$

Data calculated from (13) and the experimental results, again with $\mu_{\Delta\mathcal{G}} = 5.5386$, $\sigma_{\Delta\mathcal{G}}^2 = 0.6651$ and $\beta = 1.6774$, show good correlation as seen in Figure 5.

4.3 The amount of work apportioned to the MG and MD stages

Given the number of molecular dynamics steps required for the MG step is n_{MG} , and the number for the MD step is n_{MD} we can find the minimum average total steps based on the value of c . For the 216 water molecule model using the $H_{[8]}$ shadow Hamiltonian we used 10 steps for each MG attempt and 400 steps for each MD attempt. If we further assume that failed MD steps are lost even though the original data is retained for that point then we can find the average total number of steps n_s as

$$n_s = \frac{n_{MG}}{R_{MG}} + \frac{n_{MD}}{P_H}. \quad (14)$$

In Figure 6 the average number of steps is plotted against c using both the MD stage probability bounds P_H . The value of $c = 4.7$ is the minimum requiring 570 steps.

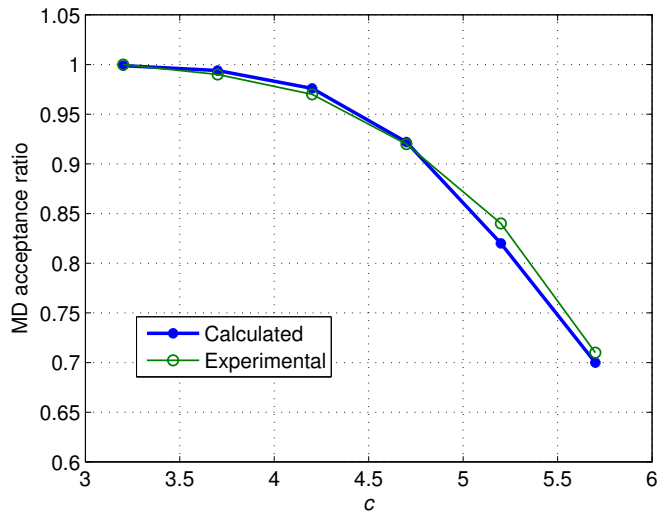


Figure 5: Predicted and experimental MD acceptance ratio data for a 216 molecule box of water model.

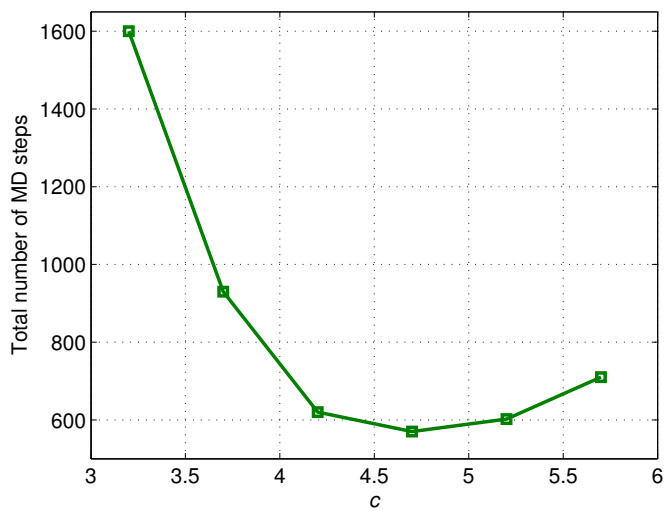


Figure 6: The average number of MD steps required using the MD stage probability bound P_H with varying c , with minima at $c = 4.7$.

5 Comparison of HMC and SHMC efficiency

Using the analysis above we can now compare the efficiency of the HMC and SHMC methods. For the HMC method the MD step is accepted with probability

$$PA_{HMC} = \min \{1, \exp(-\beta (\mathcal{H}(\Gamma') - \mathcal{H}(\Gamma)))\},$$

where Γ are the phase-space variables before the MD stage and Γ' the resulting phase-space variables after the MD stage.

If we assume that \mathcal{H} , with $\mu_{\mathcal{H}} = \mu_{\Delta\mathcal{H}}$ and $\sigma_{\mathcal{H}} = \sigma_{\Delta\mathcal{H}}$, is normally distributed then the average acceptance ratio $R_{HMC} = \langle PA_{HMC} \rangle$ is

$$R_{HMC} = \frac{1}{\sqrt{2\pi}\sigma_{\Delta\mathcal{H}}} \int_{-\infty}^{\infty} \min\{\exp(-\beta x), 1\} \exp\left(\frac{-(x - \mu_{\Delta\mathcal{H}})^2}{2\sigma_{\Delta\mathcal{H}}^2}\right) dx. \quad (15)$$

From normalization and the symplectic property of the numerical integrator [5] we have

$$\langle \exp(-\beta\Delta\mathcal{H}) \rangle = 1,$$

which is equivalent to saying that the moment-generating function for random variable $\Delta\mathcal{H}$ evaluated at $-\beta$ is equal to 1 and hence can be expanded into cumulants k_1, k_2, \dots

$$1 = \exp\left(\sum_{n=1}^{\infty} \frac{(-\beta)^n k_n}{n!}\right).$$

Since $\Delta\mathcal{H}$ is normally distributed we have $k_1 = \mu_{\Delta\mathcal{H}}$, $k_2 = \sigma_{\Delta\mathcal{H}}^2$, $k_n = 0 \forall n > 2$, giving

$$\mu_{\Delta\mathcal{H}} = \frac{\beta}{2}\sigma_{\Delta\mathcal{H}}^2. \quad (16)$$

From (15) and (16)

$$R_{HMC} = \operatorname{erfc}\left(\frac{1}{2\sqrt{2}}\beta\sigma_{\Delta\mathcal{H}}\right). \quad (17)$$

In Figure 7 the acceptance ration for the HMC method is compared to the efficiency of the SHMC (where the MG stage requires 5 MD steps and the MD stage 200) for the optimal choice of c for a range of the variance $\sigma_{\Delta\mathcal{H}}^2$. The SHMC method suffers from poor MG acceptance with virtually no acceptance above a variance of 4.

To calculate the improvement in computational efficiency we can find the ‘efficiency’ of the SHMC method for a range of $\sigma_{\Delta\mathcal{H}}$ and find the corresponding set of $\sigma_{\Delta\mathcal{H}}$ for the HMC method which gives the same ‘efficiency’. The ratio of these variances represents the time-step ratio raised to the fourth power, the time-step is inversely proportional to computational efficiency for a fixed length MD stage. In addition we can calculate the associated $\sigma_{\Delta\mathcal{D}}$ and hence the increase in the observable variance. The results can be seen in Figure 8.

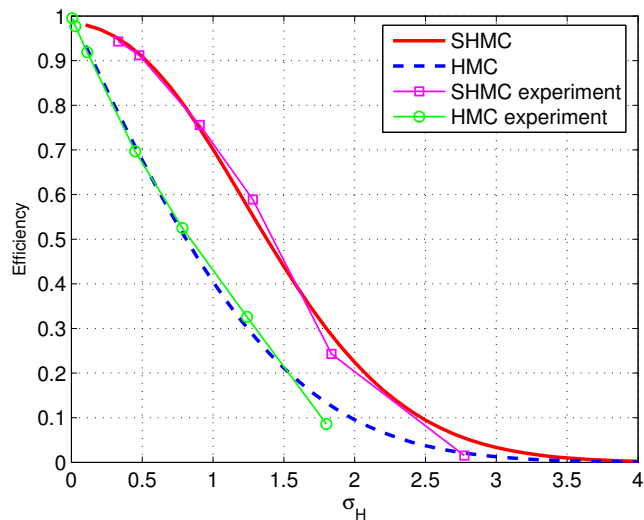


Figure 7: Comparison of the efficiency of the SHMC with 5 MG steps and 500 MD steps and HMC methods against variance, compared to experimental results.

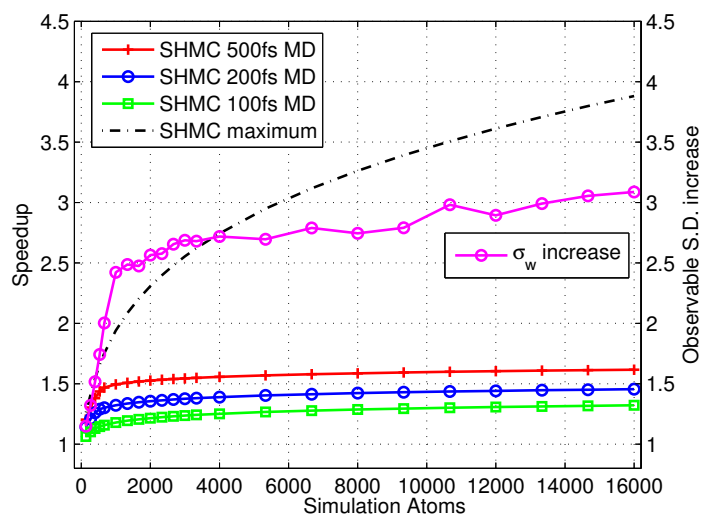


Figure 8: Comparison of the computational efficiency of the SHMC and HMC methods for 100, 200 and 500 fs MD trajectories, and the 500 fs MD trajectory SHMC observable variance.

6 Variance of observables

Since the Shadow Hybrid Monte Carlo (SHMC) method samples from the extended phase-space of the shadow Hamiltonian the observables must be re-weighted according to the following method In order to calculate unbiased values

$$\langle \hat{A} \rangle_{\rho_{NVT}} = \frac{\sum_{i=1}^m w_i A_i}{\sum_{i=1}^m w_i}, \quad (18)$$

where A is the observable of interest, m is the number of samples and the w_i are defined as

$$w_i = \exp(\beta \Delta \mathcal{H}(\Gamma_i)). \quad (19)$$

Here $\beta = 1/k_B T$ for temperature T , $\Gamma_i = (\mathbf{x}_i, \mathbf{p}_i)$ are the phase space variables when A_i was sampled and $\Delta \mathcal{H} = \tilde{\mathcal{H}} - \mathcal{H}$ where \mathcal{H} is the original Hamiltonian for which the observables are required and $\tilde{\mathcal{H}}$ is the Hamiltonian from which the data is being sampled, as described in [5]. We note that the nomenclature $\Delta \mathcal{H}$ is used to denote the values for $\Delta \mathcal{H} = \mathcal{H} - \tilde{\mathcal{H}}$ which have been accepted at the end of the MD stage of the method and the mean and variance of $\Delta \mathcal{H}$ are related to the mean and variance of $\tilde{\mathcal{H}} - \mathcal{H}$, the set of $\tilde{\mathcal{H}} - \mathcal{H}$ for all generated momenta, as shown in Section 4.

It was observed by Izaguirre and Hampton [5] that the variance of observables generated in this fashion was greater than that when using HMC. To understand this we look at the Hamiltonian and shadow Hamiltonian for a 216 water molecule model using periodic boundary conditions, integrated with the Leapfrog algorithm, as shown in Figure 1.

We can rewrite (19) as follows

$$\begin{aligned} w_i &= \exp(\beta \Delta \mathcal{H}(\Gamma_i)) \frac{\exp(\beta \langle \Delta \mathcal{H} \rangle)}{\exp(\beta \langle \Delta \mathcal{H} \rangle)} \\ &= \exp(\beta(\Delta \mathcal{H}(\Gamma_i) - \langle \Delta \mathcal{H} \rangle)) \exp(\beta \langle \Delta \mathcal{H} \rangle). \end{aligned} \quad (20)$$

Let $\hat{w}_i = \exp(\beta(\Delta \mathcal{H}(\Gamma_i) - \langle \Delta \mathcal{H} \rangle))$, $\bar{w} = \exp(\beta \langle \Delta \mathcal{H} \rangle)$ then we can write (18) as

$$\langle \hat{A} \rangle_{\rho_{NVT}} = \frac{\sum_{i=1}^m \hat{w}_i \bar{w} A_i}{\sum_{i=1}^m \hat{w}_i \bar{w}} = \frac{\sum_{i=1}^m \hat{w}_i A_i}{\sum_{i=1}^m \hat{w}_i}, \quad (21)$$

since \bar{w} is a common factor. Hence the re-weighting is only dependent on the variations in \mathcal{H} (i.e. its variance), not the separation between the Hamiltonian and the shadow Hamiltonian ($\Delta \mathcal{H}$) for accepted MD values.

We might reasonably expect that the distribution of the Hamiltonian for large systems would be gaussian with a known variance, as we can see from Figure 2 for our water model over 50 ps with a time-step of 1.0 fs (where the variance is approximately 0.7).

Given an observable A which has a gaussian distribution with a know variance σ_A , we can easily calculate the variance of the re-weighted observable using (21) since $\Delta\mathcal{H} - \langle\Delta\mathcal{H}\rangle$ is just the variation in the original Hamiltonian. We introduce

$$\delta A_i = A_i - \mu_A, \quad (22)$$

where μ_A is the average value of A , and re-write (21) as

$$\begin{aligned} \langle\hat{A}\rangle_{\rho_{NVT}} &= \frac{\sum_{i=1}^m \hat{w}_i (\mu_A + \delta A_i)}{\sum_{i=1}^m \hat{w}_i} \\ &= \mu_A + \frac{\sum_{i=1}^m \hat{w}_i A_i}{\sum_{i=1}^m \hat{w}_i} \\ &= \mu_A + \left\langle \frac{\hat{w}_i A_i}{\langle\hat{w}\rangle} \right\rangle. \end{aligned} \quad (23)$$

From this the variance of the re-weighted observable σ_w will be equal to the variance of the $A_i \hat{w}_i / \langle\hat{w}\rangle$ and is not dependent on μ_A .

Proposition 6.1 *The re-weighted observable A has variance σ_w^2 given by*

$$\sigma_w^2 = \exp(\beta^2 \sigma_{\Delta\mathcal{H}}^2) \sigma_A^2, \quad (24)$$

where σ_A^2 is the variance of the original observable and $\sigma_{\Delta\mathcal{H}}^2$ is the variance of $\Delta\mathcal{H}$.

Proof: Given a normally distributed variable Z with mean μ_Z and variance σ_Z^2 we can transform to the lognormal distribution [1] by $X = \exp(Z)$ with median $\exp(\mu_Z)$, mean $\exp(\mu_Z + \sigma_Z^2/2)$ and variance

$$\sigma_{\ln X}^2 = \exp(2\mu_Z) \exp(\sigma_Z^2) (\exp(\sigma_Z^2) - 1). \quad (25)$$

If the term $\Delta\mathcal{H} - \langle\Delta\mathcal{H}\rangle$ has variance $\sigma_{\Delta\mathcal{H}}^2$ and (clearly) mean $\mu = 0$ then the variance of the re-weighting factor \hat{w}_i , from (25), is

$$\sigma_f^2 = \exp(\beta^2 \sigma_{\Delta\mathcal{H}}^2) (\exp(\beta^2 \sigma_{\Delta\mathcal{H}}^2) - 1). \quad (26)$$

To find the variance of the product of the A_i and the re-weighting factor we use the ‘variance of products’ equation [4]

$$\sigma_{Af}^2 = \mu_A^2 \sigma_f^2 + \mu_f^2 \sigma_A^2 + \sigma_f^2 \sigma_A^2. \quad (27)$$

The identities $\text{var}(aX) = a^2 \text{var}(X)$ and $\langle\hat{w}\rangle = \exp(\mu + \beta^2 \sigma_{\Delta\mathcal{H}}^2/2)$, for a lognormal distribution, allows us to find the variance of the $\delta A \hat{w}_i / \langle\hat{w}\rangle$

$$\begin{aligned} \sigma_{Af/\langle\hat{w}\rangle}^2 &= \frac{1}{\exp(\beta^2 \sigma_{\Delta\mathcal{H}}^2)} (\mu_A^2 \sigma_f^2 + \mu_f^2 \sigma_A^2 + \sigma_f^2 \sigma_A^2) \\ &= \mu_A^2 (\exp(\beta^2 \sigma_{\Delta\mathcal{H}}^2) - 1) + \exp(\beta^2 \sigma_{\Delta\mathcal{H}}^2) \sigma_A^2. \end{aligned} \quad (28)$$

From (22) we have that δA has mean $\mu_A = 0$ giving our re-weighted observable variance as

$$\sigma_w^2 = \exp(\beta^2 \sigma_{\Delta\mathcal{H}}^2) \sigma_A^2. \quad (29)$$

□

Simulations of variables with differing mean and variance are re-weighted at 300K by Hamiltonians with different variance as depicted in Figure 9. The variance of the re-weighted variables have been determined using the block averaging method of Flyvbjerg and Peterson [3] and are compared to the exact result computed from (24) with good correlation.

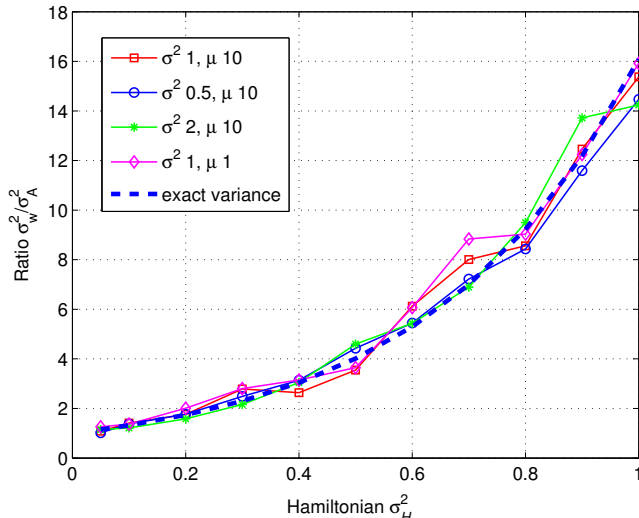


Figure 9: Change in variance for observables for different Hamiltonian variance.

If the maximum permissible ratio $R = \sigma_w^2 / \sigma_A^2$ is known we can determine the maximum target variance σ_T^2 in the Hamiltonian from (24)

$$\sigma_T^2 = \frac{\ln(R)}{\beta^2}. \quad (30)$$

This target variance can then be used to determine the maximum simulation time-step Δt_T , given the Hamiltonian variance $\sigma_{\Delta\mathcal{H}}^2$ for some known time-step Δt_H . For the Leapfrog method (or any second order method) we have

$$\Delta t_T = \Delta t_H \sqrt{\frac{\sigma_T}{\sigma_{\Delta\mathcal{H}}}}. \quad (31)$$

Since the value of c effects the variance of the Hamiltonian data that is accepted at the MD stage it could be used to adjust the variance of the observables. However, since we are already selecting c for to gain the maximum efficiency for the method this is not desirable.

7 Algorithm

The analysis from Sections 4 and 6 can be used to define an algorithm for the explicit determination of the parameters c and Δt given a user defined maximum ratio between the variance of an observable and its re-weighted result from the simulation, R .

Algorithm 1

- (1) **σ parameter Stage:** Set $\Delta t = 1$ fs and $c = 0$, repeat until Δt unchanged.
 - (a) Generate n MG steps with $\Delta t = 1$ fs, recording \mathcal{H} , $\hat{\mathcal{H}}$.
 - (b) Find $\mu_{\Delta\mathcal{H}}$, $\sigma_{\Delta\mathcal{H}}$, $\sigma_{\Delta\mathcal{G}}$ from the data.
 - (c) Calculate new Δt from user defined R and (30), (31).
 - (d) Repeat step.
- (2) **c parameter Stage:**
 - (a) Calculate c for minimum average steps from (11), (12) and (13) from last set of data from stage (1).
- (3) **SHMC stage:**
 - (a) Run SHMC steps for the desired simulation length using the above parameters.
- (3) **Post processing stage:**
 - (a) Re-weight observables.

We envisage that step (1) would not be repeated more than once. The number of MG steps in step (1), n , is user defined and should be a reasonable large number, but small in relation to the total simulation length, for example $n = 1000$ was used for the 216 water molecule simulation which is 2-3 times the number of steps in the MD stage.

8 Conclusion

It is clear from Sections 4 and 6 that if we know the mean and variance of the difference between the shadow Hamiltonian and the Hamiltonian then we can find the optimal values for both c and Δt . The proposed method shown in Section 7 should allow the optimally efficient SHMC method with only a small initial overhead in calculating the time-step and c parameter. The efficiency gains over

HMC are dependent on the length of the MD trajectories and for 100, 200 and 500 fs the ‘speedups’ are 1.32, 1.45 and 1.62 respectively.

It is also clear that, although modest gains in efficiency are available through the use of SHMC, the difficulty in generating momenta with a p.d.f. corresponding to the shadow Hamiltonian is a severe restriction on the efficiency of the method, especially with respect to MD trajectory length. This is compounded by the need to evaluate several MD steps to calculate the shadow Hamiltonian using the interpolated scheme. Although there is still an upper bound on the efficiency of the method due to the increase in the variance of the observables a more efficient method is required, in place of the MG stage, for the advantages of SHMC to be realized.

References

- [1] J. Aitchison and J. A. C. Brown. *The log-normal distribution*. University Press, Cambridge, 1957.
- [2] R. D. Engle, R. D. Skeel, and M. Drees. Monitoring energy drift with shadow Hamiltonians. *J. Comput. Phys.*, 206(2):432–452, 2005.
- [3] H. Flyvbjerg and H. G. Petersen. Error estimates on averages of correlated data. *J. Chem. Phys.*, 91:461–466, 1989.
- [4] L. A. Goodman. On the exact variance of products. *J. Am. Stat. Assoc.*, 55:708–713, 1960.
- [5] J. A. Izaguirre and S. S. Hampton. Shadow hybrid Monte Carlo: An efficient propagator in phase space of macromolecules. *J. Comput. Phys.*, 200(2):581–604, 2004.
- [6] H. Yoshida. Recent progress in the theory and application of symplectic integrators. *Celest. Mech. Dynam. Astron.*, 56:27–43, 1993.