Approach to Thermal Equilibrium in Biomolecular Simulation

Eric Barth[1], Ben Leimkuhler[2], and Chris Sweet[2]

[1] Department of Mathematics
Kalamazoo College
Kalamazoo, Michigan, USA 49006

[2] Centre for Mathematical Modelling
University of Leicester
University Road
Leicester LE1 7RH, UK

Summary. The evaluation of molecular dynamics models incorporating temperature control methods is of great importance for molecular dynamics practitioners. In this paper, we study the way in which biomolecular systems achieve thermal equilibrium. In unthermostatted (constant energy) and Nosé-Hoover dynamics simulations, correct partition of energy is not observed on a typical MD simulation timescale. We discuss the practical use of numerical schemes based on Nosé-Hoover chains, Nosé-Poincaré and recursive multiple thermostats (RMT) [8], with particular reference to parameter selection, and show that RMT appears to show the most promise as a method for correct thermostatting. All of the MD simulations were carried out using a variation

of the CHARMM package in which the Nosé-Poincaré, Nosé-Hoover Chains and RMT methods have been implemented.

1 Introduction

Molecular dynamics (MD) is an increasingly popular tool in chemistry, physics, engineering and biology. In many molecular simulations, the dynamics trajectory is used as a method of sampling a desired ensemble, for example to compute the average of some function of the phase space variables. In such cases it is important that the trajectory produce a representative collection of phase points for all variables of the model. A common ensemble used in biomolecular simulation is the NVT ensemble, which weights points of phase space according to the Gibbs density

$$\rho \propto e^{-\beta H}, \quad \beta = (k_B T)^{-1},$$

where H is the system Hamiltonian, k_B is Boltzmann's constant, and T is temperature. In normal practice, MD samples from the isoenergetic (microcanonical) ensemble, so some device must be employed to generate points from the NVT ensemble. The methods discussed in this article are based on construction of extended Hamiltonians whose microcanonical dynamics generate canonical sampling sequences (Nosé dynamics). Nosé [5] proposed a Hamiltonian of the form:

$$H^{Nose} = H\left(q, \frac{p}{s}\right) + \frac{p_s^2}{2Q} + gkT \ln s,$$

where Q is the Nosé mass, s is the thermostatting variable and p_s is its conjugate momentum. The system is often simulated in a time-reversible but non-Hamiltonian formulation

(Nosé-Hoover, [7]) that incorporates a correction of timescale (this time-transformation has some important implications for the stability of numerical methods). In a 1998 paper [4], a Hamiltonian time-regularized formulation was introduced along with reversible *and* symplectic integrators (see also [16, 15]). These methods show enhanced long term stability compared to Nosé-Hoover schemes. The Nosé-Poincaré schemes, as they are termed because of the use of a Poincaré time transformation, have been extended to NPT and other ensembles in several recent works [9, 11, 10].

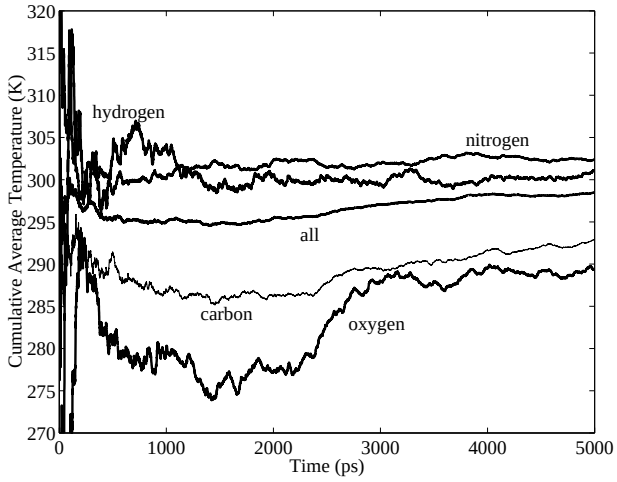

Fig. 1. A 5ns trajectory for alanine dipeptide using the Verlet integrator clearly shows that equilibrium is not achieved on the indicated timescale. The plot shows the cumulative time-averaged temperatures for the entire system (all), and for each type of atom separately.

In classical models of biomolecules, when thermostatting with schemes derived from Nosé's method, trapping of energy in subsystems can result in long equilibration times. The presence of many strongly coupled harmonic components of not too different fre-

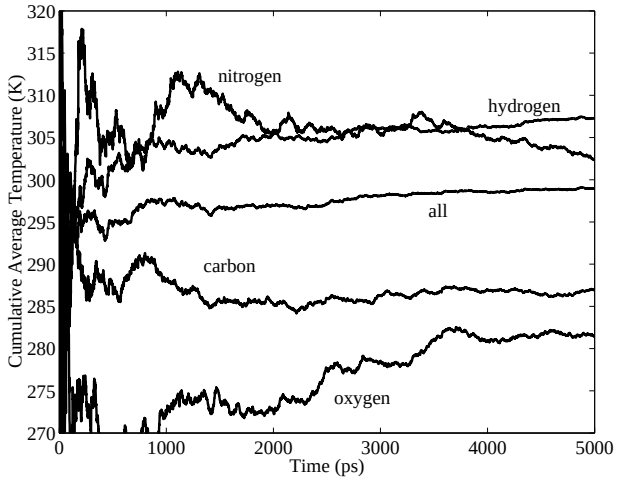

Fig. 2. Cumulative average temperature, and temperature of subsystems computed by a 5ns trajectory for alanine dipeptide using the Nosé-Hoover option in CHARMM with Q=0.3 — as we report later, this is the optimal value for Q. Correct thermalization is clearly not achieved on this time scale.

quency means that the systems should eventually equilibrate, but the equilibration time in all-atom models (including bond vibrations) nonetheless greatly exceeds the time interval on which simulation is performed (a few nanoseconds, in typical practice). The only way to be sure that an initial sample is properly equilibrated is to check that in subsequent runs, the individual momentum distributions associated to each degree of freedom are Maxwellian. This is typically not done in practice. To bring a given molecular system rapidly to equilibrium and maintain the system in that state to ensure good sampling of all degrees of freedom, it is necessary to employ a suitable thermostatting mechanism.

To illustrate the primary challenge that we will attempt to address in this paper, we have performed a molecular dynamics simulation of an unsolvated alanine dipeptide molecule using a representative molecular dynamics software package (CHARMM [1]). We used the Verlet method to perform a microcanonical simulation on the system and examined the convergence to thermal equilibrium in the "light" (H) and "heavy" (C,N,O) atoms. Note that because of the presence of conserved quantities (total linear momenta) the usual equipartition of energy does not hold; the modified formulas are given in Section 2. The details regarding the setup of this simulation can be found in Section 3.4. It is clear from these experiments that the adiabatic localization of energy is a significant cause for concern as seen in Figures 1 and 2.

It might be thought that the energy trapping is a result of performing these simulations in vacuo, but this is not the case: similar problems have been verified by the authors for solvated models.[3] It might also be thought that the Nosé dynamics technique, in introducing a "global demon" which couples all degrees of freedom, would successfully resolve this issue. In fact, as seen in Figure 2, this is not the case: although such methods successfully control the overall temperature of the system, the thermal distributions observed in light and heavy degrees of freedom using Nosé-Hoover (and Nosé-Poincaré) are incorrect, as can be seen in Figure 2. The system evidently does not have sufficient ergodicity to provide the correct energetic distribution on the timescale of interest.

Several techniques have been proposed to improve ergodicity in molecular simulations. In [12] a Nosé-Hoover chain method was developed which coupled additional thermostat-

[3] The use of solvated models raises some additional issues regarding bond thermalization and the selection of parameters for some of our methods. These results will be reported elsewhere.

ting variables to the system degrees of freedom, retaining the property that integration over the auxiliary variables reduced sampling of the extended microcanonical phase space to canonical sampling of H. As this extension is based on Nosé-Hoover, it also sacrifices the Hamiltonian structure: the additional variables are introduced in such a way that the extended system is only time-reversible, so that methods based on this scheme cannot be reversible-symplectic. In [14, 13, 8] several new Hamiltonian-based multiple thermostat schemes have been developed. Nosé-Poincaré chains, described in [13] are the natural analogue of Nosé-Hoover chains. The more recent recursive multiple thermostat (RMT) schemes of [8] are a new departure, obtaining thermalization from a more complicated interaction of thermostat variables with the physical variables. A careful analysis of Nosé dynamics and RMT schemes for harmonic models was performed in [8]; arguments presented there and numerical evidence suggest that the formulation is potentially superior to other dynamical alternatives, including Nosé-Hoover chains, in obtaining well-equilibrated sampling sequences for the canonical ensemble. However, the results of [8] have so far only been verified for harmonic oscillators and coupled harmonic models.

The method of Gaussian moment thermostatting [6] also attempts to address incorrect thermalization of Nosé-Hoover methods, but was not considered here. In this paper we study the convergence to ensemble for chains and recursive methods applied to biomolecular models. We first discuss problem formulation and computation of temperature in all-atom biomolecular models.

2 Molecular dynamics formulation

In this article, we treat a classical all-atom N-body model. The Hamiltonian is of the form

$$H(q_1, q_2, \dots q_N, p_1, p_2, \dots p_N) \ = \sum_{i=1}^{N} \frac{p_i^2}{2m_i} + V(q_1, q_2, \dots q_N).$$

Here m_i represents the mass of the ith atom, $q_i \in \mathbf{R}^3$ and $p_i \in \mathbf{R}^3$ are Cartesian position and momentum vectors of the atomic point masses, and K and V represent kinetic and potential energy, respectively. The potential energy function can be decomposed into a sum of terms, including pairwise (distance dependent) short-ranged Lennard-Jones potentials V_{LJ}, Coulombic potentials due to charges on the atoms V_C, and potential energies that describe the covalent bonding structure of the molecule, including $V_B^{(2)}, V_B^{(3)}, V_B^{(4)}$, representing 2-atom (length bond), 3-atom (angle bond), and 4-atom (dihedral angle) terms, respectively. In vacuum, with internal potentials only, the system described above is invariant under translations and rotations, thus it would admit six conserved quantities (linear and angular momentum). In practice, most biomolecular simulations also incorporate a collection of water molecules and are performed using periodic boundary conditions, meaning that the system is allowed to interact with copies of itself extended along the cubic lattice vectors. In this setting, the angular momentum conservation is broken, although the translational symmetry (and with it linear momentum conservation) is still present.

Although it is a slight abuse of language, we define the instantaneous temperature as the average kinetic energy per degree of freedom:

$$T_{\text{inst}}(p_1, p_2, \dots p_N) = \left(\frac{1}{(3N-d)kT}\right) \sum_{i=1}^{N} \frac{p_i^2}{2m_i},$$

where N is the number of atoms, k_B is the Boltzmann constant and $g \equiv 3N - d$ is the number of degrees of freedom (d is the number of conserved quantities in the dynamics). For example, in a simulation without periodic boundary conditions, common microcanonical integrators such as Verlet conserve both linear and angular momentum, so $d = 6$. In the presence of periodic boundary conditions we have $d = 3$. Standard Langevin dynamics, in which no quantities are conserved, has $d = 0$. The ergodic hypothesis of statistical mechanics states that time averages will (eventually) converge to ensemble averages:

$$\lim_{t\to\infty} \langle T_{\text{inst}} \rangle_t = \langle T \rangle_{\text{ensemble}}.$$

In this article we will need to discuss convergence to ensemble in different variables. When conserved quantities are present in a Hamiltonian system, the usual equipartition result must be adjusted. Indeed, we no longer observe equipartition, but there is still an appropriate partition of kinetic energies. Assuming that the trajectories sample from the canonical ensemble (in the microcanonical case, this is essentially equivalent to assuming ergodicity and a sufficiently large system) we have, assuming linear momentum conservation [18],

$$\left\langle \frac{p_{ix}^2}{2m_i} \right\rangle = \frac{M - m_i}{M} \frac{k_B T}{2}. \tag{1}$$

A simple, direct proof of this result can be constructed by integrating the Nosé Hamiltonian partition function. In anticipation of work on solvated systems, we introduced a weak anisotropic interaction potential between a *single pair* of backbone atoms (the C-C bond at the N-terminus of alanine dipeptide) in our constant energy simulations of the

form

$$\phi_{\text{anisotropy}}(q_i, q_j) = \frac{k_1}{2}(x_i - x_j)^2 + \frac{k_2}{2}(y_i - y_j)^2 + \frac{k_3}{2}(z_i - z_j)^2.$$

(Modest values of the three constants were used: $k_1 = 0.1$, $k_2 = 0.15$, $k_3 = 0.2$.) This has the effect of breaking angular momentum in vacuum simulations, while leaving linear momentum invariant, so that (1) can be used to compute subsystem temperatures.[4]

For trajectories generated according to the Nosé Hamiltonian, linear and angular momentum are conserved only weakly, in the sense that these quantities are conserved if initialized to zero, but will vary from any nonzero initial condition. In the simulations reported here, the linear momentum was set to be zero initially, and was conserved throught the trajectories. The angular momentum was initially nonzero (but small, on the order of 10^{-3}), and was observed to vary over the course of the simulation. Hence the number of conserved quantities for all variants of Nosé-Hoover and Nosé-Poincaré was $d = 3$.

[4] The method for calculation of instantaneous kinetic temperature for a subsystem has not been widely reported in the literature, but is needed in practice. In particular, both old and new variants of Nose-Hoover dynamics implemented in CHARMM have the option of thermostatting subsystems separately, thus necessitating the calculation of the temperature of subsystems and correct handling of the degrees of freedom. However, this seems to be done incorrectly in CHARMM (version c31b1): the first specified subsystem s, with N^s atoms, is assigned $g^s = 3N^s - 6$ degrees of freedom, with each successive subsystem m being assigned the full $g^m = 3N^m$.

3 Thermostatting using Nosé-Hoover chains, Nosé-Poincaré and RMT methods

Because correct thermalization is not achieved by the use of the Verlet or Nosé-Hoover methods, as shown in Figures 1 and 2, we studied the Hamiltonian Nosé-Poincaré method and some methods which are designed to achieve enhanced thermalization: Nosé-Hoover chains and RMT. It is well known [5, 8, 12] that the correct choice of parameters is essential if these methods are to thermostat the model correctly, and we here consider the different methods proposed for choosing them.

3.1 Nosé-Poincaré and RMT methods

The Nosé-Poincaré method [4] involves direct symplectic discretization of the Hamiltonian

$$H^{NP} = s\left[H - H_0^{Nose}\right].$$

Here H_0^{Nose} is the initial value of the Nosé Hamiltonian and leads to the Nosé-Poincaré Hamiltonian

$$H^{NP}(q, s, p, p_s) = s\left(H\left(q, \frac{p}{s}\right) + \frac{p_s^2}{2Q} + N_f kT \ln s - H_0\right). \tag{2}$$

The numerical method proposed in [4] was the generalized leapfrog method. Although this method is typically implicit, in the case of the Nosé-Poincaré system, the numerical challenge reduces to solving a scalar quadratic equation, and it is this variation which we implemented in CHARMM.

The RMT method [8] is produced by applying additional thermostats to the Nosé-Poincaré method recursively, with each additional thermostat acting on the previous one

in addition to the original system. As discussed in [8], the stability of the numerical implementation of Nosé-Poincaré chains is not as good as the underlying Nosé-Poincaré method. The RMT method [8] corrects this deficiency while introducing a stronger coupling between bath and phase variables. With M thermostats, the recommended RMT formulation is:

$$H^{RMT} = s_1 s_2 \cdots s_M \left[\sum_{i=1}^{N} \frac{p_i^2}{2m_i s_1^2 \cdots s_M^2} + V(q) + \sum_{j=1}^{M-1} \frac{p_{s_j}^2}{2Q_j s_{j+1}^2 \cdots s_M^2} + \frac{p_{s_M}^2}{2Q_M} + gkT \ln s_1 + \sum_{j=2}^{M} ((N_f + j - 1)kT \ln s_j + f_i(s_i)) - H_0 \right],$$

where $g = N_f$ and H_0 is chosen so that the initial value of H^{RMT} is zero. This formulation introduces a number of additional parameters, some of which are dependent on the choice of the $f_i(s_i)$. A recommended choice for these functions is

$$f_i(s_i) = \frac{(a_j - s_j)^2}{2C_j}. \tag{3}$$

The value a_i is chosen as the required average value of s_i, generally 1, as the additional term will operate as a negative feedback loop to minimize $(a_i - s_i)$, as can be seen from the equations of motion. In the context of Nosé-Poincaré chains the value of $C_i,\ i \geq 2$ can be estimated by considering the equations of motion for the ith thermostat. From this we see that s_i is driven by the changes in $p_{s_{i-1}}$. The purpose of the auxiliary function is to limit the excursions of s_i, which can be achieved if $ds_i/dp_{s_{i-1}}$ is a maximum at $s_i = a_i$. It was shown in [13] that C_i should satisfy,

$$C_i \leq \frac{a_i^2}{8kT}. \tag{4}$$

For RMT, the choice of C_i is less clear but numerical experiments indicate that a similar upper bound on C_i. Experiments also suggest that the precise choice of C_i is not as critical to effectiveness of RMT as is selection of thermostat mass parameters.

We note that the generalized leapfrog RMT discretization, which was used in the numerical experiments reported here, is limited by numerical stability to a substantially smaller timestep than Verlet and the Nosé-Hoover methods are able to use. Work on improving the stable timestep (and hence numerical efficiency) by enhancement of the RMT numerical integrator is ongoing.

3.2 Choice of the Nosé mass: ideas from the literature

All methods based on the work of Nosé [5] have parameters. Although the proofs of canonical sampling (which assume ergodicity) are not strictly speaking dependent on their values, a careful choice is needed in order for good sampling to be observed in practice. In standard Nosé dynamics the only parameter is the Nosé mass Q in (1).

For the purpose of this discussion we will consider the time reparameterized variation of Nosé's scheme, the Nosé-Poincaré method [4, 16], which produces trajectories in real time and has the advantage of being Hamiltonian-based (2). Various attempts have been made to identify the optimal value for Q [5, 9, 8], and these are examined below.

Choice of Q based on total kinetic energy

In molecular dynamics the temperature of a system, at equilibrium, can be defined as,

$$T = \frac{1}{N_f k} \left\langle \sum_{i=1}^{N} \frac{\tilde{p}_i^2}{m_i} \right\rangle. \tag{5}$$

From this it is tempting to assume that if this criterion is met then the system must be at equilibrium. However thermostatting methods derived from Nosé's scheme are based on a "negative feedback loop" which controls the average kinetic energy such that (5) is satisfied. This can be seen by considering the equations of motion for the thermostat conjugate momentum,

$$\dot{p}_s = \sum_{i=1}^{N} \frac{p_i^2}{m_i s^2} - N_f kT. \tag{6}$$

Taking averages, assuming time averages of time derivatives disappear and substituting $\tilde{p}_i = p_i/s$ gives (5). From this we see that (5) is satisfied for all values of Q, whereas it is known that the correct sampling is not obtained unless Q is chosen correctly.

If we study (6) carefully we see that the method only guarantees that the total average kinetic energy is fixed, it does not indicate what happens to subsystems. From the equipartition theorem we have that,

$$\left\langle \frac{\tilde{p}_i^2}{m_i} \right\rangle = kT \quad \text{all } i. \tag{7}$$

If we look at individual sub systems we find that (7) is not satisfied for all Q, as seen in Figure 3, and hence it is not possible to choose Q in this manner.

Choice of Q based on self-oscillation frequency

By utilizing linearization methods Nosé determined [5] that, for time re-parameterized thermostatting methods, the thermostat subsystem has a natural frequency

$$\omega_N = \sqrt{\frac{2N_f kT}{Q}}. \tag{8}$$

It has been proposed [5] that setting Q so that this frequency is close to some natural frequency in the original system would induce resonance between the system and thermo-

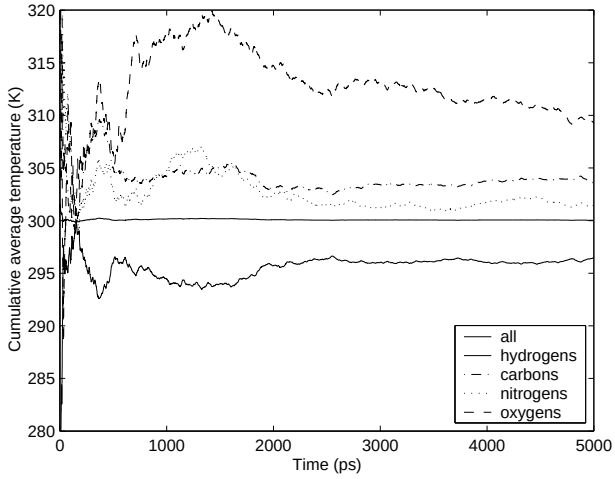

Fig. 3. Average kinetic energy for subsystems of alanine dipeptide consisting of the hydrogen atoms, oxygen atoms, nitrogen atoms and carbon atoms for a 5ns simulation time with $Q = 2.0$ using the Nosé-Poincaré method.

stat, and hopefully ergodic behavior. Although the need to increase Q with the number of degrees of freedom N_f and the temperature T can be verified empirically, setting the value of Q according to (8) generally gives poor results. In fact it is observed that if the thermostat subsystem oscillates at its resonant frequency then it becomes decoupled from the system [12, 8], and poor sampling is obtained. For systems of harmonic oscillators it can be shown [8] that the optimum choice of Q does not coincide with this estimate. As an example, for a single harmonic oscillator of frequency 1 it has been found [8] that the optimum value of Q is around 0.3 where Nosé's estimate predicts $Q = 2$.

Choice of Q based on control of average thermostat kinetic energy

It can be shown [8], under the assumption of ergodicity, that the following holds,

$$\left\langle \frac{p_s^2}{Q} \right\rangle = kT, \tag{9}$$

where p_s is the thermostat conjugate momentum. For harmonic oscillators the value of this quantity is a good indicator of the optimum choice of Q [8].

The value of $\langle p_s^2/Q \rangle$ was examined for Q in the range 0.1 to 1000 for a model of alanine dipeptide, using the CHARMM molecular dynamics package. In this case the value of $\langle p_s^2/Q \rangle$ was found to be close to kT for all values of Q tested, over a sufficiently long integration time, indicating that this method fails for complex systems of biomolecules.

Choice of Q based on kinetic energy distribution

In [9] it was proposed that the optimum value of Q could be obtained by comparing the resulting total energy distribution with the expected canonical distribution. For crystalline Aluminum it was found that a broad range of Q, 100 to 10000, gave good convergence to the correct distribution and it was concluded that, for this model, the value of Q was not critical.

Applying the same method to the alanine dipeptide produced similar results with Q in the range 0.1 to 1000 giving good total energy distributions. In this case we know that, for most of these values of Q, good sampling is not obtained when considering the average kinetic energy of subsystems.

3.3 A new approach to Nosé masses based on minimization of $\langle s \rangle$

We now propose a new scheme for choosing thermostat masses. This new method, which has produced the promising results reported below, will be examined in fuller detail in

a related publication. It has been shown [8] that for systems of harmonic oscillators the average value of the thermostatting variable s, when sampling from the canonical ensemble, is

$$\langle s \rangle_c = \exp\left(\frac{H_0}{N_f kT}\right)\left(\frac{N_f}{N_f+1}\right)^{\frac{2N_f+1}{2}} \tag{10}$$

In the limit of large N_f, this is equivalent to

$$\langle s \rangle_c = \exp\left(\frac{H_0}{N_f kT} - 1\right), \tag{11}$$

and this is found to be a good approximation for all N_f.

We can rearrange (2), given that H_0 is chosen such that $H_{NP} = 0$, as follows:

$$\ln s = \left(\frac{H_0 - H\left(q, \frac{p}{s}\right) - \frac{p_s^2}{2Q}}{N_f kT}\right). \tag{12}$$

As $Q \to \infty$ we expect that $\left\langle p_s^2/Q \right\rangle \to 0$ and $s \to \langle s \rangle$, and hence $\langle \ln s \rangle \to \ln\langle s \rangle$. We also note that, for harmonic oscillators, the following holds:

$$\langle KE \rangle = \langle PE \rangle, \tag{13}$$

where KE is the kinetic energy and PE is the potential energy. Taking averages of (12) and rearranging then gives,

$$\langle s \rangle_{Qlim} = \exp\left(\frac{H_0}{N_f kT} - 1\right). \tag{14}$$

Since (11) is the same as (14) the average value of s in the limit of large Q is the same as the average value of s when sampling from the canonical ensemble for systems of harmonic oscillators.

It is of interest to examine the average value of s as we decrease Q and the system is somewhere between these two regimes. We can highlight the difference between the two

regimes by taking averages in (12), for the canonical ensemble we have,

$$\begin{aligned}\langle \ln s \rangle_c &= \left(\frac{H_0 - \langle H\left(q, \frac{p}{s}\right)\rangle - \left\langle \frac{p_s^2}{2Q} \right\rangle}{N_f kT} \right), \\ &= \left(\frac{H_0 - (N_f + \frac{1}{2})kT}{N_f kT} \right). \end{aligned} \tag{15}$$

For the limit of large Q,

$$\langle \ln s \rangle_{Qlim} = \left(\frac{H_0 - N_f kT}{N_f kT} \right). \tag{16}$$

We note that that, although $\langle s \rangle_c = \langle s \rangle_{Qlim}$, we have $\langle \ln s \rangle_c < \langle \ln s \rangle_{Qlim}$. This is due to the probability distribution of s, which in the canonical ensemble is log-normal (in the limit of $N_f \to \infty$) giving rise to the difference in the $\ln s$ averages and funding the additional energy required by the $\langle p_s^2/2Q \rangle$ term.

If we let Q_c be the optimum value of Q for canonical sampling then our intermediate regime occurs when $\infty > Q > Q_c$, and here we expect that $\langle p_s^2/Q \rangle > 0$. Since the system is not sampling from the canonical ensemble s will not have the required log-normal distribution and $\langle \ln s \rangle$ will not provide the additional energy for $\langle p_s^2/2Q \rangle$. If we assume that, away from equilibrium, $\langle \ln s \rangle_* = \langle \ln s \rangle_{Qlim}$, then

$$\langle V(q) \rangle_* = \langle V(q) \rangle_{Qlim} - \langle p_s^2/2Q \rangle_* , \tag{17}$$

where the $*$ subscript represents the averages in the intermediate regime.

Reducing the average potential energy for harmonic oscillators, is equivalent to reducing the average kinetic energy from (13) and hence the average value of s will have to increase to accommodate this (the average kinetic energy in the re-scaled momenta is fixed due to the feedback loop, but the reduction can be seen in the original momenta). From this we would expect that the average value of s would have minima where the

system is sampling from the canonical ensemble and in the limit of large Q. Experiments with a harmonic oscillator show that this is indeed the case.

Although some of the assumptions above do not hold for typical molecular dynamics models, the critical requirement that s has a log-normal probability distribution is nonetheless verified in practice. From this we would expect that the average of s will be minimized when sampling from the canonical ensemble. To verify this $\langle s \rangle$ was measured for various Q for a simulation of an alanine dipeptide molecule in the popular CHARMM package with the results in Figure 4. The figure shows a well defined minimum, and sug-

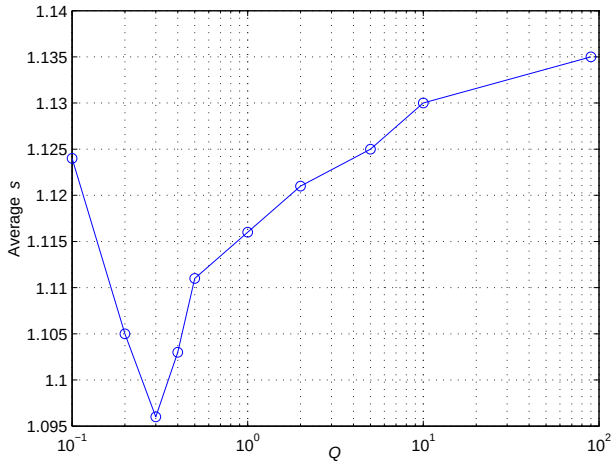

Fig. 4. Average s against Q showing minimum at $Q_c = 0.3$.

gests that the optimal value is around $Q = 0.3$. Simulations of the alanine dipeptide model were conducted at 300K with $Q = 0.3$. Average energies for hydrogen and heavy atoms are shown in Figure 5. Although exact equipartition does not occur in this experiment, the results are the best obtained from numerous Nosé-Poincaré simulations.

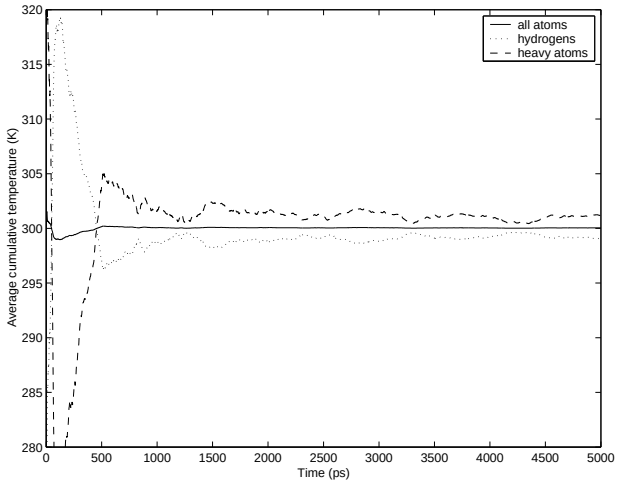

Fig. 5. Average kinetic energy for subsystems of alanine dipeptide consisting of the hydrogen atoms and the heavy atoms for a 5ns simulation time with $Q = 0.3$ using the Nosé-Poincaré method.

3.4 Preliminary experiments with RMT and Nosé-Hoover Chains

The results we describe here are obtained in all atom CHARMM simulations [1, 2] of the alanine dipeptide in vacuum, with nonbonded terms computed without cutoffs. The starting structures were equilibrated with a 50 ps heating cycle, followed by 200ps of equilibration at 300K. The Verlet, Nosé-Hoover and Nosé-Hoover Chain trajectories were generated with a 1 fs timestep, the accepted standard for biomolecules. We used a somewhat conservative timestep of 0.02 fs for RMT. To ensure that the comparison between RMT and Nosé-Hoover methods is valid, we have verified that the the smaller timesteps used in RMT did not improve the thermostatting properties of the Nosé-Hoover methods.

In Section 3.3, Figure 5, we saw that, for the correct choice of parameters, the Nosé-Poincaré method produces much better results than the Verlet or Nosé-Hoover method.

With chains and RMT methods additional parameters need to be selected. Our methodology is to select the value of the first Nosé mass by the "minimum average thermostat variable" scheme and subsequent masses to be close to, but greater than, the previous mass

$$Q_{i+1} \approx \frac{3}{2} Q_i. \tag{18}$$

For the RMT method the C_i were selected as $C_i = 1/16kT$ in accordance with (4). The results for the RMT method are seen in Figure 6, where good results are obtained for the suggested choice of parameters.

We also implemented Nosé-Hoover chains in CHARMM, using an explicit integrator described by Jang and Voth [3]. The results for the Nosé-Hoover chains are seen in Figure 7 where, even with optimum values for the Nosé masses, the thermalization is extremely slow. Results for many other values of the parameters were either similar or worse.

4 Conclusions

It is clear from the preceding experiments that the choice of the parameters applicable to Nosé methods is critical in obtaining good sampling. We observe that, even for methods which are expected to enhance the ergodicity, the range of parameter values which yield good results is small. Experiments with Nosé-Hoover chains indicate that equipartition does not occur far from the optimal choice of Nosé mass, and similar results are seen for RMT.

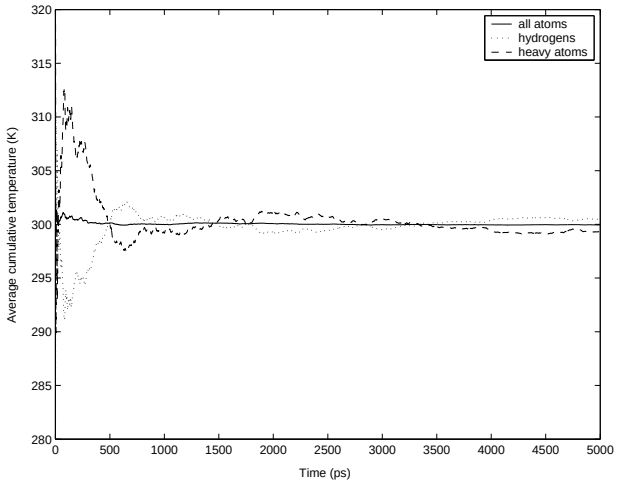

Fig. 6. Cumulative average temperature, and temperature of subsystems computed by a 5ns trajectory for alanine dipeptide using the RMT method implemented in CHARMM, with optimum $Q = 0.3$.

The timestep limitation which we have encountered illustrates the need for more careful numerical treatment of the thermostat variables in RMT. It is generally believed that solvated models do not display the equipartition problems reported here, but initial experiments by the authors indicate otherwise. These issues will be addressed in a subsequent article.

Many methods for determining the optimum value of the thermostat mass in Nosé dynamics have been proposed. Our examination of these methods has shown that, at best, most provide a poor indicator of the optimum value in the setting of biomolecules. Further study of the method presented here based on the "minimum average thermostat value" will be undertaken to provide a better understanding of this technique in the molecular dynamics setting.

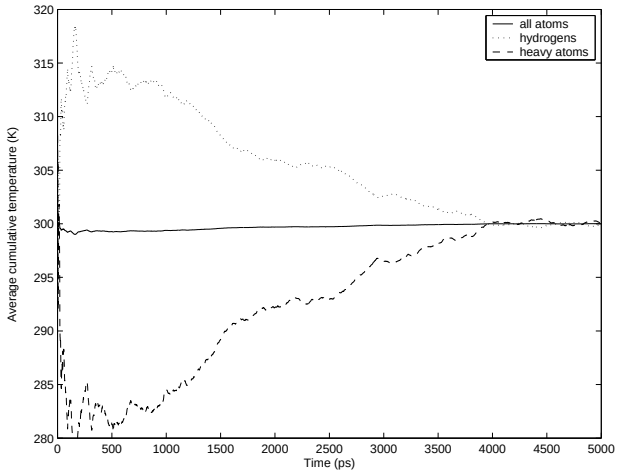

Fig. 7. Cumulative average temperature, and temperature of subsystems computed by a 5ns trajectory for alanine dipeptide using the Nosé-Hoover chains in CHARMM with optimum $Q = 0.3$. The system requires a long time (4ns) to reach equiparon.

5 Acknowledgments

This work was undertaken during visits by EB to the Mathematical Modelling Centre at the University of Leicester, and by all authors to the Institute for Mathematics and its Applications at the University of Minnesota. EB acknowledges the NIH under grant 1 R15 GM065126-01A1 as well as the donors to the Petroleum Research Fund, administered by the ACS, for partial support of this project. BL acknowledges the support of the Leverhulme Foundation and EPSRC Grant GR-R24104. The authors gratefully acknowledge helpful discussions with Frederic Legoll (ENPC).

References

1. B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan and M. Karplus, CHARMM: A program for macromolecular energy, minimization, and dynamics calculations, J. Comp. Chem, 4 (1983), pp. 187–217
2. A.D. MacKerell Jr., D. Bashford, M. Bellott, R.L. Dunbrack Jr., J. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher III, B. Roux, M. Schlenkrich, J. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin and M. Karplus, An all-atom empirical potential for molecular modeling and dynamics of proteins, J. Phys. Chem., 102 (1998), pp. 3586–3616
3. S. Jang and G. Voth, Simple reversible molecular dynamics algorithms for Nos'e-Hoover chain dynamics, J. Chem. Phys. 110: 3263 (1999)
4. S.D. Bond and B.B. Laird and B.J. Leimkuhler, The Nosé-Poincaré method for constant temperature molecular dynamics, J. Comp. Phys., 151, 114, 1999
5. S. Nosé, A molecular dynamics method for simulation in the canonical ensemble, Mol. Phys., 52, 255, 1984
6. Yi Liu and Mark E. Tuckerman, Generalized Gaussian moment thermostatting: A new continuous dynamical approach to the canonical ensemble, J. Chem. Phys, 112, 1685-1700, 2000
7. W. G. Hoover, Canonical dynamics: equilibrium phase-space distributions, Phys. Rev. A 31, 1695 (1985)
8. B. J. Leimkuhler and C. R. Sweet, Hamiltonian formulation for recursive multiple thermostats in a common timescale, SIADS, 4, No. 1, pp. 187216, 2005
9. B.B. Laird and J.B. Sturgeon, Symplectic algorithm for constant-pressure molecular dynamics using a Nosé-Poincaré thermostat, J. Chem. Phys., 112, 3474, 2000

10. Hernandez, E. Metric-tensor flexible-cell algorithm for isothermal-isobaric molecular dynamics simulations J. Chem. Phys., 115, 10282-10291, 2001
11. Dahlberg, Laaksonen and Leimkuhler, in preparation
12. G.J. Martyna and M.L. Klein and M. Tuckerman, Nosé-Hoover chains: The canonical ensemble via continuous dynamics, J. Chem. Phys., 97, 2635, 1992
13. B.J. Leimkuhler and C.R. Sweet, The canonical ensemble via symplectic integrators using Nosé and Nosé-Poincaré chains, J. Chem. Phys., 121, 1, 2004
14. B.B. Laird and B.J. Leimkuhler, A generalized dynamical thermostatting Technique, Phys. Rev. E, 68, 16704, 2003
15. S. Nosé, An improved symplectic integrator for Nosé-Poincaré thermostat, J. Phys. Soc. Jpn, 70, 75, 2001
16. C.P. Dettmann, Hamiltonian for a restricted isoenergetic thermostat, Phys. Rev. E, 60, 7576, 1999
17. C.R. Sweet, Hamiltonian thermostatting techniques for molecular dynamics simulation, Ph.D. Dissertation, University of Leicester, 2004
18. B. Laird, private communication.