Full Paper

The G-JF Thermostat for Accurate Configurational Sampling in Soft-Matter Simulations

Evyatar Arad,[a] Oded Farago,[a, b] and Niels Grønbech-Jensen*[c, d]

Abstract: We implement the statistically sound G-JF thermostat for Langevin dynamics simulations into the ESPResSo molecular package for large-scale simulations of soft-matter systems. The implemented integration method is tested against the integrator currently used by the molecular package in simulations of a fluid bilayer membrane. While the latter exhibits deviations in the sampling statistics that increase with the integration time step dt, the former reproduces near-correct configurational statistics for all dt within the stability range of the simulations. We conclude that, with very modest revisions to existing codes, one can significantly improve the performance of statistical sampling using Langevin thermostats.

Keywords: computational chemistry · molecular dynamics · simulations · soft matter

1. Introduction

Discrete-time molecular dynamics (MD) is one of the most common methods for simulating molecular systems.[1] MD simulations are performed by numerically integrating Newton's equations of motion to advance the coordinates of the particles in discrete time. The most frequently used numerical integrator for performing MD is based on the Størmer-Verlet algorithm,[2] which, for a closed system, can be written:

$$r^{n+1} = r^n + dt\, v^n + \frac{dt^2}{2m} f^n \qquad (1)$$

$$v^{n+1} = v^n + \frac{dt}{2m}(f^n + f^{n+1}), \qquad (2)$$

where r^n, v^n, and $f^n = f(r^n)$ denote the position, velocity, and force of a particle, respectively, at time t_n. The Størmer-Verlet algorithm is a second-order integrator in dt. It is considered favorable over other, higher order in dt, integrators due to its simplicity, computational efficiency, and global conservation properties. For a closed system, these properties ensure optimal stability, time reversibility, and e.g., effective energy conservation over long time integrations.[3] Despite these attractive features, a little-appreciated fact is that the parameter we assign to represent velocity (or momentum) in discrete-time dynamics is not exactly the conjugated variable to the simulated position. The conjugated relationship between the position and velocity coordinates is recovered only in continuous time (see Appendix A in Ref. [4] and references therein). The consequences of this fundamental artifact are significant, as one must accept that kinetic and configurational meas-

ures cannot be obtained correctly from the same simulation, unless a simulation is conducted with a very small time step.

The Størmer-Verlet algorithm addresses dynamics in closed (microcanonical) systems, characterized by conservation of the total energy. However, the microcanonical ensemble is less relevant for most MD applications than its canonical counterpart, where the temperature, rather than the energy, is constant. This is especially true for the relatively small systems that are often simulated as a proxy for thermodynamically large ensembles. A number of methods for constraining the temperature of a simulated system ("thermostats") exists; two representative ones are the deterministic (e.g., Nosé-Hoover)[5,6] and stochastic (Langevin) thermostats.[7] Here we focus

[a] E. Arad, O. Farago
Department of Biomedical Engineering
Ben Gurion University of the Negev
Be'er Sheva 84105 (Israel)

[b] O. Farago
Ilse Katz Institute for Nanoscale Science and Technology
Ben Gurion University of the Negev
Be'er Sheva 84105 (Israel)

[c] N. Grønbech-Jensen
Department of Mechanical and Aerospace Engineering
University of California
Davis CA 95616 (USA)
e-mail: ngjensen@ucdavis.edu

[d] N. Grønbech-Jensen
Department of Mathematics
University of California
Davis CA 95616 (USA)

on integration methods for Langevin dynamics (LD). In LD, two terms are added to Newton's equations of motion: 1) friction proportional and opposite to the velocity; and 2) an accompanying delta-function correlated ("white") thermal noise. Langevin's equation is thus given by:[8]

$$\dot{r} = v \tag{3}$$

$$m\dot{v} = f(r,t) - \alpha v + \beta(t), \tag{4}$$

where $f(r,t)$ is the deterministic force acting on the particle, $\alpha > 0$ is a constant friction coefficient, and $\beta(t)$ denotes the thermal noise. To satisfy Einstein's fluctuation-dissipation theorem, it can be assumed that the noise is Gaussian-distributed, with the following statistical properties:[9]

$$\langle \beta(t) \rangle = 0 \tag{5}$$

$$\langle \beta(t)\beta(t') \rangle = 2\alpha k_B T \delta(t - t'), \tag{6}$$

where k_B is Boltzmann's constant and T is the thermodynamic temperature.

Developing an accurate numerical integrator for Langevin's equation is not trivial, due to the nonanalytic nature of the thermal noise and the fact that the friction force is velocity-dependent. If the friction and noise terms are treated on equal footing with $f(r,t)$, one obtains the frequently used BBK (Brünger, Brooks, Karplus) integrator, which is simple, yet known to be inaccurate when employed with a moderate to large integration time step, dt.[10] Specifically, the BBK, as well as most other existing integrators (including Nosé-Hoover), tend to exhibit increasing artificial changes in the configurational sampling statistics as the time step is enlarged. This is rooted, in part, in the above-mentioned discrete-time artifact that momentum and position are not strictly mutually conjugated variables for $dt > 0$. Recently, a new and improved thermostat (a temporal discrete-time propagator of the Langevin equation) was introduced by Grønbech-Jensen and Farago (G-JF),[11] which reads:

$$r^{n+1} = r^n + b[dt\,v^n + \frac{dt^2}{2m}f^n + \frac{dt}{2m}\beta^{n+1}] \tag{7}$$

$$v^{n+1} = a\,v^n + \frac{dt}{2m}(af^n + f^{n+1}) + \frac{b}{m}\beta^{n+1}, \tag{8}$$

where

$$a = \frac{1 - \frac{\alpha dt}{2m}}{1 + \frac{\alpha dt}{2m}} \tag{9}$$

$$b = \frac{1}{1 + \frac{\alpha dt}{2m}}. \tag{10}$$

The discrete-time noise is:

$$\beta^{n+1} = \int_{t_n}^{t_{n+1}} \beta(t')\,dt', \tag{11}$$

which results in an uncorrelated Gaussian random number with zero mean and a variance given by the temperature and friction coefficient:

$$\langle \beta^n \rangle = 0 \tag{12}$$

$$\langle \beta^n \beta^l \rangle = 2\alpha k_B T dt \delta_{n,l}. \tag{13}$$

Notice that the limiting case, $\alpha = 0$, of the G-JF method outlined in Eqs. (7–13) reduces the method to the standard Størmer-Verlet algorithm of Eqs. (1) and (2).

The core of the G-JF method is that the fluctuation-dissipation relationship is intact in discrete-time with respect to the balance between the energy lost by friction over the actual distance traveled and the accumulated noise over the time step.[11] This implies that the resulting discrete-time trajectory is thermodynamically sound. It therefore enables simulations of diffusion and configurational space without compromising the sampling statistics, as the time step is varied throughout the numerical stability range.[11,12] The objective of this paper is to illuminate the statistical performance of the method for both low-dimensional systems, as well as complex, soft-matter systems for which we have implemented the G-JF algorithm in the simulation suite, ESPResSo, to demonstrate the resulting improvements that can be attained by modest revisions to existing molecular dynamics codes.

2. Application to Simple Oscillators

To appreciate the sampling strength of the G-JF thermostat, we first study a particle moving in one-dimensional space with potential energy $U(r)$. We investigate the equilibrium statistics of the system by integrating Eqs. (3) and (4) with

$$f(r) = -\frac{\partial U}{\partial r}. \tag{14}$$

We can regard Eqs. (3) and (4) as normalized, along with all variables, if we assume that r is normalized to a characteristic displacement r_0, m is measured in units of m_0, energy U is in units of E_0, time t is in units of $t_0 = r_0\sqrt{m_0/E_0}$, velocity v is in units of $v_0 = r_0/t_0$, and normalized temperature is given by $\theta = k_B T/E_0$. Our simulation results shown in this paper are for $m = \theta = 1$.

In what follows, we consider confining potentials, where an object with coordinate r has a localized equilibrium distribution function $\rho_{eq} \propto exp[-U(r)/k_B T]$. We simulate the Langevin dynamics with three discrete-time algorithms: G-JF,[11] BBK,[10] and one by Stoll and Schneider (SS)[13] (where we have set the algorithm parameter to $p = 1$). The reason for the two latter choices is that they represent commonly used methods in distributed MD suites. From the simulations, we obtain the normalized distribution function $\rho(r)$, which we use to generate the normalized potential of mean force $U_{pmf}(r) = c - \theta ln\rho(r)$, where c is a constant. A measure of the quality of the applied algorithm is then the difference $U_{pmf}(r) - U(r)$, with an appropriate choice of the constant c. We use a total of 10^{10} time steps for each acquired distribution function. From these simulations, we also derive the important normalized configurational temperature $\theta_C = k_B T_C / E_0$,[14,15]

$$T_C = \frac{E_0}{k_B} \frac{\langle (\partial U / \partial r)^2 \rangle}{\langle \partial^2 U / \partial r^2 \rangle}, \qquad (15)$$

which is a condensed measure of how well the configurational space is sampled.

We first consider a harmonic oscillator $U(r) = \frac{1}{2}\kappa r^2$ with $\kappa = 1/40$. Since this results in a linear equation of motion, the use of a Gaussian random variable will result in a Gaussian distribution $\rho(r)$ with zero mean (by symmetry, since $\langle \beta \rangle = 0$). It was shown analytically in Ref. [11] that the resulting variance of $\rho(r^n)$ is $V(\rho(r)) = \frac{1}{2}\theta$, which implies that the G-JF algorithm reproduces the correct Boltzmann distribution precisely in discrete time for any applied time step $dt < 2/\sqrt{\kappa/m} = dt_{max}$ within the stability limit of the extended Størmer-Verlet methods for Langevin dynamics.[16] This essential feature is verified by simulations, as shown in Figure 1 and Figure 2, where we display $U(r)$, along with $U_{pmf}(r)$ (Figure 1) and T_C, for different values of α and dt (Figure 2), computed using the three integration methods mentioned above. We observe the expected perfect agreement between the G-JF results for $U_{pmf}(r^n)$ and $U(r)$ (G-JF results are shown with a solid curve, while $U(r)$, which is shown dotted, is completely overlapped by the solid curve). In contrast, Figure 2 shows considerable deviations for both BBK (dashed) and SS (dash-dotted) methods as dt is increased. It is obvious that BBK consistently overestimates the configurational temperature, which is consistent with the flattening of the effective (pmf) potential seen in Figure 1. The SS method, however, has a more complex set of errors. For small dissipation, we see that this method also overestimates T_C, while large α generally underestimates the temperature. This is consistent with the hardening of the effective potential observed in Figure 1 for the SS algorithm. The results imply that both BBK and SS methods should be applied
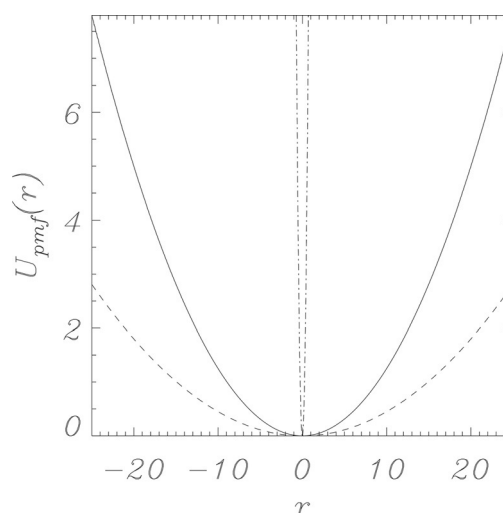

Figure 1. Potentials of mean force $U_{pmf}(r)$ from simulated harmonic oscillator using G-JF (solid), BBK (dashed), and SS (dash-dotted) methods for $\alpha = 2$ and $dt = 0.8 dt_{max}$. True potential (dotted) is precisely reproduced by G-JF.
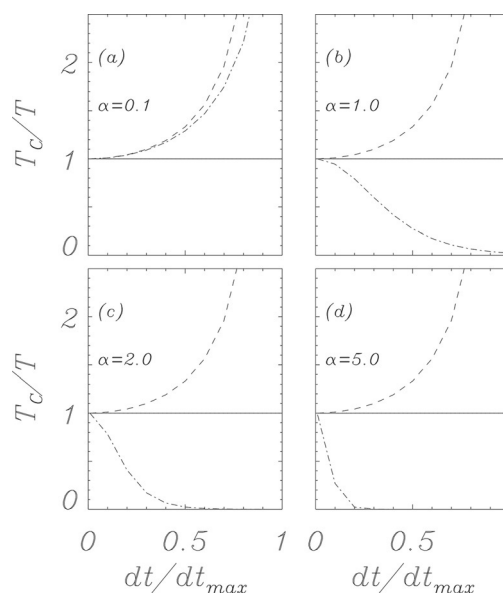

Figure 2. Configurational temperature T_C from simulated harmonic oscillator using G-JF (solid), BBK (dashed), and SS (dash-dotted) methods as a function of the applied time step (dt_{max} is the defined stability limit). True temperature (dotted) $\theta = 1$ is precisely reproduced by G-JF.

with considerable caution, and only with very small time steps, compared with the stability limit.

Second, we validate the performance of the methods for a highly nonlinear potential $U(r) = \frac{1}{2}\kappa r^2 - cos(r - \xi)$ for $\kappa = 1/40$ and $\xi = \frac{3}{4}\pi$ (which is chosen, somewhat arbitrarily, to create some asymmetry in the potential). The stability limit of the Størmer-Verlet methods is given by the maximum curvature of the potential, which in this case is $\tilde{\kappa} = \kappa + 1$. Thus, we define the stability limit for

the nonlinear problem to be $dt_{max} = 2/\sqrt{\tilde{\kappa}/m}$. The simulation results, which are displayed in Figures 3 and 4, reconfirm that the intuition from the harmonic oscillator generally translates to the strongly nonlinear case. The BBK integrator overestimates the configurational temperature by effectively lowering the local energy barriers, and where the SS method continues to have complex responses to variations in α and dt. The G-JF method is no longer exact when compared with the true (continuous time) expectations, but it is clearly superior to the reference methods. It is important to note that the discrepancies for nonlinear systems arise not from the G-JF method's implementation of dissipation and fluctuations, which are correctly balanced in discrete-time, but is an artifact of the discrete-time approximations to the behavior of the deterministic force within a single time step. This is an unavoidable feature common and inherent to all Verlet-type methods. The observed G-JF trend that the configurational temperature becomes increasingly more accurate for increasing friction coefficient α is due to the fact that the dissipation and fluctuation terms in the Langevin equation (4) become dominant for large α. Figure 4 therefore confirms the desired thermodynamic G-JF properties, since the G-JF method provides the correct configurational dissipation-fluctuation relationship in discrete-time. Figure 3 and Figure 4 also demonstrate that what may look to be minor differences in configurational temperature (see, e.g., Figure 4c for $dt = 0.8dt_{max}$) can, in fact, be masking rather large and significant deviations in the Boltzmann distribution (seen in Figure 3). This emphasizes the importance of validating the actual configurational distribution when considering if computer simulations represent the thermodynamic situation under investigation. Notice, however, that such validation is only
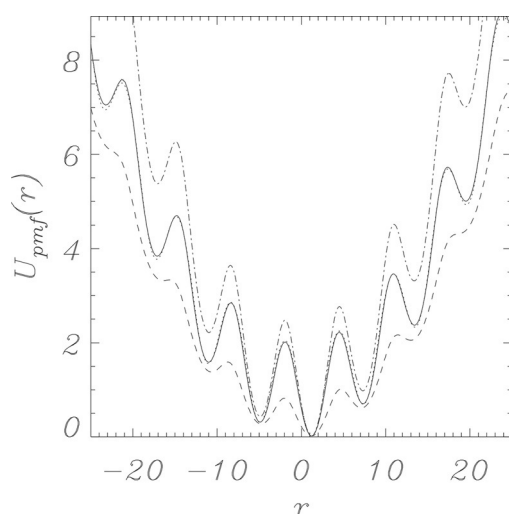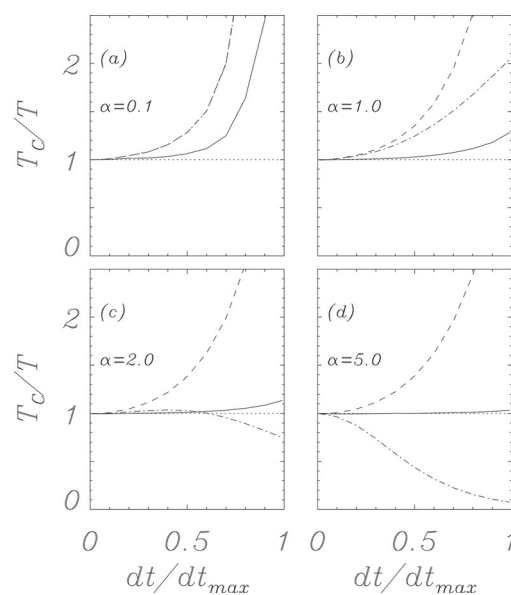

Figure 4. Configurational temperature T_C from simulated nonlinear oscillator using G-JF (solid), BBK (dashed), and SS (dash-dotted) methods as a function of the applied time step (dt_{max} is the defined stability limit). True temperature is shown as a dotted line.

possible in low-dimensional systems. It is also important to reemphasize that kinetic and configurational measures cannot be simultaneously correct, since the velocity parameter in the numerical methods is not exactly the velocity of the simulated trajectory. Thus, while the G-JF method will not provide the expected kinetic temperature, as measured by the average kinetic energy, the reason is that the velocity parameter is, in fact, not consistent with the configurational behavior in discrete time. The interesting complement to this observation is that a simulation method that provides correct kinetic behavior (such as kinetic temperature) cannot also reproduce correct configurational response unless the time step is very small.

3. Application to Soft Matter

The option to employ the G-JF thermostat has been added to the software simulation suite LAMMPS (large-scale atomic/molecular massively parallel simulator), a popular MD simulator for materials modeling, developed and maintained by Sandia National Laboratories.[17] The LAMMPS suite has the SS method as its other thermostat option (see MD comparison between G-JF and SS methods in Section 2 and in Ref. [12]). Here, we focus on another simulation package, ESPResSo (extensible simulation package for research on soft-matter systems), an open source software which has been developed at the Institute for Computational Physics of the University of Stuttgart. ESPResSo is typically used for MD simulations of large scale coarse-grained (CG) models of soft-matter

Figure 3. Potentials of mean force $U_{pmf}(r)$ from simulated harmonic oscillator using G-JF (solid), BBK (dashed), and SS (dash-dotted) methods for $\alpha = 2$ and $dt = 0.8dt_{max}$. True potential (dotted) is closely reproduced by G-JF.

systems, and it includes a BBK-type discrete-time thermostat.

To demonstrate the performance of the G-JF thermostat for more complex soft systems, we simulated a bilayer membrane of CG model lipids (see Figure 5) using the ESPResSo package. Each lipid is modeled as a trimer consisting of one hydrophilic and two hydrophobic beads of diameter σ, and the simulations are performed with no explicit solvent and with the Cooke-Kremer-Deserno force fields.[18] Specifically, all beads are subjected to a short-range repulsive potential by applying a cut-off to a standard Lennard-Jones potential, which is vertically shifted such that:

$$V_{rep}(r) = \begin{cases} 4\varepsilon\left[\left(\frac{\sigma'}{r}\right)^{12} - \left(\frac{\sigma'}{r}\right)^6 + \frac{1}{4}\right] & , \quad r < r_c \\ 0 & , \quad r \geq r_c \end{cases} \quad (16)$$

where $\sigma' = 0.95\sigma$ for head-head and head-tail interactions, $\sigma' = \sigma$ for tail-tail interactions, and $r_c = \sqrt[6]{2}\sigma'$. The bonds connecting the intra-lipid beads are described by the FENE potential:

$$V_{bond} = -\frac{1}{2}k_{bond}r_\infty^2 ln\left[1 - \left(\frac{r}{r_\infty}\right)^2\right], \quad (17)$$

with $k_{bond} = 30\varepsilon/\sigma^2$ and $r_\infty = 1.5\sigma$. Each lipid is straightened by a harmonic spring potential between the head and the second tail bead, given by:

$$V_{bend} = \frac{1}{2}k_{bend}(r - 4\sigma)^2, \quad (18)$$

where the bending stiffness $k_{bend} = 10\varepsilon/\sigma^2$. Finally, an attractive nonbonded interaction energy is introduced between any pair of hydrophobic tail beads. The attractive potential is given by:

$$V_{attr} = \begin{cases} -\varepsilon & , \quad r < r_c \\ -\varepsilon\,cos^2\frac{\pi(r-r_c)}{2\omega_c} & , \quad r_c \leq r < r_c + \omega_c \\ 0 & , \quad r \geq r_c + \omega_c \end{cases} \quad (19)$$

We choose the parameters $\varepsilon = k_B T$ and $\omega_c = 1.35\sigma$, which produce a bilayer membrane in the fluid state.[18] Normalized friction coefficients, masses, and temperature are chosen as unity. A snapshot of the bilayer membrane, taken from one of the simulations, is presented in Figure 5.

The bilayer membrane was simulated for 43,200 simulation time units, with integration time step dt varying from $dt = 0.001$ up to the stability limit of the system ($dt \approx 0.016$) in increments of $\Delta dt = 0.001$. System-wide energy measurements were taken every 12 time units, and included the kinetic temperature $T_k = \frac{2}{9}\langle E_k \rangle/N$, where N
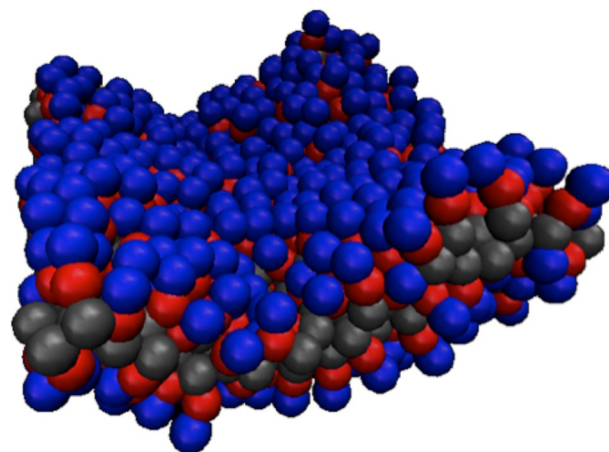

Figure 5. An equilibrium snapshot consisting of 500 lipids. Head, first tail, and second tail beads are, respectively, depicted in blue, red, and grey.

is the number of lipids (each modeled with three beads), total potential energy $\langle E_p \rangle$, and the separate contributions to $\langle E_p \rangle$ due to the FENE bonds (17), bond-bending energy (18), and nonbonded (NB) energy [sum of Eqs. (16) and (19)]. The box size for the simulation was set to $(17.6\sigma)^3$ (corresponding to nearly tensionless conditions), and was subjected to periodic boundary conditions. For each dt, two sets of simulations were performed: one with the Langevin thermostat currently used by ESPResSo, and the other with the G-JF integrator, which was implemented into the ESPResSo code. The results for $\langle E_p \rangle$ and its three constituent components are depicted in Figure 6, as a function of the simulation time step, dt. The error bars shown here are evaluated by the standard deviation of the results of three-thirds of the total simulation time. Consistent with previous studies comparing the performance of various Langevin thermostats,[12] the results here also demonstrate that, unlike other methods, and over the entire stability range, the G-JF integrator does not create increasing artificial variations in the sampling statistics. We also reveal that with the thermostat currently implemented in ESPResSo, the error in the potential energy is caused mainly by the NB interactions, while a smaller error arises from the FENE bonds. The bond-bending interactions seem to be accurately evaluated for all time steps. These observations can be understood by considering the curvature of the interaction energies, which is largest for the repulsive pair potential (16), and smallest for the bond-bending interaction (18).

Figure 7 shows the computed results for the measure of kinetic temperature T_k, as a function of dt. The trends observed here are opposite to the ones shown in Figure 6. The G-JF method leads to a decrease in T_k with dt, while the ESPResSo thermostat gives $k_B T_k/E_0 = 1$ for all simulated time steps. This feature has also been previously observed and discussed.[4,11,12,19] As mentioned in the intro-
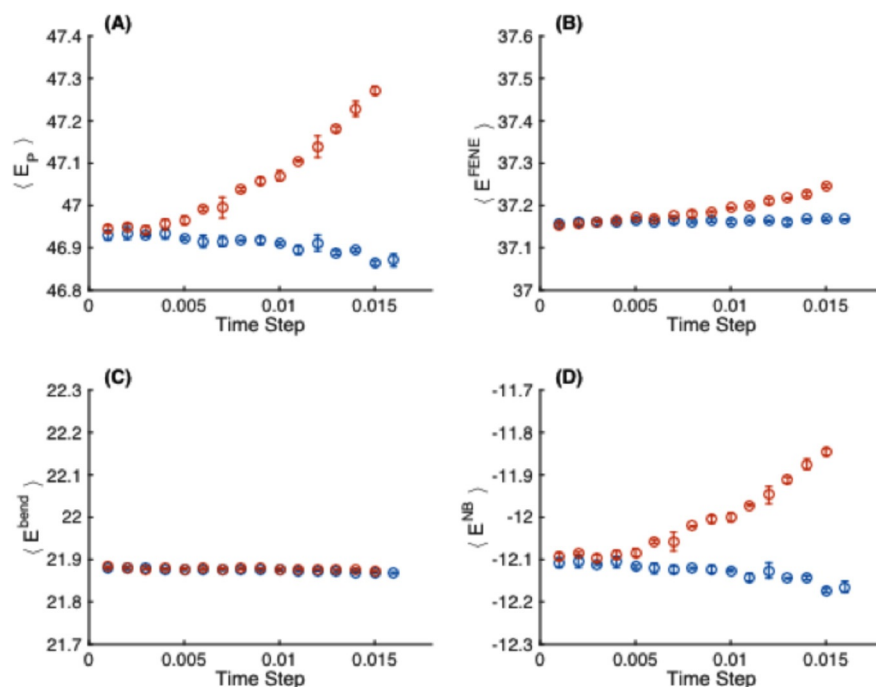
Figure 6. The mean potential energy (A) and the separate contributions of the FENE bonds (B), bond-bending energy (C), and nonbonded interactions (D), as a function of the time step of the simulation. Red and blue symbols represent, respectively, the results obtained with the Langevin thermostat currently installed in ESPResSo, and those obtained when the thermostat is replaced by G-JF. Energies are normalized per lipid.
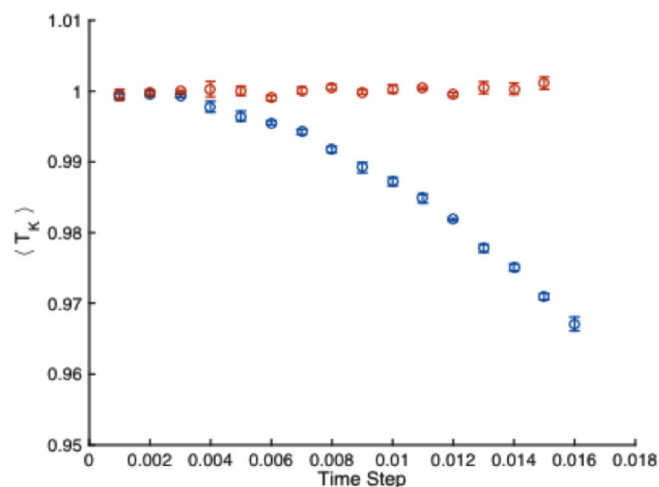

Figure 7. The measured kinetic temperature as a function of dt. Red and blue symbols represent, respectively, the results obtained with the Langevin thermostat currently installed in ESPResSo, and those obtained when the thermostat is replaced by G-JF. Energies are normalized per lipid.

duction, it stems from the fact that the discrete-time momentum mv^n is not exactly conjugated to the coordinate r^n. Consequently, the kinetic temperature is not a good measure for high-quality statistical sampling – see Ref. [4]. In general, thermostats exhibiting the correct T_k in discrete time must produce errors in computed configurational thermodynamic quantities.

4. Conclusion

In conclusion, the G-JF Langevin thermostat has been tested on both simple linear and nonlinear oscillators, and it has been demonstrated that the expected exact statistics for linear systems are obtained. For nonlinear systems, we find some deviations for large time steps. These originate from the inherent time discretization of the deterministic force – a feature common to all discrete-time numerical methods. The correct discrete-time implementation of the fluctuation relationship through the G-JF method is validated by the limit of large α, where the dynamics are dominated by noise and friction, and where G-JF gives near-perfect agreement with the continuous-time expectation. We have further implemented G-JF into the ESPResSo molecular simulation package, and it has been applied to simulations of a CG implicit-solvent bilayer membrane. The simulation results presented here demonstrate, once again, that this newly developed integrator exhibits no shift in the values of measured configurational thermodynamic quantities with increasing simulation time steps. This allows one to run a simulation with considerably larger time steps, and provides the user with peace of mind about the accuracy of the configurational results. The G-JF integrator is currently available within the LAMMPS simulation package, and we advise users of other popular suites, where older, considerably less accurate thermostats are implemented, to run simulations with caution and small time steps, dt.

Full Paper

Acknowledgments

References

[1] D. C. Rapaport, *The Art of Molecular Dynamics Simulations*, Cambridge University Press, Cambridge, **2004**.

[2] L. Verlet, *Phys. Rev.* **1967**, *159*, 98.

[3] D. Frenkel, B. Smit, *Understanding Molecular Simulations: From Algorithms to Applications*, Academic Press, San Diego, **2002**.

[4] N. Grønbech-Jensen, O. Farago, *J. Chem. Phys.* **2014**, *141*, 194108.

[5] S. Nosé, *J. Chem. Phys.* **1984**, *81*, 511.

[6] W. G. Hoover, *Phys. Rev. A* **1985**, *31*, 1695.

[7] E. Paquet, H. L. Viktor, *BioMed Res. Int.* **2015**, 183918.

[8] P. Langevin, *C.R. Acad. Sci. Paris* **1908**, *146*, 530.

[9] G. Parisi, *Statistical Field Theory*, Addison-Wesley, Menlo Park, **1988**.

[10] A. Brünger, C. L. Brooks, M. Karplus, *Chem. Phys. Lett.* **1984**, *105*, 495.

[11] N. Grønbech-Jensen, O. Farago, *Mol. Phys.* **2013**, *111*, 983.

[12] N. Grønbech-Jensen, N. R. Hayre, O. Farago, *Comput. Phys. Commun.* **2014**, *185*, 524.

[13] T. Schneider, E. Stoll, *Phys. Rev. B* **1978**, *17*, 1302.

[14] J. O. Hirschfelder, *J. Chem. Phys.* **1960**, *33*, 1462.

[15] G. Rickayzen, J. G. Powles, *J. Chem. Phys.* **2001**, *114*, 4333.

[16] We note that another recent method [B. Leimkuhler, C. Matthews, *Appl. Math. Res. Express* **2013**, *2013*, 34] also provides exact Boltzmann distributions in the linear Hooke's law case. However, this method does not reproduce the correct Einstein diffusion. The G-JF method reproduces diffusion correctly, in addition to the correct Boltzmann distribution. Please see Refs. [11,12].

[17] S. Plimpton, *J. Comp. Phys.* **1995**, *117*, 1.

[18] I. R. Cooke, K. Kremer, M. Deserno, *Phys. Rev. E: Stat. Nonlinear Soft Matter Phys.* **2005**, *72*, 011506.

[19] M. P. Eastwood, K. A. Stafford, R. A. Lippert, M. Ø. Jensen, P. Maragakis, C. Predescu, R. O. Dror, D. E. Shaw, *J. Chem. Theory Comput.* **2010**, *6*, 2045.