

# Molecular Physics

An International Journal at the Interface Between Chemistry and Physics

ISSN: 0026-8976 (Print) 1362-3028 (Online) Journal homepage: <https://www.tandfonline.com/loi/tmph20>

## Comparison of modern Langevin integrators for simulations of coarse-grained polymer melts

J. Finkelstein, G. Fiorin & B. Seibold

To cite this article: J. Finkelstein, G. Fiorin & B. Seibold (2020) Comparison of modern Langevin integrators for simulations of coarse-grained polymer melts, *Molecular Physics*, 118:6, e1649493, DOI: [10.1080/00268976.2019.1649493](https://doi.org/10.1080/00268976.2019.1649493)

To link to this article: <https://doi.org/10.1080/00268976.2019.1649493>



Published online: 02 Aug 2019.



Submit your article to this journal 



Article views: 183



View related articles 



View Crossmark data 



Citing articles: 1 [View citing articles](#) 

RESEARCH ARTICLE



## Comparison of modern Langevin integrators for simulations of coarse-grained polymer melts

J. Finkelstein<sup>a</sup>, G. Fiorin<sup>b,c</sup> and B. Seibold<sup>a</sup>

<sup>a</sup>Department of Mathematics, Temple University, Philadelphia, PA, USA; <sup>b</sup>Institute for Computational Molecular Science, Temple University, Philadelphia, PA, USA; <sup>c</sup>National Heart, Lung and Blood Institute, Bethesda, MD, USA

### ABSTRACT

For a wide range of phenomena, current computational ability does not always allow for atomistic simulations of high-dimensional molecular systems to reach time scales of interest. Coarse-graining (CG) is an established approach to alleviate the impact of computational limits while retaining the same algorithms used in atomistic simulations. It is important to understand how algorithms such as Langevin integrators perform on non-trivial CG molecular systems, and in particular how large of an integration time step can be used without introducing unacceptable amounts of error into averaged quantities of interest. To investigate this, we examined three different Langevin integrators on a CG polymer melt: the recently developed BAOAB method by Leimkuhler and Matthews [J. Chem. Phys. **138** (17), 05B601\_1 (2013)], the Grønbech-Jensen and Farago method [Mol. Phys. **111** (8), 983-991 (2013)], or G-JF, and the frequently used Brünger–Brooks–Karplus integrator [Chem. Phys. Lett. **105** (5), 495-500 (1984)], known as BBK. We compute and analyse key statistical properties for each. Our results indicate that the integrators perform similarly for a small friction parameter; however outside this regime, the use of large integration steps produces significant deviations from the predicted diffusivity and steady-state distributions for all methods examined with the exception of G-JF.

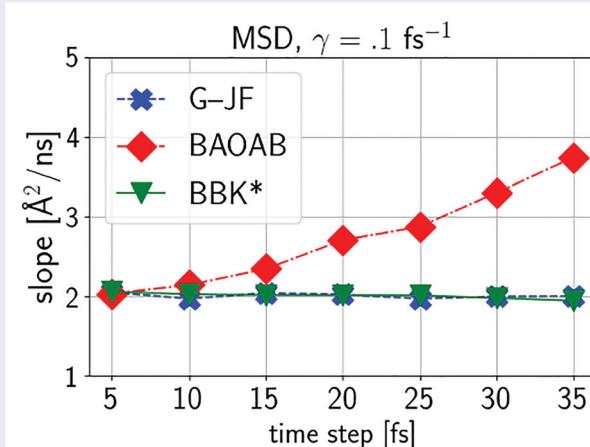
### ARTICLE HISTORY

Received 2 April 2019

Accepted 2 July 2019

### KEYWORDS

Langevin integrator;  
coarse-grained; polymer  
melt; molecular dynamics



## 1. Introduction

A central obstacle in using molecular dynamics (MD) simulations for quantitative predictions in material science and molecular biology is the presence of a wide range of time scales that are not well-separated. To investigate phenomena in such systems that occur over

large time scales, while accurately resolving smaller ones, a large number of integration steps is required. For example, in fully atomistic simulations, the size of the integration time step is constrained by the fastest physical time scales and is typically on the order of femtoseconds. It becomes computationally inefficient, yet necessary, to

use these relatively small time steps for integrating the medium to long-range time scale portion of the force field.

There exist several methods for working around this bottleneck. One such method is Nose–Hoover chains with or without the reference system propagator algorithm, or more succinctly, RESPA [1–3]. Such an algorithm allows one to propagate each degree of freedom forward in time with a time step size appropriate for that degree of freedom's characteristic time scale. These resulting solution operators are then composed via a symmetric Trotter splitting. Unfortunately, the Nose–Hoover chain equations result in the introduction of many undetermined parameters and, depending on the complexity of the system being simulated, may require different atoms to be coupled to different thermostats, thus complicating implementation. Other deterministic approaches which attempt to constrain the fast degrees of freedom are also used [4].

A Langevin thermostat is a simple and efficient way of simulating constant temperature conditions and has the intuitive physical interpretation of describing a molecular system in the presence of an implicit solvent or heat bath. The interaction between the heat bath and the system is collapsed into the friction parameter  $\gamma$ , thus avoiding the need to represent the heat bath as a set of particles altogether. In the context of coarse-grained (CG) models with explicit solvent, the missing microscopic degrees of freedom can be described by such stochastic forces, at least to a first approximation. For example, the dissipative-particle-dynamics (DPD) methodology [5] relies on this description to systematically build coarse-grained models with low computational cost.

Because improvements in computer hardware are being introduced more slowly in recent years, the usefulness of MD models that run at speeds higher than atomistic simulations will also increase. As CG models become more commonplace, there is a need to systematically understand the numerical performance and the limitations of current temporal integration schemes on CG systems, avoiding the reliance on rules of thumb that were derived primarily for atomistic simulations.

In this work, we examine methods used to numerically solve the Langevin equation, such as the ones by Leimkuhler and Matthews [6] and Grønbech-Jensen and Farago [7], known as G-JF, with particular concern toward their respective sampling properties in CG simulations. Among the family of integrators described by Leimkuhler and Matthews [6], the BAOAB method is the one characterised by the smallest configurational sampling error, and the only one here considered. Both the G-JF and BAOAB methods are weakly second-order accurate [6, 8] and produce the exact

configurational mean, variance and co-variance of the harmonic oscillator. This important property is not produced by many other Langevin schemes [9, 10]. To compare the two schemes with a representative of more traditional integration methods, our analysis includes the well-established Brünger–Brooks–Karplus method [11], or BBK. Out of the many formulations of BBK [12] we chose one (indicated here as BBK\*) that performs well for the system under study: comparisons to the classical formulation are also made.

Several numerical studies have been conducted on the performance of these integrators in atomistic simulations (e.g. see [13, 14]). However, CG models tend to use smoother potential energy functions than their fully atomistic counterparts, and thus allow for much larger integration time steps. A key question that arises then is: how large can the time step  $h$  be made without introducing unacceptable levels of error into averaged static and dynamic quantities? In [15], G-JF is used to simulate a CG lipid bilayer in implicit solvent and averaged energy terms (both potential and kinetic) are examined: however, the system was simulated for relatively short MD trajectories ( $< 50,000$  steps) and distributions were not examined. Also included in that study was the Schneider–Stoll Langevin integrator [16], the default option in LAMMPS [17] and ESPResSo [18]. We elected not to include this integrator in our work, because small  $\gamma h$  values are explicitly required in its derivation. Our present work considers a wide range of values for  $\gamma h$ . To date, we are not aware of any CG studies for BAOAB.

There are several choices of CG-model resolutions to choose from: a recent survey of several CG models [19] suggested that an adequate representation of the phase behaviour seen in atomistic simulations is given by models of polyethylene chains with three or four methylene groups per CG particle. The model by Klein and coworkers [20, 21], the MARTINI [22] and Salerno–Grest [19] models are significant examples of this level of resolution, and many other CG models featuring a wide-variety of resolutions applied to a broad range of systems also exist (e.g. see [23–29]). It is also recognised that technological requirements will motivate further effort to develop accurate CG models at lower levels of resolution (mapping into fewer CG particles for the same system). Transitioning into such models affects significantly the balance between Hamiltonian, stochastic and inertial terms in the equations of motion, and may require the use of existing Langevin methods outside of their typical range of parameter values.

In fact, Langevin parameters near the high-friction limit are of particular concern. Here, the acceleration may be neglected and the second-order integrator can safely be substituted with a first-order one for improved

numerical stability. However, due to large differences in how existing CG models are formulated, as well as differences between potential energy terms in the same model, it is far from unusual to see applications of a Langevin thermostat that approach this region at least transiently. Unfortunately, any resulting biases in sampling are often difficult to detect due to the heterogeneous nature of the physical system examined, or its proximity to a phase transition.

Here we used a polyethylene melt ( $C_{48}$ ) as a benchmark system modelled with three methylene groups per CG particle. All three schemes, BAOAB, BBK and G-JF, were compared for a wide range of friction parameter values and time step sizes by examining relevant statistical quantities from the simulations. The key finding of our study is that in the high-friction ( $\gamma \approx 0.1 \text{ fs}^{-1}$ ) regime, the G-JF method performs measurably better than BAOAB and BBK in reproducing molecular diffusivity and configurational distributions. The results obtained provide indications that Langevin integrators with similar properties to G-JF should be considered for use in CG simulations that aim to preserve dynamic properties and stationary distributions equally accurately. Though, diffusivity notwithstanding, BAOAB and G-JF sample equally well the configurational distributions of the system considered here.

## 2. Background and theory

We consider an  $N$ -particle system with potential energy  $U$ , immersed in a heat bath with the constant temperature  $T$ , modelled by the Langevin equation:

$$\begin{aligned} d\mathbf{Q} &= \mathbf{M}^{-1} \mathbf{P} dt, \\ d\mathbf{P} &= -\nabla U(\mathbf{Q}) dt - \gamma \mathbf{P} dt \\ &\quad + \sigma \mathbf{M}^{1/2} d\mathbf{W}. \end{aligned} \quad (1)$$

Here  $\sigma = \sqrt{2k_b T \gamma}$  is the noise coefficient,  $k_b$  is Boltzmann's constant,  $\gamma$  is the (spatially independent) collision rate parameter (measured in units of  $\text{fs}^{-1}$ ),  $\mathbf{M}$  a diagonal mass matrix,  $\mathbf{W}$  is  $3N$ -dimensional Brownian motion, and  $\mathcal{H}(\mathbf{Q}, \mathbf{P}) = \frac{1}{2} \mathbf{P}^T \mathbf{M}^{-1} \mathbf{P} + U(\mathbf{Q})$  is the Hamiltonian. The Langevin equation is a stochastic differential equation (SDE), so we use capital letters to remind us that position and velocity are stochastic processes.

Usually in MD, it is not the exact dynamics generated by (1) that are of interest, but rather an accurate sampling of the equilibrium distributions in phase space. Assuming that  $\mathcal{H}(q, p)$  is such that  $e^{-\mathcal{H}(q, p)}$  is integrable, one expects (1) to be ergodic and have the stationary

distribution

$$\mu(dq dp) = \mathcal{Q}^{-1} e^{-\mathcal{H}(q, p)/k_b T} dq dp,$$

where  $\mathcal{Q}$  is a normalisation constant and  $\mu$  is the Boltzmann–Gibbs, or canonical, distribution. For realistic MD potentials, such as Lennard-Jones and/or Coulombic interaction forces, the solution to (1) needs to be approximated numerically. Moreover, in many applications, such as the one here,  $U$  is non-globally Lipschitz and singular. Consequently, many standard results in SDE theory do not apply, thus limiting the possibilities of a complete formal analysis of numerical schemes for (1). One must therefore study these schemes computationally.

The fidelity to which numerical schemes reproduce the Boltzmann–Gibbs distribution  $\mu$  is our principal interest. It is the hope that infinite time-averaged observables obtained from these numerical schemes would reproduce correct statistical averages with respect to  $\mu$ . However, even with ergodicity typically assumed, the steady states produced by each numerical scheme,  $\mu_{\text{G-JF}}$ ,  $\mu_{\text{BAOAB}}$ , and  $\mu_{\text{BBK}}$ , respectively, will in general differ from  $\mu$ , and depend on the friction parameter  $\gamma$  and time step  $h$ . Therefore we expect two distinct sources of error in the calculation of distributions and observables: the use of a finite trajectory instead of an infinite one and the error due to the numerical scheme's steady state distribution differing from the canonical distribution.

Although, in general, statistical averages generated by numerical approximations of (1) cannot be derived analytically, exact formulas for mean, variance and correlation can be calculated in the flat and harmonic potential case, which is enough to characterise any stationary distribution when starting with Gaussian initial conditions.

### 2.1. Numerical methods studied

In this work, we examined three different Langevin integration schemes: G-JF, BAOAB, and the Brünger–Brooks–Karplus method [11], known as BBK, on a CG polymer melt. Both G-JF and BAOAB are included as options in LAMMPS and NAMD, respectively [30, 31]. The G-JF thermostat is a stochastic two-stage partitioned Runge–Kutta method [8, 32] and was shown to have highly desirable configurational properties [7]; particularly, Einstein's diffusion relation holds exactly and the configurational averages for the harmonic oscillator are independent of both the time step  $h$  and the friction parameter  $\gamma$ .

As described in [6], BAOAB is but one of many splitting schemes obtained by composing solution operators in various orderings which evolve the A, B and O portions

of the Langevin vector field (1):

$$\begin{aligned} \left( \begin{array}{c} d\mathbf{Q} \\ d\mathbf{P} \end{array} \right) &= \underbrace{\left( \begin{array}{c} \mathbf{M}^{-1}\mathbf{P} \\ 0 \end{array} \right)}_A dt + \underbrace{\left( \begin{array}{c} 0 \\ -\nabla U(\mathbf{Q}) \end{array} \right)}_B dt \\ &+ \underbrace{\left( \begin{array}{c} 0 \\ -\gamma \mathbf{P} dt + \sigma \mathbf{M}^{1/2} d\mathbf{W} \end{array} \right)}_O. \end{aligned}$$

BAOAB is the result of a Strang splitting with a full step for O in the middle. That is, taking half steps for B and then A, a full step for O and then half steps again for A and then B.

Similar to G-JF, BAOAB also reproduces exact sampling for the harmonic oscillator. Moreover, it possesses an additional favourable configurational sampling property, termed ‘super-convergence’ [6]: when  $\gamma$  is sufficiently large, the leading order error terms of averaged phase space quantities will exhibit a 4th order error scaling in  $h$  for typical time step values, thereby yielding more accurate averages with the same computational effort. This property provides some motivation as to why BAOAB is of interest to practitioners.

BBK has been a well-known Langevin discretisation method for the last three decades and is the default Langevin integrator in the popular MD suite, NAMD [31]. Similar to BAOAB, BBK is also a splitting method. It is weakly first-order accurate [10] and in the free particle case reproduces the Einstein relation. However, exact statistics are not recovered in the case of the harmonic oscillator, in sharp contrast to G-JF and BAOAB. Equation (1) is often re-formulated in terms of position and velocity, instead of momentum:

$$\begin{aligned} d\mathbf{Q} &= \mathbf{V} dt, \\ d\mathbf{V} &= -\mathbf{M}^{-1}\nabla U(\mathbf{Q}) dt - \gamma \mathbf{V} dt + \sigma \mathbf{M}^{-1/2} d\mathbf{W}. \end{aligned} \quad (2)$$

This form will serve as the governing equation for our forthcoming analysis and discussion. Set  $a := (1 - \gamma h/2)(1 + \gamma h/2)^{-1}$  and  $b := (1 + \gamma h/2)^{-1}$ , with  $h$  being the time step used for discretisation. Note that, for  $\gamma h$  sufficiently small, one has

$$a := \frac{1 - \gamma h/2}{1 + \gamma h/2} = e^{-\gamma h} + O((\gamma h)^3),$$

which leads to  $a^n = e^{-\gamma t} + O((\gamma h)^2)$  when  $nh = t$ . The quantities  $a$  and  $a^n$  appear several times in the subsequent paragraphs and sections.

We start by displaying the recursion formulas for the considered numerical schemes discretising (2). For a fixed time step  $h > 0$  and initial configuration

of  $(\mathbf{Q}_0, \mathbf{V}_0)$ , the position and velocity at time  $t = nh$  for each scheme are displayed below. Define  $\tilde{\sigma} = \sqrt{k_b T(1 - e^{-2\gamma h})}$ . Then, the G-JF update rule is:

$$\begin{aligned} \mathbf{Q}_n &= \mathbf{Q}_{n-1} + bh\mathbf{V}_{n-1} - \frac{bh^2}{2}\mathbf{M}^{-1}\nabla U(\mathbf{Q}_{n-1}) \\ &+ \frac{b\sigma h^{3/2}}{2}\mathbf{M}^{-1/2}\xi_{n-1}, \\ \mathbf{V}_n &= a\mathbf{V}_{n-1} - \frac{h}{2}\mathbf{M}^{-1}(a\nabla U(\mathbf{Q}_{n-1}) + \nabla U(\mathbf{Q}_n)) \\ &+ b\sigma\sqrt{h}\mathbf{M}^{-1/2}\xi_{n-1}, \end{aligned} \quad (3)$$

the BAOAB method is:

$$\begin{aligned} \mathbf{Q}_n &= \mathbf{Q}_{n-1} + \frac{h}{2}(1 + e^{-\gamma h})\mathbf{V}_{n-1} - \frac{h^2}{4}(1 + e^{-\gamma h}) \\ &\times \mathbf{M}^{-1}\nabla U(\mathbf{Q}_{n-1}) + \frac{\tilde{\sigma}h}{2}\mathbf{M}^{-1/2}\xi_{n-1}, \\ \mathbf{V}_n &= e^{-\gamma h}\mathbf{V}_{n-1} - \frac{h}{2}\mathbf{M}^{-1}(e^{-\gamma h}\nabla U(\mathbf{Q}_{n-1}) \\ &+ \nabla U(\mathbf{Q}_n)) + \tilde{\sigma}\mathbf{M}^{-1/2}\xi_{n-1}, \end{aligned} \quad (4)$$

and finally the BBK method is:

$$\begin{aligned} \mathbf{Q}_n &= \mathbf{Q}_{n-1} + h(1 - \gamma h/2)\mathbf{V}_{n-1} - \frac{h^2}{2}\mathbf{M}^{-1}\nabla U(\mathbf{Q}_{n-1}) \\ &+ \frac{\sigma h^{3/2}}{2}\mathbf{M}^{-1/2}\xi_{n-1}, \\ \mathbf{V}_n &= a\mathbf{V}_{n-1} - \frac{bh}{2}\mathbf{M}^{-1}(\nabla U(\mathbf{Q}_{n-1}) + \nabla U(\mathbf{Q}_n)) \\ &+ \frac{b\sigma\sqrt{h}}{2}\mathbf{M}^{-1/2}(\xi_{n-1} + \xi_n). \end{aligned} \quad (5)$$

In its original formulation [11], the BBK numerical scheme is given for only the position, leaving some ambiguity as to how the velocities are defined. Using the second-order approximation  $\mathbf{V}_n \approx (\mathbf{Q}_{n+1} - \mathbf{Q}_{n-1})/2h$ , the splitting formulation of BBK for both position and velocity is obtained (e.g. as seen in [33]). This is the same substitution one employs in transforming the position-only Verlet integrator to velocity Verlet:

$$\begin{aligned} \mathbf{V}_{n-1/2} &= \mathbf{V}_{n-1} + \frac{h}{2}\mathbf{M}^{-1} \left( -\nabla U(\mathbf{Q}_{n-1}) - \gamma \mathbf{M}\mathbf{V}_{n-1} \right. \\ &\quad \left. + \frac{\sigma}{\sqrt{h}}\mathbf{M}^{1/2}\xi_{n-1} \right), \\ \mathbf{Q}_n &= \mathbf{Q}_{n-1} + h\mathbf{V}_{n-1/2}, \\ \mathbf{V}_n &= \mathbf{V}_{n-1/2} + \frac{h}{2}\mathbf{M}^{-1} \\ &\quad \times \left( -\nabla U(\mathbf{Q}_n) - \gamma \mathbf{M}\mathbf{V}_n + \frac{\sigma}{\sqrt{h}}\mathbf{M}^{1/2}\xi_n \right). \end{aligned} \quad (6)$$

It is then a simple exercise to derive (5) from (6). In particular, this method requires two independent random variables  $\xi_{n-1}$  and  $\xi_n$ , where  $\xi_n$  is then re-used in the next step. However, [12] suggests the use of several variations of BBK which vary in how these random variables are selected. One such variation, which we denote by BBK\*, is obtained by taking  $\xi_{n-1} = \xi_n$  in (6), and not conducting any re-use in the next step. This BBK\* variant is not equivalent to the original version of BBK; and in fact, numerical tests indicated better all-around performance with our particular CG molecular system of interest in the commonly used regime of  $\gamma h \leq .01$  (see Figure 7).

## 2.2. Analytical properties of the methods in one dimension

Before moving to the computational results, we first summarise the key statistical properties for each of the three schemes in one dimension. This section attempts to highlight some key structural differences (and similarities) of the three schemes. Some schemes reproduce certain statistical quantities exactly whereas others reproduce such quantities only approximately.

We consider the standard examples of a single particle diffusing in a heat bath, and a standard harmonic oscillator, modelling for instance a covalent bond between two particles. In these cases, we can directly compare statistical quantities generated by the numerical schemes with those generated by the true analytical solution of (2). Though simplistic, these examples illustrate some important properties. Most of the calculations for these examples have, in parts, been previously exposited [6, 7, 9, 10, 32, 34].

### 2.2.1. Harmonic potential

We consider a potential function of the form  $U(Q) = \omega Q^2/2$ , with  $\omega > 0$ , so that  $\mathcal{H}(Q, P) = P^2/2m + \omega Q^2/2$ , and (2) becomes

$$\begin{aligned} dQ &= V dt, \\ dV &= -\omega Q dt - \gamma V dt \\ &\quad + \sigma m^{-1/2} dW. \end{aligned} \quad (7)$$

The following calculations show that the variance of position for the BAOAB and G-JF integrators is independent of  $\gamma$  and  $h$ . This is not true for BBK\*. For the linear system (7), stationary distributions can be analytically derived for the three methods, denoted by  $\mu_{\text{BAOAB}}$ ,  $\mu_{\text{BBK}^*}$ , and  $\mu_{\text{G-JF}}$ , respectively. We re-write (3) applied to (7) into matrix

form:

$$\begin{bmatrix} Q_{n+1} \\ V_{n+1} \end{bmatrix} = \begin{bmatrix} 1 - \frac{b\omega h^2}{2m} & bh \\ \frac{-h\omega b}{m} \left(1 - \frac{h^2\omega}{4m}\right) & a - \frac{h^2\omega b}{2m} \end{bmatrix} \begin{bmatrix} Q_n \\ V_n \end{bmatrix} + \begin{bmatrix} \frac{b\sigma h^{3/2}}{2\sqrt{m}} \\ \frac{b\sigma \sqrt{h}}{\sqrt{m}} \left(1 - \frac{h^2\omega}{4m}\right) \end{bmatrix} \xi_n. \quad (8)$$

This is a two-dimensional ergodic Markov chain with a unique stationary measure,  $\mu_{\text{G-JF}}$ . We can then calculate a corresponding matrix equation for  $Q_{n+1}^2$ ,  $V_{n+1}^2$  and  $Q_{n+1}V_{n+1}$ . Taking expectations on both sides of this equation, and taking  $n \rightarrow \infty$ , yields a subsequent  $3 \times 3$  linear system for  $\mathbb{E}(Q_\infty^2)$ ,  $\mathbb{E}(V_\infty^2)$  and  $\mathbb{E}((QV)_\infty)$ , the vector of steady-state averages. Solving the resultant linear system yields the G-JF stationary distribution:

$$\begin{aligned} \mu_{\text{G-JF}}(dq dv) &\propto \exp \left( -\beta \left( \frac{mv^2}{2 \left(1 - \frac{h^2\omega}{4m}\right)} + \frac{\omega q^2}{2} \right) \right) dq dv. \end{aligned}$$

Using (4) and (5) one can derive analogous linear equations and expressions for BBK\* and BAOAB, as follows.

- BBK\*:

$$\begin{bmatrix} Q_n \\ V_n \end{bmatrix} = \begin{bmatrix} 1 - \frac{h^2\omega}{2m} & h \left(1 - \frac{\gamma h}{2}\right) \\ \frac{-h\omega b}{m} \left(1 - \frac{h^2\omega}{4m}\right) & a \left(1 - \frac{h^2\omega}{2m}\right) \end{bmatrix} \times \begin{bmatrix} Q_{n-1} \\ V_{n-1} \end{bmatrix} + \begin{bmatrix} \frac{\sigma h^{3/2}}{2\sqrt{m}} \xi_{n-1} \\ \frac{\sigma b\sqrt{h}}{\sqrt{m}} \left(1 - \frac{h^2\omega}{4m}\right) \xi_{n-1} \end{bmatrix} \quad (9)$$

and

$$\begin{aligned} \mu_{\text{BBK}^*}(dq dv) &\propto \exp \left( -\beta \left( \frac{mv^2}{2} + \frac{(1 - \frac{h^2\omega}{4m})}{(1 - \gamma h/2 + (\gamma h/2)^2)} \frac{\omega q^2}{2} \right) \right) dq dv. \end{aligned}$$

- **BAOAB:**

$$\begin{bmatrix} Q_{n+1} \\ V_{n+1} \end{bmatrix} = \begin{bmatrix} 1 - (1 + e^{-\gamma h}) \frac{h^2 \omega}{4m} & (1 + e^{-\gamma h}) \frac{h}{2} \\ -\left(1 - \frac{h^2 \omega}{4m}\right) & e^{-\gamma h} \left(1 - \frac{h^2 \omega}{4m}\right) \end{bmatrix} \times \begin{bmatrix} Q_n \\ V_n \end{bmatrix} \quad (10)$$

$$+ \begin{bmatrix} \frac{h}{2} \sqrt{\frac{k_b T}{m} (1 - e^{-2\gamma h})} \\ \left(1 - \frac{h^2 \omega}{4m}\right) \sqrt{\frac{k_b T}{m} (1 - e^{-2\gamma h})} \end{bmatrix} \xi_n, \quad (11)$$

and

$$\begin{aligned} \mu_{\text{BAOAB}}(dq dv) \\ \propto \exp\left(-\beta\left(\frac{mv^2}{2(1 - h^2 \omega/4m)} + \frac{\omega q^2}{2}\right)\right) dq dv. \end{aligned}$$

The variances and co-variances for position and velocity are listed in Table 1. These expressions show the dependence on the dimensionless quantity  $\gamma h$  and the time step  $h$ . In particular, both BAOAB and G-JF produce the exactly correct configurational variance and co-variance.

Table 1 also shows that neither G-JF or BAOAB produce the correct velocity statistics for the harmonic potential. Two very recent papers [35, 36] extend the G-JF method (the resulting variants are called GJF-2GJ and GJF-F, respectively) by updating velocities on the half-step instead of at integral increments of the time step  $h$ . The authors show that this slight modification allows for *both* correct position and velocity statistics in the case of a one-dimensional harmonic oscillator. Moreover, the trajectories in both methods are the same as in G-JF and in fact for GJF-2GJ, no modification to an already

**Table 1.** Numerical stationary averages, as functions of  $h$  and  $\gamma$ , for the three numerical methods applied to the one-dimensional harmonic oscillator.

Method	$\langle Q_n^2 \rangle_{h,\gamma}$	$\langle V_n^2 \rangle_{h,\gamma}$	$\langle Q_n V_n \rangle_{h,\gamma}$
Exact	$\frac{k_b T}{\omega}$	$\frac{k_b T}{m}$	0
G-JF	$\frac{k_b T}{\omega}$	$\frac{k_b T}{m} (1 - h^2 \omega/4m)$	0
BBK*	$\frac{k_b T}{\omega} \left( \frac{1 - \gamma h/2 + (\gamma h/2)^2}{1 - \frac{h^2 \omega}{4m}} \right)$	$\frac{k_b T}{m}$	0
BAOAB	$\frac{k_b T}{\omega}$	$\frac{k_b T}{m} (1 - h^2 \omega/4m)$	0

Note: Both G-JF and BAOAB are exact for position variance, while BBK\* is not.

implemented G-JF algorithm is necessary. This is due to the velocity update,  $V_{n+1}$ , being constructed from the same  $Q_{n+1}$  and  $Q_n$  as the original G-JF, so one can compute half-step velocities either during the simulation or in post-processing.

The GJF-2GJ and GJF-F corrections alleviate a known issue: because temperature statistics are calculated in most MD software packages by using averaged kinetic energies from the variance of atomic velocities, numerical schemes with inaccurate velocity statistics yield a biased estimate of the resulting temperature at finite  $h$ . In these cases, adherence to equipartition should not be used as the main criterion to whether the system has correct temperature. Instead, the practitioner should estimate the system's temperature by less direct routes, such as computing diffusion profiles (the approach taken here) or steady-state positional distributions. Although the latter is typically a difficult task for realistic systems, methods such as the one in [37] can be used with any integrator to implement temperature calculation using only positional degrees of freedom.

### 2.2.2. Thermal diffusion

The simplest possible case for (2) is when  $F \equiv 0$ , i.e. free diffusion. Albeit simple, it is insightful to understand the behaviour of the integrators in this case. In thermal diffusion, (2) reduces to

$$\begin{aligned} dQ &= V dt, \\ dV &= -\gamma V dt + \sigma m^{-1/2} dW, \end{aligned}$$

which can be solved analytically. The velocity  $V_t$  is an Ornstein–Uhlenbeck process and has solution:

$$V_t = e^{-\gamma t} V_0 + \frac{\sigma}{\sqrt{m}} \int_0^t e^{-\gamma(t-s)} dW_s. \quad (12)$$

The particle position,  $Q_t$ , is then

$$\begin{aligned} Q_t &= Q_0 + \frac{1}{m} \int_0^t V_u du = Q_0 + \frac{1}{\gamma} (1 - e^{-\gamma t}) V_0 \\ &+ \frac{\sigma}{\sqrt{m}} \int_0^t \left( \int_0^u e^{-\gamma(u-s)} dW_s \right) du. \end{aligned} \quad (13)$$

Equation (13) is used to find the mean position and mean squared position. Assuming  $\mathbb{E}(Q_t) = 0$ ,

$$\begin{aligned} \mathbb{E}(Q_t^2) &= \mathbb{E}(Q_0^2) + 2 \overbrace{\mathbb{E}(Q_0)}^{=0} \mathbb{E} \left( \int_0^t V_u du \right) \\ &+ \mathbb{E} \left( \int_0^t V_u du \right)^2 \end{aligned}$$

$$= \mathbb{E}(Q_0^2) + \frac{1}{\gamma^2} (1 - e^{-\gamma t})^2 \mathbb{E}(V_0^2) \\ + 2D \left( t - \frac{2}{\gamma} (1 - e^{-\gamma t}) + \frac{1}{2\gamma} (1 - e^{-2\gamma t}) \right),$$

where  $D := k_b T/m\gamma$  is the diffusion coefficient. The large time asymptotic behaviour of the mean squared position for mean zero initial position is then

$$\mathbb{E}(Q_t^2) \sim 2Dt. \quad (14)$$

The velocity autocorrelation function and the covariance can also be computed from (12) and (13). These quantities are displayed in Table 2.

### 2.2.3. Diffusive behaviour of the numerical schemes

We set  $\omega = 0$  in (8)–(11) to obtain the update rules for each numerical scheme in the zero potential case. A key quantity of interest is the mean square displacement for the particle position. For BBK\*,

$$Q_n = Q_{n-1} + h(1 - \gamma h/2)V_{n-1} + \frac{\sigma h}{2\sqrt{m}}\xi_{n-1}, \\ V_n = aV_{n-1} + \frac{b\sigma\sqrt{h}}{\sqrt{m}}\xi_{n-1}. \quad (15)$$

Given a fixed time step size  $h > 0$ , iterate the above recursive formula backwards to write the position as a finite sum of independent Gaussians and the initial conditions:

$$Q_n = Q_0 + \frac{\left(1 - \frac{\gamma h}{2}\right)h(1 - a^{n+1})}{1 - a}V_0 \\ + \frac{\sigma h}{\sqrt{m}} \sum_{k=0}^{n-1} \xi_k \left( \frac{a(1 - a^{k+1})}{1 - a} + \frac{1}{2} \right).$$

Therefore,

$$\mathbb{V}(Q_n) = \mathbb{V}(Q_0) + (1 - \gamma h/2)^2 h^2 \left( \frac{1 - a^{(n+1)}}{1 - a} \right)^2 \mathbb{V}(V_0) \\ + 2D \left( t - \frac{2}{\gamma} (1 - a^n)(1 - \gamma h/2)^2 \right. \\ \left. + \frac{1}{2\gamma} (1 - a^{2n})(1 - \gamma h/2)^4 \right).$$

where  $t = nh$ . Sending  $n \rightarrow \infty$  shows that the scheme preserves the Einstein diffusion relation in the limit. As with BBK\*, G-JF preserves the Einstein diffusion relation in the long time limit as well. The position and its

variance at time  $t = nh$  are:

$$Q_n = Q_0 + \frac{bh(1 - a^{n+1})}{1 - a}V_0 + \frac{\sigma bh^{3/2}}{\sqrt{m}} \sum_{k=0}^{n-1} \xi_k \\ \times \left( \frac{(1 - a^{k+1})b}{1 - a} + \frac{1}{2} \right), \\ \mathbb{V}(Q_n) = \mathbb{V}(Q_0) + \frac{b^2 h^2 (1 - a^{(n+1)})^2}{(1 - a)^2} \mathbb{V}(V_0) \\ + 2D \left( t - \frac{2}{\gamma} (1 - a^n)a + \frac{1}{2\gamma} (1 - a^{2n})a^2 \right),$$

which has the same limiting behaviour as in [7]. Additionally, both BBK\* and G-JF generate the correct steady-state behaviour for velocity. For a fixed time step  $h$ , the BAOAB scheme approximates thermal diffusion via:

$$Q_n = Q_{n-1} + \frac{h}{2}(1 + e^{-\gamma h})V_{n-1} + \frac{h}{2}\sqrt{\frac{k_b T(1 - e^{-2\gamma h})}{m}}\xi_{n-1} \\ V_n = e^{-\gamma h}V_{n-1} + \sqrt{\frac{k_b T(1 - e^{-2\gamma h})}{m}}\xi_{n-1}.$$

An important distinction here is that  $V_{n+1}$  is given by the exact Ornstein–Uhlenbeck flow (in law), whereas the velocity updates are only approximate for BBK\* and G-JF. As is done with BBK\* and G-JF, by iterating backwards, we can write the position as a finite sum of i.i.d random variables:

$$Q_n = Q_0 + \frac{h}{2} \left( \frac{1 + e^{-\gamma h}}{1 - e^{-\gamma h}} \right) (1 - e^{-\gamma t})V_0 \\ + \frac{h k_b T}{2 m} (1 - e^{-2\gamma h}) \\ \times \sum_{k=0}^{n-1} \xi_k \left( \frac{1 + e^{-\gamma h}}{1 - e^{-\gamma h}} (1 - e^{-\gamma h(k+1)}) + 1 \right). \quad (16)$$

Then  $\mathbb{V}(Q_n)$  is

$$\mathbb{V}(Q_0) + \frac{\gamma h}{4} \left( \frac{1 + e^{-\gamma h}}{1 - e^{-\gamma h}} \right)^2 (1 - e^{-\gamma t})^2 \mathbb{V}(V_0) \\ + 2D \left( \frac{\gamma h}{2} \frac{(1 + e^{-\gamma h})^2}{1 - e^{-2\gamma h}} t \right. \\ \left. - \frac{\gamma h^2}{2} \frac{(1 - e^{-2\gamma h})(1 + e^{-\gamma h})}{(1 - e^{-\gamma h})^3} (1 - e^{-\gamma t})e^{-\gamma h} \right. \\ \left. + \frac{1}{2\gamma} (1 - e^{-2\gamma t})e^{-2\gamma h} \right).$$

This appears to be vastly different than the other schemes, but one can check that (14) is recovered when sending  $\gamma h \rightarrow 0$ . So whenever  $\gamma$  and  $h$  are fixed such that  $\gamma h$  is sufficiently small,  $\lim_{n \rightarrow \infty} \mathbb{V}(Q_n)/nh \approx 2D$ , and

**Table 2.** The analytic and numerical schemes' averages in the case of Brownian motion (Einstein diffusion) with  $\delta_0$ -distributed  $Q_0$  and Maxwell–Boltzmann-distributed  $V_0$ .

Method	$\langle Q_n^2 \rangle_{h,\gamma}$	$\langle V_n^2 \rangle_{h,\gamma}$	$\langle Q_n V_n \rangle_{h,\gamma}$	$C_{h,\gamma}(nh)$
Exact	$\sim 2Dnh$	$k_b T/m$	$(1 - e^{-\gamma h n})D$	$e^{-\gamma h n} \frac{k_b T}{m}$
G-JF	$\sim 2Dnh$	$k_b T/m$	$(1 - a^n)D$	$a^n \frac{k_b T}{m}$
BBK*	$\sim 2Dnh$	$k_b T/m$	$(1 - a^n)D(1 + \frac{\gamma h}{2})$	$a^n \frac{k_b T}{m}$
BAOAB	$\sim (\gamma h) \left( \frac{1 + e^{-\gamma h}}{1 - e^{-\gamma h}} \right) Dnh$	$k_b T/m$	$(1 - e^{-\gamma h n})D \left( \frac{\gamma h}{2} \frac{1 + e^{-\gamma h}}{1 - e^{-\gamma h}} \right)$	$e^{-\gamma h n} \frac{k_b T}{m}$

Notes: The averages for the numerical schemes are given as functions of  $h$  and  $\gamma$ . Both G-JF and BBK\* are exact for position variance, while BAOAB is not.  $C_{h,\gamma}(nh)$  is the velocity autocorrelation function of the numerical scheme with time step  $h$  and friction parameter  $\gamma$  at time  $nh$ .

BAOAB produces an acceptable approximation to the correct diffusive behaviour. More particularly, the time evolution of the variance for BAOAB at large times evolves according to  $2\tilde{D}t$  where

$$\tilde{D} = D \left( \frac{\gamma h}{2} \frac{(1 + e^{-\gamma h})^2}{1 - e^{-2\gamma h}} \right),$$

is the effective BAOAB diffusion coefficient, showing how the calculated diffusion for BAOAB deviates from theory when  $\gamma h$  is sufficiently large. The quantity inside the parentheses tends to 1 as  $\gamma h \rightarrow 0$  but exhibits linear behaviour as  $\gamma h$  is increased. Key statistical quantities for each numerical scheme are tabulated in Table 2 for simple initial conditions. In fact, among all methods within the aforementioned A,B,O family of splitting schemes, when a single random sample per time step is desired, the incorrect diffusive behaviour elucidated above is universal. In the free particle case, B induces the identity operator, so that the number of possible lettered combinations reduce to AO, OA, and AOA. A quick calculation reveals that none of these methods reproduces the Einstein relation.

### 3. Computational methodology

In our computational study we considered a collection of 128 polyethylene chains ( $C_{48}H_{98}$ ), simulated at a fixed temperature of 450 K (i.e. well above the melting point) in a box with side lengths of 58.065 Å and periodic boundary conditions. This resulted in a density of 0.4415 amu/Å<sup>3</sup> = 0.7331 g/ml. We used a coarse-grained model for the polymer melt obtained directly from liquid-phase physical properties [21]. The hydrocarbon chains are modelled in coarse-grained resolution, where the  $-\text{CH}_2\text{CH}_2\text{CH}_2-$  and  $\text{CH}_3\text{CH}_2\text{CH}_2-$  groups are mapped to spherical ‘beads’, called CM and CT, respectively [21]. Compared to atomistic resolution, this reduces the system from a total of 18,432 atoms to 2048 CG beads. Initial positions of the CM and CT particles

are taken from their respective centres-of-mass, and are evolved according to the following interaction energy:

$$U(\mathbf{q}) = \sum_{a \in \text{angles}} k_a (\theta - \theta_0)^2 + \sum_{b \in \text{bonds}} k_b (r - r_0)^2 + U_{LJ}^{(\alpha, \beta)}(\mathbf{q}),$$

where  $(\alpha, \beta) \in \{(\text{CT}, \text{CT}), (\text{CT}, \text{CM}), (\text{CM}, \text{CM})\}$  and  $U_{LJ}^{(\alpha, \beta)}$  is the 9–6 Lennard-Jones potential [21]

$$U_{LJ}^{(\alpha, \beta)}(\mathbf{q}) = \sum_{i \neq j}^N \frac{27}{4} \epsilon_{\alpha, \beta} \left( \left( \frac{\sigma_{\alpha, \beta}}{|\mathbf{q}_i - \mathbf{q}_j|} \right)^9 - \left( \frac{\sigma_{\alpha, \beta}}{|\mathbf{q}_i - \mathbf{q}_j|} \right)^6 \right) \times 1_{\{|\mathbf{q}_i - \mathbf{q}_j| < \delta\}},$$

with  $\epsilon_{\text{CT,CT}} = 0.42$  kcal/mol and  $\sigma_{\text{CT,CT}} = 4.506$  Å for the CT–CT interaction,  $\epsilon_{\text{CT,CM}} = 0.444$  kcal/mol and  $\sigma_{\text{CT,CM}} = 4.5455$  Å for the CT–CM interaction, and  $\epsilon_{\text{CM,CM}} = 0.469$  kcal/mol and  $\sigma_{\text{CM,CM}} = 4.585$  Å for the CM–CM interaction. Here  $\delta$  represents the Lennard-Jones cut-off distance of 15 Å. The function  $1_{\{|\mathbf{q}_i - \mathbf{q}_j| < \delta\}}$  is defined to be 1 for all pairs  $i, j$  satisfying  $|\mathbf{q}_i - \mathbf{q}_j| < \delta$  and 0 otherwise. The CM–CM bonds have force field constant  $k_b = 6.16$  kcal/mol and equilibrium length  $r_0 = 3.64$  Å, and the CM–CT bonds have force field constant  $k_b = 6.16$  kcal/mol and equilibrium length  $r_0 = 3.65$  Å. The force field parameters for the CM–CM–CM and CM–CM–CT angles were the same value:  $k_a = 1.19$  kcal/mol·rad<sup>2</sup>, equilibrium angles were  $\theta_0 = 173^\circ$  and  $\theta_0 = 175^\circ$ , respectively.

To construct a reference ensemble, we first identified a value for the integration time step that was guaranteed not to introduce artefacts. The fastest CG bond oscillation is of the type CM–CM. The mass of a CM particle is 42.7097 amu, giving a frequency of oscillation between two CM particles of

$$\nu_{\text{bond}} = \frac{1}{2\pi} \sqrt{\frac{0.03175286}{\mu}} \text{ fs}^{-1} = 0.00613710132 \text{ fs}^{-1}$$

with reduced mass  $\mu = 21.35485$  amu and a numerical units conversion factor of 0.002577344. So, the CG bond oscillation period is  $1/\nu_{\text{bond}} \approx 162.94$  fs. Thus a time step of 5 fs resolves well the time evolution of the CG bond vibration forces, more than 30 time steps per oscillation. Meanwhile, the Lennard-Jones forces have a characteristic frequency of about

$$\nu_{\text{LJ}} = \tau^{-1} = \sqrt{\frac{\epsilon}{m\sigma^2}} \approx \frac{1}{2205} \text{ fs}^{-1},$$

showing that the LJ forces are extremely well-resolved for all choices of time step. Both  $\nu_{\text{bond}}$  and  $\nu_{\text{LJ}}$  suggest that our choice of  $h = 5$  fs is sufficiently small.

We used the molecular dynamics engine LAMMPS [17], and implemented the integrators considered here using its Python interface (`fix python/move`) to the underlying data structures. The BAOAB, BBK and BBK\* integrators are not available as packages in LAMMPS and were instead implemented using this Python interface. Although a G-JF option is available with the `fix langevin` command, the version of LAMMPS at the time of writing does not implement G-JF in same way as given in [7]. Instead, it uses uniform random variables to approximate the Gaussian noise, in an effort to increase computational speed. However, such an approximation is only valid for small enough time steps [38]. As we are interested in the large time step regime, we implemented the original version of G-JF with Gaussian noise using the Python wrapper based on equations (20) and (21) in [7].

For each choice of  $\gamma$  and  $h$ , 100 independent simulations were conducted to reduce the error associated to finite length simulations in approximating phase space averages. Each simulation was performed from an identical spatial configuration for approximately 250 ns. This starting configuration was obtained by an initialisation run using LAMMPS' `fix npt` command, which implements an MTK thermostat/barostat [39], for 100 ns with a temperature of 450 K and pressure set to 1 bar. All simulations used to benchmark the Langevin integrators were run in the NVT ensemble.

## 4. Simulation results

The goal of our study is to understand how faithfully the different integrators reproduce relevant statistical averages of the coarse-grained model, particularly in the regime of large time steps and  $\gamma$  values. To that end, numerical experiments were performed using a range of friction parameters and time steps. We examined the cases of  $\gamma \in \{0.1, 0.01, 0.001, 0.0001\}$  and  $h \in \{5, 10, 15, 20, 25, 30, 35\}$  with units of  $\text{fs}^{-1}$  and fs, respectively.

In the simple case of a Brownian particle in a fluid,  $\gamma$  represents the rate of collision of the Brownian particle with bath particles. So the dimensionless quantity  $\gamma h$  gives a measure as to how many collisions occur over the length of the time step  $h$ . Similarly here, the number  $\gamma h$  determines the strength of the interaction between the system and the heat bath and is a fundamental quantity of the dynamics.

In applications of Langevin dynamics with atomistic models, friction rate parameters of the order of  $0.01 \text{ fs}^{-1}$  or less are typically used, as they give a reasonable approximation of the experimental diffusion coefficients of small molecules. Indeed in [11], the authors used BBK in an atomistic water simulation as their benchmark test, and  $\gamma$  was chosen quite small,  $0.0000196 \text{ fs}^{-1}$ . In [6], the BAOAB method was tested and compared to other integrators based on its performance on an atomistic alanine dipeptide molecule in water with  $\gamma = 0.001 \text{ fs}^{-1}$ .

However, in CG simulation models, much of the magnitude of the inter-atomic forces shifts from the conservative to the stochastic terms, and a higher friction rate  $\gamma$  may be needed to retain the same diffusivity. Alternatively, high-friction rates are also used simply to improve numerical stability of MD simulations near particular conditions (for example, near phase transitions). Therefore, we consider here a relatively broad range of  $\gamma$  values,  $0.0001 \text{ fs}^{-1}$  to  $0.1 \text{ fs}^{-1}$ : given the choice of integration time step  $h$  used in the following, the upper end of this interval may result in values of  $\gamma h$  larger than 1.

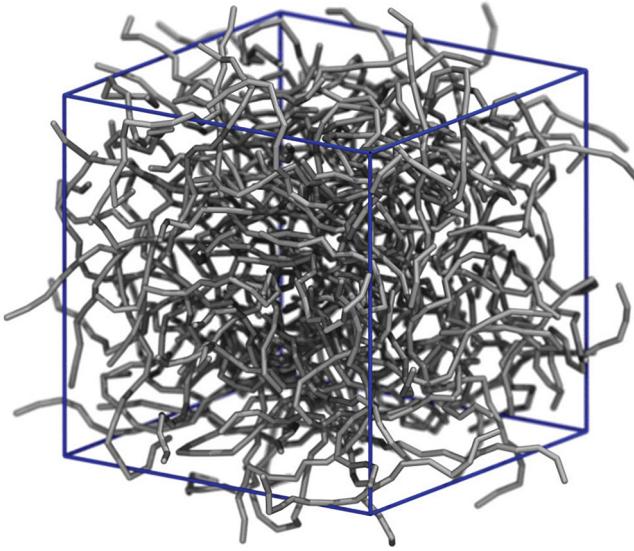
For smaller  $\gamma$  values, i.e.  $\gamma \leq 0.001 \text{ fs}^{-1}$ , the three integrators become numerically unstable around  $h \approx 38$  fs. Hence, 35 fs was chosen as the upper bound for our range of time step values. This stability limit for G-JF and BAOAB increases significantly for the largest choice of  $\gamma = 0.1 \text{ fs}^{-1}$  to slightly more than 50 fs. In contrast, this larger value for  $\gamma$  seemed to not have as much of an effect on BBK\* – simulations still exhibited instability at 40 fs.

### 4.1. Diffusive behaviour

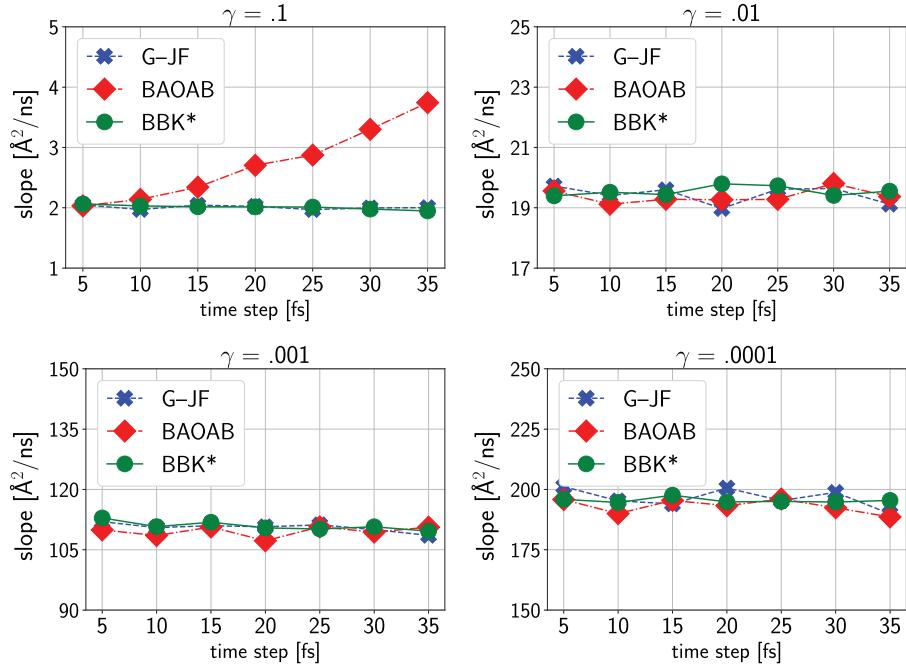
To characterise diffusion, the mean squared displacement (MSD) of individual molecules was computed as a function of the simulation time. After each time step in the simulation, we computed the MSD of the centre-of-mass for each polymer chain over that time step, using the LAMMPS command `compute msd`, and then averaged this result over all chains and added this to the same calculation from the previous step, i.e. we computed:

$$\mathcal{D}_h(t_k) := \frac{1}{N} \sum_{i=1}^N |Q_i^{\text{CM}}(t_k) - Q_i^{\text{CM}}(t_{k-1})|^2 + \mathcal{D}_h(t_{k-1}),$$

where  $t_k = kh$ ,  $1 \leq k \leq N$ , and  $Q_i^{\text{CM}}$  represents the centre-of-mass position for the  $i$ -th polymer chain ( $i = 1, \dots, 128$ ). The centre-of-mass drift of the entire system was subtracted from the MSD data at each time step before computing the MSD. The resulting MSD data was then block-averaged [40] over the 100 independent runs for each choice of  $\gamma$  and  $h$ , yielding 5000 independent samples (each block was taken long enough to allow for correlations to die off). This was done for each of the



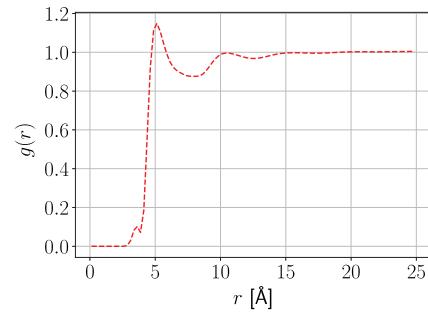
**Figure 1.** Graphical representation of the polymer melt inside the unit cell (blue): 128 CG C<sub>48</sub> polymer chains. This rendering was made using VMD after an initial thermalisation run.



**Figure 2.** Diffusion coefficients (with units of  $\text{\AA}^2/\text{ns}$ ) were calculated from numerical simulation for different  $\gamma$  values as a function of the time step  $h$ . The plot in the upper left displays the same kind of linear behaviour for BAOAB as the method exhibits in the one-dimensional free particle. In the remaining cases, all three methods produce the same calculated slopes within statistical error as a function of  $h$ .

three integrators. The MSD was observed to be linear with respect to time within statistical error. These plots were fitted with regression lines and the means of the slopes of these lines are plotted as a function of the time step  $h$  in Figure 2.

As discussed earlier, the diffusion coefficient for BAOAB in the case of the one-dimensional single particle with zero net external potential does not adhere to the Einstein diffusion relation (see Table 2). In that same spirit, the diffusion coefficient increases linearly in the upper left plot of Figure 2 as  $h$  (and thus  $\gamma h$ ) increases. The computed diffusion coefficients for the G-JF and BBK\* integrators are relatively unchanged as a function



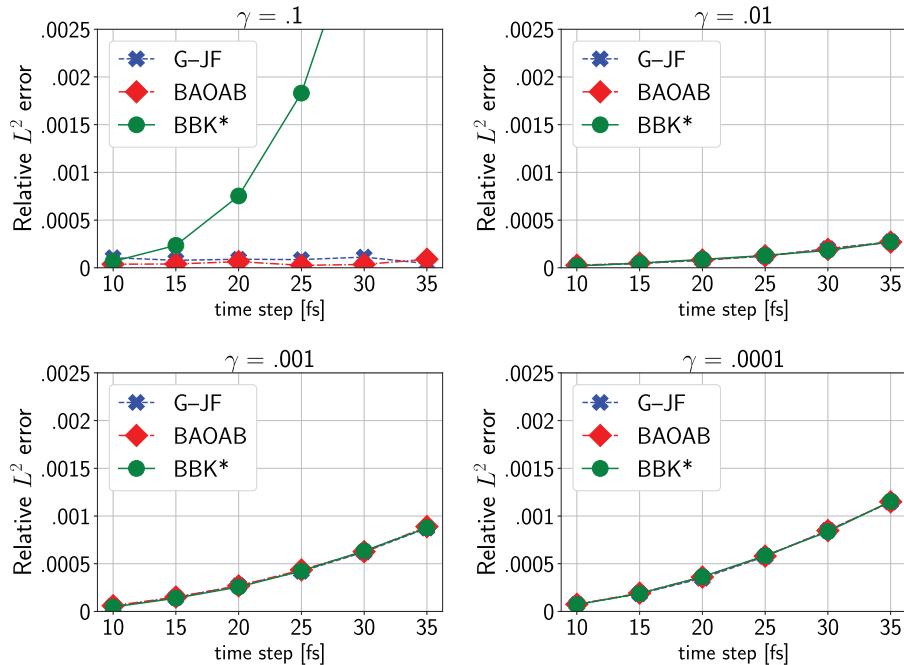
**Figure 3.** For each  $\gamma$  value and each integrator, the CM–CM radial distribution function for  $h = 5 \text{ fs}$  was calculated. These radial distribution functions were then used as a proxy for the true CG radial distribution function. For a given  $\gamma$  value, we denote these reference RDF's as  $g_{\text{G-JF}}^{\gamma}(r)$ ,  $g_{\text{BAOAB}}^{\gamma}(r)$  and  $g_{\text{BBK}^*}^{\gamma}(r)$ . These were visually indistinguishable and so we only show  $g_{\text{G-JF}}^{\gamma}(r)$  here.

of the time step, again, in line with the behaviour in the simple one-dimensional case. This is a desirable property, as it provides evidence that using larger time steps with G-JF and BBK\* does not corrupt the system's diffusive behaviour for any choice of  $\gamma$ . For  $\gamma \leq 0.01 \text{ fs}^{-1}$ , all integrators exhibit statistically similar diffusive behaviour, giving confidence that the choice of integrator should not influence diffusion in this regime.

We would like to stress the fact that in this study, we observe only classical diffusion, unlike some previous studies. In [41] and [19], polymer melts with CG particles composed of three  $\text{CH}_2$  monomers, as considered here, were studied. Sub-diffusion was observed for the quantity  $\mathcal{D}_h$  for simulation times up to and exceeding our simulation time of 250 ns (although the polymer lengths were at least double ours). Therefore we initially considered the possibility of anomalous diffusion during our simulation, however no such power law was observed. Further examination of error residuals with MSD data and regression lines did not provide evidence of non-linear relationships between MSD and simulation time.

#### 4.2. Configurational averages

An important quantity typically used in statistical thermodynamics is the radial distribution function (RDF). To discriminate the distributions of intermolecular contacts from intra-molecular ones, we restricted the



**Figure 4.** Relative error of radial distribution functions for the three numerical schemes for four different values of  $\gamma$ . For the largest  $\gamma$  value, we notice minimal change in error as the time step is increased for G-JF and BAOAB; but a more significant error for BBK\*. Although outside the range of the panel (a), for the largest integration step, we recorded an error of approximately 0.75% in the RDF, which can make an impact on the quality of results. For smaller choices of  $\gamma$ , we observe behaviour consistent with Hamiltonian dynamics: an increasing time step leads to increased error.

computation of the RDF to pairs of particles from distinct chains. The CM–CM distributions are here examined: because the system is mainly made up of CM particles, this RDF is the most fully sampled.

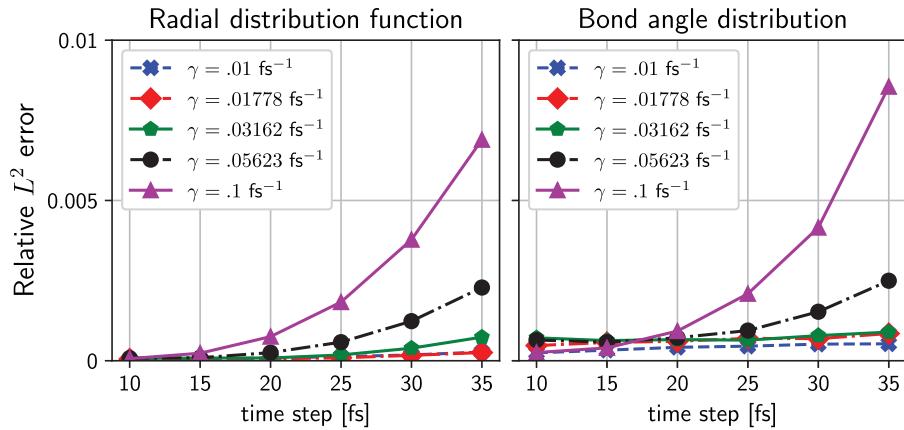
For each friction parameter  $\gamma$  and time step  $h$ , the RDF for the intermolecular CM–CM particle pairing was calculated every 1000 time steps, and these calculated distributions were then averaged over time. This was done for each of the 100 independent simulations, followed by a final averaging over these 100 simulations in order to reduce sampling error.

Lacking analytical expressions for the true RDF, computations conducted with time step  $h = 5 \text{ fs}$  are used as a reference solution. Given how small  $h = 5 \text{ fs}$  is relative to the time scale of the overall processes, it is reasonable to assume that the true intermolecular CM–CM RDF for the CG system will be well-approximated by the one calculated for  $h = 5 \text{ fs}$ . For each  $\gamma$  value, we calculate the reference RDF's:  $g_{\text{G-JF}}^\gamma(r)$ ,  $g_{\text{BAOAB}}^\gamma(r)$  and  $g_{\text{BBK}^*}^\gamma(r)$ . Then the  $L^2$  relative differences between the reference RDFs and the RDFs obtained for the other choices of  $h$ , denoted here as  $g_{\text{G-JF}}^{\gamma,h}(r)$ ,  $g_{\text{BAOAB}}^{\gamma,h}(r)$  and  $g_{\text{BBK}^*}^{\gamma,h}(r)$ , are computed and plotted as functions of  $h$ . These results are displayed in Figure 4. Given both  $\gamma$  and  $h$ , this relative error for a given method is defined to be:

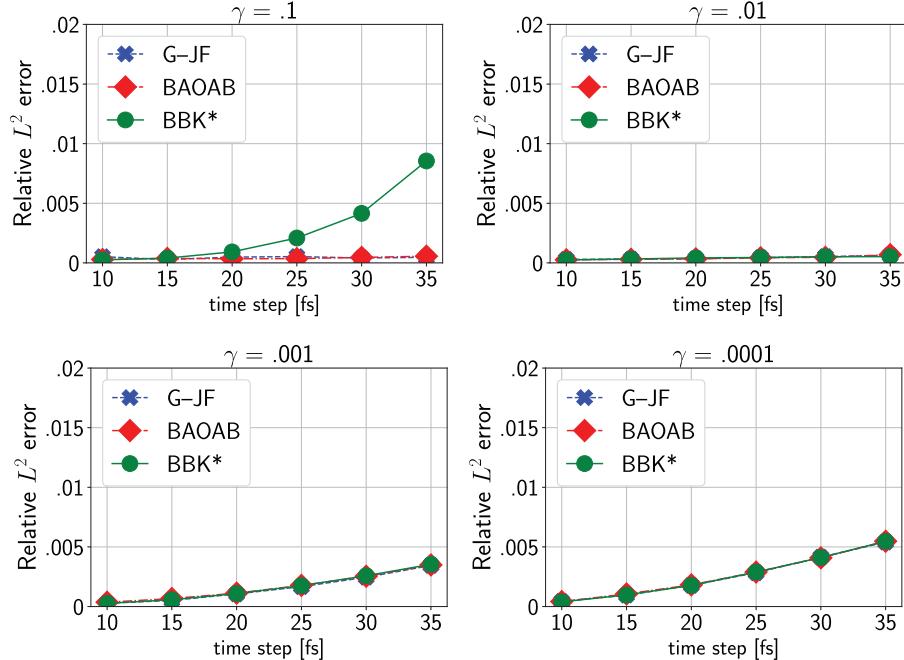
$$\text{rel.error} = \frac{\|g_{\text{method}}^\gamma(r) - g_{\text{method}}^{\gamma,h}(r)\|_{L^2}}{\|g_{\text{method}}^\gamma(r)\|_{L^2}}.$$

In the large  $\gamma$  regime, BBK\* exhibits more and more deviation from its baseline RDF as  $h$  is increased, while the other two methods remain unchanged. In fact, although not displayed in the plot, for  $h = 35$  fs, the BBK\* error is larger than that of BAOAB and G-JF by an order of magnitude. If one considered here the original BBK formulation, the resulting error would be slightly lower than BBK\* (Figure 7), but still much higher than BAOAB and G-JF. When  $\gamma \leq 0.01$  fs $^{-1}$ , all three integrators (BAOAB, BBK\* and G-JF) perform almost identically. This fact should be useful to the practitioner

working in the small  $\gamma$  regime when trying to strike a balance between diffusivity and adherence to the canonical distribution. Note also that for BBK\*, this is a distinguishing property of the formulation used in this paper: the original BBK formulation has a larger error in this regime (Figure 7). Some intuition is available. If we consider the harmonic potential, the mean square position for BBK is given by  $k_b T \omega^{-1} (1 - \frac{h^2 \omega}{4m})^{-1}$ ; therefore in the simple linear case, the configurational statistics depend only on the size of the time step, much like what is seen in Figure 7.



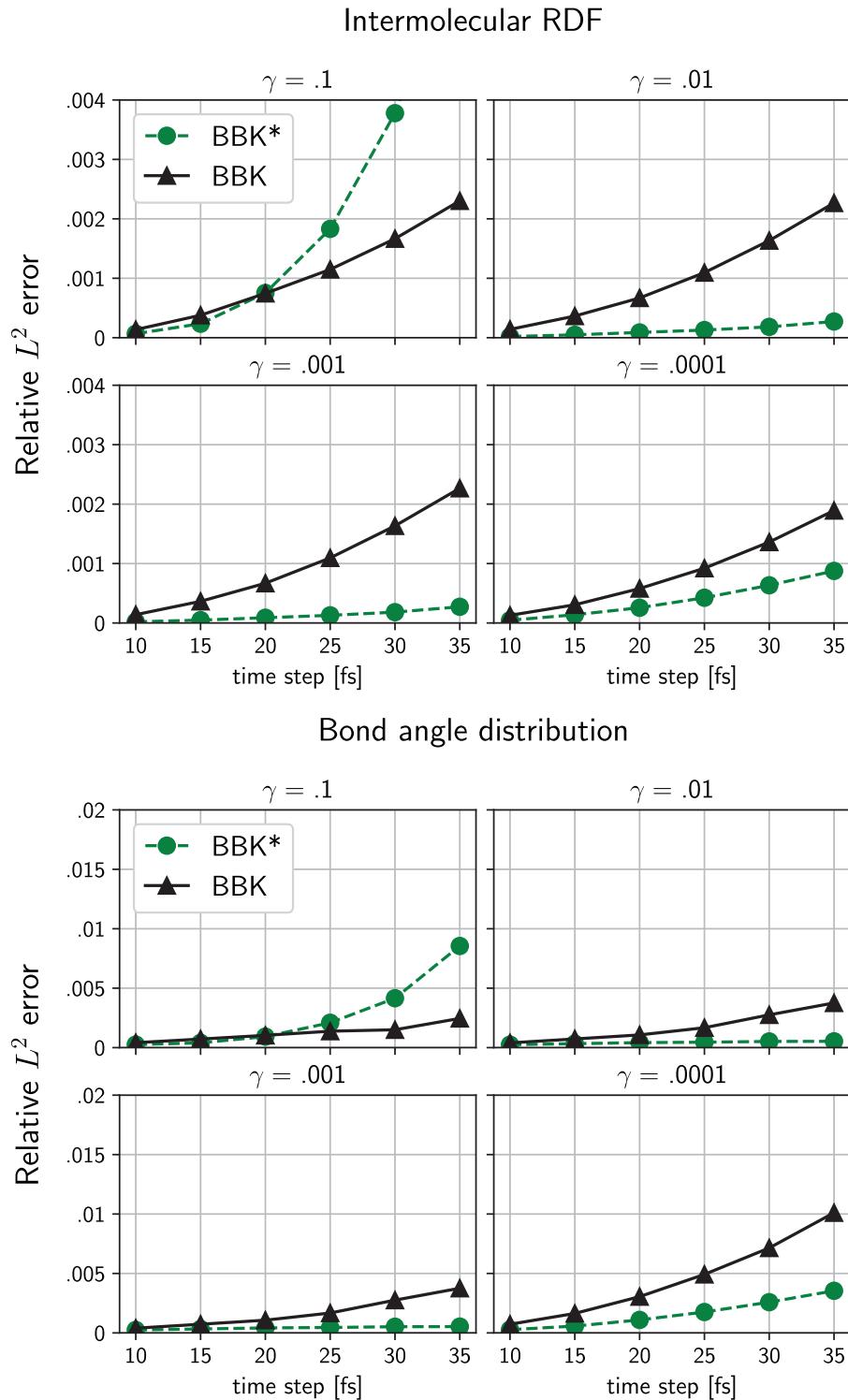
**Figure 5.** Relative  $L^2$  error for radial and bond angle distribution functions for the BBK\* method as  $\gamma$  varies between  $10^{-2}$  fs $^{-1}$  and  $10^{-1}$  fs $^{-1}$ , taking on the values  $10^{-2}, 10^{-1.75}, 10^{-1.5}, 10^{-1.25}, 10^{-1}$  fs $^{-1}$ . We see a smooth transition as  $\gamma$  becomes smaller, not an abrupt phase transition, further suggesting dependence of the steady-state distribution on the dimensionless quantity  $\gamma h$ .



**Figure 6.** Relative error of the bond distribution for the three numerical schemes when  $\gamma = 0.1$  fs $^{-1}, 0.01$  fs $^{-1}, 0.001$  fs $^{-1}, 0.0001$  fs $^{-1}$ . The reference is taken to be  $h = 5$  fs. As with the RDF, minimal change in error occurs when the time step is increased for G-JF and BAOAB, but for BBK\* it is much larger. At 35 fs, BBK\* experiences an almost 1% error in its bond angle distribution. Convergence to Hamiltonian dynamics is seen in the bottom row.

The divergence of the BBK\* RDF from the other RDFs when  $\gamma$  is largest, led us to question what happens when an intermediate  $\gamma$  value is used. Figure 5 displays these results. A smooth transition of error occurs between  $\gamma = .01 \text{ fs}^{-1}$  and  $\gamma = .1 \text{ fs}^{-1}$ . This suggests that the BBK\* RDF exhibits a  $\gamma h$  dependence that is not present in the other RDFs – which is not

surprising, given that a similar behaviour occurs in the simple 1-d harmonic oscillator case. Again, we stress that in most atomistic applications,  $\gamma$  is simply not large enough for the  $\gamma h$  dependence to be noticeable. However, this may no longer be the case for CG dynamics where the larger  $\gamma$  regime becomes more relevant.



**Figure 7.** A comparison of radial and angle distribution errors for BBK and BBK\*.

Another configurational quantity of interest for polymer chains is the bond angle distribution. Many properties of polymer melts, as well as their transition between liquid and solid phases, are determined by the propensity of the polymer chains to align. Failing to accurately reproduce the angle distribution can dramatically influence the accuracy of the simulation's thermodynamic properties.

Of the two possible triplets CM–CM–CT and CM–CM–CM, the latter was the best sampled and so was more amenable to statistical inference. For larger  $\gamma$ , G-JF and BAOAB produce minimal to no variation in error as  $h$  is increased. Figure 6 displays the results. BBK\* again exhibits systematic errors for larger  $\gamma$  values. The top row in Figure 6 indicates a large relative distortion of the bond angle distribution for BBK\*. This evidence may lead one to use caution in applying BBK\* to liquid-phase CG systems with larger  $\gamma$  values, especially since this behaviour for BBK\* is also observed with the CM–CM intermolecular RDF. For smaller  $\gamma$ , that is as the system gets closer to pure Hamiltonian dynamics, again all three integrators perform similarly, indicating no significant preference of integrator for this regime. Table 1 showed that in a harmonic potential, the configurational statistics were dependent on  $\gamma h$  for BBK\*. So it is not unreasonable to expect that a similar  $\gamma h$  dependence for other configurational quantities may occur in more complicated situations when using BBK\*, as seen in our simulations.

We observe very good agreement of BBK\* with G-JF and BAOAB for  $\gamma \leq .01 \text{ fs}^{-1}$  in Figures 4 and 6. However, for all choices of  $\gamma$ , the original BBK exhibits a clear trend: the relative error increases as the time step  $h$  is increased.

## 5. Conclusions

In this paper, we systematically studied three different Langevin integrators on a CG polymer melt. For the ideal cases of the Brownian motion and harmonic oscillator, key statistical properties were calculated analytically for each integrator, which provided guiding insights into diffusive and statistical behaviour of realistic molecular systems. In particular, for pure Brownian motion, both BBK\* and G-JF capture the true diffusive behaviour exactly for all choices of  $\gamma$  and  $h$ ; but BAOAB is only approximate, with the diffusion coefficient depending on the dimensionless parameter  $\gamma h$ . This carried over to the CG polymer simulation results. In Section 4.1, the calculated diffusion coefficient as a function of time step,  $D_h$ , was found to be statistically independent of the time step  $h$  for BBK\* and G-JF for all  $\gamma$  values, whereas BAOAB displayed the same type of linear behaviour for  $\gamma h$  values of  $O(1)$  as in the free particle case.

The computational results indicate that G-JF is the only integrator among those considered here that

describes equally well configurational distributions and diffusive behaviour over all  $\gamma$  choices. As expected, BBK\* and BBK perform poorly for the largest choice of  $\gamma$ . BAOAB samples equally well the configurational distributions, but exhibits a spurious dependence of the diffusivity on the integration time step near the high-friction regime. These conclusions have implications for CG MD simulations, where large ratios between friction and Hamiltonian forces are more frequently encountered than in atomistic ones. The evidence presented in this paper should be useful to the practitioner in supporting the use of the G-JF thermostat for CG simulations.

## Acknowledgments

We would like to thank Richard Berger for technical assistance.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

Part of this research was funded by the US Army Research Laboratory under contract number W911NF-16-2-0189. B. Seibold wishes to acknowledge support by National Science Foundation (NSF) [grant number DMS-1719640]. Calculations were carried out on Temple University's HPC resources and thus were supported in part by the National Science Foundation through major research instrumentation [grant number 1625061].

## Data availability statement

The Python files used for the LAMMPS implementation of each numerical scheme can be found in the repository [https://github.com/finkeljos/pylammps\\_langevin\\_integrators](https://github.com/finkeljos/pylammps_langevin_integrators).

## References

- [1] M. Tuckerman, *Statistical Mechanics and Molecular Simulations* (Oxford University Press, Oxford, 2010).
- [2] M. Tuckerman, B. Berne and G. Martyna, *J. Chem. Phys.* **97** (3), 1990–2001 (1992).
- [3] G.J. Martyna, M. Klein and M. Tuckerman, *J. Chem. Phys.* **97** (4), 2635–2643 (1992).
- [4] R. Jean-Paul, *J. Comput. Phys.* **23** (3), 327–341 (1977).
- [5] I. Pivkin, B. Caswell, and G. Karniadakis, *Reviews in Computational Chemistry*, Chap. 2, 27 vols. (Wiley, Hoboken, NJ, 2011).
- [6] B. Leimkuhler and C. Matthews, *J. Chem. Phys.* **138** (17), 05B601\_1 (2013).
- [7] N. Grønbech-Jensen and O. Farago, *Mol. Phys.* **111** (8), 983–991 (2013).
- [8] S. Anmarkrud, K. Debrabant and A. Kværnø, *BIT Numer. Math.* **58** (2), 257–280 (2018).
- [9] W. Wang and R. Skeel, *Mol. Phys.* **101** (14), 2149–2156 (2003).

[10] R. Pastor, B. Brooks and A. Szabo, *Mol. Phys.* **65** (6), 1409–1419 (1988).

[11] A. Brünger, C. Brooks III and M. Karplus, *Chem. Phys. Lett.* **105** (5), 495–500 (1984).

[12] B. Leimkuhler and C. Matthews, *Molecular Dynamics* (Springer, Cham, Switzerland, 2015).

[13] N. Grønbech-Jensen, N. Hayre and O. Farago, *Comput. Phys. Commun.* **185** (2), 524–527 (2014).

[14] X. Shang, M. Kröger and B. Leimkuhler, *Soft. Matter.* **13** (45), 8565–8578 (2017).

[15] E. Arad, O. Farago and N. Grønbech-Jensen, *Isr. J. Chem.* **56** (8), 629–635 (2016).

[16] T. Schneider and E. Stoll, *Phys. Rev. B* **17** (3), 1302 (1978).

[17] S. Plimpton, *J. Comput. Phys.* **117** (1), 1–19 (1995).

[18] H.-J. Limbach, A. Arnold, B. Mann and C. Holm, *Comput. Phys. Commun.* **174** (9), 704–727 (2006).

[19] K.M. Salerno, A. Agrawal, D. Perahia and G.S. Grest, *Phys. Rev. Lett.* **116** (5), 058302 (2016).

[20] J. Shelley, M. Shelley, R. Reeder, S. Bandyopadhyay and M. Klein, *J. Phys. Chem. B* **105** (19), 4464–4470 (2001).

[21] W. Shinoda, R. DeVane and M. Klein, *Mol. Simulat.* **33** (1–2), 27–36 (2007).

[22] S. Marrink, H. Risselada, S. Yefimov, D. Tieleman and A. DeVries, *J. Phys. Chem. B* **111** (27), 7812–7824 (2007).

[23] I. Cooke, K. Kremer and M. Deserno, *Phys. Rev. E* **72** (1), 011506 (2005).

[24] R. Goetz and R. Lipowsky, *J. Chem. Phys.* **108** (17), 7397–7409 (1998).

[25] S. Izvekov and G. Voth, *J. Phys. Chem. B* **109** (7), 2469–2473 (2005).

[26] G. Brannigan, L. Lin and F. Brown, *Eur. Biophys. J.* **35** (2), 104–124 (2006).

[27] M. Orsi, D. Haubertin, W. Sanderson and J. Essex, *J. Phys. Chem. B* **112** (3), 802–815 (2008).

[28] T. Murtola, M. Karttunen and I. Vattulainen, *J. Chem. Phys.* **131** (5), 08B601 (2009).

[29] K. Hadley and C. McCabe, *Biophys. J.* **99** (9), 2896–2905 (2010).

[30] LAMMPS Source Code (2017). <<http://lammps.sandia.gov/>> .

[31] Sequencer.c, NAMD Source Code (2018). <<https://www.ks.uiuc.edu/Research/namd/>> .

[32] K. Burrage and G. Lythe, *SIAM J. Numer. Anal.* **47** (3), 1601–1618 (2009).

[33] J.A. Izaguirre, D.P. Catarello, J.M. Wozniak and R. Skeel, *J. Chem. Phys.* **114** (5), 2090–2098 (2001).

[34] B. Leimkuhler and C. Matthews, *J. Chem. Phys.* **138** (17), 05B601\_1 (2013).

[35] L.F. Grønbech-Jensen and N. Grønbech-Jensen, *Mol. Phys.* **117**(18):2511–2526, (2019).

[36] O. Farago, Langevin Thermostat for Robust Configurational and Kinetic Sampling. Preprint arXiv:1903.02674 (2019).

[37] B.D. Butler, G. Ayton, O.G. Jepps and D.J. Evans, *J. Chem. Phys.* **109** (16), 6519–6522 (1998).

[38] B. Dünweg and W. Paul, *Int. J. Mod. Phys. C* **2** (3), 817–827 (1991).

[39] G.L. Martyna, D. Tobias and M. Klein, *J. Chem. Phys.* **101** (5), 4177–4189 (1994).

[40] H. Flyvbjerg and H.G. Petersen, *J. Chem. Phys.* **91** (1), 461–466 (1989).

[41] B. Peters, K.M. Salerno, A. Agrawal, D. Perahia and G. Grest, *J. Chem. Theory Comput.* **13** (6), 2890–2896 (2017).