SC8 Mini Project: Molecular Dynamics for Modelling of Dimers

Candidate Number 1007177

May 2019

1 Introduction

All-atom molecular dynamics (MD) has grown tremendously as a field in recent years, aided by the equally rapid improvements in computational capabilities which have roughly followed Moore's Law. We are now at the stage of having the capability to model sufficiently large domains with sufficiently small timesteps to aid discovery in biological fields, particularly in cell biology.

In this mini project, we introduce an analytically tractable MD model of Brownian motion which is used frequently in the literature. The popularity of this model resides in its convergence to the Langevin description in a certain limit, facilitating the creation of multi-resolution regimes for more efficient modelling.

Throughout the project, we focus on a toy problem of modelling a dimer in a solvent. The beauty of using such a simple molecule is that we can make tremendous analytical progress in solving the coupled SDEs that describe the motion of the two constituent monomers. Were we to consider a longer polymer and apply the Rouse or the Zimm model, we would not be able to make such headway and thus could not demonstrate the consistency of our MD model with the underlying dynamics we wish to emulate.

We treat a dimer as a pair of identical spheres which interact via a potential. This could represent a simple diatomic molecule such as H_2 or an ultra-coarse-grained representation of a more complicated structure (e.g. a protein), as has proven popular in recent years [2]. A key macrostatistic of the dimer is its *bond length* - the separation between the two monomers.

Through perturbation analysis, we are able to estimate the value of this bond length predicted by the coarser Langevin model and demonstrate consistency in the observed value of the bond length when conducting MD simulations.

2 Mesoscopic Dimer Description

In this section, we consider a coarse-grained representation of a dimer as two identical spheres of mass M and radius r_0 which interact via a potential. If $\mathbf{X}_1(t)$ and $\mathbf{X}_2(t)$ are the positions of the monomers at time t, then we denote by $\mathbf{R}$ the vector of the separation between the two, i.e. $\mathbf{R} = \mathbf{X}_2 - \mathbf{X}_1$. We denote its magnitude (the *dimer length*) by $R = |\mathbf{R}|$, which provides the input to our potential $\Phi(R)$.

For the time being, we describe the surrounding solvent as a noise term from Brownian motion. Thus, we can describe the time evolution of the position and velocity of each monomer by the Langevin descriptions

$$\mathbf{X}_1(t + \mathrm{d}t) = \mathbf{X}_1(t) + \mathbf{V}_1(t)\mathrm{d}t, \tag{1}$$

$$\mathbf{V}_1(t + \mathrm{d}t) = \mathbf{V}_1(t) + \frac{\Phi'(R)}{M}\frac{\mathbf{R}}{R}\mathrm{d}t - \gamma\mathbf{V}_1\mathrm{d}t + \gamma\sqrt{2D}\mathrm{d}\mathbf{W}_1, \tag{2}$$

$$\mathbf{X}_2(t + \mathrm{d}t) = \mathbf{X}_2(t) + \mathbf{V}_2(t)\mathrm{d}t, \tag{3}$$

$$\mathbf{V}_2(t + \mathrm{d}t) = \mathbf{V}_2(t) - \frac{\Phi'(R)}{M}\frac{\mathbf{R}}{R}\mathrm{d}t - \gamma\mathbf{V}_2\mathrm{d}t + \gamma\sqrt{2D}\mathrm{d}\mathbf{W}_2, \tag{4}$$

where $\mathrm{d}\mathbf{W}_i$ is a vector of 3 independent Wiener processes and $\gamma = \frac{6\pi\eta r_0}{M}$ is a mass-normalized drag coefficient given by Stokes' theorem.

Taking (4)−(2) and implementing the overdamped assumption $\gamma \to \infty$ gives

$$0 = -\frac{2\Phi'(R)}{\gamma M}\frac{\mathbf{R}}{R}\mathrm{d}t - (\mathbf{V}_2(t) - \mathbf{V}_1(t))\mathrm{d}t + 2\sqrt{D}\mathrm{d}\mathbf{W}. \tag{5}$$

Now taking (3)−(1) and substituting in our expression from (5) yields the Ornstein-Uhlenbeck process

$$\mathbf{R}(t + \mathrm{d}t) = \mathbf{R}(t) - \frac{2\Phi'(R)}{\gamma M}\frac{\mathbf{R}(t)}{R}\mathrm{d}t + 2\sqrt{D}\mathrm{d}\mathbf{W}. \tag{6}$$

We can use this to write the Fokker-Planck equation for the probability density function $f(R)$ as

$$\frac{\partial f}{\partial t} = \frac{1}{2}\frac{\partial^2}{\partial \mathbf{R}^2}\left(4Df(R)\right) - \frac{\partial}{\partial \mathbf{R}}\left(-\frac{2}{M\gamma}\frac{\mathrm{d}\Phi(R)}{\mathrm{d}\mathbf{R}}f(R)\right), \tag{7}$$

from which we can calculate the stationary solution by setting the left hand side to 0

$$\begin{aligned} 0 &= \frac{\partial^2}{\partial \mathbf{R}^2}\left(2Df(R)\right) + \frac{\partial}{\partial \mathbf{R}}\left(\frac{2}{M\gamma}\frac{\mathrm{d}\Phi(R)}{\mathrm{d}\mathbf{R}}f(R)\right), \\ \implies \quad c_1 &= \frac{\partial}{\partial \mathbf{R}}\left(2Df(R)\right) + \frac{2}{M\gamma}\frac{\mathrm{d}\Phi(R)}{\mathrm{d}\mathbf{R}}f(R), \\ \implies f(R) &= c_2 \exp\left[-\frac{\Phi(R)}{MD\gamma}\right], \end{aligned}$$

where we have set $c_1 = 0$ by imposing the boundary condition $f(\infty) = 0$. We can replace c_2 by converting to spherical polar coordinates and normalizing our density function, giving

$$f(R) = \frac{\exp\left[-\frac{\Phi(R)}{MD\gamma}\right]}{4\pi\int_0^\infty r^2 \exp\left[-\frac{\Phi(r)}{MD\gamma}\right]\mathrm{d}r}. \tag{8}$$

The expected value of R and hence the dimer bond length L_d is therefore

$$L_d = 4\pi\int_0^\infty r^3 f(r)\mathrm{d}r = \frac{\int_0^\infty r^3 \exp[-\frac{\Phi(r)}{MD\gamma}]\mathrm{d}r}{\int_0^\infty r^2 \exp[-\frac{\Phi(r)}{MD\gamma}]\mathrm{d}r}. \tag{9}$$

We now formally define our potential $\Phi(R)$. There are two choices for potential predominantly used in the literature: the harmonic spring potential and the Morse potential [8]. In this project, we choose to focus on the Morse potential since it is generally viewed to be a better approximation than the harmonic oscillator as it accounts for the anharmonicity seen in a lot of real bonds. Thus, our function $\Phi(R)$ shall henceforth be assumed to be

$$\Phi(R) = D_e(1 - \exp[-a(R - \ell_0)])^2 - D_e. \tag{10}$$

If we apply the substitution $x := \frac{1}{\varepsilon}$, where we define and assume that $\varepsilon = \frac{MD\gamma}{D_e} \ll 1$, we can convert both integrals into the form of a generalized Laplace integral

$$L_d = \frac{\int_0^\infty r^3 \exp[-x\phi(r)]\mathrm{d}r}{\int_0^\infty r^2 \exp[-x\phi(r)]\mathrm{d}r},$$

where

$$\phi(r) := (1 - \exp[-a(r - \ell_0)])^2.$$

Thus, using Laplace's method, we can see that the behaviour of each integral is dominated near the rest length (corresponding to the interval $r \in [\ell_0 - \delta, \ell_0 + \delta]$, where $\delta \ll 1$)

$$L_d \approx \frac{\int_{1-\delta}^{1+\delta} r^3 \exp[-xa^2(r - \ell_0)^2]\mathrm{d}r}{\int_{1-\delta}^{1+\delta} r^2 \exp[-xa^2(r - \ell_0)^2]\mathrm{d}r}.$$

Applying the substitution $s := \sqrt{x}a(r - \ell_0)$ and noting that taking our limits off to infinity introduces only exponentially small errors, we end up with a fraction involving Gaussian integrals which we can simplify greatly by using the general result that $\int_{-\infty}^{+\infty} s^n e^{-s^2}\mathrm{d}s = 0$ for odd n:

$$\begin{aligned} L_d &\approx \frac{\frac{\sqrt{\pi}}{a\sqrt{x}}[\ell_0^3 + \frac{3\ell_0}{2a^2x}]}{\frac{\sqrt{\pi}}{a\sqrt{x}}[\ell_0^2 + \frac{1}{2a^2x}]} \\ &\approx \left(\ell_0 + \frac{3}{2a^2x\ell_0}\right)\left(1 - \frac{1}{2a^2x\ell_0^2} + \frac{1}{4a^4x^2\ell_0^4} - ...\right) \\ &= \ell_0 - \frac{1}{2a^2x\ell_0} + \frac{3}{2a^2x\ell_0} + O\left(\frac{1}{x^2}\right), \end{aligned}$$

where we have used the Binomial Theorem to approximate the behaviour of our quotient. Thus we can say that for small ε, L_d can be approximated with good accuracy as

$$L_d \approx \ell_0 \left[1 + \frac{MD\gamma}{a^2 D_e \ell_0^2}\right]. \tag{11}$$

We now seek to provide a more explicit consideration of the surrounding solvent by using molecular dynamics. We shall verify the efficacy of our method by comparing the output of our simulations with the analytically derived result in (11).

3 The Hard Sphere Heat Bath

The hard sphere heat bath is a theoretical model of a solvent which considers exclusively short-range interactions with the diffusing particle. The model has many similarities with the model of an ideal gas proposed by Bernoulli [1].

We treat the two monomers as spheres of radius r_0 and mass M, immersed in a solvent represented by a collection of light point particles whose only interactions are perfectly elastic collisions with the monomers. Between collisions, the monomers obey Newton's second law of motion

$$M\frac{d\mathbf{V}_1}{dt} = \Phi'(R)\frac{\mathbf{R}}{R}, \tag{12}$$

$$M\frac{d\mathbf{V}_2}{dt} = -\Phi'(R)\frac{\mathbf{R}}{R}, \tag{13}$$

where our notation is the same as the notation used in Section 2. The heat bath particles are described in terms of their positions $\mathbf{x}^j$ and velocities $\mathbf{v}^j$, where $j = 1, 2, 3, \ldots$ is the number of the particle. We assume that all particles have the same mass m and define the ratio of mass of the heavy to light particles μ by

$$\mu = \frac{M}{m}.$$

Heat bath particles are assumed to be distributed throughout space with density λ_μ given by

$$\lambda_\mu = \frac{3}{8r_0^2}\sqrt{\frac{(\mu+1)\gamma}{2\pi D}}, \tag{14}$$

where γ is the mass-normalized Stokes drag coefficient which we saw in the previous section, so depends on the viscosity η of the solvent we wish to represent. We control the temperature of our system by distributing each component of the velocities of the heat bath particles with mean 0 and variance σ_μ^2 where

$$\sigma_\mu = \sqrt{(\mu+1)D\gamma}. \tag{15}$$

If we then consider the expected average kinetic energy per particle, we find

$$\begin{aligned} E_{\mathrm{KE}} &= 3 \times \frac{1}{2} \times m \times \sigma_\mu^2 \\ &= \frac{3}{2} m(\mu+1) \times \frac{k_B T}{6\pi\eta r_0} \times \frac{6\pi\eta r_0}{M} \\ &= \frac{3}{2}\left(\frac{1}{\mu}+1\right) k_B T \\ &\to \frac{3}{2} k_B T \qquad\qquad \text{as } \mu \to \infty, \end{aligned}$$

which is the average kinetic energy of a particle in 3 degrees of freedom given by the equipartition theorem. As we will be considering this model for large values of μ, this explains our choice of parameter in (15). We choose a Maxwell-Boltzmann distribution to initialize the velocities of each heat bath particle

$$f_\mu(\mathbf{v}) = \frac{1}{\sigma_\mu^3 (2\pi)^{3/2}} \exp\left[-\frac{v_1^2 + v_2^2 + v_3^2}{2\sigma_\mu^2}\right]. \tag{16}$$

In the event of a collision between a monomer and heat bath particle, we update their velocities by preserving kinetic energy and linear momentum (i.e. assuming frictionless collisions). It is easily shown that this yields [4]

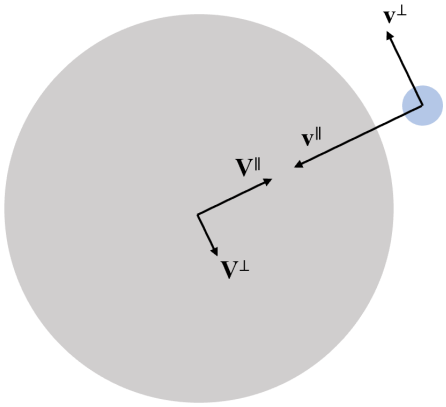

Figure 1: Schematic of collision between diffusing monomer (grey) and heat bath particle (blue). Components of velocities are shown. Not to scale.

$$\tilde{\mathbf{V}}_i = [\mathbf{V}_i]^{\perp} + \frac{\mu - 1}{\mu + 1} [\mathbf{V}_i]^{\parallel} + \frac{2}{\mu + 1} \left[\mathbf{v}^j\right]^{\parallel}, \tag{17}$$

$$\tilde{\mathbf{v}}^j = \left[\mathbf{v}^j\right]^{\perp} + \frac{1 - \mu}{\mu + 1} \left[\mathbf{v}^j\right]^{\parallel} + \frac{2\mu}{\mu + 1} [\mathbf{V}_i]^{\parallel}, \tag{18}$$

where tildes denote post-collision velocities and the superscripts $\parallel$ and $\perp$ denote the projections of the pre-collision velocities on the line through the centre of the monomer and the heat bath particle (the *collision axis*), as shown in Figure (1).

3.1 Domain and boundary conditions

Whilst the properties of the heat bath we have defined are assumed to hold for all of $\mathbb{R}^3$, it would be impossible to track infinitely many heat bath particles. Hence, we can restrict our attention to a small subset of $\mathbb{R}^3$ by just considering the heat bath particles in the local vicinity of each monomer. Formally speaking, we define a cubic *co-moving frame* of side length L around each monomer

$$\Omega_i = \mathbf{X}_i(t) + \left[-\frac{L}{2}, \frac{L}{2}\right] \times \left[-\frac{L}{2}, \frac{L}{2}\right] \times \left[-\frac{L}{2}, \frac{L}{2}\right] \qquad i = 1, 2. \tag{19}$$

At each timestep, we reposition each frame to keep them centered on the monomers and apply suitable boundary conditions to preserve (14) and (15) by introducing new particles with suitably sampled velocities to replace the particles which have now left the domain.

To derive these boundary conditions, we consider one of the heat baths and let $\Omega = [0, L]\times[0, L]\times[0, L]$. We consider how many particles to introduce across the surface at $x = 0$.

In order to introduce a new particle with coordinates (x, y, z) across this face, it must have sufficient velocity relative to the monomer in the x direction to have travelled from outside the domain to inside it over the timestep Δt. Thus

$$v_x > V_x + \frac{x}{\Delta t}, \tag{20}$$

where v_x and V_x are the x-components of velocity of the heat bath particle and monomer respectively. With this in mind, the expected density, $h\,(x)$, of new particles with x-coordinate x can be evaluated as

$$\begin{aligned} h\,(x) &= \int_{-\infty}^{\infty} H\,(v_x\Delta t - x - V_x\Delta t)\,\frac{\lambda_\mu}{\sigma_\mu\sqrt{2\pi}}\exp\left(-\frac{v_x^2}{2\sigma_\mu^2}\right)\mathrm{d}v_x \\ &= \frac{\lambda_\mu}{2}\mathrm{erfc}\left(\frac{x + V_x\Delta t}{\sqrt{2}\omega_\mu\Delta t}\right), \end{aligned} \tag{21}$$

Integrating (21) over all possible new coordinates in domain $(0, \infty)\times[0, L]\times[0, L]$ gives us the expected number of particles, m_{new}, that have entered our domain across this face as

$$\begin{aligned} m_{\mathrm{new}} &= \frac{\lambda_\mu L^2}{2}\int_0^{\infty} h(x)\mathrm{d}x \\ &= \lambda_\mu L^2\Delta t\left(\frac{\sigma_\mu}{\sqrt{2\pi}}\exp\left[-\frac{V_x^2}{2\sigma_\mu^2}\right] - \frac{V_x}{2}\mathrm{erfc}\left[\frac{V_x}{\sigma_\mu\sqrt{2}}\right]\right). \end{aligned} \tag{22}$$

To implement this in our simulations, we generate the number of particles to introduce from a Poisson distribution with mean m_{new}. For each new particle, we then initialize its x coordinate by sampling from the distribution $\sqrt{\pi}\mathrm{erfc}\,(\alpha)$ using the algorithm in Appendix A and setting

$$x_{\text{new}} = \sqrt{2}\sigma_\mu \Delta t \alpha.$$

The particle's y and z coordinates are sampled uniformly in [0,L]. Finally, we need to initialize the new particle's velocity. We wish for the velocities to still be distributed according to a normal distribution with variance σ_μ^2 but need the particle's x-component of velocity to be sufficient to allow it to have travelled from outside the domain to inside it, being careful to consider the fact that the frame has also moved in the previous timestep. In other words, we need

$$v_x \geq \frac{x_{\text{new}}}{\Delta t} + V_x = \sqrt{2}\sigma_\mu \alpha + V_x \tag{23}$$

Hence, we sample from a truncated normal distribution

$$\zeta \sim H(\xi - \sqrt{2}\sigma_\mu \alpha - V_f)\xi, \tag{24}$$

where $\xi \sim \mathcal{N}(0,1)$, using the acceptance-rejection algorithm in Appendix A and set $v_x = \sigma_\mu \zeta$. The particle's y and z velocities are sampled from a normal distribution $\mathcal{N}(0, \sigma_\mu^2)$.

We do this across each of the six faces of each co-moving frame Ω_i at each timestep and, in this way, guarantee that the conditions specified by (14) and (15) are preserved.

3.2 Overcounting

When introducing new heat bath particles to our computational domain, we have sampled from the entirety of the half-space outside the frame. In doing so, we have considered solvent particles from some regions twice or even three times, leading to an overestimate in the number of heat bath particles that should be introduced.

In order to rectify this overestimate, we define an acceptance probability, p_{acc} given by

$$p_{\text{acc}}\left(\mathbf{x}_{\text{new}}, \mathbf{v}_{\text{new}}\right) = \begin{cases} 1, & \text{for } \mathbf{y} - \mathbf{X} \in \mathcal{Y}_1; \\ 1/2, & \text{for } \mathbf{y} - \mathbf{X} \in \mathcal{Y}_2; \\ 1/3, & \text{for } \mathbf{y} - \mathbf{X} \in \mathcal{Y}_3, \end{cases}$$

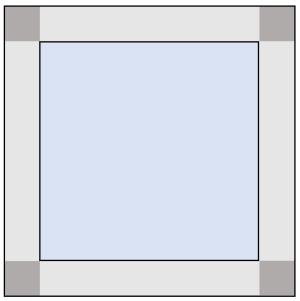

Figure 2: 2D illustration of overcounting regions. Double counted regions are shown in light grey whilst triple counted regions are shown in dark grey.

where $\mathbf{y} = \mathbf{x}_{\text{new}} - \mathbf{v}_{\text{new}}\Delta t$ is the previous position of the newly introduced particle and $\mathcal{Y}_i$ is the region of space containing points with exactly i of their coordinates outside the interval $[-L/2, L/2]$.

Then, once we have initialized the position and velocity of a new heat bath particle, we generate a random number in the interval [0, 1] and only keep the new particle if p_{acc} is less than this.

4 Microscopic Dimer Modelling

In Section 3 we have presented a theoretical heat bath model for a solvent. We shall now replace the Brownian dynamics of the solvent used in Section 2 to empirically demonstrate that this model is consistent with the overdamped analytical prediction of the dimer bond length (11).

We place each monomer in its own heat bath. One could well question whether or not this is a reasonable modelling choice, since in doing this we will lose any correlations between the monomers which one would expect to exist in the real world. However, it is shown in the literature that the impact of doing this is has an insignificant impact on accuracy whilst offering a big improvement in efficiency by reducing size of the computational domain [7]. Hence, we opt to do this in our simulations.

We stray from the literature somewhat in using an event-driven model rather than a fixed timestep. When using a fixed timestep, an investigator has to be cautious not to set Δt too large since this could then allow multiple collisions to happen in a single timestep which would go undetected by our algorithm. Using an event-driven algorithm essentially enables one to use

the maximal timestep with one collision and is thus more efficient than its counterpart. In comparative simulations, the event-driven model was found to be 2.3 times quicker than using a fixed timestep of $\Delta t = 10^{-6}$.

We use the Euler-Maruyama method to rewrite the Langevin description from our mesoscopic dimer description, noting that we no longer need to consider a noise term as our solvent is now represented explicitly by the heat bath:

$$\mathbf{X}_1(t+\Delta t) = \mathbf{X}_1(t) + \mathbf{V}_1(t)\Delta t, \tag{25}$$

$$\mathbf{V}_1(t+\Delta t) = \mathbf{V}_1(t) + \frac{\Phi'(R)}{M}\frac{\mathbf{R}}{R}\Delta t - \gamma \mathbf{V}_1 \Delta t, \tag{26}$$

$$\mathbf{X}_2(t+\Delta t) = \mathbf{X}_2(t) + \mathbf{V}_2(t)\Delta t, \tag{27}$$

$$\mathbf{V}_2(t+\Delta t) = \mathbf{V}_2(t) - \frac{\Phi'(R)}{M}\frac{\mathbf{R}}{R}\Delta t - \gamma \mathbf{V}_2 \Delta t. \tag{28}$$

Our algorithm is summarised in pseudocode in Table (1) and our parameters in Table (4).

1) Place both monomers with initial separation ℓ_0 and initialize separate heat baths.
2) For each monomer, calculate time to next collision. Choose the minimum of the two times and update system at this time according to (25)-(28).
3) Update velocities of the heat bath particle and monomer involved in the collision.
4) Reposition co-moving frames and introduce new solvent where necessary.
5) Calculate and record bond length, R.
6) Return to 2 unless total time is greater than 100, in which case terminate simulation.

Table 1: Summary of microscopic dimer modelling algorithm

Symbol	Name	Value
r_0	Monomer radius	0.08
γ	Mass-normalized drag coefficient	10
μ	Ratio of masses	10^3
M	Mass of monomer	1
D	Diffusion coefficient	1
L	Frame length	0.32
a	Width of Morse potential	22.4
D_e	Well depth	1000
ℓ_0	Equilibrium bond length	$2.25 - 5 \times r_0$

Table 2: Parameters for our simulations, chosen where possible to aid comparison to the literature

5 Results

The results of long-time simulations (code written in Fortran 90) for range of values of ℓ_0 are shown in Figure (3). The average bond length L_d for a range of rest lengths ℓ_0 is shown using the theoretical heat bath model and compared to the predicted result from (11).

As expected, the results from the simulations using the heat bath presented in Section 3 are consistent with the predictions of our perturbation analysis from Chapter 2, with no points showing a significant difference from the prediction.

Thus, we have demonstrated the efficacy of using this MD model to increase the level of detail of our solvent.

6 Discussion

One of the primary motivations behind presenting this particular MD model is its potential application in models with multiple levels of resolution. It is shown in the literature that the hard-sphere heat bath model described above converges to the Langevin description in the limit $\mu \to \infty$ [5]. This

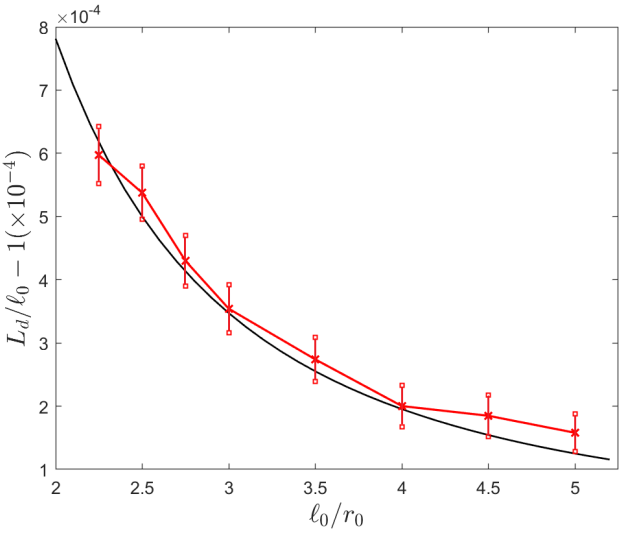

Figure 3: The extension of the average length of a dimer from its separation distance ℓ_0. Equilibrium data was collected from long-time simulations lasting 100 dimensionless time units, as is done in [7]

enables one to couple molecular dynamics with Langevin dynamics and even Brownian dynamics by dividing up the computational domain into regions of varying resolution.

Multiple resolution regimes are extremely desirable for investigators wishing to consider large domains (e.g. a eukaryotic cell) in which MD levels of detail would not be feasible using current computers for typical metabolic timescales. A high level of detail can be retained in the region of interest whilst the rest of the domain can be modelled more coarsely without a significant impact on accuracy in results. In the literature, this is utilised effectively to model actin dynamics in filopodia [6].

In this project, we have contributed to the study of the hard sphere heat bath by proposing an event-driven version which outperforms the fixed timestep model. Furthermore, we have replaced the harmonic potential used in the literature by the more physically accurate Morse potential.

Further work would benefit from considering a wider range of values of ℓ_0 since we have only considered a fairly narrow range.

A Random number generation

We utilize two different acceptance-rejection sampling algorithms in order to introduce new heat bath particles with the correct boundary conditions. Acceptance-rejection sampling works because of the following lemma from [3], presented without proof:

Lemma A.1. *Let $h(x)$ and $g(x)$ be two densities such that $h(x) \leq Mg(x)$ for all x in the support of $h(x)$. Then the random variable x given as the output of the following algorithm*

1. *Generate $z \sim g(z)$*
2. *Generate $u \sim \mathcal{U}(0,1)$. If $u \leq h(z)/Mg(z)$, take $x = z$. Else, repeat*

is distributed according to $h(x)$

A.1 Complementary error function

With this lemma in mind, we use acceptance-rejection sampling with $g(x) = \exp(-x/a_1)/a_1$ (i.e. the pdf of an exponential distribution with mean a_1) to sample from the complementary error function. This is presented in the literature and analytically shown to have the highest acceptance rate when $a_1 = 0.5316$ and $a_2 = 0.8316$ [5]. The algorithm is summarised in Table (3).

- Generate a random number $\eta_1 \sim \mathcal{U}(0,1)$
- Compute an exponentially distributed random number η_2 by setting $\eta_2 = a_1 \log(\eta_1)$
- Generate a random number $\eta_3 \sim \mathcal{U}(0,1)$
- If $\eta_3 < a_2 \operatorname{erfc}(\eta_2) \exp(\eta_2/a_1)$, choose η_2 as a sample from our distribution. Otherwise, repeat algorithm.

Table 3: Algorithm for sampling from complementary error function.

A.2 Truncated normal distribution

We use another acceptance-rejection algorithm for sampling from the truncated normal distribution proposed in the literature [9]. The underlying principle of the algorithm is that the translated exponential distribution is similar in shape to the truncated normal distribution whilst always being greater than it, making it a suitable choice for the acceptance-rejection sampling method, particularly if we choose it to have mean a, where

$$a = \frac{A + \sqrt{A^2 + 4}}{2} \tag{29}$$

and A is the value at which our Gaussian distribution is truncated.

- Generate a random number $\eta_1 \sim \mathcal{U}(0, 1)$
- Compute a translated exponentially distributed random number η_2 by setting $\eta_2 = A - \log(\eta_1)/a$
- Generate a random number $\eta_3 \sim \mathcal{U}(0, 1)$
- If $\eta_3 < \exp(-(\eta_2 - a)^2/2)$, choose η_2 as a sample from our distribution. Otherwise, repeat algorithm.

Table 4: Algorithm for sampling from truncated normal distribution.

References

[1] Daniel Bernoulli. 1738 hydrodynamica. *Argentorati (Strassburg), Germany: Johann Reinhold Dulsecker*, 1940.

[2] James F Dama, Anton V Sinitskiy, Martin McCullagh, Jonathan Weare, Benoît Roux, Aaron R Dinner, and Gregory A Voth. The theory of ultra-coarse-graining. 1. general principles. *Journal of chemical theory and computation*, 9(5):2466–2480, 2013.

[3] Luc Devroye. Sample-based non-uniform random variate generation. In *Proceedings of the 18th conference on Winter simulation*, pages 260–265. ACM, 1986.

[4] D Dürr, S Goldstein, and JL Lebowitz. A mechanical model of brownian motion. *Communications in Mathematical Physics*, 78(4):507–530, 1981.

[5] Radek Erban. From molecular dynamics to brownian dynamics. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 470(2167):20140036, 2014.

[6] Radek Erban, Mark B Flegg, and Garegin A Papoian. Multiscale stochastic reaction–diffusion modeling: application to actin dynamics in filopodia. *Bulletin of mathematical biology*, 76(4):799–818, 2014.

[7] Ravinda S Gunaratne, Daniel B Wilson, Mark B Flegg, and Radek Erban. Multi-resolution dimer models in heat baths with short-range and long-range interactions. *Interface Focus*, 9(3):20180070, 2019.

[8] Philip M Morse. Diatomic molecules according to the wave mechanics. ii. vibrational levels. *Physical Review*, 34(1):57, 1929.

[9] Christian P Robert. Simulation of truncated normal variables. *Statistics and computing*, 5(2):121–125, 1995.