COARSE-GRAINING MOLECULAR DYNAMICS: STOCHASTIC MODELS WITH NON-GAUSSIAN FORCE DISTRIBUTIONS

[1]B.Umadevi,[2]K.Rammohan
Department of Humanities & Science
Priyadarshini Institute of Science and Technology for Women Khammam

Abstract

Incorporating atomistic and molecular information into models of cellular behaviour is challenging because of a vast separation of spatial and temporal scales between processes happening at the atomic and cellular levels. Multiscale or multi-resolution methodologies address this difficulty by using molecular dynamics (MD) and coarse-grained models in different parts of the cell. Their applicability depends on the accuracy and properties of the coarsegrained model which approximates the detailed MD description. A family of stochastic coarse-grained (SCG) models, written as relatively low-dimensional systems of nonlinear stochastic differential equations, is presented. The nonlinear SCG model incorporates the non-Gaussian force distribution which is observed in MD simulations and which cannot be described by linear models. It is shown that the nonlinearities can be chosen in such a way that they do not complicate parametrization of the SCG description by detailed MD simulations. The solution of the SCG model is found in terms of gamma functions.

Keywords: multiscale modelling · coarse-graining · molecular dynamics · Brownian dynamics
DOI Number:10.48047/NQ.2022.20.21.NQ99160 **Neuroquantology 2022; 20(21):1511-1525**

1 Introduction

With increased experimental information on atomic or near-atomic structure of biomolecules and intracellular components, there has been a growing need to incorporate such microscopic data (coming from X-ray crystallography, NMR spectroscopy or cryo-electron microscopy) into dynamical models of intracellular processes. A common approach is to use molecular dynamics (MD) simulations based on classical molecular mechanics. Such MD models are written as relatively large systems of ordinary or stochastic differential equations for the positions and velocities of individual atoms, which can also be subject to algebraic constraints (Leimkuhler and Matthews, 2015; Lewars, 2016). Although all-atom MD simulations of systems consisting of a million of atoms have been reported in the literature (Tarasova et al., 2017; Farafanov and Nerukh, 2019), such simulations are restricted to relatively small computational domains, which are up to tens of nanometres long. It is beyond the reach of state-of-the-art computers to simulate intracellular processes which include transport of molecules over micrometers, because this would require simulations of trillions of atoms (Erban and Chapman, 2019).

An example is modelling of calcium (Ca2+) dynamics. On one hand, at the macroscopic level, Ca2+ waves can propagate between cells over hundreds of micrometres and Kang and Othmer (2009) developed a model of Ca 2+ waves in a network of astrocytes. It builds

on previous modelling work by Kang and Othmer (2007) describing intracellular Ca2+ dynamics as a system of differential equations for concentrations of chemical species involved, including inositol 1,4,5-trisphosphate (IP3), a chemical signal that binds to the IP3 receptor to release Ca2+ ions from the endoplasmic reticulum. On the other hand, at the atomic level, Hamada et al. (2017) recently solved IP3-bound and unbound structures of large cytosolic domains of the IP3 receptor by X-ray crystallography and clarified the IP3-dependent gating mechanism through a unique leaflet structure. Although it is not possible to incorporate such a detailed information into Ca2+ modelling by using all-atom MD in the entire intracellular space, there is still potential to design multiscale (multi-resolution) models which compute Ca2+ dynamics with the resolution of individual Ca2+ ions. Dobramysl et al. (2016) implement such a methodology at the Brownian dynamics (BD) level to study Ca2+ puff statistics stemming from IP3 receptor channels. Denoting the position of an individual Ca2+ ion by X ≡ (X1, X2, X3), its diffusive BD trajectory is given by

$$\mathrm{d}X_i = \sqrt{2D}\,\mathrm{d}W_i, \qquad \text{for } i = 1, 2, 3, \tag{1}$$

where D is the diffusion constant and Wi , i = 1, 2, 3, are three independent Wiener processes. Since individual positions of Ca2+ ions are only needed in the vicinity of channel sites, Dobramysl et al. (2016) model diffusion of ions far away of the channel by a coarser model, utilizing the two-regime method developed by Flegg et al. (2012). This method enables efficient simulations with the BD level of resolution by coarse-graining the BD model in those parts of the simulation domain, where the coarse-grained model can be safely used without introducing significant numerical errors (Flegg et al., 2014, 2015; Robinson et al., 2015).

Although BD models or their multi-resolution extensions simulate individual molecules of chemical species involved, the binding of Ca2+ ions to channel sites or other interactions between molecules are only described using relatively coarse probabilistic approaches. For example, the BD model of Dobramysl et al. (2016) describes interactions in terms of reaction radii and binding probabilities as implemented by Erban and Chapman (2009) and Lipkov´a et al. (2011). Atomic-level information is not included in BD models. In order to use this information, multi-resolution methodologies have to consider MD simulations in parts of the simulation domain. In the case of ions, such a multi-resolution scheme has been developed by Erban (2016), where an all-atom MD model of ions in water is coupled with a stochastic coarsegrained (SCG) description of ions in the rest of the computational domain. The accuracy and efficiency of such multi-resolution methodologies depend on the quality of the SCG description of the underlying MD model. In this paper, we present and analyze a class of SCG models which can be used to fit non-Gaussian distributions estimated from all-atom MD simulations. While the velocity distribution of the coarse-grained particle can be well approximated by a Gaussian (normal) distribution in our MD simulations, this is not the case of the force distribution. Non-Gaussian force distributions have also been reported by Shin et al. (2010) and Carof et al. (2014) in their MD simulations of particles in Lennard-Jones fluids. Thus our SCG model is formulated in a way which incorporates a Gaussian distribution for the velocity and a non-Gaussian distribution for the force (acceleration).

Given an integer N ≥ 1, a coarse-grained particle (for example, an ion) will be described by (2N + 2) three-dimensional variables: its position X, velocity V and 2N auxiliary variables Uj and Zj , where j = 1, 2, . . . , N. Denoting X ≡ (X1, X2, X3), V ≡ (V1, V2, V3), Uj ≡ (Uj,1, Uj,2, Uj,3) and Zj ≡ (Zj,1, Zj,2, Zj,3), the time evolution of the SCG model is given by

$$\mathrm{d}X_i = V_i\,\mathrm{d}t, \qquad \text{for } i = 1, 2, 3, \tag{2}$$

$$\mathrm{d}V_i = \sum_{j=1}^{N} U_{j,i}\,\mathrm{d}t, \tag{3}$$

$$\mathrm{d}U_{j,i} = \left(-\eta_{j,1}V_i + h_j(Z_{j,i})\right)\, g_j'(g_j^{-1}(U_{j,i}))\,\mathrm{d}t, \qquad \text{for } j = 1, 2, \ldots, N, \tag{4}$$

$$\mathrm{d}Z_{j,i} = -\left(\eta_{j,2}\, h_j(Z_{j,i}) + \eta_{j,3} U_{j,i}\right)\mathrm{d}t + \eta_{j,4}\,\mathrm{d}W_{j,i}, \qquad (5)$$

where gj : R → R is an increasing differentiable function, g ′ j is its derivative, g −1 j is its inverse, hj : R → R is a continuous function and ηj,k are positive constants for j = 1, 2, . . . , N and k = 1, 2, 3, 4. We note that some of our assumptions on gj can be relaxed as long as g ′ j (g −1 j (Uj,i)) appearing in equation (4) can be suitably defined.

The SCG description (2)–(5) includes 2N functions gj and hj and 4N additional parameters ηj,k, which can be all adjusted to fit properties of the detailed all-atom MD model. In particular the SCG model (2)–(5) can better match the MD trajectories of ions than the BD description given by equation (1), which only has one parameter, diffusion constant D, to fit to the MD results.

One of the shortcomings of equation (1) is that its derivation from the underlying MD model requires us to consider the limit of sufficiently large times. In particular, we need to discretize equation (1) with a relatively large.

time step, say a nanosecond, to use it as a description of the trajectory of an ion. Since the typical time step of an all-atom MD model is a femtosecond, it is difficult to design a multi-resolution scheme which would replace all-atom MD simulations by equation (1) in parts of the computational domain. The SCG model (2)–(5) can be used to fit not only the diffusion constant D but other important properties of all-atom MD models, which improves the accuracy of the SCG model at time steps comparable with the MD timestep.

SCG models can be constructed using a relatively automated procedure by postulating that an ion interacts with additional 'fictitious particles'. Such a methodology has been applied to coarse-grained modelling of biomolecules by Davtyan et al. (2015, 2016) to improve the fit between an MD model and the dynamics on a coarse-grained potential surface. They use fictitious particles with harmonic interactions with coarse-grained degrees of freedom (i.e. they add quadratic terms to the potential function of the system and linear terms to equations of motions) and each fictitious particle is also subject to a friction force and noise. An application of such an approach to ions leads to systems of linear stochastic differential equations (SDEs) and can be used, after some transformation, to obtain a simplified version of the SCG model (2)–(5), where functions gj and hj are given as identities, i.e. gj (y) = hj (y) = y for y ∈ R and j = 1, 2, . . . , N. Using this simplifying assumption in the SCG model (2)–(5), we obtain

$$\mathrm{d}X_i = V_i\,\mathrm{d}t, \qquad \text{for } i = 1,2,3, \qquad (6)$$

$$\mathrm{d}V_i = \sum_{j=1}^{N} U_{j,i}\,\mathrm{d}t, \qquad (7)$$

$$\mathrm{d}U_{j,i} = (-\eta_{j,1} V_i + Z_{j,i})\,\mathrm{d}t, \qquad \text{for } j = 1,2,\ldots,N, \qquad (8)$$

$$\mathrm{d}Z_{j,i} = -\left(\eta_{j,2} Z_{j,i} + \eta_{j,3} U_{j,i}\right)\mathrm{d}t + \eta_{j,4}\,\mathrm{d}W_{j,i}. \qquad (9)$$

This is a linear system of SDEs with 4N parameters. It has been shown by Erban (2016) that such models can fit an increasing number of properties of all-atom MD simulations as we increase N. For example, the linear SCG model (6)–(9) can be used to fit the diffusion constant D and second moments of the velocity and the force for N = 1, while the velocity autocorrelation function can better be fitted for larger values of N, e.g. for N = 3. However, there are other properties of MD simulations which cannot be captured by linear models even if consider arbitrarily large N. They include, for example, all distributions which are not Gaussian. This motivates the introduction of general functions hj and gj in the SCG model (2)–(5).

Considering the SCG model (2)–(5) in its full generality, it can capture more interesting dynamics. However, coarse-grained models can only be useful if they can be easily parametrized. Thus in our analysis, we focus on choices of functions gj and hj which both improve the properties of the SCG description and do not complicate its analysis and parametrization. The rest of the paper is organized as follows. In Section 2, we consider the linear SCG model (6)– (9) for N = 1, which is followed in Section 3 with the analysis of the linear model for general values of N. To get some further insights into the properties of this model, we study its connections with the corresponding generalized Langevin equation. In Section 4, we consider the

nonlinear SCG model (2)– (5) for N = 1. We consider specific choices of nonlinearity g1, for which the model can be solved in terms of incomplete gamma functions. This helps us to design three approaches to parametrize the nonlinear SCG model, which are applied to data obtained from MD simulations. We conclude with the analysis of the nonlinear SCG model (2)–(5) for general values of N in Section 5.

2 Linear model for N = 1 and the generalized Langevin equation

We begin by considering the linear SCG model (6)–(9) for N = 1. To simplify our notation in this section, we will drop some subscripts and denote X = Xi , V = Vi , U = U1,i, Z = Z1,i, W = W1,i and ηk = η1,k for k = 1, 2, 3, 4. Then equations (6)–(9) read as follows

$$\mathrm{d}X = V\,\mathrm{d}t, \tag{10}$$
$$\mathrm{d}V = U\,\mathrm{d}t, \tag{11}$$
$$\mathrm{d}U = (-\eta_1 V + Z)\,\mathrm{d}t, \tag{12}$$
$$\mathrm{d}Z = -(\eta_2 Z + \eta_3 U)\,\mathrm{d}t + \eta_4\,\mathrm{d}W, \tag{13}$$

where X is (one coordinate of) the position of the coarse-grained particle (ion), V is its velocity, U is its acceleration, Z is an auxiliary variable, dW is white noise and ηj , j = 1, 2, 3, 4, are positive parameters. In order to find the values of four parameters ηj suitable for modelling ions, Erban (2016) estimates the diffusion constants D and three second moments hV 2 i, hU 2 i and hZ 2 i from allatom MD simulations of ions (K+, Na+, Ca2+ and Cl−) in aqueous solutions. The four parameters of the SCG model (10)–(13) can then be chosen as

$$\eta_1 = \frac{\langle U^2\rangle}{\langle V^2\rangle},\quad \eta_2 = \frac{\langle Z^2\rangle}{D}\left(\frac{\langle V^2\rangle}{\langle U^2\rangle}\right)^2,\quad \eta_3 = \frac{\langle Z^2\rangle}{\langle U^2\rangle},\quad \eta_4 = \sqrt{\frac{2}{D}}\frac{\langle V^2\rangle\langle Z^2\rangle}{\langle U^2\rangle}. \tag{14}$$

Then the SCG model (10)–(13) gives the same values of D, hV 2 i, hU 2 i and hZ 2 i as obtained in all-atom MD simulations.

Since the model (10)–(13) only has four parameters, we can only hope to get the exact match of four quantities estimated from all-atom MD. To get some insights into what we are missing, we will derive the corresponding generalized Langevin equation and study its consequences. The generalized Langevin equation can be written in the form

$$\frac{\mathrm{d}V}{\mathrm{d}t} = -\int_0^t K(\tau)\,V(t-\tau)\,\mathrm{d}\tau + R(t), \tag{15}$$

where K : [0,∞) → R is a memory kernel and random term R(t) satisfies the generalized fluctuation-dissipation theorem, given below in equation (21). To derive the generalized Langevin equation (15), consider the two-variable subsystem (12)–(13) of the SCG model. Denoting y = (U, Z) T, where T stands for transpose, equations (12)–(13) can be written in vector notation as follows

$$\mathrm{d}\mathbf{y} = B\,\mathbf{y}\,\mathrm{d}t + \mathbf{b}_1 V\,\mathrm{d}t + \mathbf{b}_2\,\mathrm{d}W, \tag{16}$$

where matrix B ∈ R 2×2 and vectors bj∈ R 2 , j = 1, 2, are given as

$$B = \begin{pmatrix} 0 & 1 \\ -\eta_3 & -\eta_2 \end{pmatrix},\qquad \mathbf{b}_1 = \begin{pmatrix} -\eta_1 \\ 0 \end{pmatrix}\qquad \text{and}\qquad \mathbf{b}_2 = \begin{pmatrix} 0 \\ \eta_4 \end{pmatrix}.$$

Let us denote the eigenvalues and eigenvectors of B as λj and νj = (1, λj) T, j = 1, 2, respectively. The eigenvalues of B are the solutions of the characteristic polynomial λ 2 + η2 λ + η3 = 0. They are given by

$$\lambda_1 = -\frac{\eta_2}{2} + \mu \quad \text{and} \quad \lambda_2 = -\frac{\eta_2}{2} - \mu \quad \text{where} \quad \mu = \sqrt{\frac{\eta_2^2}{4} - \eta_3}. \tag{17}$$

Since η2 and η3 are positive parameters, we conclude that real parts of both eigenvalues are negative. In what follows, we will assume η 2 2 6= 4η3. Then we have two distinct eigenvalues and the general solution of the SDE system (16) can be written as follows

$$\mathbf{y}(t) = \Phi(t)\,\mathbf{c} + \Phi(t)\int_0^t \Phi^{-1}(s)\,\mathbf{b}_1 V(s)\,\mathrm{d}s + \Phi(t)\int_0^t \Phi^{-1}(s)\,\mathbf{b}_2\,\mathrm{d}W, \tag{18}$$

where c ∈ R 2 is a constant vector determined by initial conditions and matrix Φ(t) ∈ R 2×2 is given as

$$\Phi(t) = (\exp(\lambda_1 t)\boldsymbol{\nu}_1 \mid \exp(\lambda_2 t)\boldsymbol{\nu}_2) = \begin{pmatrix} \exp(\lambda_1 t) & \exp(\lambda_2 t) \\ \lambda_1\exp(\lambda_1 t) & \lambda_2\exp(\lambda_2 t) \end{pmatrix},$$

i.e. each column is a solution of the ODE system dy = B y dt. Calculating the inverse of Φ(t) and considering long-time behaviour, equation (18) simplifies to

$$U(t) = -\int_0^t K(\tau)\, V(t-\tau)\, \mathrm{d}\tau + R(t), \qquad (19)$$

where memory kernel K(τ) is given by

$$K(\tau) = \frac{\eta_1}{\lambda_1 - \lambda_2}\left(\lambda_1 \exp(\lambda_2\, \tau) - \lambda_2 \exp(\lambda_1\, \tau)\right) \qquad (20)$$

and noise term R(t) is Gaussian with zero mean and the equilibrium correlation function satisfying the generalized fluctuation-dissipation theorem in the form

$$\langle R(t_1) R(t_2)\rangle = \frac{\eta_4^2}{2\eta_1\eta_2\eta_3}\, K(t_2 - t_1). \qquad (21)$$

Using (17), memory kernel (20) can be rewritten as

$$K(\tau) = \eta_1 \exp\left(-\frac{\eta_2\, \tau}{2}\right)\left(\cosh(\mu\, \tau) + \frac{\eta_2}{2\mu}\sinh(\mu\, \tau)\right), \qquad (22)$$

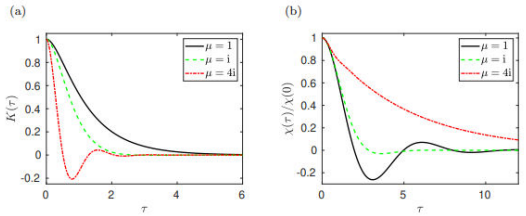

Fig. 1 (a) Memory kernel K(τ) given by equation (22) for η1 = 1, η2 = 4 and three different values of η3, namely η3 = 3 (solid line, μ = 1), η3 = 5 (dashed line, μ = i) and η3 = 20 (dot-dashed line, μ = 4i). (b) Normalized velocity autocorrelation function χ(τ)/χ(0) computed by using equation (25) for the same parameter values as in panel (a).

where μ = p η 2 2 /4 − η3. We note that the auxiliary coefficient μ is a square root of a real negative number for η 2 2 < 4η3. However, formula (22) is still valid in this case: for η 2 2 < 4η3 it can be rewritten in terms of sine and cosine functions, taking into account that μ = i |μ| is pure imaginary, sinh(i |μ| τ) = i sin(|μ|) τ and cosh(i |μ| τ) = cos(|μ| τ). The memory kernel K(τ), given by equation (22), is plotted in Figure 1(a) for different values of parameter μ. For simplicity, we use non-dimensionalized versions of our equations with dimensionless parameters η1 = 1 and η2 = 4. We choose three different values of η3 so that the values of μ are 1, i and 4i. In Figure 1(b), we plot the equilibrium velocity autocorrelation function which is defined as

$$\chi(\tau) = \lim_{t\to\infty}\langle V(t)\, V(t-\tau)\rangle,$$

for τ ∈ [0,∞). More precisely, we plot χ(τ)/χ(0) which is normalized so that its value at τ = 0 is equal to 1. It is related to the memory kernel by

$$\frac{\chi(\tau)}{\chi(0)} = \mathscr{L}^{-1}\left(\frac{1}{s + \mathscr{L}[K](s)}\right), \qquad (23)$$

where L K (s) = R ∞ 0 K(τ) exp(−sτ) dτ is the Laplace transform of the memory kernel K(τ) and L −1 denotes Laplace inversion. Following Erban and Chapman (2019), we evaluate the right hand side of equation (23) as follows. Substituting equation (22) into (23), we obtain

$$\frac{\chi(\tau)}{\chi(0)} = \mathscr{L}^{-1}\left(\frac{s^2 + \eta_2 s + \eta_3}{s^3 + \eta_2 s^2 + (\eta_1 + \eta_3)s + \eta_1\eta_2}\right). \qquad (24)$$

The polynomial in the denominator, p(s) = s 3 + η2s 2 + (η1 + η3)s + η1η2, has positive coefficients. Since p(−η2) < 0 < p(0), it has one negative real root in interval (−η2, 0), which we denote by a1. The other two roots (a2 and a3 say) may be real or complex, but if they are complex they will be complex conjugates since p(s) has real coefficients. Assuming that the real part of each root is negative, we first find the partial fraction decomposition of the rational function in (24) as

$$\frac{s^2 + \eta_2 s + \eta_3}{s^3 + \eta_2 s^2 + (\eta_1 + \eta_3)s + \eta_1\eta_2} = \frac{c_1}{s - a_1} + \frac{c_2}{s - a_2} + \frac{c_3}{s - a_3},$$

where ci ∈ C are constants (which depend on η1, η2 and η3). Then we can rewrite (23) as

$$\frac{\chi(\tau)}{\chi(0)} = c_1 \exp(a_1\tau) + c_2 \exp(a_2\tau) + c_3 \exp(a_3\tau)\,. \qquad (25)$$

The results computed by (25) are shown in Figure 1(b). We note that although equation (25) may include complex exponentials, the resulting χ(τ) is always real. Since the diffusion constant, D, and the second moment of the equilibrium velocity distribution, hV 2 i, are related to χ by

$$D = \int_0^\infty \chi(\tau)\,\mathrm{d}\tau = \frac{\eta_4^2}{2\,\eta_1^2\,\eta_2^2} \quad \text{and} \quad \langle V^2\rangle = \chi(0) = \frac{\eta_4^2}{2\,\eta_1\,\eta_2\,\eta_3}\,,$$

the parametrization (14) guarantees that both the value of χ(0) and the integral of χ(τ) are captured accurately. However, the simplified SCG description (10)–(13) is not suitable to perfectly fit the velocity autocorrelation function or the memory kernel for all values of τ ∈ [0,∞). In order to do this, we have to consider the SCG model (6)–(9) for larger values of N as it is done in the following section.

3 General linear SCG model and autocorrelation functions

Considering the linear SCG model (6)–(9) for general values of N, we can solve equations (8)–(9) for each value of j = 1, 2, . . . , N to generalize our previous result (19) as

$$U_{j,i}(t) = -\int_0^t K_j(\tau)\,V_i(t-\tau)\,\mathrm{d}\tau + R_{j,i}(t)\,, \qquad (26)$$

where kernel Kj (τ) is given by (compare with (22))

$$K_j(\tau) = \eta_{j,1}\,\exp\left(-\frac{\eta_{j,2}\,\tau}{2}\right)\left(\cosh(\mu_j\,\tau) + \frac{\eta_{j,2}}{2\mu_j}\sinh(\mu_j\,\tau)\right) \qquad (27)$$

With

$$\mu_j = \sqrt{\frac{\eta_{j,2}^2}{4} - \eta_{j,3}} \qquad (28)$$

and noise term Rj,i(t) is Gaussian with zero mean and the equilibrium correlation function satisfying

$$\langle R_{j,i}(t_1)R_{j,i}(t_2)\rangle = \frac{\eta_{j,4}^2}{2\,\eta_{j,1}\,\eta_{j,2}\,\eta_{j,3}}\,K_j(t_2-t_1).$$

Substituting (26) to (7), we obtain the generalized Langevin equation

$$\frac{\mathrm{d}V_i}{\mathrm{d}t} = -\int_0^t K(\tau)\,V_i(t-\tau)\,\mathrm{d}\tau + R_i(t)\,, \qquad (29)$$

Where

$$K(\tau) = \sum_{j=1}^{N} K_j(\tau) \quad \text{and} \quad R_i(t) = \sum_{j=1}^{N} R_{j,i}(t). \qquad (30)$$

In particular, we have 3N parameters to fit memory kernel K(τ), which can be estimated from all-atom MD simulations. There have been a number of approaches developed in the literature to estimate the memory kernel from MD simulations. Shin et al. (2010) use an integral equation with relates memory kernel K(τ) with the autocorrelation function for the force and the correlation function between the force and the velocity. Estimating these correlation functions from long time MD simulations and solving the integral equation, they obtain memory kernel K(τ). Other methods to estimate the memory kernel, K(τ), of the corresponding generalized Langevin equation (29) have been presented by Gottwald et al. (2015) and Jung et al. (2017). An alternative approach to parametrize the linear SCG model (6)–(9) is to estimate the velocity autocorrelation function, χ(τ), from all-atom MD simulations. This can be done by computing how correlated is the current velocity (at time t) with velocity at previous times. Since equations (10)–(13) are linear SDEs, we can follow Mao (2007) to solve them analytically, using eigenvalues and eigenvectors of matrices appearing in their corresponding matrix formulation. Using this analytic solution, Erban (2016) use an acceptance-rejection algorithm to fit the parameters of linear SCG model (6)–(9) for N = 3 to match the velocity autocorrelation functions of ions estimated from all-atom MD simulations of Na+ and K+ in the SPC/E water. Since the parameter μj given by (28) is a square root of a real number, it can be both positive or purely imaginary. In particular, kernelsKj (τ) given by equation (27) can include both exponential, sine and cosine functions as illustrated in Figure 1(a). Since memory kernel K(τ) is given as the sum

of Kj (τ) in equation (30), typical memory kernels and correlation functions estimated from all-atom MD simulations can be successfully matched by linear SCG models for relatively small values of N. However, as shown by Mao (2007), analytic solutions of linear SDEs also imply that the process is Gaussian at any time t > 0, provided that we start with deterministic initial conditions. Thus the linear SCG model (6)–(9) for abtitrary values of N can only fit distributions which are Gaussian. This motivates our investigation of the nonlinear SCG model in the next two sections.

4 Nonlinear SCG model for N = 1

We begin by considering the nonlinear SCG model (2)–(5) for N = 1. As in Section 2, we simplify our notation by dropping some subscripts and denoting X = Xi , V = Vi , U = U1,i, Z = Z1,i, W = W1,i, g = gj , h = hj and ηk = η1,k for k = 1, 2, 3, 4. Then equations (2)–(5) read as follows

$$\mathrm{d}X = V\,\mathrm{d}t, \tag{31}$$
$$\mathrm{d}V = U\,\mathrm{d}t, \tag{32}$$
$$\mathrm{d}U = (-\eta_1 V + h(Z))\; g'(g^{-1}(U))\,\mathrm{d}t, \tag{33}$$
$$\mathrm{d}Z = -\left(\eta_2\, h(Z) + \eta_3\, U\right)\,\mathrm{d}t + \eta_4\,\mathrm{d}W, \tag{34}$$

where X denotes (one coordinate of) the position of the coarse-grained particle, V is its velocity, U is its acceleration, Z is an auxiliary variable, dW is white noise, ηj , for j = 1, 2, 3, 4, are positive parameters and functions g : R → R and h : R → R are yet to be specified.

Equation (31) describes the time evolution of the position, while equations (32)–(34) admit a stationary distribution. We denote it by p(v, u, z). Then p(v, u, z) dv du dz gives the probability that V (t) ∈ [v, v+dv), U(t) ∈ [u, u+du) and Z(t) ∈ [z, z + dz) at equilibrium. The stationary distribution, p(v, u, z), of SDEs (32)–(34) can be obtained by solving the corresponding stationary Fokker-Planck equation

$$\frac{\eta_4^2}{2}\frac{\partial^2 p}{\partial^2 z}(v,u,z) = \frac{\partial}{\partial v}\Big(u\,p(v,u,z)\Big) + \frac{\partial}{\partial u}\Big(\big(-\eta_1 v + h(z)\big)g'(g^{-1}(u))\,p(v,u,z)\Big) + \frac{\partial}{\partial z}\Big(\big(-\eta_2 h(z) - \eta_3 u\big)p(v,u,z)\Big),$$

which give

$$p(v,u,z) = \frac{C}{g'(g^{-1}(u))}\,\exp\left[-\frac{2\eta_2}{\eta_4^2}\left(\eta_1\eta_3\,\frac{v^2}{2} + \eta_3\,G\big(g^{-1}(u)\big) + H(z)\right)\right], \tag{35}$$

) where C is the normalization constant, and functions G and H are integrals of functions g and h, respectively, which are given

$$G(y) = \int_0^y g(\xi)\,\mathrm{d}\xi \qquad \text{and} \qquad H(y) = \int_0^y h(\xi)\,\mathrm{d}\xi. \tag{36}$$

) We note that for the special case where g and h are given as identities, i.e. g(y) = h(y) = y for y ∈ R, the nonlinear SCG model (31)–(34) is equal to the linear SCG model (10)–(13) and functions G and H are G(y) = H(y) = y 2/2. Then the stationary distribution (35) is product of Gaussian distributions in v, u and z variables. In particular, we can easily calculate the second moments of these distributions in terms of parameters ηj . Estimating these moments from all-atom MD simulations, we can parametrize the resulting linear SCG model (10)–(13) as shown in equation (14). However, if we want to match a non-Gaussian force distribution, we have to consider nonlinear models. A simple one-parameter example is studied in the next section.

4.1 One-parameter nonlinear function

Consider that g is a function depending on one additional positive parameter η5 as follows

$$g(y) = |y|^{1/\eta_5}\,\operatorname{sign} y, \tag{37}$$

where we use sign to denote the sign (signum) function

$$\operatorname{sign} y = \begin{cases} -1, & \text{for } y < 0, \\ 0, & \text{for } y = 0, \\ 1, & \text{for } y > 0. \end{cases} \tag{38}$$

The function defined by (37) only satisfies our assumptions on g for η5 ∈ (0, 1] as it is not differentiable at y = 0 for η5 > 1, but we will proceed with our analysis for any positive η5 > 0. Consider that function h is an identity, i.e. h(y) = y for y ∈ R, then equations (31)–(34) reduce to

$$\mathrm{d}X = V\,\mathrm{d}t, \tag{39}$$
$$\mathrm{d}V = U\,\mathrm{d}t, \tag{40}$$
$$\mathrm{d}U = (-\eta_1 V + Z)\;\eta_5^{-1}\,|U|^{1-\eta_5}\,\mathrm{d}t, \tag{41}$$
$$\mathrm{d}Z = -\left(\eta_2\, Z + \eta_3\, U\right)\,\mathrm{d}t + \eta_4\,\mathrm{d}W, \tag{42}$$

where we would have to be careful, if we used this model to numerically simulate trajectories for η5 > 1, because of possible division by zero for U = 0 in equation (41). If η5 ∈ (0, 1], then we do not have such technical issues. Using equation (35), the stationary distribution is equal to

$$p(v,u,z) = C|u|^{\eta_5-1}\exp\left[-\frac{\eta_2}{\eta_4^2}\left(\eta_1\eta_3\, v^2 + \frac{2\eta_3\eta_5}{1+\eta_5}|u|^{1+\eta_5} + z^2\right)\right], \quad (43)$$

where the normalization constant is given by

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(v,u,z)\,\mathrm{d}v\,\mathrm{d}u\,\mathrm{d}z = 1.$$

Integrating (43), we get

$$C = \frac{\eta_2\sqrt{\eta_1\eta_3}}{\pi\eta_4^2}\left(\frac{\eta_2\eta_3\eta_5}{\eta_4^2}\right)^{\eta_5/(1+\eta_5)}\left(\frac{1+\eta_5}{2}\right)^{1/(1+\eta_5)}\frac{1}{\Gamma\left(\frac{\eta_5}{1+\eta_5}\right)},$$

where Γ is the gamma function defined as

$$\Gamma(s) = \int_0^{\infty}\xi^{s-1}\exp(-\xi)\,\mathrm{d}\xi. \quad (44)$$

Let α ≥ 0. Integrating (43), we get

$$\langle|U|^{\alpha}\rangle = \left(\frac{\eta_4^2(1+\eta_5)}{2\eta_2\eta_3\eta_5}\right)^{\alpha/(1+\eta_5)}\frac{\Gamma\left(\frac{\alpha+\eta_5}{1+\eta_5}\right)}{\Gamma\left(\frac{\eta_5}{1+\eta_5}\right)}. \quad (45)$$

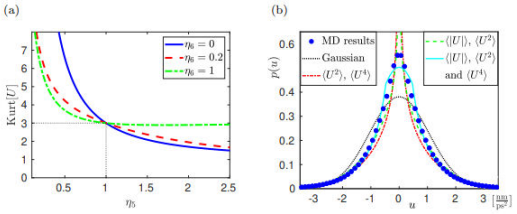

Fig. 2 (a) Kurtosis Kurt[U] given by equation (59) as a function of parameter η5 for three different values of parameter η6. The result for η6 = 0 (blue solid line) corresponds to the case of one-parameter function g, defined by (37), where the kurtosis is given by (46). (b)Distribution of U estimated from a long-time MD simulation (blue circles) compared with the results obtained by the linear SCG model (10)–(13) (black dotted line), nonlinear SCG models (31)–(34) with one-parameter function g, defined by (37), fitting hU2 i and hU4 i (red dot-dashed line) and h|U|i and hU2 i (green dashed line), and the nonlinear SCG model (31)–(34) with two-parameter function g defined by (52), matching all three moments h|U|i, hU2 i and hU4 i (cyan solid line).

Using (45) for α = 2 and α = 4, we obtain the following expression for kurtosis

$$\mathrm{Kurt}[U] = \frac{\langle U^4\rangle}{\langle U^2\rangle^2} = \Gamma\left(\frac{\eta_5}{1+\eta_5}\right)\Gamma\left(\frac{4+\eta_5}{1+\eta_5}\right)\left(\Gamma\left(\frac{2+\eta_5}{1+\eta_5}\right)\right)^{-2}. \quad (46)$$

In particular, the kurtosis is only a function of one parameter, η5. It is plotted in Figure 2(a) as the blue solid line, together with the kurtosis obtained for a more general two-parameter SCG model studied in Section 4.2. We observe that the distribution of U is leptokurtic for η5 < 1 and platykurtic for η5 > 1. If η5 is equal to 1, then our SCG model given by equations (31)–(34) reduces to the linear SCG model given by equations (10)–(13), i.e. the stationary distribution is Gaussian and its kurtosis is 3. This is shown by the dotted line in Figure 2(a).

Since equation (46) only depends on parameter η5, we can use the kurtosis of the acceleration distribution (which is equal to the kurtosis of the force distribution) esimated from MD simulations to find the value of parameter η5. To calculate the kurtosis, we estimate the fourth moment hU 4 i in addition to the second moment, hU 2 i, used before in our estimating proceduce (14) for the linear model. In particular, we not only get equation (46) for calculating the value of parameter η5, but also a restriction on other parameters η2, η3 and η4. Using (45) for α = 2, it can be stated as follows

$$\frac{\eta_4^2}{2\eta_2\eta_3} = \frac{\eta_5}{1+\eta_5}\left(\frac{1+\eta_5}{\pi}\sin\left(\frac{\pi}{1+\eta_5}\right)\langle U^2\rangle\right)^{(1+\eta_5)/2}\left(\Gamma\left(\frac{\eta_5}{1+\eta_5}\right)\right)^{1+\eta_5}, \quad (47)$$

where we have used properties of the gamma function, including Γ(1 + y) = y Γ(y) and Euler's reflection formula, Γ(1−y)Γ(y) sin(πy) = π, to simplify the right hand side. We note that in the Gaussian case, η5 = 1, the right hand side of equation (47) further simplifies to

$$\frac{\eta_4^2}{2\,\eta_2\,\eta_3} = \langle U^2 \rangle, \qquad (48)$$

which is indeed the formula for the second moment of U given by the linear SCG model (10)–(13). Equation (47) provides one restriction on four remaining parameters, η1, η2 η3 and η4, which need to be specified. This can be done by estimating three additional statistics from MD simulations, as in the case of the linear SCG model (10)–(13) in equation (14). Indeed, the stationary distributions of V and Z are Gaussian with mean zero. Their second moments and the diffusion constant, D, for the nonlinear SCG model (31)–(34) can be calculuted as

$$D = \frac{\eta_4^2}{2\,\eta_1^2\,\eta_2^2}, \qquad \langle V^2 \rangle = \frac{\eta_4^2}{2\,\eta_1\,\eta_2\,\eta_3} \qquad \text{and} \qquad \langle Z^2 \rangle = \frac{\eta_4^2}{2\,\eta_2}. \qquad (49)$$

Therefore, assuming that D, hV 2 i, hZ 2 i are obtained from MD simulations and η 2 4/(2η2η3) is given by (47), we can calculate parameters ηk by

$$\eta_1 = \frac{1}{\langle V^2 \rangle}\left(\frac{\eta_4^2}{2\,\eta_2\,\eta_3}\right), \qquad \eta_2 = \frac{\langle Z^2 \rangle\,\langle V^2 \rangle^2}{D}\left(\frac{\eta_4^2}{2\,\eta_2\,\eta_3}\right)^{-2}, \qquad (50)$$

$$\eta_3 = \langle Z^2 \rangle\left(\frac{\eta_4^2}{2\,\eta_2\,\eta_3}\right)^{-1}, \qquad \eta_4 = \sqrt{\frac{2}{D}}\,\langle Z^2 \rangle\,\langle V^2 \rangle\left(\frac{\eta_4^2}{2\,\eta_2\,\eta_3}\right)^{-1}. \qquad (51)$$

We note that in the Gaussian case, η5 = 1, we can substitute equation (48) for η 2 4/(2η2η3) and the parametrization approach (50)–(51) simplifies to equation (14) used in the case of the linear SCG model (10)–(13). In the next subsection, we generalize formula (37) to a two-parameter function and show that the parametrization approach (50)–(51) is still applicable to the case of more general SCG models.

4.2 Two-parameter nonlinear function

Consider that g is a function depending on two positive parameters η5 and η6 as follows

$$g(y) = \begin{cases} 0, & \text{for } |y| \le \eta_6^{\eta_5}(1-\eta_5), \\ \left(\eta_6\left(1-\frac{1}{\eta_5}\right) + \frac{\eta_6^{1-\eta_5}}{\eta_5}|y|\right)\operatorname{sign} y, & \text{for } \eta_6^{\eta_5}(1-\eta_5) < |y| \le \eta_6^{\eta_5}, \\ |y|^{1/\eta_5}\operatorname{sign} y, & \text{for } |y| > \eta_6^{\eta_5}, \end{cases} \qquad (52)$$

where sign function is defined by (38). In particular, our expression for function g is equal to the formula (37) for sufficiently large values of |y|. As discussed in the previous section, if we used formula (37), there would be some issues for y close to zero (for example, the division by zero for U = 0 and η5 > 1 in equation (41)), so our generalized formula (52) replaces (37) with a linear function for smaller values of |y|. On the face of it, it looks that there could also be some issues with the generalized formula (52), because it is not strictly increasing for |y| ≤ η η5 6 (1 − η5). However, function (52) is increasing and invertible away of this region with its inverse given by

$$g^{-1}(u) = \begin{cases} \eta_5\eta_6^{\eta_5-1}\left(|u| - \eta_6\left(1-\frac{1}{\eta_5}\right)\right)\operatorname{sign} u, & \text{for } 0 < |u| \le \eta_6, \\ |u|^{\eta_5}\operatorname{sign} u, & \text{for } |u| > \eta_6. \end{cases}$$

Moreover, what we really need in equations (31)–(34) is g ′ (g −1 (u)) which can be defined as the following continuous function

$$g'(g^{-1}(u)) = \frac{1}{\eta_5} \times \begin{cases} \eta_6^{1-\eta_5}, & \text{for } |u| \le \eta_6, \\ |u|^{1-\eta_5}, & \text{for } |u| > \eta_6, \end{cases} \qquad (53)$$

where the removable discontinuity at u = 0 has disappeared because we have defined g ′ (g −1 (0)) = η 1−η5 6 /η5. Integrating (52) and substituting (53), we get

$$G(g^{-1}(u)) = \begin{cases} \frac{\eta_5\eta_6^{\eta_5-1}}{2}u^2, & \text{for } |u| \le \eta_6, \\ \frac{\eta_5(\eta_5-1)\eta_6^{1+\eta_5}}{2(1+\eta_5)} + \frac{\eta_5}{1+\eta_5}|u|^{1+\eta_5}, & \text{for } |u| > \eta_6, \end{cases} \qquad (54)$$

where G is the integral of function g defined by (36). Consider again that h is an identity, i.e. h(y) = y for y ∈ R. Then the stationary distribution (35) is again Gaussian in V and Z variables with their second moments given by equation (49). Let us denote the marginal stationary distribution of U by

$$p_u(u) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(v,u,z)\,\mathrm{d}v\,\mathrm{d}z.$$

Using (35) and (54), we hav

$$p_u(u) = \begin{cases} C_u\,\eta_6^{\eta_5-1}\exp\left[-\dfrac{\eta_2\eta_3\eta_5\eta_6^{1+\eta_5}}{\eta_4^2}\left(\dfrac{u^2}{\eta_6^2}+\dfrac{1-\eta_5}{1+\eta_5}\right)\right], & \text{for } |u|\le\eta_6, \\ C_u\,|u|^{\eta_5-1}\exp\left[-\dfrac{2\eta_2\eta_3\eta_5}{\eta_4^2(1+\eta_5)}|u|^{1+\eta_5}\right], & \text{for } |u|>\eta_6, \end{cases} \tag{55}$$

where Cu is the normalization constant given by

$$\int_{-\infty}^{\infty} p_u(u)\,\mathrm{d}u = 1.$$

Let us define

$$\kappa_1 = \frac{\eta_2\eta_3\eta_5\eta_6^{1+\eta_5}}{\eta_4^2} \quad \text{and} \quad \kappa_2 = \frac{1}{1+\eta_5}. \tag{56}$$

Integrating (55), we get, for any α ≥ 0,

$$\frac{\langle|U|^\alpha\rangle}{\eta_6^\alpha} = \frac{F(\kappa_1,\kappa_2,\alpha)}{F(\kappa_1,\kappa_2,0)}, \tag{57}$$

where function F(κ1, κ2, α) is defined by

$$F(\kappa_1,\kappa_2,\alpha) = (2\kappa_1\kappa_2)^{(1-\alpha)\kappa_2}\exp(2\kappa_1\kappa_2)\,\Gamma\big(1+(\alpha-1)\kappa_2, 2\kappa_1\kappa_2\big) + \kappa_1^{(1-\alpha)/2}\exp(\kappa_1)\,\gamma\left(\frac{\alpha+1}{2},\kappa_1\right) \tag{58}$$

and Γ (resp. γ) is the upper (resp. lower) incomplete gamma function defined by

$$\Gamma(s,y) = \int_y^\infty \xi^{s-1}\exp(-\xi)\,\mathrm{d}\xi, \qquad \gamma(s,y) = \int_0^y \xi^{s-1}\exp(-\xi)\,\mathrm{d}\xi.$$

Substituting α = 2 and α = 4 in equation (57), we get

$$\mathrm{Kurt}[U] = \frac{\langle U^4\rangle}{\langle U^2\rangle^2} = \frac{F(\kappa_1,\kappa_2,4)\,F(\kappa_1,\kappa_2,0)}{(F(\kappa_1,\kappa_2,2))^2}. \tag{59}$$

This formula for the kurtosis is visualized in Figure 2(a) as a function of parameter η5 for three different values of parameter η6. We note that the case η6 = 0 corresponds to the case studied in Section 4.1. If η6 = 0, then equation (56) implies κ1 = 0. Since γ(s, 0) = 0 and Γ(s, 0) = Γ(s), where Γ(s) is the standard gamma function given by (44), we can confirm that equation (59) converges to our previous result (46) as η6 → 0. Substituting α = 1 into (58), we obtain F(κ1, κ2, 1) = exp (κ1). Consequently, using α = 1 in equation (57), we obtain

$$\frac{\langle|U|\rangle}{\eta_6} = \frac{\exp(\kappa_1)}{F(\kappa_1,\kappa_2,0)}. \tag{60}$$

Using α = 2 in equation (57), we get

$$\frac{\langle U^2\rangle}{\langle|U|\rangle^2} = \frac{F(\kappa_1,\kappa_2,2)\,F(\kappa_1,\kappa_2,0)}{\exp(2\kappa_1)}. \tag{61}$$

Consequently, if we use MD simulations to estimate not only the second and fourth moments, hU 2 i and hU 4 i, but also the first absolute moment h|U|i, we can substitute the estimated MD values into equations (59) and (61) to obtain two equations for two unknowns κ1 and κ2. Solving these two equations numerically, we can get κ1 and κ2. Then we can use (56) and (60) to get the original parameters η5 and η6 by

$$\eta_5 = \frac{1-\kappa_2}{\kappa_2} \quad \text{and} \quad \eta_6 = \frac{\langle|U|\rangle\,F(\kappa_1,\kappa_2,0)}{\exp(\kappa_1)}. \tag{62}$$

Moreover, equation (56) also implies the following restriction on other parameters η2, η3 and η4

$$\frac{\eta_4^2}{\eta_2\eta_3} = \frac{1-\kappa_2}{\kappa_1\,\kappa_2\,\exp(\kappa_1/\kappa_2)}\Big(\langle|U|\rangle\,F(\kappa_1,\kappa_2,0)\Big)^{1/\kappa_2}. \tag{63}$$

This restriction is equivalent to restriction (47). Therefore, assuming again that D, hV 2 i, hZ 2 i are obtained from MD simulations and η 2 4/(2η2η3) is given by (63), we can calculate parameters η1, η2, η3 and η4 by equations (50)–(51). We note that the two additional parameters η5 and η6 can be used to satisfy both equations (59) and (61), while in Section 4.1 we could only use one equation (equation (46) for kurtosis) to fit one parameter η5. However, in the case of one-parameter function (37), we could (instead of fitting the kurtosis) match the quantity hU 2 i/h|U|i2 with MD simulations, i.e. we could replace equation (46) by equation (61) simplified to the one-parameter case corresponding to function (37). Passing to the limit η6 → 0 in equation (61) and using Euler's reflection formula, Γ(1 − y)Γ(y) sin(πy) = π, we obtain that the one-parameter nonlinearity (37) implies the following formula

$$\frac{\langle U^2\rangle}{\langle |U|\rangle^2} = \frac{\pi}{1+\eta_5}\left(\sin\left(\frac{\pi}{1+\eta_5}\right)\right)^{-1}. \qquad (64)$$

Thus, in Section 4.1, we could use h|U|i and hU 2 i estimated from long-time MD simulations to calculate the left hand side of equation (64), which could then be used to select parameter η5. Other parameters could again be chosen by equations (50)–(51).

5 Nolinear SCG model for general values of N

We have already observed in Sections 2 and 3 that the linear SCG model (6)– (9) can match the MD values of a few moments for N = 1, while we need to consider larger values of N to match the entire velocity autocorrelation function. Considering the nonlinear SCG model (2)–(5), we have two options to capture more details of the non-Gaussian force distribution observed in MD simulations. We could either keep N = 1, as in Section 4, and introduce additional parameters into nonlinearity g = g1, or we could consider larger values of N. In Section 4, we have shown that by going from one-parameter to two-parameter function g, we improve the match with MD results. In this section, we will discuss the second option: we will use larger values of N. Consider equations corresponding to the i-coordinate, i = 1, 2, 3, of the nonlinear SCG model (2)–(5). Let us denote the stationary distribution of equations (3)–(5) by

$$p(v,\mathbf{u},\mathbf{z}) \equiv p(v,u_1,u_2,\ldots,u_N,z_1,z_2,\ldots,z_N).$$

Then p(v, u, z) dv du1 du2 . . . duN dz1 dz2 . . . dzN gives the probability that Vi(t) ∈ [v, v + dv), Uj,i(t) ∈ [uj , uj + duj) and Zj,i(t) ∈ [zj , zj + dzj), for j = 1, 2, . . . , N, at equilibrium. The stationary distribution can be obtained by solving the corresponding stationary Fokker-Planck equation

$$\begin{aligned}\frac{\eta_{j,4}^2}{2}\frac{\partial^2 p}{\partial^2 z_j}(v,\mathbf{u},\mathbf{z}) &= \frac{\partial}{\partial v}\left(p(v,\mathbf{u},\mathbf{z})\sum_{j=1}^{N}u_j\right)\\ &+\sum_{j=1}^{N}\frac{\partial}{\partial u_j}\Big(\big(-\eta_{j,1}v+h_j(z_j)\big)g_j'(g_j^{-1}(u_j))\,p(v,\mathbf{u},\mathbf{z})\Big)\\ &+\sum_{j=1}^{N}\frac{\partial}{\partial z_j}\Big(\big(-\eta_{j,2}h_j(z_j)-\eta_{j,3}u_j\big)p(v,\mathbf{u},\mathbf{z})\Big). \qquad (65)\end{aligned}$$

Our analysis in Section 4.1 shows that parameters ηj,2, ηj,3 and ηj,4 appear on the left hand side of equation (47) as a suitable fraction, which in the Gaussian case corresponds to the second moment of the acceleration (see equation (48)). Considering general N, we define this fraction as new parameters.

$$\sigma_j = \frac{\eta_{j,4}^2}{2\,\eta_{j,2}\,\eta_{j,3}}, \qquad \text{for} \qquad j=1,2,\ldots,N,$$

and we again assume that the second moment of the velocity distribution, hV 2 i = hV 2 ii, can be estimated from long-time MD simulations. In order to find the stationary distribution, we will require that parameters ηj,1, ηj,2, ηj,3 and ηj,4 satisfy (compare with equation (49) for N =1)

$$\langle V^2\rangle = \frac{\eta_{j,4}^2}{2\,\eta_{j,1}\,\eta_{j,2}\,\eta_{j,3}} = \frac{\sigma_j}{\eta_{j,1}}, \qquad \text{for all} \qquad j=1,2,\ldots,N.$$

Then the stationary distribution, obtained by solving (65), is given by

$$p(v,\mathbf{u},\mathbf{z}) = C\left(\prod_{j=1}^{N}\frac{1}{g_j'(g_j^{-1}(u_j))}\right)\exp\left[-\frac{v^2}{2\langle V^2\rangle}-\sum_{j=1}^{N}\frac{1}{\sigma_j}G_j(g_j^{-1}(u_j))\right.$$
$$\left.-\sum_{j=1}^{N}\frac{2\eta_{j,2}}{\eta_{j,4}^2}H_j(z_j)\right], \qquad (66)$$

where C is the normalization constant and functions Gj and Hj are integrals of functions gj and hj , respectively, which are given by

$$G_j(y)=\int_0^y g_j(\xi)\,\mathrm{d}\xi\,, \qquad H_j(y)=\int_0^y h_j(\xi)\,\mathrm{d}\xi\,, \qquad \text{for} \quad j=1,2,\ldots,N.$$

Following (37), we assume that hj (zj) = zj and each gj is a function of one additional positive parameter ηj,5, j = 1, 2, . . . , N, given as

$$g_j(y)=|y|^{1/\eta_{j,5}}\,\mathrm{sign}\,y\,. \qquad (67)$$

Then we have,

$$g_j'(g_j^{-1}(u_j))=\frac{|u_j|^{1-\eta_{j,5}}}{\eta_{j,5}} \qquad \text{and} \qquad G_j(g_j^{-1}(u_j))=\frac{\eta_{j,5}}{1+\eta_{j,5}}|u_j|^{1+\eta_{j,5}}.$$

Then the stationary distribution (66) is Gaussian in Vi and Zj,i variables and we can integrate (66) to calculate the marginal distribution of Uj,i by

$$p_j(u_j)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}p(v,\mathbf{u},\mathbf{z})\,\mathrm{d}v\,\mathrm{d}u_1\,\mathrm{d}u_2\,\ldots\,\mathrm{d}u_{j-1}\,\mathrm{d}u_{j+1}\,\ldots\,\mathrm{d}u_N\,\mathrm{d}\mathbf{z}\,.$$

Consequently,

$$p_j(u_j)=C_j|u_j|^{\eta_{j,5}-1}\exp\left[-\frac{\eta_{j,5}}{\sigma_j(1+\eta_{j,5})}|u_j|^{1+\eta_{j,5}}\right], \qquad (68)$$

where the normalization constant Cj is given

$$\int_{-\infty}^{\infty}p_j(u_j)\,\mathrm{d}u_j=1.$$

Integrating (68), we can calculate

$$\langle|U_{j,i}|^{\alpha}\rangle=\int_{-\infty}^{\infty}|u_j|^{\alpha}p_j(u_j)\,\mathrm{d}u_j, \qquad \text{for any } \alpha\geq 0,$$

As

$$\langle|U_{j,i}|^{\alpha}\rangle=\left(\frac{\sigma_j(1+\eta_{j,5})}{\eta_{j,5}}\right)^{\alpha/(1+\eta_{j,5})}\frac{\Gamma\left(\frac{\alpha+\eta_{j,5}}{1+\eta_{j,5}}\right)}{\Gamma\left(\frac{\eta_{j,5}}{1+\eta_{j,5}}\right)}. \qquad (69)$$

The acceleration of the coarse-grained particle is given by

$$U_i=\sum_{j=1}^{N}U_{j,i}.$$

Using the symmetry of (68), odd moments of Uj,i are equal to zero. In particular, hUj,ii = 0 and hU 3 j,ii = 0 for j = 1, 2, . . . , N. Consequently,

$$\langle U_i^2\rangle=\sum_{j=1}^{N}\langle U_{j,i}^2\rangle\,, \qquad (70)$$

$$\langle U_i^4\rangle=3\langle U_i^2\rangle^2+\sum_{j=1}^{N}\langle U_{j,i}^4\rangle-3\langle U_{j,i}^2\rangle^2\,, \qquad (71)$$

which gives

$$\mathrm{Kurt}[U_i]=\frac{\langle U_i^4\rangle}{\langle U_i^2\rangle^2}=3+\frac{\sum_{j=1}^{N}\langle U_{j,i}^4\rangle-3\langle U_{j,i}^2\rangle^2}{\sum_{j=1}^{N}\langle U_{j,i}^2\rangle}. \qquad (72)$$

Substituting equation (69) for moments on the right hand side of equation (72), we can express the kurtosis of Ui in terms of 2N parameters σj and ηj,5, where j = 1, 2, . . . , N. For example, if we choose the values of dimensionless parameters ηj,5 equal to given numbers and define new parameters

$$\kappa_j=\left(\sigma_j\right)^{2/(1+\eta_{j,5})},$$

then equation (69) implies that hU 2 j,ii is a linear function of κj and hU 4 j,ii is a quadratic function of κj . Equations (70) and (71) can then be rewritten as the following system of two equations for κ1, κ2, . . . , κ

$$\sum_{i=1}^{N} c_{1,j}\kappa_j = \langle U_i^2\rangle, \qquad \sum_{i=1}^{N} c_{2,j}\kappa_j^2 = \langle U_i^4\rangle - 3\langle U_i^2\rangle^2,$$

where c1,j and c2,j are known constants, which will depend on our initial choice of values of ηj,5. Thus, using N > 2, we still have an opportunity to not only fit the second and fourth moments of the force distribution, but other moments as well. For example, the 6-th moment, hU 6 ii, would include the linear combination of the third powers of κj . We could also fit other properties of the force distribution estimated from MD simulations. For example, we could generalize one-parameter nonlinearities (67) to two-parameter nonlinear functions, as we did in equation (52). Then we could match the value of the distribution at u = 0, if our aim was to get a better fit of the MD acceleration distribution obtained in the illustrative example in Figure 2(b). Another possible generalization is to consider nonlinear functions hj , provided that we estimate more statistics on the auxiliary variable Z from MD simulations.

6 Discussion and conclusions

We have presented and analyzed a family of SCG models given by equations (2)–(5), which can be parametrized to fit properties of detailed all-atom MD models. A special choice of functions gj and hj in equations (2)–(5) leads to the linear SCG model (6)–(9) which is used in a multiscale (multi-resolution) method developed by Erban (2016) as an intermediate description between all-atom MD simulations and BD models. The linear SCG model is studied in more detail in Sections 2 and 3, where we highlight that 4N parameters of this model can match some statistics estimated from all-atom MD simulations with increased accuracy as we increase N, but there are also statistics which cannot be matched for any value of N. They include non-Gaussian force distributions.

In Sections 2 and 3, we show that the linear SCG model (6)–(9) corresponds to the generalized Langevin equation with the stochastic driving force being Gaussian. Such systems have been analysed since the work of Kubo (1966). One approach to match non-Gaussian MD force distributions could be to use the non-Gaussian generalized Langevin equation which was analyzed by Fox (1977) using methods of multiplicative stochastic processes. However, if we want to generalize the linear SCG model (6)–(9) while keeping its structure as a relatively low-dimensional system of SDEs, then it can be done by introducing nonlinear functions gj and hj as shown in equations (2)–(5). The advantage of the presented approach is that we can directly replace the linear model by equations (2)–(5) in multiscale methods which use all-atom MD simulations in parts of the computational domain and (less detailed) BD simulations in the remainder of the domain. Coupling MD and BD models is a possible approach to incorporate atomic-level information into models of intracellular processses which include transport of molecules between different parts of the cell (Erban, 2014, 2016; Gunaratne et al., 2019).

The nonlinear SCG model (2)–(5) is studied in Section 4 for N = 1. Describing the nonlinearity as the one-parameter function given by (37), we can use its dimensionless parameter η5 to match the kurtosis of the force distribution estimated from all-atom MD simulations. Although the one-parameter case is easy to analyze in terms of the gamma function, it has some undesirable properties for small forces. If η5 > 1, we can obtain large terms in the dynamical equation (41) for small values of U; this corresponds to the zero value of stationary probability distribution (43) for u = 0. If η5 < 1, we have small terms in the dynamical equation (41), but the stationary probability distribution (43) is unbounded for u = 0. In Section 4.2, we show that these issues can be avoided if the two-parameter nonlinear function (52) is used instead of the one-parameter function (37). The resulting equations are solved in terms of incomplete

gamma functions. In Section 5, we study the nonlinear model for general values of N where each gj is a one-parameter nonlinearity given by equation (67). However, we could also consider two-parameter functions gj , like we did in equation (52) for N = 1, to improve the properties of the SCG model for general values of N.

Acknowledgements.

I would like to thank the Royal Society for a University Research Fellowship.

References

Carof A, Vuilleumier R, Rotenberg B (2014) Two algorithms to compute projected correlation functions in molecular dynamics simulations. Journal of Chemical Physics 140(12):124103

Davtyan A, Dama J, Voth G, Andersen H (2015) Dynamic force matching: A method for constructing dynamical coarse-grained models with realistic time dependence. Journal of Chemical Physics 142:154104

Davtyan A, Voth G, Andersen H (2016) Dynamic force matching: Construction of dynamic coarse-grained models with realistic short time dynamics and accurate long time dynamics. Journal of Chemical Physics 145:224107

Dobramysl U, R¨udiger S, Erban R (2016) Particle-based multiscale modeling of calcium puff dynamics. Multiscale Modelling and Simulation 14(3):997–1016

Erban R (2014) From molecular dynamics to Brownian dynamics. Proceedings of the Royal Society A 470:20140036

Erban R (2016) Coupling all-atom molecular dynamics simulations of ions in water with Brownian dynamics. Proceedings of the Royal Society A 472:20150556

Erban R, Chapman SJ (2009) Stochastic modelling of reaction-diffusion processes: algorithms for bimolecular reactions. Physical Biology 6(4):046001

Erban R, Chapman SJ (2019) Stochastic Modelling of Reaction-Diffusion Processes. Cambridge Texts in Applied Mathematics. ISBN 9781108498128. Cambridge University Press

Farafonov V, Nerukh D (2019) MS2 bacteriophage capsid studied using all-atom molecular dynamics. Interface Focus 9:20180081

Flegg M, Chapman SJ, Erban R (2012) The two-regime method for optimizing stochastic reaction-diffusion simulations. Journal of the Royal Society Interface 9(70):859–868 Flegg M, Chapman SJ, Zheng L, Erban R (2014) Analysis of the two-regime method on square meshes. SIAM Journal on Scientific Computing 36(3):B561–B588

Flegg M, Hellander S, Erban R (2015) Convergence of methods for coupling of microscopic and mesoscopic reaction-diffusion simulations. Journal of Computational Physics 289:1– 17

Fox R (1977) Analysis of nonstationary, Gaussian and non-Gaussian, generalized Langevin equations using methods of multiplicative stochastic processes. Journal of Statistical Physics 16(3):259–279

Gottwald F, Karsten S, Ivanov S, K¨uhn O (2015) Parametrizing linear generalized Langevin dynamics from explicit molecular dynamics simulations. Journal of Chemical Physics 142:244110

Gunaratne R, Wilson D, Flegg M, Erban R (2019) Multi-resolution dimer models in heat baths with short-range and long-range interactions. Interface Focus 9:20180070

Hamada K, Miyatake H, Terauchi A, Mikoshiba K (2017) IP3-mediated gating mechanism of the IP3 receptor revealed by mutagenesis and X-ray crystallography. Proceedings of the National Academy of Sciences 114(18):4661–4666

Hoover W (1985) Canonical dynamics: Equilibrium phase-space distributions. Physical Review E 31(3):1695–1697

Jung G, Hanke M, Schmid F (2017) Iterative reconstruction of memory kernels. Journal of Chemical Theory and Computation 13:2481–2488

Kang M, Othmer H (2007) The variety of cytosolic calcium responses and possible roles of PLC and PKC. Physical Biology 4:325–343

Kang M, Othmer H (2009) Spatiotemporal characteristics of calcium dynamics in astrocytes. Chaos 19:037116

Kubo R (1966) The fluctuation-dissipation theorem. Reports on Progress in Physics 29:255– 284

Leimkuhler B, Matthews C (2015) Molecular Dynamics, Interdisciplinary Applied Mathematics, vol 39. Springer