Highlighting research performed in the group of Prof. Wenxiao Pan from the Department of Mechanical Engineering at the University of Wisconsin-Madison.

Implicit-solvent coarse-grained modeling for polymer solutions *via* Mori-Zwanzig formalism

Based on the Mori-Zwanzig projection, atomistic representations of polymers in solution are projected to a reduced-dimension space of coarse-grained coordinates while the dynamic properties of polymers are conserved. Provided significantly reduced degrees of freedom and larger timesteps, the established implicit-solvent coarse-grained model is much more efficient than full atomistic simulations, and hence, grants larger accessible length scales and renders tractable simulating long-time effects.

ROYAL SOCIETY OF **CHEMISTRY** | Celebrating IYPT 2019

rsc.li/soft-matter-journal

Registered charity number: 207890

Soft Matter

Implicit-solvent coarse-grained modeling for polymer solutions *via* Mori-Zwanzig formalism

Shu Wang,[a] Zhen Li [ID][b] and Wenxiao Pan [ID]*[a]

We present a bottom-up coarse-graining (CG) method to establish implicit-solvent CG modeling for polymers in solution, which conserves the dynamic properties of the reference microscopic system. In particular, tens to hundreds of bonded polymer atoms (or Lennard-Jones beads) are coarse-grained as one CG particle, and the solvent degrees of freedom are eliminated. The dynamics of the CG system is governed by the generalized Langevin equation (GLE) derived *via* the Mori-Zwanzig formalism, by which the CG variables can be directly and rigorously linked to the microscopic dynamics generated by molecular dynamics (MD) simulations. The solvent-mediated dynamics of polymers is modeled by the non-Markovian stochastic dynamics in GLE, where the memory kernel can be computed from the MD trajectories. To circumvent the difficulty in direct evaluation of the memory term and generation of colored noise, we exploit the equivalence between the non-Markovian dynamics and Markovian dynamics in an extended space. To this end, the CG system is supplemented with auxiliary variables that are coupled linearly to the momentum and among themselves, subject to uncorrelated Gaussian white noise. A high-order time-integration scheme is used to solve the extended dynamics to further accelerate the CG simulations. To assess, validate, and demonstrate the established implicit-solvent CG modeling, we have applied it to study four different types of polymers in solution. The dynamic properties of polymers characterized by the velocity autocorrelation function, diffusion coefficient, and mean square displacement as functions of time are evaluated in both CG and MD simulations. Results show that the extended dynamics with auxiliary variables can construct arbitrarily high-order CG models to reproduce dynamic properties of the reference microscopic system and to characterize long-time dynamics of polymers in solution.

1 Introduction

For polymers or macromolecules in solution, atomistic simulation techniques, such as all-atom molecular dynamics (MD), have been well established to precisely represent molecular structures and to accurately predict static and dynamic properties of the system by tracking individual atoms of molecules and solvent. In practical applications,[1–4] first-principles MD simulations can be computationally prohibitive to simulate large-scale polymer solution systems and/or to capture long-time effects. If only the mesoscopic properties and collective dynamics of polymers are of particular interest, there is no need to simulate all the atomistic details of the system, and some degrees of freedom (DOFs) may be eliminated appropriately for saving computational cost. To this end, coarse-grained (CG) modeling, which represents a group of atoms as a single CG particle, becomes attractive.[5–9]

Owing to the larger characteristic length scale of the CG system, larger time steps are permitted in CG simulations. If the solvent DOFs are further eliminated, it leads to the so-called implicit-solvent CG modeling.[10–13] Provided significantly reduced DOFs and larger time steps, the implicit-solvent CG modeling can be much more efficient than full atomistic simulations, and hence, grants larger accessible length scales and renders tractable simulating long-time effects.

Numerous methods have been developed for CG modeling to correctly capture the static properties (*e.g.*, pressure, compressibility, radial distribution function) of polymers in solution, including iterative Boltzmann inversion,[14] inverse Monte Carlo,[15] force matching method,[16] minimization of relative entropy,[17,18] *etc.* In the present work, our focus is on the dynamic properties, *e.g.*, the velocity autocorrelation function (VACF) and diffusivity. In a canonical ensemble, thermostats are required in CG modeling to maintain a constant temperature (on average). In general, the dynamic properties of a system relate closely to the applied thermostat, which is associated with a dissipative term and a random term correlated by the fluctuation-dissipation theorem (FDT).[19] For polymers in solution, the dynamics of polymers not

a Department of Mechanical Engineering, University of Wisconsin–Madison, Madison, WI 53706, USA. E-mail: wpan9@wisc.edu

b Department of Mechanical Engineering, Clemson University, Clemson, SC 29634, USA

only depends on the interactions between polymer molecules but also the interactions with the solvent.

Explicitly coarse-graining solvent by clustering multiple molecules into a CG representation can be challenging because the solvent molecules are not interconnected and can move apart in time, and hence, no simple mapping relates their atomistic and CG configurations.[20–23] Thus, complicated dynamical clustering methods, such as the K-means clustering algorithm,[24] SWINGER algorithm,[25] trajectory correspondence clustering scheme,[26] and position-dependent CG mapping,[27] have to be applied. A popular explicit-solvent CG modeling method is dissipative particle dynamics (DPD), which has been applied for modeling polymers or lipid/cell membranes in solution.[28–31] In DPD, the dissipative and random forces are not rigorously derived from the microscopic data but rather calibrated through a post-processing or optimization procedure to match a target dynamic property.[32] As a result, correctly capturing other dynamic properties beyond the target one cannot be guaranteed.[32–34] Another issue of explicit-solvent CG modeling is its computational cost. For modeling dilute solutions, the simulations with many solvent DOFs can be still costly.

Therefore, implicit-solvent CG modeling can be an attractive alternative, which eliminates solvent DOFs and implicitly incorporates the solvent-mediated effects on polymers. In fact, the interactions with solvent can lead to strong memory effects in the dynamics of polymers. Thus, elimination of solvent DOFs results in a non-Markovian memory in the equation of motion (EOM) of the CG variables. However, in some efforts of implicit-solvent CG modeling for polymer solutions, the dissipative and random terms were assumed Markovian; *i.e.*, no memory effect was considered in the CG dynamics. As a result, such CG modeling can only reproduce certain averaged dynamic properties, *e.g.*, the average diffusion coefficient, by fine-tuning or optimizing the parameters in the dissipative and random terms.[13,35,36] With Markovian approximation in the implicit-solvent CG modeling, the VACF of CG particles follows strictly an exponential decay, which is unlikely able to reproduce the entire complex transition dynamics and to characterize long-time behaviors of polymers in solution. The non-Markovian memory in implicit-solvent CG modeling can play an important role in producing correct long-time dynamics.[37] To this end, the Mori-Zwanzig formalism[38–40] provides a forward path to construct implicit-solvent CG models directly from microscopic dynamics in a bottom-up fashion. The dissipative and random terms in the resulting generalized Langevin equation (GLE) can be directly constructed and evaluated from the microscopic dynamics by mapping the microscopic system to the CG system using the Mori-Zwanzig projection. Therefore, we employ the GLE derived *via* the Mori-Zwanzig formalism to govern the dynamics of the CG system. We note that in literature, Lyubimov *et al.* derived from the GLE the analytical factor for dynamical rescaling of the friction coefficient in CG modeling to correctly capture the long-time diffusion of polymers, which is applicable to melt systems.[41,42] The derived rescaling factor is transferable for different polymer systems and thermodynamic conditions, with the temperature and radius-of-gyration as the input parameters.[41,42] Jung *et al.* employed the GLE with iteratively reconstructed

non-Markovian memory kernels to reproduce the VACF of nanocolloids in dilute solution,[43,44] where the self-memory of a single colloid immersed in solvent was augmented with pairwise correlations.[44] Davtyan *et al.*[45] developed a dynamic force matching technique to study a system that consists of a Lennard-Jones (LJ) particle dissolved in a LJ solvent. To reproduce the non-Markovian dynamics of the LJ particle, the implicit-solvent CG modeling was supplemented by a set of fictitious particles with harmonic interactions among themselves and a special coupling to the LJ particle.

The non-Markovian memory term in GLE involves a convolution of the memory kernel with the velocities of CG particles. Direct evaluation of this memory term can be expensive because it requires to record the history of the CG variables at every time step and to numerically solve the convolution. In the meanwhile, colored noise must be generated for the random term to satisfy the second FDT that ensures correct equilibrium statistics in the system. Instead of direct evaluation of the memory term and generating colored noise, an alternative approach is considered in this work. The non-Markovian dynamics described by the GLE is mapped to a Markovian process extended in a higher dimensional space. To this end, the memory kernel is approximated by the exponentially decaying oscillatory functions, and auxiliary variables are introduced and coupled to the momenta of CG particles. With the second FDT still satisfied, only inexpensive white-noise terms are needed in the extended dynamics. In so doing, no non-Markovian memory term needs to be evaluated, and no colored noise needs to be sampled. By such, this approach renders tractable the practical implementations of the proposed CG modeling for polymer solutions in real applications. We note that this approach of mapping the non-Markovian dynamics to an extended Markovian process using auxiliary variables has been applied in others' work to treat the memory term in GLE.[46–50] The present work represents a different attempt to apply it for implicit-solvent CG modeling of polymers in solution. An additional benefit of employing the extended dynamics with auxiliary variables is that we can control the accuracy of CG modeling by the number of auxiliary variables used and hence achieve the desired tradeoff between accuracy and computational efficiency.

The extended dynamics was previously solved by the velocity-Verlet temporal integrator.[50,51] In this work, we introduce a higher-order integrator scheme for solving it, which is based on the G-JF integrator originally derived by Grønbech-Jensen and Farago for solving Langevin dynamics in MD simulations.[52,53] By numerical experiments, we compared the performance of the G-JF and velocity-Verlet schemes with respect to both accuracy and efficiency. The proposed implicit-solvent CG modeling was applied to study four types of polymer solution systems, including homogeneous star polymers, inhomogeneous mixture of non-monosized star polymers, and branched-chain polymers in solution and also tri-*n*-butyl phos-phate polymers in chloroform. In each system, we computed the dynamic properties of polymers, including the VACF, diffusion coefficient, and mean-square displacement (MSD) as functions of time in a wide spectrum of time scales. The predictions by the CG simulations were compared with the results of reference MD simulations.

The rest of the paper is organized as follows. In Section 2, we first describe the formulation of GLE derived from the Mori-Zwanzig formalism in Section 2.1, next give the equation of the extended dynamics in Section 2.2, and then provide in Section 2.3 the schemes of two temporal integrators, the velocity-Verlet and G-JF, for solving the extended dynamics. Section 3 presents the results of simulating four different types of polymer solution systems employing the proposed implicit-solvent CG modeling. The dynamic properties of polymers predicted by the CG modeling are compared with the reference MD simulations. The accuracy and efficiency of the two temporal integrators are compared in Section 3.1. Finally, we conclude and summarize our main findings and contributions in Section 4.

2 Formulation

2.1 GLE *via* Mori-Zwanzig formalism

Mori-Zwanzig projection can extract reduced-dimension variables from a full-dimension system.[38,40] Here, the full-dimension system corresponds to the (all-atom) microscopic system, which is the MD system in this work; the reduced-dimension system corresponds to the CG system. Consider a MD system that contains n atoms with coordinates r_i and momenta p_i, $i = 1, 2,\ldots,n$. In its corresponding CG system, these n atoms are coarse-grained as N clusters (referred to as CG particles), in which each cluster contains n_c atoms. The variables associated in the CG setting include the position $\mathbf{R}$ and momentum $\mathbf{P}$ of the center-of-mass (COM) of each CG particle, which can be defined as:

$$\begin{aligned} \mathbf{R}_I &= \frac{1}{M_I}\sum_{i=1}^{n_c} m_{Ii}\mathbf{r}_{Ii} \\ \mathbf{P}_I &= \sum_{i=1}^{n_c} \mathbf{p}_{Ii}, \end{aligned} \quad (1)$$

with $M_I = \sum_{i=1}^{n_c} m_{Ii}$ the mass of the I-th CG particle and m_{Ii} the mass of the i-th atom in the I-th CG particle. To be consistent in notation, we use the lowercase m, $\mathbf{r}$, and $\mathbf{p}$ to represent the mass, position, and momentum of an atom, while the uppercase M, $\mathbf{R}$, and $\mathbf{P}$ denote the mass, position, and momentum of a CG particle.

Following the Mori-Zwanzig formalism,[39,40] the EOM of CG particles can be derived in the form of GLE as:

$$\begin{aligned} \mathbf{P}_I &= \langle \mathbf{F}_I \rangle - \beta \sum_{J=1}^{N} \int_0^t \mathrm{d}t' \left\langle \left[\delta\mathbf{F}_I^Q(t-t') \right] \left[\delta\mathbf{F}_J^Q(0) \right]^{\mathrm{T}} \right\rangle \\ &\quad \times \frac{\mathbf{P}_J(t')}{\mathrm{M}_J} + \delta\mathbf{F}_I^Q(t), \end{aligned} \quad (2)$$

with the inverse temperature $\beta = 1/(k_\mathrm{B}T)$, Boltzmann constant k_B, and thermodynamic temperature T. On the right-hand side of eqn (2), the first term represents the ensemble average force on the I-th CG particle: $\langle \mathbf{F}_I \rangle = \frac{1}{\beta}\frac{\partial}{\partial \mathbf{R}_I}\ln\omega(\mathbf{R})$, where $\mathbf{R} = \{\mathbf{R}_1, \mathbf{R}_2,\ldots,\mathbf{R}_N\}$ is a point in the CG phase space, and $\omega(\mathbf{R})$ denotes a normalized partition function of all the microscopic configurations at phase point $\mathbf{R}$. As the present work concerns the

dynamic properties and diffusion process and does not consider the structural properties or free energy, $\langle F_I \rangle$ is regarded as the average for CG particles over all phase points. Thus, without external force fields, the mean force exerted on a CG particle is approximated to be zero; *i.e.*, $\langle \mathbf{F}_I \rangle = 0$. The other two terms on the right-hand side of eqn (2) correspond to the dissipative and random forces, respectively, which compensate for the lost DOFs as a consequence of coarse-graining.[39,40] The random force $\delta\mathbf{F}_I^Q(t) = \mathrm{e}^{-iQLt}\delta\mathbf{F}_I$, where $\delta\mathbf{F}_I = \mathbf{F}_I - \langle\mathbf{F}_I\rangle$ with $\mathbf{F}_I = \dot{\mathbf{P}}_I$; Q is the orthogonal operator in Mori-Zwanzig projection, and L the Liouville operator.[54]

In eqn (2), the dissipative force is determined by time convolution of the memory kernel and momenta of CG particles. The memory kernel is defined as the autocorrelation function of random forces; *i.e.*, $K_{IJ}(t) = \beta\langle[\delta\mathbf{F}_I^Q(t)][\delta\mathbf{F}_J^Q(0)]^{\mathrm{T}}\rangle$. This definition ensures the second FDT is satisfied.[19] Direct evaluation of the dissipative force is challenging since it depends on the random forces and momenta of all CG particles. Assume there is no correlation between the random forces on different CG particles;[50] *i.e.*, $\beta\langle[\delta\mathbf{F}_I^Q(t)][\delta\mathbf{F}_J^Q(0)]^{\mathrm{T}}\rangle = \delta_{IJ}K(t)$. Eqn (2) can then be simplified to:

$$\dot{\mathbf{P}}_I = -\int_0^t K(t-t')\mathbf{V}_I(t')\mathrm{d}t' + \delta\mathbf{F}_I^Q(t), \quad (3)$$

where $\mathbf{V}_I(t') = \frac{\mathbf{P}_I(t')}{M_I}$. To determine the memory kernel $K(t)$, we rely on the property that the velocity $\mathbf{V}$ and random force $\delta\mathbf{F}^Q$ come from two orthogonal subspaces and hence are not correlated to each other; *i.e.*, $\langle\delta\mathbf{F}^Q\mathbf{V}^{\mathrm{T}}\rangle = 0$. Thus, multiplying both sides of eqn (3) by $\mathbf{V}(0)^{\mathrm{T}}$ leads to:

$$\langle\mathbf{F}(t)\mathbf{V}(0)^T\rangle = -\int_0^t K(t-t')\langle\mathbf{V}(t')\mathbf{V}(0)^T\rangle\mathrm{d}t', \quad (4)$$

where $\langle\mathbf{F}(t)\mathbf{V}(0)^{\mathrm{T}}\rangle$ defines the force-velocity correlation function (FVCF); the VACF is defined as $\langle\mathbf{V}(t)\mathbf{V}(0)^{\mathrm{T}}\rangle$. These two correlation functions can be directly evaluated in the reference MD system. In turn, the memory kernel $K(t)$ can be computed *via* deconvolution of eqn (4). In practice, the deconvolution was done numerically. By discretizing the integral in eqn (4) using the midpoint quadrature rule, the discrete form of eqn (4) in matrix notation reads:

$$\begin{pmatrix} \mathrm{FVCF}^1 \\ \mathrm{FVCF}^2 \\ \vdots \\ \mathrm{FVCF}^n \end{pmatrix}$$

$$= -\Delta t \begin{pmatrix} \mathrm{VACF}^{\frac{1}{2}} & 0 & \cdots & 0 \\ \mathrm{VACF}^{1+\frac{1}{2}} & \mathrm{VACF}^{\frac{1}{2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{VACF}^{n-\frac{1}{2}} & \mathrm{VACF}^{n-1-\frac{1}{2}} & \cdots & \mathrm{VACF}^{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} K^{\frac{1}{2}} \\ K^{1+\frac{1}{2}} \\ \vdots \\ K^{n-\frac{1}{2}} \end{pmatrix},$$

$$(5)$$

where superscripts denote the values at different discrete times, e.g., $FVCF^n = FVCF(n\Delta t)$. By solving this linear system, we obtained K at discrete times. $K(t)$ at $t = 0$ and other integer time steps can be determined by linear or second-order extrapolation or interpolation. If the data of VACF and FVCF are noisy, regularization can be employed, e.g., via truncated SVD or Tikhonov regularization.[55] However, in this work, we took sufficient ensemble averages for the VACF and FVCF until they were smooth enough, and hence, no regularization was employed.

Given the memory kernel $K(t)$ determined, the GLE defined by eqn (3) is closed. Directly solving this equation requires to evaluate the time convolution of the memory kernel and velocity and to generate color noise for the random force, which needs to store the historical information and can be prohibitively expensive. One way to reduce the computational cost is to truncate the memory kernel to a much shorter time scale. However, this truncation strategy may not be effective for the systems studied herein because the solvent-mediated kinetics could result in long-tailed memory kernels.

2.2 Extended dynamics

To circumvent the difficulty of directly solving eqn (3), we followed the literature[46,50] and introduced auxiliary variables $\mathbf{s} = \{s_{l1}, s_{l2}, s_{l3}, s_{l4}, s_{l5}, s_{l6}\}_{l=1,\ldots,\mathcal{N}}$ to replace the GLE (with non-Markovian terms) in the form of eqn (3) with equivalent Markovian Langevin equations extended in higher dimensions as given by:

$$\begin{pmatrix} \dot{\mathbf{P}} \\ \dot{\mathbf{s}} \end{pmatrix} = -\begin{pmatrix} 0 & \mathbf{A}_{ps} \\ \mathbf{A}_{sp} & \mathbf{A}_{ss} \end{pmatrix}\begin{pmatrix} \mathbf{P} \\ \mathbf{s} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{B}_s \end{pmatrix}\begin{pmatrix} 0 \\ \xi \end{pmatrix}. \quad (6)$$

In eqn (6), $\mathbf{A}_{ps} = -\mathbf{A}_{sp}^T$; ξ is a vector of uncorrelated Gaussian random variables with $\langle \xi(t) \rangle = \mathbf{0}$ and $\langle \xi_{I,\mu}(t)\xi_{J,\nu}(0) \rangle = \delta_{IJ}\delta_{\mu\nu}\delta(t)$, where ξ_ν and ξ_μ denote the different elements of ξ. The parameter matrices $\mathbf{A} = [0, \mathbf{A}_{ps}; \mathbf{A}_{sp}, \mathbf{A}_{ss}]$ and $\mathbf{B} = \text{diag}(0, \mathbf{B}_s)$.

From eqn (6), we can get:

$$\dot{\mathbf{P}} = \mathbf{A}_{ps}\mathbf{s}$$

$$\mathbf{s} = \int_{-\infty}^t e^{-(t-t')\mathbf{A}_{ss}}\left[-\mathbf{A}_{sp}\mathbf{P}(t') + \mathbf{B}_s\xi(t') \right]dt'. \quad (7)$$

Comparing eqn (7) with eqn (3), the memory kernel $K(t)$ and random force $\delta F^Q(t)$ can be expressed as:

$$K(t) = -M\mathbf{A}_{ps}e^{t\mathbf{A}_{ss}}\mathbf{A}_{sp}, \quad (8)$$

$$\delta\mathbf{F}^Q(t) = -\int_{-\infty}^t \mathbf{A}_{ps}e^{-(t-t')\mathbf{A}_{ss}}\mathbf{B}_s\xi(t')dt'. \quad (9)$$

To satisfy the second FDT, $\mathbf{B}_s$ is given by

$$\mathbf{B}_s\mathbf{B}_s^T = k_B T(\mathbf{A}_{ss} + \mathbf{A}_{ss}^T)M. \quad (10)$$

In eqn (10), $\mathbf{B}_s$ is a real matrix only if $(\mathbf{A}_{ss} + \mathbf{A}_{ss}^T)$ is positive-semidefinite and can be determined via Cholesky factorization.

A general, reasonable approximation for the memory kernel may be an expansion of exponentially damped oscillators; i.e.,

$$K(t) \approx \sum_{l=1}^{\mathcal{N}} \exp\left(-\frac{a_l}{2}t\right)[b_l \cos(w_l t) + c_l \sin(w_l t)]. \quad (11)$$

Thus, the drift matrix $\mathbf{A}$ in eqn (8) must be a real matrix that has complex eigenvalues with positive real parts. In eqn (11), $\mathcal{N}$ is the total number of oscillators truncated to approximate $K(t)$. Correspondingly, the dimension of $\mathbf{s}$ is $6\mathcal{N}$. Using more terms of exponentially damped oscillators, $K(t)$ can be approximated more accurately, leading to a higher dimensional extended dynamics. In eqn (11), a_l, b_l, c_l, and w_l are the parameters in each oscillator and comprise the elements of $\mathbf{A}$. Assembling these parameters determines the matrix $\mathbf{A}$ as:[50]

$$\mathbf{A}_l = \begin{bmatrix} 0 & \sqrt{\frac{b_l}{2} - \frac{\omega_l c_l}{a_l}} & \sqrt{\frac{b_l}{2} + \frac{\omega_l c_l}{a_l}} \\ -\sqrt{\frac{b_l}{2} - \frac{\omega_l c_l}{a_l}} & a_l & \frac{1}{2}\sqrt{4\omega_l^2 + a_l^2} \\ -\sqrt{\frac{b_l}{2} + \frac{\omega_l c_l}{a_l}} & -\frac{1}{2}\sqrt{4\omega_l^2 + a_l^2} & 0 \end{bmatrix}$$

$$(12)$$

In eqn (12), the top right block corresponds to $\mathbf{A}_{ps}$; the bottom left is $\mathbf{A}_{sp}$; and, $\mathbf{A}_{ss}$ is on the bottom right. We note that eqn (12) can be regarded as a specification of the drift matrix $\mathbf{A}$ used in the work of Ceriotti et al.,[46] which reduces the total number of unknown parameters used to approximate the memory kernel and accelerates the optimization process to determine the unknown parameters.

2.3 Temporal integrators

Given eqn (6) determined, we numerically solved it to generate trajectories of the CG system. To this end, two temporal integrators were considered and compared.

The first one is the so-called velocity-Verlet scheme, which has previously been used for solving the extended dynamics.[50,51] By denoting discrete times with the integer time step superscript, such as $\mathbf{R}^n = \mathbf{R}(t_n)$, the velocity-Verlet scheme modified for solving eqn (6) reads:

$$\mathbf{U}^{n+\frac{1}{2}} = \mathbf{U}^n + \mathbf{M}^{-1}(-\mathbf{A}\mathbf{M}\mathbf{U}^n + \mathbf{B}\xi^n)\frac{\Delta t}{2}$$

$$\mathbf{R}^{n+1} = \mathbf{R}^n + \mathbf{V}^{n+\frac{1}{2}}\Delta t \quad (13)$$

$$\mathbf{U}^{n+1} = \mathbf{U}^{n+\frac{1}{2}} + \mathbf{M}^{-1}\left(-\mathbf{A}\mathbf{M}\mathbf{U}^{n+\frac{1}{2}} + \mathbf{B}\xi^{n+1}\right)\frac{\Delta t}{2}.$$

Here, the vector $\mathbf{R}$ denotes the position of a CG particle; the vector $\mathbf{U}$ combines the velocity and auxiliary variables: $\mathbf{U} = [\mathbf{V}, \mathbf{s}]^T$; the matrix $\mathbf{M}$ consists of the CG particle's mass and can be written as: $\mathbf{M} = \text{diag}\{M, 1,\ldots,1\}$; $\mathbf{A}$ and $\mathbf{B}$ are the parameter matrices in eqn (6); ξ^n is a vector of uncorrelated Gaussian random numbers with $\langle \xi^n \rangle = 0$ and $\langle \xi^n \xi^{n+1} \rangle = \delta_{n,n+1}\Delta t$. The velocity-Verlet scheme approximates the delta-function correlated noise $\xi(t)$ with a set of rectangular pulses of mean-squared size

$\sqrt{1/\Delta t}$, each of which acting over the time interval $(t_n - \Delta t/2, t_n + \Delta t/2)$.[52] Employing this discretized approximation for $\xi(t)$ yields $\mathcal{O}(\Delta t)$ error in the simulated statistical quantities, such as temperature.[52] Thus, to ensure accuracy and stability, the velocity-Verlet scheme requires small timestep size Δt.

The second temporal integrator is the G-JF scheme, derived by Grønbech-Jensen and Farago (G-JF) originally for Langevin dynamics simulations.[52,53] We adapted the G-JF scheme for solving the extended dynamics in eqn (6) as below:

$$\mathbf{U}^{\frac{1}{2}} = \mathbf{b}\mathbf{U}^0 + \frac{1}{2}\mathbf{b}\mathbf{M}^{-1}\mathbf{B}\boldsymbol{\xi}^1$$

$$\mathbf{U}^{n+\frac{1}{2}} = \mathbf{a}\mathbf{U}^{n-\frac{1}{2}} + \frac{1}{2}\mathbf{b}\mathbf{M}^{-1}\mathbf{B}\left(\boldsymbol{\xi}^n + \boldsymbol{\xi}^{n+1}\right)$$

$$\mathbf{R}^{n+1} = \mathbf{R}^n + \mathbf{V}^{n+\frac{1}{2}}\Delta t \tag{14}$$

$$\mathbf{U}^n = \frac{1}{2}\mathbf{b}^{-1}\left(\mathbf{U}^{n+\frac{1}{2}} + \mathbf{a}\mathbf{U}^{n-\frac{1}{2}}\right) + \frac{1}{4}\mathbf{M}^{-1}\mathbf{B}\left(\boldsymbol{\xi}^n - \boldsymbol{\xi}^{n+1}\right),$$

where $\mathbf{a}$ and $\mathbf{b}$ are two constant matrices defined as:

$$\mathbf{b} = \left[\mathbf{I} + \frac{1}{2}\mathbf{M}^{-1}\mathbf{A}\mathbf{M}\Delta t\right]^{-1}$$

$$\mathbf{a} = \mathbf{I} - \mathbf{M}^{-1}\mathbf{A}\mathbf{M}\mathbf{b}\Delta t. \tag{15}$$

The G-JF scheme implies $\boldsymbol{\xi}^{n+1} \equiv \int_{t_n}^{t_{n+1}}\boldsymbol{\xi}(t')\mathrm{d}t'$, which does not introduce discretized approximation for the random variable $\xi(t)$.[52] The only approximation made in eqn (14) is employing the trapezoidal quadrature rule to approximate:

$$\int_{t_n}^{t_{n+1}}\dot{\mathbf{R}}\mathrm{d}t' = \mathbf{R}^{n+1} - \mathbf{R}^n = \int_{t_n}^{t_{n+1}}\mathbf{V}\mathrm{d}t', \tag{16}$$

with

$$\mathbf{R}^{n+1} - \mathbf{R}^n \approx \left(\mathbf{V}^{n+1} + \mathbf{V}^n\right)\frac{\Delta t}{2}, \tag{17}$$

which introduces an error of $\mathcal{O}(\Delta t^3)$ and controls the accuracy of the G-JF scheme. The G-JF scheme was previously shown more stable and accurate than the velocity-Verlet scheme for solving the GLE.[43,44] In this work, we compared the performance of these two schemes for solving the extended dynamics (see Section 3.1.2).

3 Results

We examined four different types of polymer solution systems to demonstrate the proposed implicit-solvent CG modeling and to assess its ability to conserve the dynamic properties of the reference atomistic systems. The systems studied include homogeneous or inhomogeneous star polymers and branched polymers in solution. And both model and real polymers were considered. For each system, we first performed MD simulations and then constructed the CG model following the formulation described in Section 2.1 and 2.2. The extended dynamics of GLE was solved using the velocity-Verlet or G-JF temporal integrator described in Section 2.3. For each system,

the dynamic properties (VACF and diffusion) of polymers predicted by the CG modeling were compared with the results of MD simulations. In-house computer codes were developed to implement the extended dynamics and temporal integrators for the CG modeling in the framework of LAMMPS.[56] By such, both MD and CG simulations were performed using LAMMPS.

3.1 Homogeneous solution of star polymers

First, we examined a system composed of homogeneous star polymers in solution. In the CG modeling of this system, we compared the accuracy and efficiency of the velocity-Verlet and G-JF temporal integrators for solving the extended dynamics.

3.1.1 Microscopic model.
Each star polymer consists of 10 identical arms each with 3 monomers, as illustrated in Fig. 1. In the microscopic model, the monomers are represented as Lennard-Jones (LJ) beads connected by FENE bonds; the solvent is composed of LJ beads identical to the monomers. The details of the MD force fields are provided in Appendix A. The microscopic representation of this polymer solution system is depicted in Fig. 2(a). Since each star polymer has totally 31 monomers, it's referred to as N31 star polymer. In total, there are 1000 N31 star polymers and 5000 solvent beads filled into a periodic cubic box of length 37.188σ with the number density of 0.7, defined as the total number of LJ beads divided by the volume of the cubic box. The system size was chosen large enough such that the finite size effect on the VACF can be neglected.[57]

With this microscopic model, MD simulations were performed under the canonical ensemble (NVT) using Nośe–Hoover thermostat with $k_BT = 1.0$ and the time step $\Delta t = 0.001\tau$. All values herein
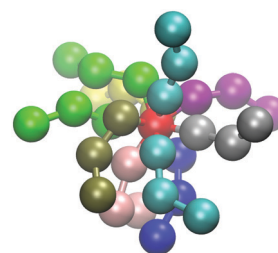

Fig. 1 Microscopic model of a star polymer with 10 identical arms and 3 monomers per arm.

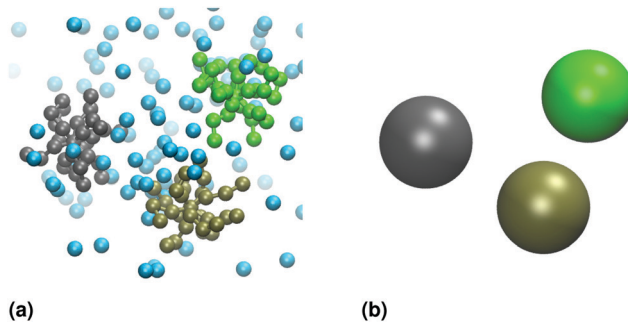(a)   (b)

Fig. 2 (a) Microscopic model of N31 star polymers in solvent; (b) CG model, where each N31 star polymer is coarse-grained as a single CG particle, and the solvent DOFs are removed.
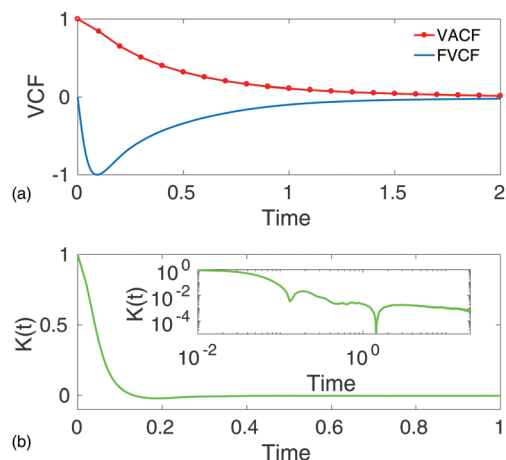
Fig. 3 (a) Normalized VACF and FVCF computed from MD simulations; (b) normalized memory kernel $K(t)$ solved from eqn (4), where the inset shows the global view of $|K(t)|$ in log–log scale.

Fig. 4 Approximation of the memory kernel $K(t)$ with $\mathcal{N}$ = 2, 3, or 5 terms of exponentially damped oscillators in eqn (11). The inset shows the global view of $|K(t)|$ in log–log scale.

are referred to the reduced LJ units; *i.e.*, the mass, length, energy, and time units were set as: $m = 1$, $\sigma = 1$, $\varepsilon = 1, \tau = \sigma(m/\varepsilon)^{0.5} = 1$. To obtain accurate ensemble prediction from the noisy data of MD simulations, 30 independent simulations were conducted, and each was run for 10^6 time steps after the thermal equilibrium state was reached, from which we computed the ensemble-averaged quantities of interest.

3.1.2 CG model. The CG model is illustrated in Fig. 2(b), where the solvent DOFs were eliminated, and the solvent-mediated kinetic effect was incorporated in the GLE *via* the non-Markovian memory; each star polymer was coarse-grained as a single CG particle whose dynamics is governed by eqn (6). We note that as CG particles do not interact with each other in the present framework, the CG system may consist of only one CG particle. However, including many particles in the CG simulations provide large samples for accurate ensemble predictions; and, keeping the CG system the same size as the microscopic system allows for consistent comparison of computational cost with the MD simulations. Thus, in practice the CG system consists of 1000 identical CG particles in the same periodic cubic box of length 37.188σ.

Memory kernel. Using the data from the MD simulations, we first determined the memory kernel $K(t)$. To this end, the ensemble-averaged VACF and FVCF of the star polymers' COMs were computed from the MD simulations, as depicted in Fig. 3(a). We then solved for $K(t)$ by numerical deconvolution of eqn (4), whose solution is shown in Fig. 3(b). As noted, the kinetics of polymers mediated by solvent can display a slow-decay memory.

To determine the extended dynamics (eqn (6)), we next approximated $K(t)$ by a linear combination of exponentially damped oscillators as in eqn (11). With the fitting parameters a_l, b_l, c_l, and w_l ($l = 1, \ldots, \mathcal{N}$), the matrix **A** can be assembled according to eqn (12), and **B** can then be determined from eqn (10). To this end, the memory kernel $K(t)$ was fitted by a linear combination of 2, 3, or 5 terms of exponentially damped
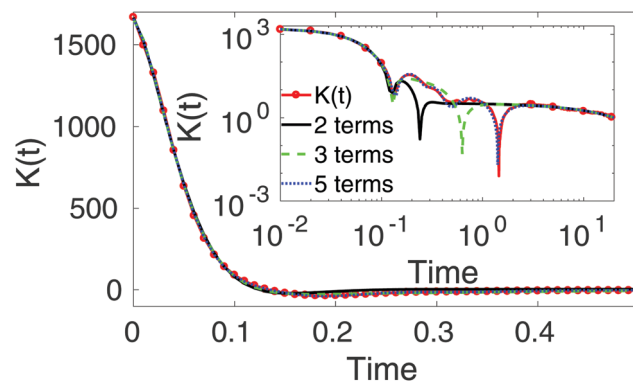
oscillators, as shown in Fig. 4. Here, we only computed and fitted $K(t)$ up to $t = 20$ when the magnitude of $K(t)/K(0)$ is smaller than 10^{-3}. It can be seen that the approximation with 2 or 3 terms displays noticeable discrepancy from the "exact" $K(t)$ (obtained from the MD simulations) at around $t = 0.1$ or $t = 0.6$, respectively; the approximation with 5 terms agrees with the "exact" $K(t)$ very well.

Dynamic properties. Given the extended dynamics determined, we performed the CG simulations by numerically solving eqn (6) using the G-JF or velocity-Verlet temporal integrator. The dynamic properties characterized by the VACF, diffusion coefficient $D(t)$, and mean square displacement (MSD) of CG particles were evaluated from the CG simulations and compared with those computed from the MD simulations. $D(t)$ is a time integral of VACF(t); *i.e.*, $D(t) = \frac{1}{3}\int_0^t \text{VACF}(t')\mathrm{d}t'$. The results are presented in Fig. 5 and 6, where the G-JF integrator was employed with $\Delta t = 0.01$. The ensemble average was taken over 10 independent CG
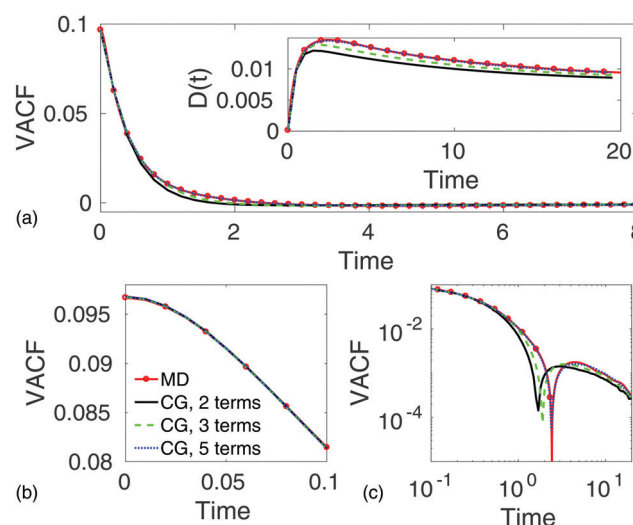

Fig. 5 (a) VACF predicted by both CG and MD simulations, where the inset shows the diffusion coefficient $D(t)$; (b) zoom-in view of short-time VACF; (c) long-time $|\text{VACF}|$ in log–log scale.
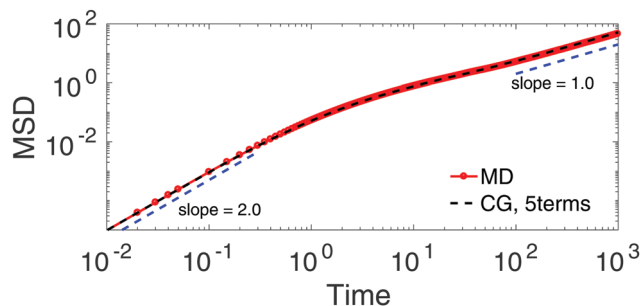
Fig. 6 MSD predicted by both CG and MD simulations.

simulations each conducted for 10^5 time steps after the thermal equilibrium state was reached.

We note that the predicted VACF at $t = 0$ agrees with its theoretical value; *i.e.*, VACF(0) = $3k_BT/M$ = 0.0968, which confirms that the second FDT was accurately satisfied in the CG simulations. Also, all CG predictions are able to correctly reproduce the zero slope of VACF at $t = 0$, as can be seen in Fig. 5(b). As time evolved, the CG simulation using 5 terms to fit the memory kernel is more accurate for reproducing VACF(t) and $D(t)$ than the CG simulation using 2 or 3 terms, as shown in Fig. 5(a and c). It concludes that more accurate approximation of the memory kernel leads to more accurate predictions of dynamic properties, especially for long-time dynamics. Overall, the CG simulation using 5 terms and the G-JF temporal integrator achieved good accuracy in conserving all dynamic properties, particularly in reproducing the entire curves of VACF, D, and MSD as functions of time up to 10^3. At short time scales, the star polymers experienced a supper-diffusion regime, where MSD(t) $\propto t^2$, then a sub-diffusion regime with a decreasing $D(t)$, and finally reached a normal diffusion at long time scales, where MSD(t) $\propto t$. The CG modeling is able to reproduce all stages of the diffusion process (across 5 orders in time) of the star polymers in solution, as shown in Fig. 6.
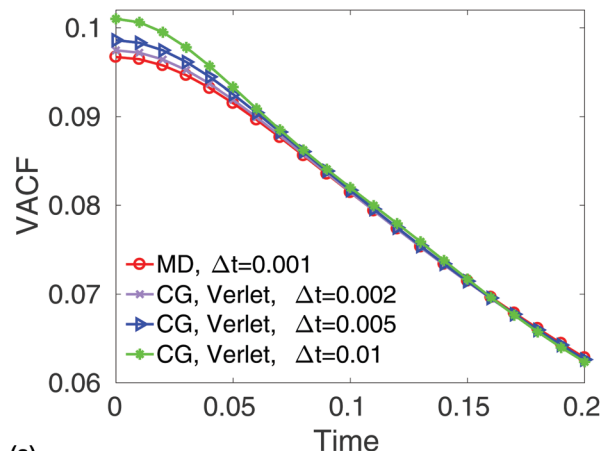
Velocity-Verlet vs. G-JF temporal integrator. We further compared the velocity-Verlet and G-JF schemes for solving the extended dynamics (eqn (6)) with respect to both accuracy and efficiency.

To examine the accuracy, we computed the VACF using both schemes with different timestep sizes Δt and compared with the VACF obtained from the MD simulation.

The errors were calculated for $t \leq 20$ using the normalized Euclidean norm and summarized in Table 1. Note that the velocity-Verlet scheme was not stable for larger timesteps, hence, no data is shown for $\Delta t > 0.01$. The G-JF scheme

Table 1 Errors of the VACF computed from the CG simulations using the velocity-Verlet and G-JF schemes with different Δt

$\Delta t(\tau)$	Velocity-Verlet (error%)	G-JF (error%)
0.005	1.24	0.82
0.01	2.10	0.79
0.05	—	0.89
0.1	—	1.84

(a)
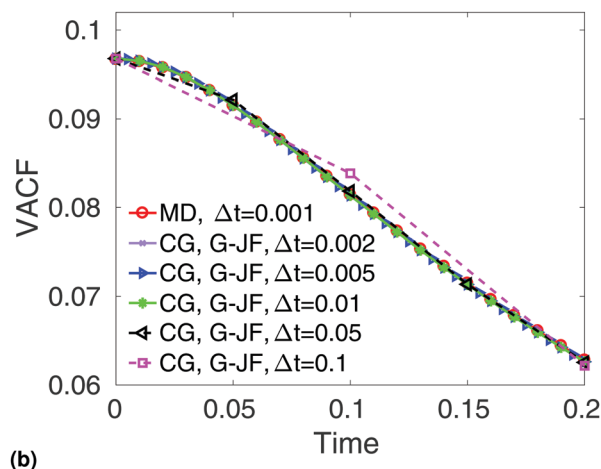
(b)

Fig. 7 Comparison of the short-time VACF computed from the CG simulations using the (a) velocity-Verlet and (b) G-JF temporal integrators with different Δt.

achieved higher-order accuracy than the velocity-Verlet for any timestep size compared and was stable for larger timesteps when the velocity-Verlet already failed. By further comparing the VACF at a short time scale, as depicted in Fig. 7, we find that the G-JF scheme is robust in reproducing the VACF at $t = 0$, accurately satisfying the second FDT with any Δt. In contrast, the velocity-Verlet scheme exhibits noticeable errors in reproducing VACF(0), which becomes more pronounced with larger Δt.

The efficiency of the two schemes was assessed according to the computer time used by each scheme to achieve the same accuracy on the prediction of VACF. From the results in Table 1, we chose to compare the velocity-Verlet using $\Delta t = 0.01$ with the G-JF using $\Delta t = 0.1$, which achieved similar accuracy for the VACF. In each case, the simulation was conducted for $10^5\tau$ on a single core of Intel i5-6500 CPU. The computer time used by each scheme was reported in Table 2. In this comparison, the computer time spent for constructing the CG model is not included. Compared with the MD simulation, while both CG simulations achieved significant speedup by one to two orders, the G-JF scheme was about one order faster than the velocity-Verlet. It implies that the more accurate G-JF scheme does not

Table 2 Comparison of computational cost of CG simulations using the velocity-Verlet and G-JF schemes

Simulation	$\Delta t (\tau)$	Cost (s)	Speedup factor
MD	0.001	$3.44e^5$	—
CG velocity-Verlet	0.01	$1.76e^4$	19.55
CG G-JF	0.1	$3.07e^3$	112.05

introduce significant computational overhead compared with the velocity-Verlet scheme.

Due to the superior performance of the G-JF scheme in both accuracy and efficiency, it was chosen as the temporal integrator for all the CG simulations presented in the rest of this paper.

3.2 Dilute solution of homogeneous star polymers

To further examine the applicability of the proposed implicit-solvent CG modeling to dilute solutions, we next studied a dilute solution of homogeneous star polymers.

3.2.1 Microscopic model. In the microscopic model, 125 N31 star polymers and 125 500 solvent beads are filled into a periodic cubic box of length 56.9619σ with the number density of 0.7. The concentration of N31 star polymers in the solution is 3%, defined as the total number of polymer beads divided by the total number of all beads in the cubic box. The force fields and other settings of the MD simulations were the same as described in Section 3.1.1.

3.2.2 CG model. Same as in Section 3.1.2, in the CG model, each N31 star polymer is coarse-grained as a single CG particle; the solvent DOFs are eliminated, and the solvent-mediated kinetic effect is incorporated in the GLE *via* the non-Markovian memory. Following the same CG procedure, we employed eqn (6) to govern the CG particle's dynamics. First, the memory kernel $K(t)$ computed from the MD simulations was approximated as in eqn (11) with $\mathcal{N} = 4$, 6, or 8 terms of exponentially damped oscillators. $K(t)$ was computed and fitted up to $t = 10$ when $|K(t)|/K(0) < 10^{-3}$. As shown in Fig. 8, more fitting terms lead to more accurate approximation of $K(t)$; the approximation with 8 terms agrees with the "exact" $K(t)$ reasonably well. Next, eqn (6) was solved using the G-JF integrator with $\Delta t = 0.01$. The computed
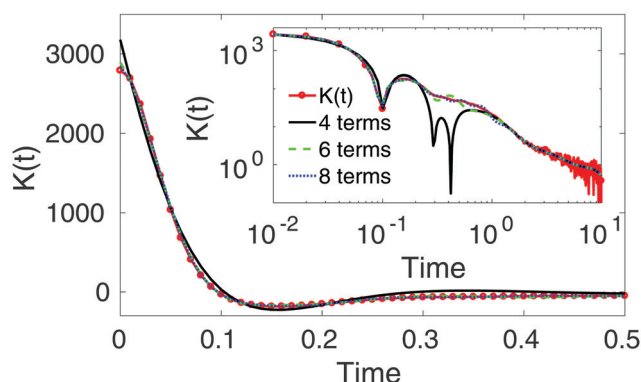
Fig. 9 Dilute solution of homogeneous star polymers: (a) VACF predicted by both CG and MD simulations, where the inset shows the diffusion coefficient $D(t)$; (b) zoom-in view of short-time VACF; (c) long-time |VACF| in log–log scale.
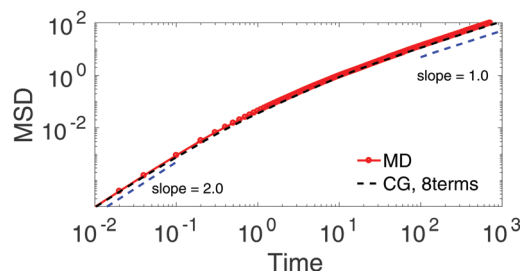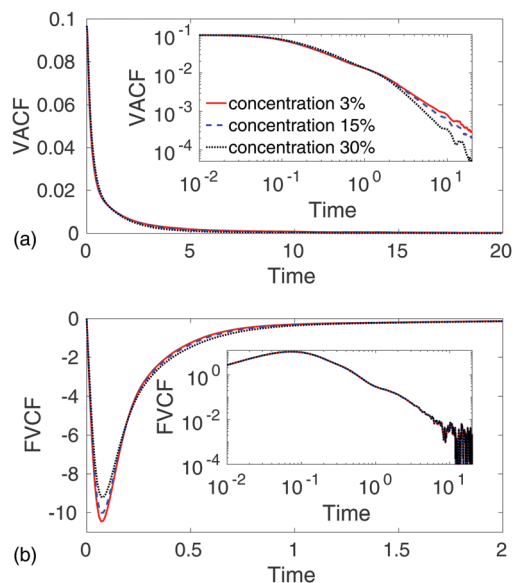
Fig. 10 Dilute solution of homogeneous star polymers: MSD predicted by both CG and MD simulations.

VACF and $D(t)$ are presented in Fig. 9; and, Fig. 10 shows the computed MSD as a function of time until $t = 10^3$. In a very dilute solution, the N31 star polymers display much longer memory in dynamics. By comparison with the results of MD simulations, the CG simulation using more terms to fit the memory kernel is more accurate for reproducing VACF(t) and $D(t)$. In particular, the CG simulation using 8 terms accurately captured all dynamic properties of the N31 star polymers in a dilute solution, at both short and long time scales.

Finally, we examined the dilute solutions of N31 star polymers at different concentrations. Up to the concentration of 15%, the polymers' VACF and FVCF do not exhibit significant changes, as depicted in Fig. 11. Thus, the CG model constructed at the concentration of 3% can be transferable to other concentrations up to 15%. More concentrate solutions, *e.g.*, at 30%, display noticeably different VACF and FVCF, for which the CG model must be reconstructed.

3.3 Inhomogeneous mixture of various star polymers in solution

Next, we studied an inhomogeneous system with different kinds of star polymers mixed in solution. In particular, it consists of three kinds of star polymers: N5 (4 arms with 1 monomer per arm), N21 (10 arms with 2 monomers per arm), and N101 (10 arms with 10 monomers per arm). Its microscopic

Fig. 8 Dilute solution of homogeneous star polymers: memory kernel $K(t)$ computed from MD simulations and its approximations *via* eqn (11) with $\mathcal{N} = 4$, 6, or 8 fitting terms. The inset shows the global view of $|K(t)|$ in log–log scale.

Fig. 11 Dilute solution of homogeneous star polymers: VACF and FVCF predicted by MD simulations at different concentrations. The solution at 15% concentration consists of 626 N31 star polymers and 109 969 solvent beads; and, the solution at 30% concentration consists of 1252 N31 star polymers and 90 563 solvent beads. The inset shows the global view of |VACF| or |FVCF| in log–log scale.
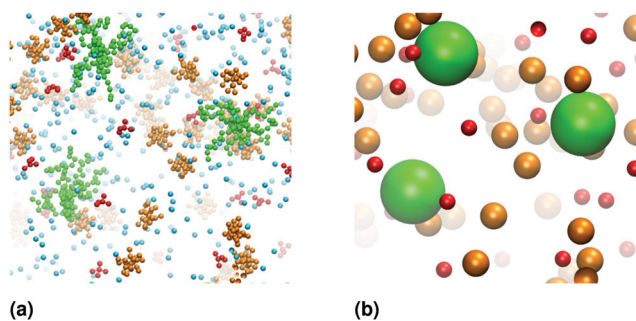

Fig. 12 (a) Microscopic and (b) CG representations of a mixture of three kinds of star polymers (N5, N21, and N101) in solution.

and CG representations are illustrated in Fig. 12. In the CG modeling, each star polymer was coarse-grained as a single CG particle, and the solvent DOFs were eliminated. Thus, the CG model consists of a collection of three kinds of CG particles.

3.3.1 Microscopic model. In the microscopic model, the three kinds of star polymers are composed of 5, 21, and 101 identical monomers, respectively. The solution consists of 1000 N5, 1000 N21, and 216 N101 star polymers. The monomers are represented as LJ beads connected by FENE bonds; the solvent contains 5000 LJ beads. The star polymers and solvent were filled into a periodic cubic box of length 41.8206σ with the number density of 0.7. The force fields and other settings of the MD simulations were the same as in Section 3.1.1.
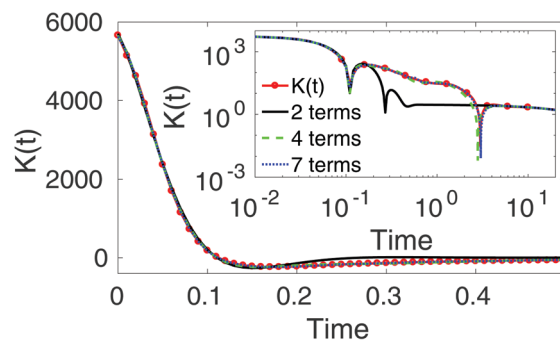
3.3.2 CG model. The CG system consists of three kinds of CG particles, in total, 1000 + 1000 + 216 CG particles in the periodic cubic box of length 41.8206σ. For each kind of CG particles, we determined its extended dynamics following the
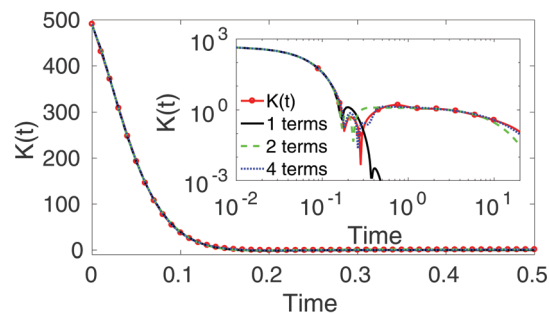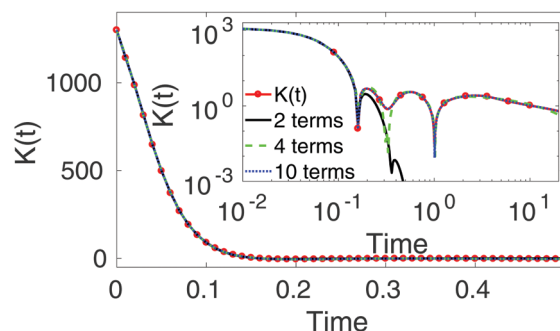
Fig. 13 Inhomogeneous star polymers in solution: memory kernel $K(t)$ computed from MD simulations and its approximations *via* eqn (11) with different numbers ($\mathcal{N}$) of fitting terms for the three kinds of star polymers: (a) N5, (b) N21, and (c) N101. The inset shows the global view of $|K(t)|$ in log–log scale.

same procedure as in Section 3.1.1. First, the memory kernel $K(t)$ was determined for each kind of star polymers from the MD simulations. Next, $K(t)$ was approximated by a linear combination of exponentially damped oscillators, where $K(t)$ was computed and fitted up to $t = 20$ when $|K(t)|/K(0) < 10^{-3}$. Fig. 13 summarizes the "exact" $K(t)$ computed from the MD simulations and its approximations using different numbers of fitting terms for each kind of star polymers. We note that the fitting using 4, 10, and 7 terms for N5, N21, and N101 star polymers, respectively, could achieve satisfactorily accurate approximation of $K(t)$.

Given the fitting parameters in eqn (11), the matrices **A** and **B** could be obtained and in turn eqn (6) was determined for each kind of CG particles corresponding to N5, N21, and N101
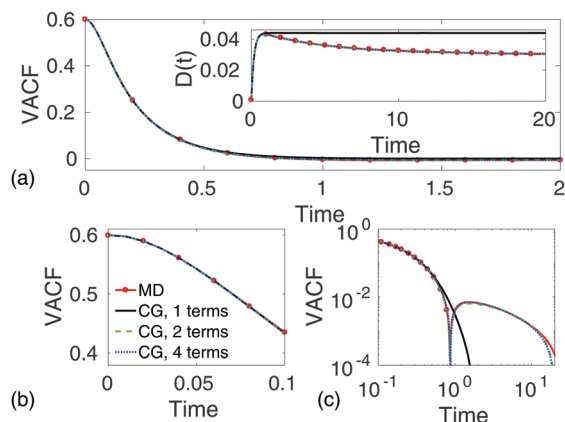
Fig. 14 Inhomogeneous star polymers in solution: (a) VACF predicted by both CG and MD simulations for N5 star polymers, where the inset shows the diffusion coefficient $D(t)$; (b) zoom-in view of short-time VACF; (c) long-time |VACF| in log–log scale.

Fig. 16 Inhomogeneous star polymers in solution: (a) VACF predicted by both CG and MD simulations for N101 star polymers, where the inset shows the diffusion coefficient $D(t)$; (b) zoom-in view of short-time VACF; (c) long-time |VACF| in log–log scale.

star polymers, respectively. Thus, the CG simulations solved eqn (6) using the G-JF temporal integrator with $\Delta t = 0.01$ and computed the VACF, $D(t)$, and MSD for each kind of CG particles, as presented in Fig. 14–17. By comparison with the MD simulation results, we demonstrate that the proposed implicit-solvent CG modeling is able to conserve both the short-time and long-time dynamic properties of star polymers even in an inhomogeneous solution system. We also note that more accurate approximation of the memory kernel using more fitting terms led to more accurate prediction of VACF and thereby its time integral $D(t)$, especially for long-time predictions, *e.g.*, $t > 5$. If the short-time ($t \leq 5$) dynamics was of interest, the approximations using less fitting terms already enabled reasonably accurate predictions on the VACF, *e.g.*, using only 2 terms for the N5 and 4 terms for both N21 and N101 star polymers, as shown in Fig. 14–16. Using less fitting terms resulted in lower dimensional extended dynamics, and in turn, more efficient CG simulations.

Fig. 17 Inhomogeneous star polymers in solution: MSD predicted by both CG and MD simulations for all three kinds of star polymers.

3.4 Branched-chain polymers in solution

In this section, we simulated branched-chain polymers in solution. Each polymer chain is built-up from five bonded N31 star polymers. In the CG modeling, each star polymer was coarse-grained as a single CG particle; the solvent and bond connections between N31 star polymers were eliminated, and their effects on the polymers' kinetics were incorporated in the non-Markovian memory in GLE. The microscopic and CG representations of this system are illustrated in Fig. 18.
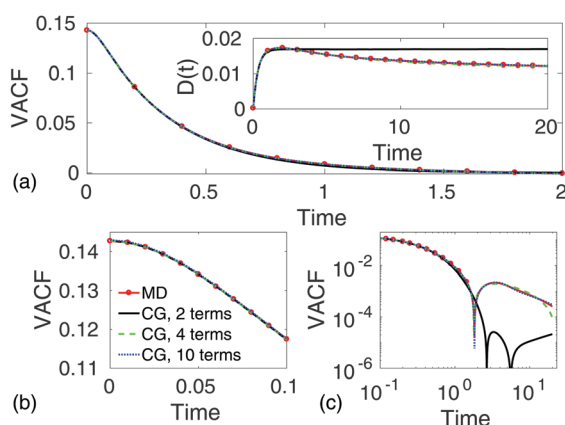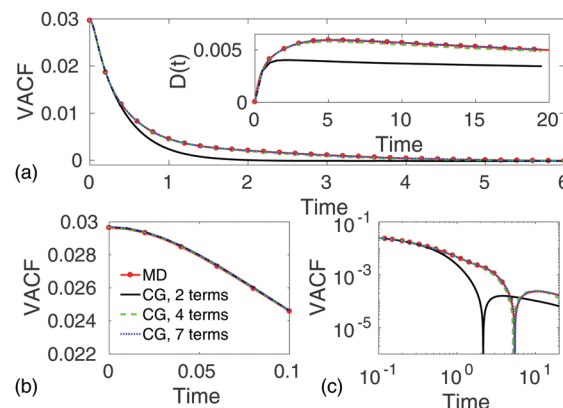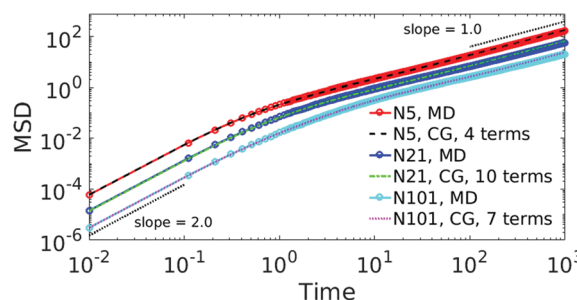

Fig. 15 Inhomogeneous star polymers in solution: (a) VACF predicted by both CG and MD simulations for N21 star polymers, where the inset shows the diffusion coefficient $D(t)$; (b) zoom-in view of short-time VACF; (c) long-time |VACF| in log–log scale.
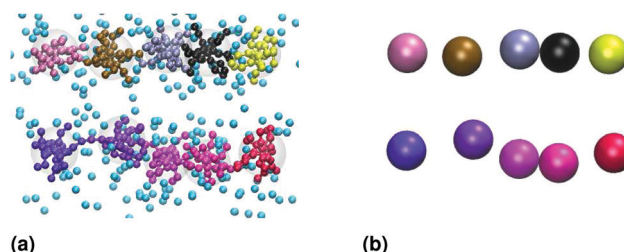

Fig. 18 (a) Microscopic model of the branched-chain polymers in solution, where each polymer chain consists of five bonded N31 star polymers; (b) CG model with each N31 star polymer coarse-grained as a single CG particle, where both solvent and bond connections are removed.
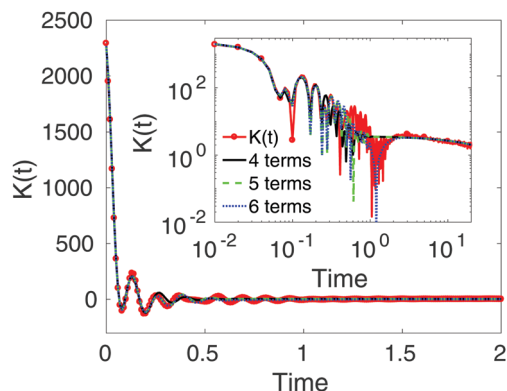
Fig. 19 Branched-chain polymers in solution: memory kernel $K(t)$ computed from MD simulations and its approximations *via* eqn (11) with $\mathcal{N}$ = 4, 5, or 6 fitting terms. The inset shows the global view of $|K(t)|$ in log–log scale.
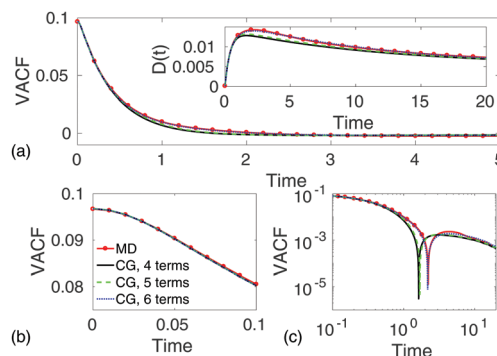

Fig. 20 Branched-chain polymers in solution: (a) VACF predicted by both CG and MD simulations, where the inset shows the diffusion coefficient $D(t)$; (b) zoom-in view of short-time VACF; (c) long-time $|$VACF$|$ in log–log scale.

3.4.1 Microscopic model. The microscopic model of this system contains 200 polymer chains and 5000 solvent beads. The N31 star polymers are the same as described in Section 3.1.1. Bonds were applied between two monomers of each two adjacent N31 star polymers in a chain. All bonds assumed the same FENE potential. The MD force fields are as described in Appendix A. The MD simulations were performed in a periodic cubic box of length 37.188σ with the number density of 0.7. Other settings of the MD simulations were the same as in Section 3.1.1.

3.4.2 CG model. Without solvent and bond connections, the CG system consists of 200×5 independent CG particles in the same periodic cubic box of length 37.188σ. The CG system was governed by the extended dynamics, which was determined following the same procedure as described in Section 3.1.2. Fig. 19 depicts the memory kernel $K(t)$ obtained from the MD simulations, which was fitted by a linear combination of 4, 5, or 6 exponentially damped oscillators. This system exhibits long and oscillating memory. Fitting it with more terms led to more accurate approximation of $K(t)$.

Given the fitting parameters in eqn (11), the extended dynamics (eqn (6)) was formulated accordingly and then solved using the G-JF temporal integrator with Δt = 0.01. Fig. 20 presents the VACF and $D(t)$ computed from both CG and MD simulations. And Fig. 21 depicts the computed MSD. It can be seen that the CG model with the memory kernel approximated by 6 terms can accurately reproduce the dynamic properties of the branched-chain polymers in solution, at both short and long time scales. Recall that both solvent and bond connections were removed in the CG model. Thus, we demonstrate that the non-Markovian memory in GLE can incorporate the kinetic effects of bond connections in chain polymers as well as the solvent.

3.5 Tri-*n*-butyl phosphate in chloroform

After studying three model polymer solution systems, we further examined the proposed CG modeling for a real polymer solution system. In this system, tri-*n*-butyl phosphate (TBP) molecules are immersed in chloroform (CHCl$_3$). The chemical
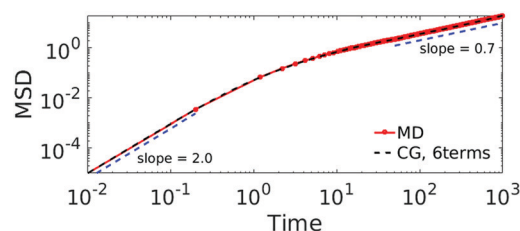

Fig. 21 Branched-chain polymers in solution: MSD predicted by both CG and MD simulations.
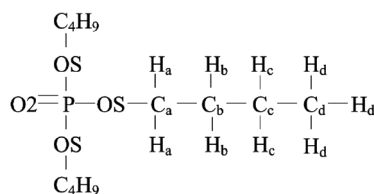
Fig. 22 Chemical composition of a TBP molecule.

composition of a TBP molecule is sketched in Fig. 22. Each TBP molecule was coarse-grained as a single CG particle in the CG modeling. The all-atom and CG representations of this system are illustrated in Fig. 23. Unlike in the previous sections, all simulations here employed real units.

3.5.1 All-atom model. The all-atom model consists of 27 TBP and 1000 CHCl$_3$ molecules filled into a periodic cubic box.
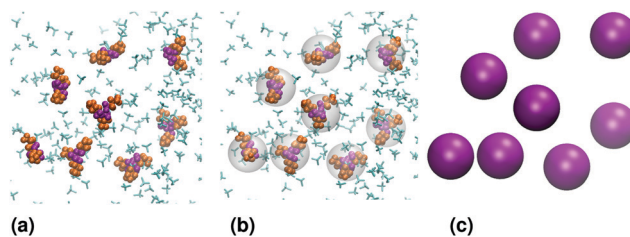
Fig. 23 (a) All-atom model of TBP molecules immersed in CHCl$_3$ solvent; (b) coarse-graining each TBP molecule as a single CG particle; (c) CG model as a collection of CG particles, each representing a TBP molecule.
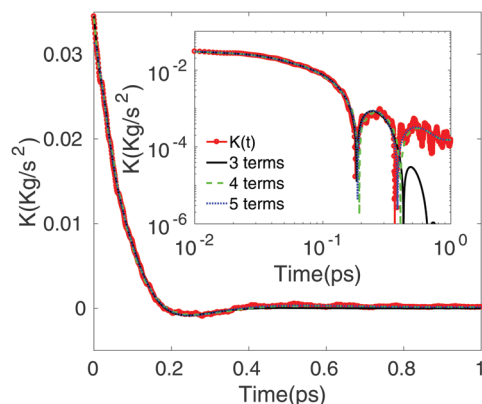
Fig. 24 TBP in $CHCl_3$: memory kernel $K(t)$ computed from MD simulations and its approximation as in eqn (11) with $\mathcal{N}$ = 3, 4, or 5 fitting terms. The inset shows the global view of $|K(t)|$ in log–log scale.

The Optimized Potentials for Liquid Simulations-All Atom (OPLS-AA) force fields[58] were employed to simulate the TBP molecules, while the all atom model without polarization[59] was used for the $CHCl_3$ molecules. (Details about the OPLS-AA force fields are provided in Appendix B.) To proceed, the volume of the system in equilibrium must be determined first, which gives the appropriate size of the periodic box to be used for modeling the system. Thus, a MD simulation was first performed in the isothermal–isobaric ensemble (NPT) using the Nośe–Hoover thermostat for 1 ns under the standard pressure 1 atm and room temperature 300 K. As a result, the length of the cubic periodic box to be used was determined as 52.8 Å. With that, the MD simulations in the canonical ensemble (NVT) were next performed under the temperature 300 K with the timestep Δt = 1 fs. Data were collected and the ensemble-averaged quantities of interest were computed after the thermal equilibrium state was reached; i.e., after 1 ns.

3.5.2 CG model. The CG model consists of 27 CG particles in a periodic cubic box of length 52.8 Å. Following the same CG

Fig. 25 TBP in $CHCl_3$: (a) VACF predicted by both CG and MD simulations, where the inset shows the diffusion coefficient $D(t)$; (b) zoom-in view of short-time VACF; (c) long-time |VACF| in log–log scale.

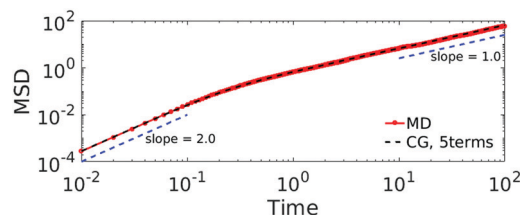**Fig. 26** TBP in $CHCl_3$: MSD predicted by both CG and MD simulations.

procedure as described in Section 3.1.2, we formulated eqn (6) to govern the dynamics of each CG particle. First, the memory kernel $K(t)$ computed from the MD simulations was approximated as in eqn (11). Here, we used 3, 4, or 5 terms of exponentially damped oscillators to fit $K(t)$, as shown in Fig. 24. Next, eqn (6) was solved using the G-JF temporal integrator with Δt = 10 fs. Fig. 25 depicts the computed VACF and $D(t)$ of TBP molecules, and Fig. 26 shows the computed MSD as a function of time. By comparison with the results of MD simulations, we demonstrate in a real polymer system that the proposed CG modeling is able to accurately capture the dynamics of polymers in solution, at both short and long time scales. Without explicit solvent in the CG model, the non-Markovian memory in GLE properly incorporated the effect of $CHCl_3$ solvent on the dynamics of TBP molecules.

4 Conclusion

We have presented a bottom-up implicit-solvent coarse-graining strategy for polymer solutions, which has been validated on homogeneous coarse-graining of monosized polymers and sub-units of chain polymers, and also heterogeneous coarse-graining for a mixture of non-monosized polymers. The CG model was constructed by grouping a cluster of bonded atoms into a single CG particle and removing the solvent DOFs. Elimination of DOFs, especially the solvent DOFs, introduced non-negligible memory effects. By employing the Mori-Zwanzig projection to the microscopic dynamics, the GLE with a non-Markovian memory was derived to govern the CG dynamics. The microscopic dynamics was generated by MD simulations, thus, the memory kernel in the GLE was computed from the MD trajectories. Directly solving the GLE with a non-Markovian memory and colored noise can be computationally expensive. Thus, we introduced auxiliary variables coupled to the momenta of CG particles to replace the GLE with an extended dynamics, which could circumvent computing the convolution and sampling colored noise and be solved more efficiently. To this end, the memory kernel was approximated by a linear combination of exponentially damped oscillators.

Previous efforts in solving the extended dynamics mainly used the velocity-Verlet temporal integrator. However, we found that the velocity-Verlet scheme limited the efficiency of CG simulations by requiring very small timestep sizes due to stability and accuracy. Therefore, in this work, we introduced the higher-order G-JF temporal integrator for solving the extended dynamics. Compared with the velocity-Verlet scheme, the G-JF integrator was found more stable and accurate given
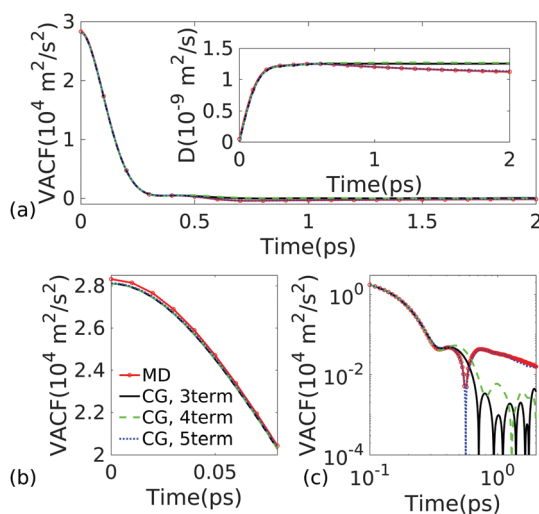
the same timestep size. For achieving the same accuracy, the G-JF integrator allowed to use much larger timesteps but did not introduce significantly more computational burden, and hence greatly improved the efficiency of the CG simulations.

We have assessed and validated the proposed CG modeling in four different polymer solution systems, including a homogeneous solution of monosized star polymers, inhomogeneous mixture of non-monosized star polymers, branched-chain polymers in solution, and also TBP polymers in chloroform. For all these types of polymer solutions, we have demonstrated that the proposed CG modeling can accurately reproduce the dynamic properties of the reference microscopic (or all-atom) systems. Although the solvent DOFs were eliminated in the CG modeling, the non-Markovian memory properly incorporated the solvent-mediated effect on the dynamics of polymers. In the system of branched-chain polymers, although the bond connections between CG clusters were also eliminated in the CG modeling, the GLE with the non-Markovian memory was still able to correctly capture the polymers' dynamics. The dynamic properties of polymers were characterized by the VACF, diffusion coefficient D, and MSD as functions of time. The predictions by the CG simulations agree well with the MD simulation results for each of these quantities over the entire curve across five orders in time. To construct the CG model, the memory kernel was computed from a MD simulation and approximated up to $t \sim \mathcal{O}(10)$ until $|K(t)|/K(0) < 10^{-3}$. The diffusion process was predicted by the constructed CG model until $t = 10^3$.

Approximating the memory kernel with more terms of exponentially damped oscillators ensured more accurate CG modeling for reproducing especially long-time dynamic properties. Thus, the extended dynamics with auxiliary variables enables to construct arbitrarily high-order CG models to characterize long-time behaviors and complex transition dynamics of polymers in solution. In practical applications, the number of terms used in the CG modeling can be decided according to the time scales and accuracy required in conserving the dynamic properties of the reference microscopic systems.

In summary, we highlight two contributions of this work. First, the GLE with a non-Markovian memory was previously employed for coarse-graining polymers in melts (without solvent) when the time scales are not well separated for the momentum and random force of a CG cluster.[50,60] And, only homogeneous coarse-graining of monosized star polymers was addressed. Here, we go beyond and have demonstrated that for polymers in solution, in addition to coarse-graining polymer molecules, the solvent DOFs may be removed in the CG modeling, and their kinetic effect on the polymers can be captured by the non-Markovian memory in the GLE derived *via* the Mori-Zwanzig formalism. We addressed various polymer systems: homogeneous or inhomogeneous solution of polymers and star or chain polymers. Second, we introduced the higher-order G-JF temporal integrator in the CG modeling for solving the extended dynamics, which shows improved stability and accuracy and can significantly accelerate the CG simulations, compared with the commonly used velocity-Verlet integrator.

We note that for dilute solutions, *e.g.*, the solution of homogeneous star polymers studied in Section 3.2, the polymers' VACF and FVCF do not vary significantly at different concentrations. Thus, the memory kernel and CG model constructed can be transferable for different concentrations in the dilute regime. However, in general, the CG model constructed cannot be transferred from one system to another and between different thermodynamic conditions since the VACF and FVCF can be significantly different. For polymers in solution, an analytical formulation is still absent to scale the VACF and FVCF, and thereby the memory kernel for different polymer systems and/or different thermodynamic conditions.

To capture the structural properties of polymers in solution, the present CG modeling needs to be extended to include bonded and non-bonded conservative interactions between CG sites, which is our next step. Even though it only focuses on the dynamic properties, the proposed CG modeling *via* the GLE with correct memory functions can play an important role in applications involving anomalous diffusion. One example is to understand and predict the experimental observations of anomalous diffusion using single particle tracking techniques,[61–65] where the time evolution of a single tracing particle can be tracked and recorded at high spatiotemporal resolutions. A GLE model with effective memory function can be constructed from the experimental data to correctly predict the anomalous diffusion. Note that the acceleration (*i.e.*, the total force) can be obtained from instantaneous velocities measured at high temporal resolutions, from which the FVCF and VACF can be evaluated and used to determine the memory kernel from eqn (4).

Conflicts of interest

There are no conflicts to declare.

Appendix A: MD force fields of Lennard-Jones particles

In Sections 3.1–3.4, the microscopic representations of polymers and solvent consist of LJ particles. In the MD simulations, the dynamics of these LJ systems is governed by the Hamiltonian:

$$H = \sum_{i=1}^{n} \frac{\mathbf{p}_i^2}{2m_i} + \sum_{i \neq j} E(r_{ij}), \qquad (A1)$$

where H defines the phase space trajectories of the system; $r_{ij} = \|\mathbf{r}_{ij}\| = \|\mathbf{r}_i - \mathbf{r}_j\|$ is the distance between two LJ beads; E denotes the total potential energy contributed by the inter-atomic and bonded potentials. The inter-atomic potential adopts the pure repulsive Weeks–Chandler–Andersen (WCA) potential and given by:

$$E_{WCA}(r) = \begin{cases} 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^6 + \frac{1}{4} \right] & r \leq 2^{1/6} \\ \infty & r > 2^{1/6} \end{cases}, \qquad (A2)$$

where $r_c = 2^{1/6}$ is the cutoff distance. The bonded inter-action between connected LJ beads in polymers is modeled as a spring with a finitely extensible nonlinear elastic (FENE) potential; *i.e.*,

$$E_{\text{FENE}}(r) = \begin{pmatrix} -\frac{1}{2}kr_0^2 \ln\left[1 - \left(\frac{r}{r_0}\right)^2\right] & r < r_0 \\ \infty & r \geq r_0 \end{pmatrix} \quad (A3)$$

where $k = 30\varepsilon/\sigma^2$ is the spring constant, and $r_0 = 1.5\sigma$ is the maximum length of the FENE spring. In sum, we have:

$$E(\mathbf{r}_{ij}) = E_{\text{WCA}}(\mathbf{r}_{ij}) + E_{\text{FENE}}(\mathbf{r}_{ij}). \quad (A4)$$

Appendix B: MD force fields for tri-*n*-butyl phosphate and chloroform

The dynamics of atoms in TBP and CHCl$_3$ molecules also follows the Hamiltonian in eqn (A1). The total potential energy E is given by:

$$E = \sum_{\text{bonds}} K_r (r - r_{\text{eq}})^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_{\text{eq}})^2$$

$$+ \sum_{\text{dihedrals}} \left[\frac{1}{2}V_1(1+\cos\phi)+\frac{1}{2}V_2(1-\cos2\phi)+\frac{1}{2}V_3(1+\cos3\phi)\right]$$

$$+ \sum_i \sum_j \left[4\varepsilon_{ij}\left\{\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}-\left(\frac{\sigma_{ij}}{r_{ij}}\right)^6\right\}+\frac{Cq_iq_j}{r_{ij}}\right]. \quad (B1)$$

The terms in eqn (B1), in order of appearance, correspond to the bond-length, bond-angle, dihedral-angle, LJ, and electrostatic potentials, respectively. The OPLS-AA force fields[58,66,67] with charges from Modified Neglect of Differential Overlap (MNDO)[68] were used to specify the parameters for the interaction potential of all atoms of a TBP molecule, as in Tables 3–6. The CHCl$_3$ was represented with the

Table 3 Bond-length potential parameters for TBP

Bond type	K_r (kcal mol^{-1} Å^{-2})	r_{eq} (Å)
OS–P	230	1.610
OS–C	320	1.410
P–O2	525	1.480
C–H	340	1.090
C–C	268	1.529

Table 4 Bond-angle potential parameters for TBP

Angle type	K_θ (kcal mol^{-1} rad^{-2})	θ_{eq} (deg)
C–OS–P	100	120.5
OS–P–O2	100	108.23
OS–P–OS	45	102.6
OS–C–H	35	109.5
OS–C–C	50	109.5
H–C–H	33	107.8
C–C–H	37.5	110.7
C–C–C	58.35	112.7

Table 5 Dihedral angle potential parameters for TBP

Dihedral type	V_1 (kcal mol^{-1})	V_2 (kcal mol^{-1})	V_3 (kcal mol^{-1})
C–OS–P–O2	0	0	0
C–OS–P–OS	0	0	0
H–C–OS–P	0	0	0.3
C–C–OS–P	−1.42	−0.62	0.1
OS–C–C–H	0	0	0.468
C–C–C–OS	1.3	−0.05	0.2
H–C–C–H	0	0	0.3
C–C–C–H	0	0	0.3
C–C–C–C	1.3	−0.05	0.2

Table 6 LJ and electrostatic potential parameters for TBP

Atom type	ε_{ii} (kcal mol^{-1})	σ_{ii} (Å)	q_i (e)
O2	0.2	3.15	−0.87
P	0.2	3.74	1.77
OS	0.14	2.9	−0.56
C$_a$	0.066	3.5	0.20
C$_b$	0.066	3.5	0.06
C$_c$	0.066	3.5	0.02
C$_d$	0.066	3.5	0.02
H$_a$	0.03	2.5	−0.02
H$_b$	0.03	2.5	0
H$_c$	0.03	2.5	0
H$_d$	0.03	2.5	0

Table 7 Bond/angle parameters for CHCl$_3$

Bond/angle type	r_{eq} (Å)	θ_{eq} (deg)
C–C	1.76	—
C–H	1.07	—
Cl–C–Cl	—	111.2
Cl–C–H	—	107.6

Table 8 LJ and electrostatic potential parameters for CHCl$_3$

Atom type	ε_{ii} (kcal mol^{-1})	σ_{ii} (Å)	q_i (e)
C	0.137	3.41	0.5609
Cl	0.275	3.45	−0.1686
H	0.020	2.81	−0.0551

rigid five-site all-atom model without polarization.[59] The total potential energy for CHCl$_3$ is the sum of LJ and electrostatic potentials since it was represented by a rigid model. The parameters for CHCl$_3$ are given in Tables 7 and 8. From Tables 6 and 8, the parameters σ_{ij} and ε_{ij} in the LJ potential can be determined *via* the geometric mixing rule $\sigma_{ij} = (\sigma_{ii}\sigma_{jj})^{1/2}$ and $\varepsilon_{ij} = (\varepsilon_{ii}\varepsilon_{jj})^{1/2}$.

References

1 Z. G. Mills, W. Mao and A. Alexeev, *Trends Biotechnol.*, 2013, **31**, 426–434.

2 J. Mu, R. Motokawa, C. D. Williams, K. Akutsu, S. Nishitsuji and A. J. Masters, *J. Phys. Chem. B*, 2016, **120**, 5183–5193.

3 L. Rovigatti, N. Gnan, L. Tavagnacco, A. J. Moreno and E. Zaccarelli, *Soft Matter*, 2019, **15**, 1108–1119.

4 P. Bełdowski, P. Weber, A. Dėdinaitė, P. M. Claesson and A. Gadomski, *Soft Matter*, 2018, **14**, 8997–9004.

5 M. G. Saunders and G. A. Voth, *Annu. Rev. Biophys.*, 2013, **42**, 73–93.

6 S. Kmiecik, D. Gront, M. Kolinski, L. Wieteska, A. E. Dawid and A. Kolinski, *Chem. Rev.*, 2016, **116**, 7898–7936.

7 M. Dinpajooh and M. G. Guenza, *Soft Matter*, 2018, **14**, 7126–7144.

8 K. M. Salerno, A. Agrawal, D. Perahia and G. S. Grest, *Phys. Rev. Lett.*, 2016, **116**, 058302.

9 A. Gooneie, S. Schuschnigg and C. Holzer, *Polymers*, 2017, **9**, 16.

10 J. Kleinjung and F. Fraternali, *Curr. Opin. Struct. Biol.*, 2014, **25**, 126–134.

11 T. T. Pham, M. Bajaj and J. R. Prakash, *Soft Matter*, 2008, **4**, 1196–1207.

12 R. Chudoba, J. Heyda and J. Dzubiella, *Soft Matter*, 2018, **14**, 9631–9642.

13 G. Sevink and J. Fraaije, *Soft Matter*, 2014, **10**, 5129–5146.

14 D. Reith, M. Pütz and F. Müller-Plathe, *J. Comput. Chem.*, 2003, **24**, 1624–1636.

15 A. P. Lyubartsev and A. Laaksonen, *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdisc. Top.*, 1995, **52**, 3730.

16 S. Izvekov and G. A. Voth, *J. Phys. Chem. B*, 2005, **109**, 2469–2473.

17 M. S. Shell, *J. Chem. Phys.*, 2008, **129**, 144108.

18 T. Sanyal and M. S. Shell, *J. Chem. Phys.*, 2016, **145**, 034109.

19 R. Kubo, *Rep. Prog. Phys.*, 1966, **29**, 255.

20 K. R. Hadley and C. McCabe, *Mol. Simul.*, 2012, **38**, 671–681.

21 A. V. Onufriev and S. Izadi, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2018, **8**, e1347.

22 J. Zavadlav, R. Podgornik, M. Melo, S. Marrink and M. Praprotnik, *Eur. Phys. J.-Spec. Top.*, 2016, **225**, 1595–1607.

23 W. G. Noid, *J. Chem. Phys.*, 2013, **139**, 090901.

24 S. Izvekov and B. M. Rice, *J. Chem. Phys.*, 2014, **140**, 104104.

25 J. Zavadlav, S. J. Marrink and M. Praprotnik, *J. Chem. Theory Comput.*, 2018, **14**, 1754–1761.

26 Y. Han, J. F. Dama and G. A. Voth, *J. Chem. Phys.*, 2018, **149**, 044104.

27 H. Lynn and M. Thachuk, *J. Chem. Phys.*, 2019, **150**, 024108.

28 Y.-H. Tang, Z. Li, X. Li, M. Deng and G. E. Karniadakis, *Macromolecules*, 2016, **49**, 2895–2903.

29 X. Qiang, X. Wang, Y. Ji, S. Li and L. He, *Polymer*, 2017, **115**, 1–11.

30 A. L. Blumers, Y.-H. Tang, Z. Li, X. Li and G. E. Karniadakis, *Comput. Phys. Commun.*, 2017, **217**, 171–179.

31 D. A. Fedosov, W. Pan, B. Caswell, G. Gompper and G. E. Karniadakis, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2011, **108**, 11772–11777.

32 Z. Li, X. Bian, X. Yang and G. E. Karniadakis, *J. Chem. Phys.*, 2016, **145**, 044102.

33 H. A. Karimi-Varzaneh, N. F. A. van der Vegt, F. Müller-Plathe and P. Carbone, *ChemPhysChem*, 2012, **13**, 3428–3439.

34 P. Español and P. B. Warren, *J. Chem. Phys.*, 2017, **146**, 150901.

35 Z. Liu, Y. Zhu, R. R. Rao, J. R. Clausen and C. K. Aidun, 2018, *arXiv preprint arXiv:1801.02299*.

36 J. R. Spaeth, I. G. Kevrekidis and A. Z. Panagiotopoulos, *J. Chem. Phys.*, 2011, **134**, 164902.

37 X. Bian, C. Kim and G. E. Karniadakis, *Soft Matter*, 2016, **12**, 6331–6346.

38 H. Mori, *Prog. Theor. Phys.*, 1965, **33**, 423–455.

39 R. Zwanzig, *J. Stat. Phys.*, 1973, **9**, 215.

40 R. Zwanzig, *Nonequilibrium statistical mechanics*, Oxford University Press, 2001.

41 I. Y. Lyubimov, J. McCarty, A. Clark and M. G. Guenza, *J. Chem. Phys.*, 2010, **132**, 224903.

42 I. Lyubimov and M. G. Guenza, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2011, **84**, 031801.

43 G. Jung, M. Hanke and F. Schmid, *J. Chem. Theory Comput.*, 2017, **13**, 2481–2488.

44 G. Jung, M. Hanke and F. Schmid, *Soft Matter*, 2018, **14**, 9368–9382.

45 A. Davtyan, J. F. Dama, G. A. Voth and H. C. Andersen, *J. Chem. Phys.*, 2015, **142**, 154104.

46 M. Ceriotti, G. Bussi and M. Parrinello, *J. Chem. Theory Comput.*, 2010, **6**, 1170–1180.

47 A. D. Baczewski and S. D. Bond, *J. Chem. Phys.*, 2013, **139**, 044107.

48 L. Ma, X. Li and C. Liu, *J. Chem. Phys.*, 2016, **145**, 204117.

49 H. Lei, N. A. Baker and X. Li, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 14183–14188.

50 Z. Li, H. S. Lee, E. Darve and G. E. Karniadakis, *J. Chem. Phys.*, 2017, **146**, 014104.

51 Y. Yoshimoto, Z. Li, I. Kinefuchi and G. E. Karniadakis, *J. Chem. Phys.*, 2017, **147**, 244110.

52 N. Grønbech-Jensen and O. Farago, *Mol. Phys.*, 2013, **111**, 983–991.

53 N. Grønbech-Jensen, N. R. Hayre and O. Farago, *Comput. Phys. Commun.*, 2014, **185**, 524–527.

54 T. Kinjo and S.-a. Hyodo, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2007, **75**, 051109.

55 P. C. Hansen, *Numerical Algorithms*, 2002, vol. 29, pp. 323–378.

56 S. Plimpton, *J. Comput. Phys.*, 1995, **117**, 1–19.

57 B. Choi, K. H. Han, C. Kim, P. Talkner, A. Kidera and E. K. Lee, *New J. Phys.*, 2017, **19**, 123038.

58 W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1996, **118**, 11225–11236.

59 T.-M. Chang, L. X. Dang and K. A. Peterson, *J. Phys. Chem. B*, 1997, **101**, 3413–3419.

60 Z. Li, X. Bian, X. Li and G. E. Karniadakis, *J. Chem. Phys.*, 2015, **143**, 243128.

61 R. Huang, I. Chavez, K. Taute, B. Lukic, S. Jeney, M. Raizen and E.-L. Florin, *Nat. Phys.*, 2011, **7**, 576–580.

62 T. Li, S. Kheifets, D. Medellin and M. G. Raizen, *Science*, 2010, **328**, 1673–1675.

63 S. Kheifets, A. Simha, K. Melin, T. Li and M. G. Raizen, *Science*, 2014, **343**, 1493–1496.

64 M. Hohmann, F. Kindermann, T. Lausch, D. Mayer, F. Schmidt, E. Lutz and A. Widera, *Phys. Rev. Lett.*, 2017, **118**, 263401.

65 K. Burnecki, E. Kepten, Y. Garini, G. Sikora and A. Weron, *Sci. Rep.*, 2015, **5**, 11306 EP.

66 W. L. Jorgensen, *OPLS-AA/M Parameter file*, http://zarbi.chem.yale.edu/oplsaam.html.

67 R. Paine, *Tinker Molecular Modeling*, https://dasher.wustl.edu/tinker/distribution/params/oplsaa.prm.

68 P. Beudaert, V. Lamare, J.-F. Dozol, L. Troxler and G. Wipff, *Solvent Extr. Ion Exch.*, 1998, **16**, 597–618.