

# Iterative Reconstruction of Memory Kernels

Gerhard Jung,<sup>1,2,\*</sup> Martin Hanke,<sup>3,†</sup> and Friederike Schmid<sup>1,‡</sup>

<sup>1</sup>*Institut für Physik, Johannes Gutenberg-Universität Mainz, Staudingerweg 9, 55128 Mainz, Germany*

<sup>2</sup>*Graduate School of Excellence Materials Science in Mainz, Staudingerweg 9, 55128 Mainz, Germany*

<sup>3</sup>*Institut für Mathematik, Johannes Gutenberg-Universität Mainz, Staudingerweg 9, 55128 Mainz, Germany*

In recent years it has become increasingly popular to construct coarse-grained models with non-Markovian dynamics in order to account for an incomplete separation of time scales. One challenge of a systematic coarse-graining procedure is the extraction of the dynamical properties, namely the memory kernel, from equilibrium all-atom simulations. In this paper we propose an iterative method for memory reconstruction from dynamical correlation functions. Compared to previously proposed non-iterative techniques, it ensures by construction that the target correlation functions of the original fine-grained systems are reproduced accurately by the coarse-grained system, regardless of time step and discretization effects. Furthermore, we also propose a new numerical integrator for generalized Langevin equations that is significantly more accurate than the more commonly used generalization of the Velocity-Verlet integrator.

We demonstrate the performance of the above described methods using the example of backflow induced memory in the Brownian diffusion of a single colloid. For this system we are able to reconstruct realistic coarse-grained dynamics with time steps about 200 times larger than used in the original molecular dynamics simulations.

## I. INTRODUCTION

The idea of systematic coarse-graining in general is going from a fine-grained model with many degrees of freedom to an effective model that is only described by a few coarse-grained variables. These coarse-grained variables are usually clusters of atomic particles. Popular examples are star polymers<sup>1-3</sup> or colloids<sup>4</sup> each described by one effective particle. To understand the dynamics of coarse-grained models it is necessary to derive equations of motion for these effective degrees of freedom. In equilibrium this can be done systematically by applying the Mori-Zwanzig formalism<sup>5-8</sup>. The formalism uses projection operators to map on the space of the coarse-grained dynamical variables. Independent of the choice of these projection operators, the resulting equations of motion can always be condensed into the form of a generalized Langevin equation (GLE),

$$\mathbf{F}(t) = M\dot{\mathbf{V}}(t) = \mathbf{F}_C(t) - \int_{-\infty}^t ds \mathbf{K}(t-s)\mathbf{V}(s) + \partial\mathbf{F}(t), \quad (1)$$

with the coarse-grained velocities  $\mathbf{V}(t)$ , mean conservative force  $\mathbf{F}_C(t)$  and random force  $\partial\mathbf{F}(t)$ , respectively. The memory kernel matrix  $\mathbf{K}(s)$  contains the information of the frequency dependent responses of the coarse-grained particles. There is a deep connection between the random force autocorrelation function (RACF) and the memory kernel: the fluctuation-dissipation theorem,

$$\langle \partial\mathbf{F}(t)\partial\mathbf{F}(t') \rangle = k_B T \mathbf{K}(t-t'), \quad (2)$$

with Boltzmann constant  $k_B$  and thermodynamic temperature  $T$ . This relation can be derived using the Mori-Zwanzig formalism<sup>7</sup> or by analysis of the GLE<sup>9</sup> with the requirement that at equilibrium, the coarse-grained variables should be Boltzmann-distributed with respect to the conservative forces.

Derived in the 1960s as a theoretical framework to understand the process of systematic coarse-graining, the Mori-Zwanzig formalism has become increasingly popular in recent years as a practical computational tool. The range of applications is huge: from quantum-bath oscillators<sup>10</sup> over molecular simulations<sup>1,2,11</sup> to mesoscopic modeling<sup>12</sup>. In the following, we will study the Brownian motion of a single colloid, therefore, the memory kernel can be considered to be a scalar function  $K(t)$  and the mean conservative force is zero,  $\mathbf{F}_C(t) = 0$ .

The challenge of performing non-Markovian modeling is the extraction of the memory kernel from theory or fine-grained simulations. While the former might be generally preferable there are only few cases for which theoretical predictions are possible<sup>12-14</sup>. Therefore, it is very important to develop efficient and precise methods to reconstruct memory kernels.

Over the years several direct reconstruction procedures have been proposed. In principle, one can divide the existing methods into three groups: Time-based inversion methods<sup>4,15,16</sup>, Fourier space inversion<sup>17</sup> and parametrization techniques<sup>18,19</sup>. However, these direct reconstruction techniques are sensitive to discretization errors, which may cause problems if the underlying trajectories from the microscopic model have not been generated or recorded with a very fine time resolution.

In the present paper, we propose a reconstruction technique that is based on an iterative procedure. This procedure allows us to optimize the memory kernel stepwise until the coarse-grained model recovers precisely the fine-grained correlation functions. The idea is motivated by the success of iterative methods in a similar inverse problem, the reconstruction of coarse-grained potentials from spatial correlation functions. The most prominent and increasingly popular methods in this field are Inverse Monte Carlo (IMC)<sup>20</sup> and Iterative Boltzmann Inversion (IBI)<sup>21</sup>. To the best knowledge of the authors, nobody

	IBI <sup>21</sup>	
unknown quantity $X$	pair potential $V(r)$	
matched observable $Y$	RDF $g(r)$	
mapping function $\phi(Y)$	$-\frac{1}{\beta} \ln(Y)$	
initial guess $X_0$	$\phi(Y_{\text{MD}})$	
$X = X_0$ valid in limit	$\rho \rightarrow 0$	
basic iteration step	$X_{i+1} = X_i + \phi(Y_{\text{MD}}) - \phi(Y_i)$	
optimized iteration step	$X_{i+1} = X_i + \alpha(\phi(Y_{\text{MD}}) - \phi(Y_i))$	

	IMRF	IMRV
$X$	memory kernel $K(t)$	
$Y$	FACF $\langle F(t)F(0) \rangle$	VACF $\langle v(t)v(0) \rangle$
$\phi(Y)$	$\beta Y$	$-\beta M^2 \frac{Y(t+\Delta t) - 2Y(t) + Y(t-\Delta t)}{\Delta t^2}$
$X_0$	$\phi(Y_{\text{MD}})$	
$X = X_0$	$M \rightarrow \infty$	
iteration	$X_{i+1} = X_i + \Delta\phi_i$ with $\Delta\phi_i = \phi(Y_{\text{MD}}) - \phi(Y_i)$	
op. iteration	$X_{i+1} = X_i + h_i(t) \Delta\phi_i$ (Eq. (9))	

TABLE I. Comparison of the iterative Boltzmann inversion (IBI) with the iterative memory reconstruction (IMRF, IMRV). IBI reconstructs a pair potential from the radial distribution function (RDF) of the coarse-grained variables.

has applied an iterative procedure to the problem of reconstructing memory kernels before.

The application of an iterative procedure requires the use of an efficient and accurate numerical integrator for the GLE, because every error made in the integration leads to an error in the memory kernel. Additionally, the integrator should fulfill the fluctuation-dissipation theorem as precisely as possible to simulate a well-defined thermodynamic ensemble. Based on the derivation of a Langevin integrator by Grønbech-Jensen and Farago<sup>22</sup> we propose in this paper an integration procedure for the GLE that is much more accurate than the generalization of the popular Brünger-Brooks-Karplus integrator (BBK)<sup>23</sup> to GLEs<sup>3</sup>.

Our paper is organized as follows: In Sec. II, we introduce our novel iterative reconstruction technique and present an algorithm that can be implemented straightforwardly. We then derive a discretized version of the fluctuation-dissipation theorem in Sec. III to propose a new numerical integrator for the GLE. In Sec. IV, we compare the iterative reconstruction to previously proposed methods using the example of the Brownian diffusion of a single colloid. The reference data is produced with molecular dynamics simulations (MD). In this section we will also show how to apply our method to increase the timescale of a coarse-grained simulation. We summarize and conclude in Sec. V.

## II. ITERATIVE MEMORY RECONSTRUCTION

As an initial guess for the iterative reconstruction we will use the force autocorrelation function (FACF), because in the infinite mass limit the FACF and RACF are the same<sup>24,25</sup>,

$$\lim_{M \rightarrow \infty} \langle F(t)F(0) \rangle = \langle \partial F(t)\partial F(0) \rangle = k_{\text{B}}TK(t). \quad (3)$$

Additionally, also for finite masses, the memory kernel at time  $t = 0$  is equal to the force fluctuations<sup>7</sup>,

$$\langle F(0)^2 \rangle = \langle \partial F(0)^2 \rangle = k_{\text{B}}TK(0). \quad (4)$$

These relations suggest the following iterative method, which uses the differences in the FACF between MD and GLE simulations as an additive correction to the memory kernel,

$$K_{i+1}(t) = K_i(t) + \beta(\langle F(t)F(0) \rangle_{\text{MD}} - \langle F(t)F(0) \rangle_{\text{GLE}}), \quad (5)$$

with the inverse temperature  $\beta = 1/k_{\text{B}}T$ . This procedure, the iterative memory reconstruction (IMR), is very similar to the iterative Boltzmann inversion (IBI), both of which can be seen as a simple fixed point iteration based on a zero-order approximation

$$X \approx \phi(Y),$$

where  $X$  is the unknown coarse-grained quantity and  $Y$  an appropriate observable; see table II for further details.

By Fourier transform of the iterative method and the GLE Eq. (1) one can formulate algorithm Eq. (5) also in the frequency domain:

$$\hat{K}_{i+1}(\omega) = \hat{K}_i(\omega) + \beta \hat{C}_F(\omega) - \frac{i\omega M \hat{K}_i(\omega)}{i\omega M + \hat{K}_i(\omega)}, \quad (6)$$

with the one-sided Fourier transform of the memory kernel,

$$\hat{K}(\omega) = \int_0^\infty dt e^{i\omega t} K(t), \quad (7)$$

and similarly  $\hat{C}_F(\omega)$  the one-sided Fourier transform of the MD-FACF.

In practice, the convergence of the iteration procedure described in Eq. (5) is still poor. When applying a global correction we observed that differences between the FACFs did not disappear, but were only shifted to larger times  $t$ . Introducing a constant prefactor  $\alpha < 1$  to the iteration step (as is done in IBI) was not sufficient to achieve a well-behaved convergence. Therefore, the simple solution to this problem is to introduce a "correction" time  $t_{\text{cor}} = n_{\text{cor}}\Delta t$  that localizes the time window in which the correction is applied (see Eq. (9)). This consideration leads to the iteration prescription

$$K_{i+1}(t) = K_i(t) + h_i(t)(\phi(Y_{\text{MD}}) - \phi(Y_i)) \quad (8)$$

with

$$h_i(t) = \begin{cases} 1 & t/t_{\text{cor}} \leq \frac{i}{2} \\ 1 - t/t_{\text{cor}} + i/2 & \frac{i}{2} < t/t_{\text{cor}} < \frac{i}{2} + 1 \\ 0 & t/t_{\text{cor}} \geq \frac{i}{2} + 1 \end{cases}, \quad (9)$$

mapping variable  $Y$  and mapping function  $\phi(Y)$  as defined in Tab. II. While the specific choice of  $h_i(t)$  for  $t/t_{\text{cor}} \in [i/2, i/2 + 1]$  is arbitrary, it definitely has to be nonnegative and continuous to prevent discontinuities in the memory kernel. The optimal choice of  $t_{\text{cor}}$  is strongly system dependent and should be determined individually. At small  $t_{\text{cor}}$ , the algorithm will always converge, however, the necessary number of iterations can be large. Therefore, we suggest to start with large correction times and optimize until no “shifting” can be observed anymore.

Another important aspect is the choice of the fine-grained correlation function used as input for the iteration. The integration procedure that will be derived in Sec. III integrates the velocity with an error in  $\mathcal{O}(\Delta t^3)$ . However, the force can only be calculated by finite differences,

$$F(t) = M \frac{v(t + \Delta t) - v(t)}{\Delta t}, \quad (10)$$

with an error in  $\mathcal{O}(\Delta t)$ . This leads to significant deviations in the FACF at larger time steps even though the velocity autocorrelation function  $\langle v(t)v(0) \rangle$  (VACF) is reproduced very accurately (see e.g. Fig. 2). Therefore, we also propose a procedure that uses the VACF as matched observable (see Tab. II, IMRV), motivated by the identity:

$$\langle F(t)F(0) \rangle = -M^2 \frac{\partial^2}{\partial t^2} \langle v(t)v(0) \rangle. \quad (11)$$

An exemplary series of iterations using the IMRV scheme is shown in Fig. 1. The fine-grained data were obtained from MD simulations of a single colloid in a Lennard-Jones fluid (see Sec. IV for details). The figure illustrates the progress in the iterative reconstruction of the memory kernel. One can observe a well-behaved convergence that improves stepwise. The final GLE simulations using the IMRV show no significant deviations from the MD simulations. Furthermore, one can see that for this time step, the IMRF gives accurate results as well.

To apply this procedure to GLE simulations with larger time steps we need to find an accurate numerical integrator for the GLE in Eq. (1), because every error made in the discretization will lead to a similar error in the memory kernel. Such an integrator will be proposed in the next section.

### III. NUMERICAL INTEGRATION OF THE GLE

To derive a numerical integrator for the GLE it is necessary to find discretized versions of both Eq. (1) and the

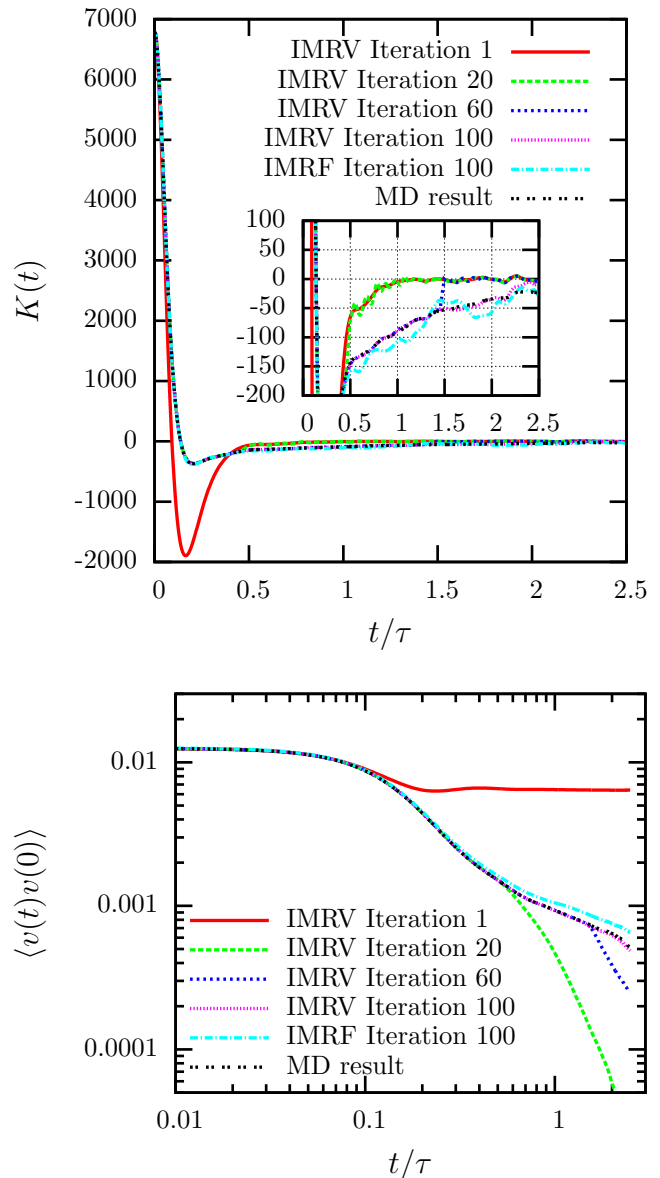


FIG. 1. Iterative memory reconstruction (IMRV, IMRF) applied to the diffusion of a single Brownian particle. In the top panel, the memory kernel is shown for several iteration steps. The “MD result” refers to the inverse Volterra method (see Fig. 3, Ref.<sup>4</sup>). The inset emphasizes the differences between the steps by zooming in. The bottom panel shows the VACF.

fluctuation-dissipation theorem Eq. (2). The first step in discretizing the GLE is to introduce a discrete memory kernel,

$$M\dot{v}(t) = - \sum_{m=0}^{N-1} K_m v(t - m\Delta t) + \partial F(t), \quad (12)$$

with a not yet specified memory sequence  $K_m$  with  $m = 0, \dots, N - 1$ , and a finite time step  $\Delta t$ .

To derive the fluctuation-dissipation theorem we first identify the frequency dependent response  $\hat{\gamma}(\omega)$  of the

discretized memory kernel by Fourier transform of Eq. 12,

$$\hat{v}(\omega) = \frac{\partial \hat{F}(\omega)}{i\omega M + \hat{\gamma}(\omega)}, \quad (13)$$

$$\text{with } \hat{\gamma}(\omega) = \sum_{m=0}^{N-1} K_m e^{-i\omega m \Delta t}. \quad (14)$$

Here,  $\hat{v}(\omega)$  denotes the Fourier transform of the velocity,

$$\hat{v}(\omega) = \int_{-\infty}^{\infty} dt e^{i\omega t} v(t), \quad (15)$$

and similarly  $\partial \hat{F}(\omega)$  the Fourier transform of the random force. Following the derivation of Hauge and Martin-Löf<sup>26</sup> for the continuous version of the GLE, we can identify the power-spectrum  $\hat{C}_{\partial F}(\omega)$  of the random force, defined as the Fourier transform of the RACF,

$$\hat{C}_{\partial F}(\omega) = 2k_B T \Re\{\hat{\gamma}(\omega)\} = 2k_B T \sum_{m=0}^{N-1} K_m \cos(\omega m \Delta t), \quad (16)$$

with the real part  $\Re\{\hat{\gamma}(\omega)\}$  of the response function. By inverse Fourier transform we finally find the fluctuation-dissipation theorem,

$$\langle \partial F(t) \partial F(t') \rangle = k_B T \sum_{m=0}^{N-1} a_m K_m \delta(t - t' - m \Delta t), \quad (17)$$

with  $a_0 = 2$  and  $a_m = 1$  for  $m \neq 0$ . The most important message of this equation is the factor of 2 that enters the instantaneous contribution of the correlation function. This factor is also present in the fluctuation-dissipation theorem for the Langevin equation without memory.

On the basis of this derivation we now develop a numerical integrator for the GLE. The scheme is inspired by the derivation of a Langevin integrator by Grønbech-Jensen and Farago<sup>22</sup>.

The first step is to integrate Eq. (12) over a time interval  $\Delta t$  between the times  $t_n$  and  $t_{n+1} = t_n + \Delta t$ :

$$\int_{t_n}^{t_{n+1}} M \dot{v} dt' = - \sum_{m=0}^{N-1} K_m \int_{t_n}^{t_{n+1}} \dot{r}(t' - m \Delta t) dt' + \int_{t_n}^{t_{n+1}} \partial F(t') dt'. \quad (18)$$

This can, without approximation, be written as,

$$v_{n+1} - v_n = - \sum_{m=0}^{N-1} \frac{K_m}{M} (r_{n+1-m} - r_{n-m}) + \frac{1}{M} \beta_n, \quad (19)$$

with the velocity  $v_n = v(t_n)$ , position  $r_n = r(t_n)$  and correlated Gaussian distributed random numbers

$$\langle \beta_{n+m} \beta_n \rangle = k_B T a_m K_m \Delta t. \quad (20)$$

To find a closed form for this integrator we need to approximate the relation  $\dot{r} = v$  by the term

$$r_{n+1} - r_n \approx \frac{\Delta t}{2} (v_{n+1} + v_n). \quad (21)$$

This introduces an error in the algorithm that scales with  $\Delta t^3$  and is therefore the limiting factor of this discretization scheme. Formerly derived schemes for Langevin or Brownian dynamics are based on similar approximations (e.g. Ref.<sup>22,27,28</sup>). With these relations we can finally write down the full integration algorithm for the discretized GLE in Eq. (12):

$$\begin{aligned} r_{n+1} &= r_n + b \Delta t v_n - \frac{b \Delta t}{2} f_n^D + \frac{b \Delta t}{2M} \beta_n \\ v_{n+1} &= a v_n - b f_n^D + \frac{b}{M} \beta_n, \end{aligned} \quad (22)$$

with

$$f_n^D = \sum_{m=1}^{N-1} \frac{K_m}{M} (r_{n+1-m} - r_{n-m}), \quad (23)$$

$$a \equiv \frac{1 - \frac{K_0 \Delta t}{2M}}{1 + \frac{K_0 \Delta t}{2M}} \quad \text{and} \quad b \equiv \frac{1}{1 + \frac{K_0 \Delta t}{2M}}. \quad (24)$$

In the special case of  $K_m = 0$  for  $m \neq 0$ , these equations reduce to the GJF-integrator proposed in Ref.<sup>22</sup>.

To test the algorithm, we use the memory kernel that was reconstructed in Sec. II. The discretization of the memory kernel will be chosen as follows

$$K_m = \begin{cases} \frac{K(0)\Delta t}{2} & \text{for } m = 0 \\ K(m\Delta t)\Delta t & \text{for } m \neq 0 \end{cases} \quad (25)$$

This choice is consistent with the continuous version of the GLE because firstly it leads to a proper discretization of the integral (trapezoidal rule) and secondly it recovers a random force autocorrelation function that is similar to the continuous memory kernel due to Eq. (17).

The colored noise was produced using the Fourier transform technique proposed by Barrat *et al.*<sup>2,29</sup> (see App. A). We will compare the results to a generalization of the popular Brünger-Brooks-Karplus integrator<sup>3,23</sup>, using the discretization

$$\begin{aligned} r_{n+1} &= r_n + v_n \Delta t + \frac{\Delta t^2}{2M} f^n \\ v_{n+1} &= v_n + \frac{\Delta t}{2M} (f_n + f_{n+1}) \\ f_n &= - \sum_{m=0}^{N-1} K_m v_{n-m} + \partial F_n, \end{aligned} \quad (26)$$

with  $\partial F_n = \partial F(t_n)$ .

The observables that will be determined are the reduced temperature  $k_B T = M \langle v(0)^2 \rangle$  and the VACF for two integration time steps  $\Delta t = 0.001$  and  $\Delta t = 0.1$ .

In Fig. 2 we can observe that with the new integration scheme the kinetic energy is correct even for very large time steps. In contrast, the BBK integrator shows huge deviations from the target temperature (Fig. 2, top panel). This observation is supported by the VACF (Fig. 2, bottom panel). While for small time steps the results of both integrators agree with MD, there are significant deviations when using the BBK integrator at larger

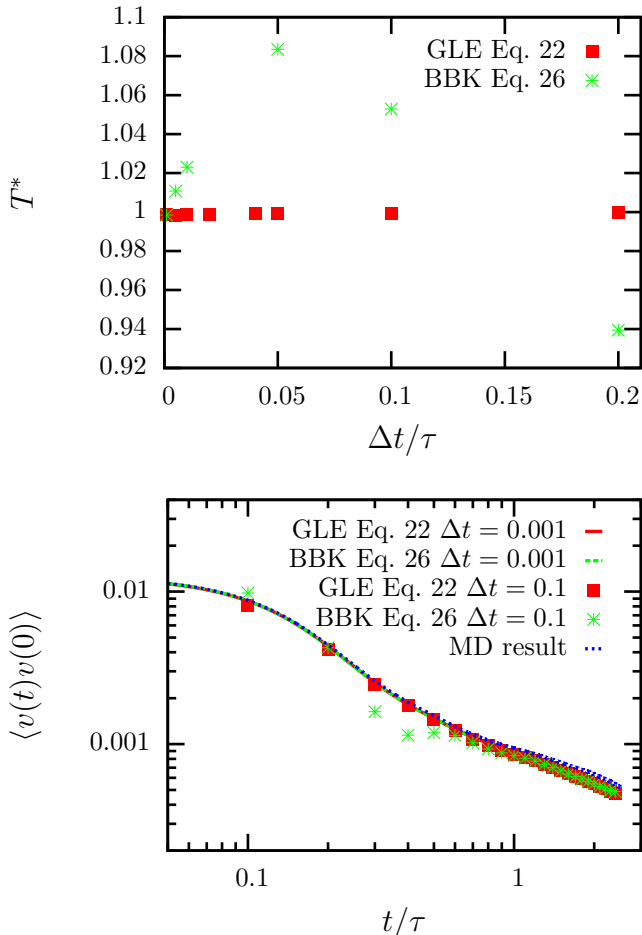


FIG. 2. Reduced temperature  $T^*$  (top panel) and VACF (bottom panel) of GLE simulations with different time steps. The target temperature is  $T_0^* = 1$ . An illustration of the inserted memory kernel can be found in Fig. 3.

time steps. On the contrary, the proposed GLE integrator still shows remarkably good agreement even for time steps that are 100 times larger than the ones used in MD.

#### IV. RESULTS

With the iterative memory reconstruction and the integrator established, we can now apply them to the diffusion of a single colloid in a Lennard-Jones (LJ) fluid.

The system we considered was created by placing LJ particles on a fcc-lattice with lattice constant  $a = 1.71\sigma$  and therefore a reduced density of  $\rho^* = \rho\sigma^3 = 0.8$ . The reduced temperature was set to  $T^* = k_B T/\epsilon = 1.0$ . The LJ diameter  $\sigma$ , energy  $\epsilon$  and time  $\tau = \sigma\sqrt{m/\epsilon} = 1$  are defining the length, energy and time units of the simulation. To calculate the interactions we chose a LJ cutoff  $r_c = 2.5\sigma$  and a particle mass  $m^* = 1m$ . The colloid was carved out off the fcc-lattice with a radius  $R = 3\sigma$ , and then defined as rigid body so that the inter-

particle distances were fixed. The resulting colloid mass was  $M = 80m$  with a hydrodynamic radius  $R_H = 2.7\sigma$  (determined by the radial distribution function). The cubic simulation box had a size of  $L = 41.04\sigma$  with periodic boundary conditions in all three dimensions. To sample at the correct temperature, we equilibrated the system using a Langevin thermostat. The system was then integrated with a time step of  $\Delta t_{MD} = 0.001$  in the  $NVE$ -ensemble. The simulations were performed with the simulation package *Lammps*<sup>30</sup>. We use this system as toy model to analyze the performance of different memory reconstruction techniques. Therefore, we will not systematically study finite size effects or the radius dependence of the memory kernel. Nevertheless, we verified that the data are in quantitative agreement with the predictions of hydrodynamic theory<sup>25</sup> (curves not shown here).

In the following, we compare different memory reconstruction techniques: (i) Backward orthogonal dynamics<sup>15,16</sup> to first and second order (see also App. B), (ii) the inverse Volterra method<sup>4</sup> both with a time discretization  $\Delta t_{MD} = 0.001$  and (iii) the IMRV method with a time step  $\Delta t_{GLE} = 0.005$  and a correction time  $t_{cor} = 0.05$ . In the numerical simulation, the memory kernel was evaluated until a cutoff time  $t_{cut} = 2.5$  resulting in a memory sequence  $K_m$  of 500 elements.

The results for the memory kernel are illustrated in Fig. 3 (top panel). The figure shows that all methods reproduce a similar memory kernel and therefore the same dynamical properties. This impression is supported when comparing the VACFs produced in GLE simulations using the different memory kernels (see Fig. 3, bottom panel). Only the first order Backward orthogonal dynamics shows small deviations from the MD results.

To test the time discretization effects in the considered methods, we reduced the discretization of the reconstruction to  $\Delta t_{MD} = 0.005$ . Fig. 4 confirms the observations we made in the previous paragraph. The first order Backward orthogonal dynamics suffers from a significant time step dependence that was already pointed out in Ref.<sup>15</sup>. However, the accuracy of the Backward orthogonal dynamics can be very much improved by applying the second order scheme. The figure also suggests a small time step dependence of the inverse Volterra method but nonetheless the results are still fairly accurate. Compared to the other methods, the IMRV methods clearly performs best, the results are almost indistinguishable from the original MD data.

We now want to test the limits of the IMR and apply it with discretization time steps  $\Delta t_{GLE} = 0.02 - 0.2$ . The correction time was chosen to be  $t_{cor} = 2\Delta t_{GLE}$ . For the largest time steps we only updated the correction window (see Eq. 9) every second or fourth iteration to approximately conserve an effective correction time  $t_{cor} = 0.05$ . Surprisingly, the IMRV converged for all considered time steps (see Fig. 5). The discretized memory kernels are naturally different for larger time steps because each value represents effectively the average of the continuous memory kernel over a broader time win-

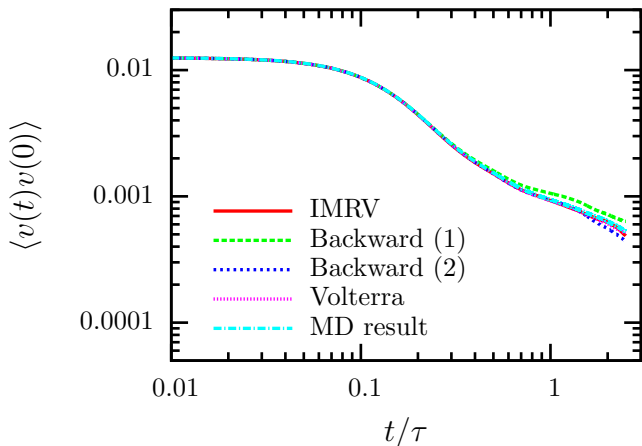
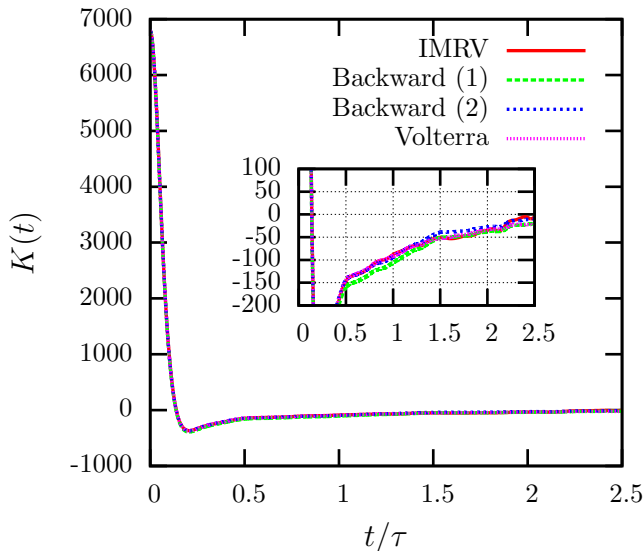


FIG. 3. Memory kernel (top panel) and VACF (bottom panel) of a single Brownian particle as described in Sec. IV. The kernel was reproduced using the iterative memory reconstruction (IMRV), the Backward orthogonal dynamics to first order<sup>15</sup> and second order<sup>16</sup> and the inverse Volterra method<sup>4</sup> with a time discretization  $\Delta t_{\text{MD}} = 0.001$ .

dow. In the above problem the IMRF did not converge for larger time steps. The reason is the already mentioned discretization error in the force determination (see Eq. 10).

These studies show the strength of the iterative reconstruction. Independent of the time step it can always find a memory kernel that reproduces precisely the desired VACF (or FACF). This is a very noteworthy and important result because it enables us to significantly increase the time step of the coarse-grained simulations. In practice, this means that the choice of the time step will not be limited by the dissipative and stochastic part of the GLE in most cases, but by the conservative part. Since the computational time of the algorithm scales approximately linear with the time step (because the number of

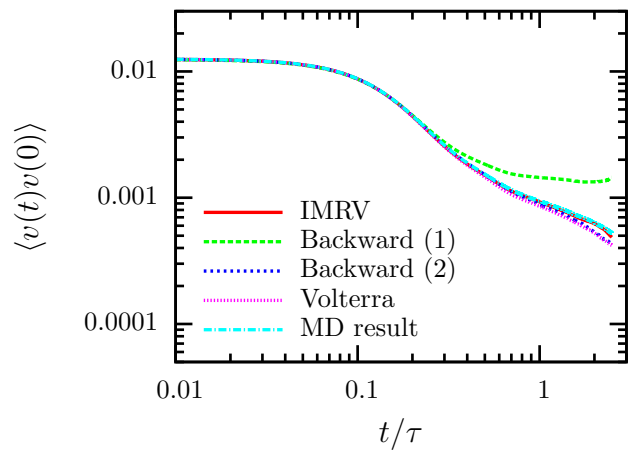


FIG. 4. VACF of a single Brownian particle as described in Fig. 3. The time discretization was  $\Delta t_{\text{MD}} = 0.005$ .

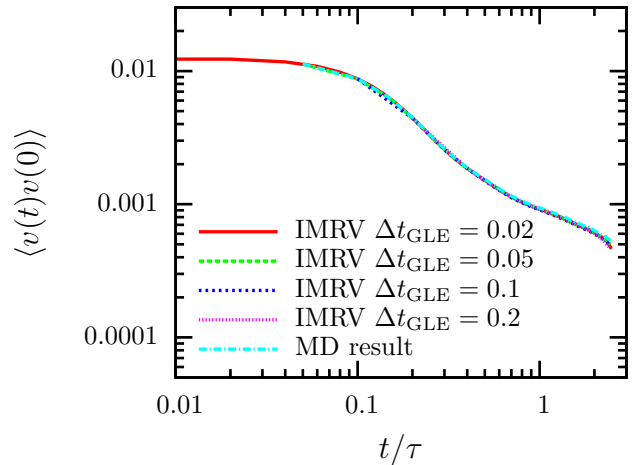
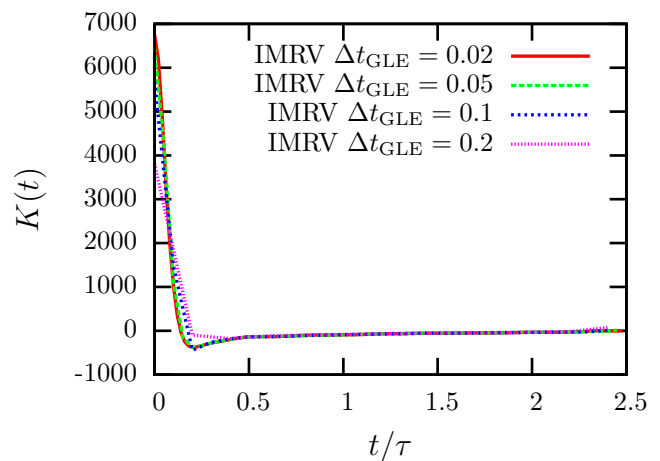


FIG. 5. Memory kernel (top panel) and VACF (bottom panel) of a single Brownian particle as described in Sec. IV. The iterative memory reconstruction (IMRV) was applied for various different integration time steps. The results of the IMRV are more accurate, than the straightforward integration of the GLE using the original memory kernel (see Fig. 2).

elements in the sum in Eq. 23 is reduced) the efficiency of the sampling scales quadratically. Therefore, the proposed coarse-graining of the timescale is essential for the applicability and efficiency of the generalized Langevin equation.

## V. CONCLUSION

In this paper we proposed a new technique to reconstruct memory kernels from molecular dynamics simulations. Instead of using direct inversion techniques, we developed an iterative algorithm to determine memory kernels that precisely reproduce the fine-grained dynamics. We showed that this procedure opens up an elegant way of finding discretizations of memory kernels that reproduce realistic dynamics even for very large time steps. The consequence is that the size of the time step for physical models will be restricted by the conservative forces, and not by the resolution of the memory kernel. Additionally, the strength of the method is its flexibility in optimizing one target observable (e.g. VACF or FCF), instead of being restricted to a predetermined inversion. This could grant some freedom to increase the representativity and transferability of the coarse-grained model, similar to related work concerning IBI (e.g. Ref.<sup>31</sup>). In particular, the IMR could allow the construction of a multi-criteria algorithm by optimizing a weighted combination of several time correlation functions.

To use the iterative reconstruction with good accuracy, we proposed a novel discretization scheme for the GLE. The tests suggest that the new integrator performs much better than a generalization of the popular BBK integrator, and that the velocity is integrated accurately even for very large time steps. Additionally, we derived a discretization of the fluctuation-dissipation theorem for the GLE. We can show that the discretized version counters intuition and introduces a scaling factor of 2 to the instantaneous fluctuations, similar to the Langevin equation. This result is often overlooked and needs to be generally considered when performing simulations with the generalized Langevin equation.

With the present work as foundation, we plan to study more complex systems of interacting colloids. This will introduce pair potentials and pairwise memory kernel, similar to Refs.<sup>2,3</sup>. It will be interesting to investigate whether we can combine iterative methods to determine pair potentials with the iterative memory reconstruction. Additionally, we plan to systematically analyze the mathematical foundations of the iterative algorithm to gain a deeper understanding of its properties and convergence.

## ACKNOWLEDGMENT

The authors want to thank Jay D. Schieber and Oded Farago for helpful discussions. This work was funded by the German Science Foundation within project A3 of the SFB TRR 146. Computations were carried out on the Mogon Computing Cluster at ZDV Mainz.

## Appendix A: Colored noise

The goal is to generate correlated Gaussian distributed random numbers  $\beta_n$ , that fulfill the FDT,

$$\langle \beta_{n+m} \beta_n \rangle = k_B T a_m K_m \Delta t, \quad (\text{A1})$$

with the discretized memory sequence  $K_m$  for  $m = 0, \dots, N-1$  and the parameter  $a_m$  defined by  $a_0 = 2$  and  $a_m = 1$  for  $m \neq 0$  (see Sec. III).

First, we introduce the real parameter  $\alpha_s$  for  $s = -N+1, \dots, N-1$ , defined by,

$$a_m K_m \equiv \sum_{s=-N+1}^{N-1} \alpha_s \alpha_{s+m}, \quad (\text{A2})$$

where  $\alpha_{s+m} = \alpha_{s+m-2N+1}$  if  $s+m \geq N$ . It can now be shown that for a sequence of uncorrelated Gaussian distributed random numbers  $W_n$ , the relation,

$$\beta_n = \sqrt{k_B T \Delta t} \sum_{s=-N+1}^{N-1} \alpha_s W_{n+s}, \quad (\text{A3})$$

generates random numbers with the target correlation function. Therefore, the challenge is to determine the parameter  $\alpha_s$ . This can be achieved by applying the discrete Fourier transform (DFT) on the memory sequence,

$$\hat{K}_k = \sum_{m=-N+1}^{N-1} a_m K_m e^{-ikm \frac{2\pi}{2N-1}}, \quad (\text{A4})$$

with  $K_{-m} = K_m$ . We will now define

$$\hat{\alpha}_k = \sqrt{\hat{K}_k} \quad (\text{A5})$$

and consequentially determine the parameter  $\alpha_s$  by inverse DFT

$$\alpha_s = \frac{1}{2N-1} \sum_{k=-N+1}^{N-1} \hat{\alpha}_k e^{iks \frac{2\pi}{2N-1}}. \quad (\text{A6})$$

It is straightforward to show that the definitions in Eqs. (A2) and (A5) are consistent.

We close with one important comment. In practical applications, a small number of coefficients  $\hat{K}_k$  are slightly below zero. While this is theoretically impossible, because the modes connected to the negative values  $\hat{K}_k$  would be instable, the problem can occur due



to discretization errors if the memory sequence  $K_m$  has “edges”. The solution used in this paper is zeroing all values  $\tilde{K}_k < 0$ . Although, this leads to very small deviations from the FDT theorem ( $\approx 0.01\%$ ), the difference did not have any physical consequence.

## Appendix B: Backward orthogonal dynamics

In Ref.<sup>15</sup> Carof *et al.* introduced the idea to calculate the memory kernel by reconstruction of the random force  $\partial\mathbf{F}(t)$  using molecular dynamics simulations. Their method is directly based on the Mori-Zwanzig formalism. In the context of the present work, we have extended their Backward orthogonal dynamics scheme to second order. The derivation is presented in the following.

The memory kernel can be calculated using the relation,

$$K(t) = \left\langle \sum_i \tilde{F}_i^{\bar{t}} F_i(t) \right\rangle, \quad (\text{B1})$$

with the force  $F_i(t) = M_i v_i(t)$  and the projected force  $\tilde{F}_i^{\bar{t}}$  on particle  $i$ . The projected force is defined by

$$\tilde{F}_i^{\bar{t}} = F_i(0) + \int_0^{\bar{t}} du F_i(u) \frac{\langle M_i v_i(u) \tilde{F}_i^{\bar{u}} \rangle}{\langle M_i^2 v_i(0)^2 \rangle}. \quad (\text{B2})$$

These equations correspond to Eqs. (B2) and (B5) in Ref.<sup>15</sup> with  $A(t) = B(t) = F_i(t)$ . To reconstruct the dynamics of the projected force  $\tilde{F}_i^{\bar{t}}$  we rewrite Eq. B2 into a discretized integration scheme:

$$\tilde{F}_i^{\bar{t}+\Delta t} = \tilde{F}_i^{\bar{t}} + \int_t^{\bar{t}+\Delta t} du F_i(u) \frac{\langle M_i v_i(u) \tilde{F}_i^{\bar{u}} \rangle}{\langle M_i^2 v_i(0)^2 \rangle}. \quad (\text{B3})$$

The first order Backward scheme can now be derived by approximating the integral using the rectangle method, leading to

$$\tilde{F}_i^{\bar{t}+\Delta t} = \tilde{F}_i^{\bar{t}} + F_i(t) \alpha(t) \Delta t + \mathcal{O}(\Delta t^2), \quad (\text{B4})$$

with

$$\alpha(t) = \frac{\langle M_i v_i(t) \tilde{F}_i^{\bar{t}} \rangle}{\langle M_i^2 v_i(0)^2 \rangle}, \quad (\text{B5})$$

and  $\tilde{F}_i^{\bar{0}} = F_i(0)$ . The algorithm to evaluate this equation is described in detail in Ref.<sup>15</sup> Eqs. (21)-(27).

However, it is also possible to approximate Eq. B3 with the trapezoidal rule, similar to the derivation of the sec-

ond order Forward scheme<sup>16</sup>:

$$\begin{aligned} \tilde{F}_i^{\bar{t}+\Delta t} &= \tilde{F}_i^{\bar{t}} + F_i(t) \frac{\langle M_i v_i(t) \tilde{F}_i^{\bar{t}} \rangle}{\langle M_i^2 v_i(0)^2 \rangle} \frac{\Delta t}{2} \\ &+ F_i(t + \Delta t) \frac{\langle M_i v_i(t + \Delta t) \tilde{F}_i^{\bar{t}+\Delta t} \rangle}{\langle M_i^2 v_i(0)^2 \rangle} \frac{\Delta t}{2} + \mathcal{O}(\Delta t^3). \end{aligned} \quad (\text{B6})$$

By taking the correlation on both sides of Eq. B6 with  $M_i v_i(t + \Delta t)$  and inserting the solution back into Eq. B6, we can finally write down the propagator for the second order Backward orthogonal dynamics:

$$\begin{aligned} \tilde{F}_i^{\bar{t}+\Delta t} &= \tilde{F}_i^{\bar{t}} + F_i(t) \alpha(t) \frac{\Delta t}{2} \\ &+ \frac{F_i(t + \Delta t)}{1 - \frac{\Delta t}{2} \kappa} \left[ \zeta(t) + \epsilon \alpha(t) \frac{\Delta t}{2} \right] \frac{\Delta t}{2} + \mathcal{O}(\Delta t^3), \end{aligned} \quad (\text{B7})$$

with

$$\zeta(t) = \frac{\langle M_i v_i(t + \Delta t) \tilde{F}_i^{\bar{t}} \rangle}{\langle M_i^2 v_i(0)^2 \rangle}, \quad (\text{B8})$$

$$\kappa = \frac{\langle M_i v_i(t) F_i(t) \rangle}{\langle M_i^2 v_i(0)^2 \rangle}, \quad (\text{B9})$$

$$\epsilon = \frac{\langle M_i v_i(t + \Delta t) F_i(t) \rangle}{\langle M_i^2 v_i(0)^2 \rangle}. \quad (\text{B10})$$

A numerical comparison between the presented extension to the Backward orthogonal dynamics and the already established Forward orthogonal dynamics can be found in Fig. 6. As expected, the second order algorithm performs much better than the first order scheme. Additionally it can be shown that the results of the two orthogonal algorithms are numerically identical. The Backward scheme, however, has the advantage to be applicable on-the-fly.

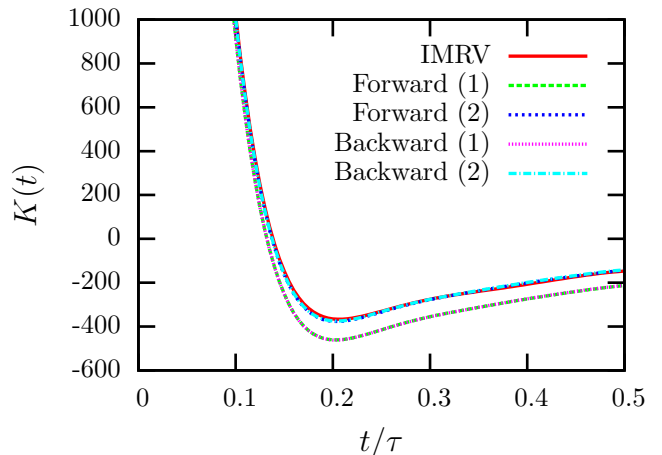


FIG. 6. Comparison of memory kernels reconstructed with orthogonal dynamics algorithms of first (1) and second (2) order. The time discretization was  $\Delta t_{\text{MD}} = 0.005$ . The IMRV result is included from Fig. 3 as reference. The Forward (1) and Backward (1) as well as Forward (2) and Backward (2) curves lie exactly on top of each other.



- \* jungge@uni-mainz.de  
† hanke@mathematik.uni-mainz.de  
‡ friederike.schmid@uni-mainz.de
- <sup>1</sup> Hijón, C.; Español, P.; Vanden-Eijnden, E.; Delgado-Buscalioni, R. Mori-Zwanzig formalism as a practical computational tool. *Faraday Discuss.* **2010**, *144*, 301–322.
  - <sup>2</sup> Li, Z.; Bian, X.; Li, X.; Karniadakis, G. E. Incorporation of memory effects in coarse-grained modeling via the Mori-Zwanzig formalism. *J. Chem. Phys.* **2015**, *143*, 243128.
  - <sup>3</sup> Li, Z.; Lee, H. S.; Darve, E.; Karniadakis, G. E. Computing the non-Markovian coarse-grained interactions derived from the Mori-Zwanzig formalism in molecular systems: Application to polymer melts. *J. Chem. Phys.* **2017**, *146*, 014104.
  - <sup>4</sup> Shin, H. K.; Kim, C.; Talkner, P.; Lee, E. K. Brownian motion from molecular dynamics. *Chem. Phys.* **2010**, *375*, 316–326.
  - <sup>5</sup> Zwanzig, R. Memory Effects in Irreversible Thermodynamics. *Phys. Rev.* **1961**, *124*, 983–992.
  - <sup>6</sup> Mori, H. Transport, Collective Motion, and Brownian Motion. *Prog. Theor. Phys.* **1965**, *33*, 423–455.
  - <sup>7</sup> Zwanzig, R. *Nonequilibrium statistical mechanics*; Oxford University Press, 2001.
  - <sup>8</sup> Kinjo, T.; Hyodo, S.-a. Equation of motion for coarse-grained simulation based on microscopic description. *Phys. Rev. E* **2007**, *75*, 051109.
  - <sup>9</sup> Kubo, R. The fluctuation-dissipation theorem. *Reports Prog. Phys.* **1966**, *29*, 306.
  - <sup>10</sup> Ceriotti, M.; Bussi, G.; Parrinello, M. Colored-Noise Thermostats à la Carte. *J. Chem. Theory Comput.* **2010**, *6*, 1170–1180.
  - <sup>11</sup> Carof, A.; Marry, V.; Salanne, M.; Hansen, J.-P.; Turq, P.; Rotenberg, B. Coarse graining the dynamics of nanoconfined solutes: the case of ions in clays. *Mol. Simul.* **2014**, *40*, 237–244.
  - <sup>12</sup> Córdoba, A.; Indei, T.; Schieber, J. D. Elimination of inertia from a Generalized Langevin Equation: Applications to microbead rheology modeling and data analysis. *J. Rheol.* **2012**, *56*, 185–212.
  - <sup>13</sup> Guenza, M. Many chain correlated dynamics in polymer fluids. *J. Chem. Phys.* **1999**, *110*, 7574–7588.
  - <sup>14</sup> Córdoba, A.; Schieber, J. D.; Indei, T. The effects of hydrodynamic interaction and inertia in determining the high-frequency dynamic modulus of a viscoelastic fluid with two-point passive microrheology. *Phys. Fluids* **2012**, *24*, 073103.
  - <sup>15</sup> Carof, A.; Vuilleumier, R.; Rotenberg, B. Two algorithms to compute projected correlation functions in molecular dynamics simulations. *J. Chem. Phys.* **2014**, *140*, 124103.
  - <sup>16</sup> Lesnicki, D.; Vuilleumier, R.; Carof, A.; Rotenberg, B. Molecular Hydrodynamics from Memory Kernels. *Phys. Rev. Lett.* **2016**, *116*, 147804.
  - <sup>17</sup> Schnurr, B.; Gittes, F.; MacKintosh, F. C.; Schmidt, C. F. Determining Microscopic Viscoelasticity in Flexible and Semiflexible Polymer Networks from Thermal Fluctuations. *Macromolecules* **1997**, *30*, 7781–7792.
  - <sup>18</sup> Lei, H.; Baker, N. A.; Li, X. Data-driven parameterization of the generalized Langevin equation. *Proc. Natl. Acad. Sci.* **2016**, *113*, 14183–14188.
  - <sup>19</sup> Fricks, J.; Yao, L.; Elston, T. C.; Forest, M. G. Time-Domain Methods for Diffusive Transport in Soft Matter. *SIAM J. Appl. Math.* **2009**, *69*, 1277–1308.
  - <sup>20</sup> Lyubartsev, A. P.; Laaksonen, A. Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach. *Phys. Rev. E* **1995**, *52*, 3730–3737.
  - <sup>21</sup> Reith, D.; Pütz, M.; Müller-Plathe, F. Deriving effective mesoscale potentials from atomistic simulations. *J. Comput. Chem.* **2003**, *24*, 1624–1636.
  - <sup>22</sup> Grønbech-Jensen, N.; Farago, O. A simple and effective Verlet-type algorithm for simulating Langevin dynamics. *Mol. Phys.* **2013**, *111*, 983–991.
  - <sup>23</sup> Brünger, A.; Brooks, C. L.; Karplus, M. Stochastic boundary conditions for molecular dynamics simulations of ST2 water. *Chem. Phys. Lett.* **1984**, *105*, 495–500.
  - <sup>24</sup> Lee, S. H.; Kapral, R. Two-particle friction in a mesoscopic solvent. *J. Chem. Phys.* **2005**, *122*, 214916.
  - <sup>25</sup> Theers, M.; Westphal, E.; Gompper, G.; Winkler, R. G. From local to hydrodynamic friction in Brownian motion: A multiparticle collision dynamics simulation study. *Phys. Rev. E* **2016**, *93*, 032604.
  - <sup>26</sup> Hauge, E. H.; Martin-Löf, A. Fluctuating hydrodynamics and Brownian motion. *J. Stat. Phys.* **1973**, *7*, 259–281.
  - <sup>27</sup> Thalmann, F.; Farago, J. Trotter derivation of algorithms for Brownian and dissipative particle dynamics. *J. Chem. Phys.* **2007**, *127*, 124109.
  - <sup>28</sup> Ricci, A.; Cicotti, G. Algorithms for Brownian dynamics. *Mol. Phys.* **2003**, *101*, 1927–1931.
  - <sup>29</sup> Barrat, J.-L.; Rodney, D. Portable Implementation of a Quantum Thermal Bath for Molecular Dynamics Simulations. *J. Stat. Phys.* **2011**, *1*, 679–689.
  - <sup>30</sup> Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117*, 1–19.
  - <sup>31</sup> Villa, A.; Peter, C.; van der Vegt, N. F. A. Transferability of Nonbonded Interaction Potentials for Coarse-Grained Simulations: Benzene in Water. *J. Chem. Theory Comput.* **2010**, *6*, 2434–2444.