Non-Gaussian distribution of displacements for Lennard-Jones particles in equilibrium

Aleksandra Pachalieva[1,2,*] and Alexander J. Wagner[3,†]

[1]*Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*
[2]*Department of Mechanical Engineering, Technical University of Munich, 85748 Garching, Germany*
[3]*Department of Physics, North Dakota State University, Fargo, North Dakota 58108, USA*

Most mesoscale simulation methods assume Gaussian distributions of velocity-like quantities. These quantities are not true velocities, however, but rather time-averaged velocities or displacements of particles. We show that there is a large range of coarse-graining scales where the assumption of a Gaussian distribution of these displacements fails, and a more complex distribution is required to adequately express these distribution functions of displacements.

I. INTRODUCTION

A key result of statistical mechanics is that in equilibrium the velocity distribution of particles is given by a Gaussian distribution function, typically referred to as the the Maxwell-Boltzmann distribution and that all the velocities are uncorrelated. Here we focus on the distribution of particle displacements which is often implicitly needed, but has received far less attention. As these displacements can be viewed as time averages of the velocities it is typically assumed that this result generalizes and the particle displacements also follow a Gaussian distribution.

We report here the surprising result that particle displacements can deviate from a Gaussian distribution. We believe that this is the first time this unexpected result has been reported in the literature. This can be of practical relevance for mesoscale simulation methods like Brownian dynamics [1], dissipative particle dynamics [2], stochastic rotation dynamics [3], and the lattice Boltzmann method [4,5] to name but a few. This is because for these methods particle velocities and particle displacements in a finite time are often implicitly equated.

The question of physical displacements of particles has not received a lot of attention but is of general interest in statistical mechanics, as the short-term displacement is often modelled by a random walk. This has been discussed recently by Masoliver *et al.* [6,7].

The paper is structured as follows: In Sec. II, we show the numerical evidence that the distribution of displacements indeed differs from a Gaussian distribution. This is followed by a detailed description of the simulation setup used to obtain the MD data given in Sec. III. In Sec. IV, we show the mismatch between the MD data and the single Gaussian distribution of displacements. We propose two novel probability distribution functions which could be adjusted to match the second- and fourth-order moments of the measured data, respectively, in Secs. V and VI. Since one could be concerned

that the good agreement between the experiments and theory was solely caused by matching these two moments, in Sec. V we show the results for a ballistic-diffusive Gaussian mixture mode, which also matches those two moments but the agreement is much poorer. This demonstrates that the agreement between the Poisson weighted sum of Gaussians probability distribution function proposed in Sec. VI and the measured distribution function is better than simply matching those two moments. Finally, some concluding remarks and future work are mentioned in Sec. VII.

II. MOTIVATION

In typical hydrodynamic systems, the locally conserved quantities are relaxed toward local equilibrium much faster than quantities that can be relaxed through collisions. For these systems the distribution of particle velocities will be close to a Maxwell-Boltzmann distribution corresponding to the local conserved quantities density, momentum, and temperature. This observation is at the core of many descriptions of nonequilibrium thermodynamics. For the Boltzmann equation it leads to an approximation which allows the two-particle collision term to be replaced by a simpler term of relaxing the velocity distribution toward the local Maxwellian distribution. This is known as the Bhatnagar-Gross-Krook (BGK) approximation [8]. In the BGK formalism, the entire local relaxation depends on the details of small deviations from the local equilibrium distribution function.

In the molecular dynamics lattice gas (MDLG) [9,10] context, we measure the distribution function of particle displacements from an underlying MD simulations of Lennard-Jones particles in equilibrium and thus, obtain an equilibrium distribution function for a specific simulation. For the particular application of measuring collisions, it is required to obtain precise measurements of the deviations from equilibrium. We noticed that the collision operator did not appear to relax toward the equilibrium distribution function predicted by Parsa *et al.* [9], but instead it relaxes to a distribution that deviates by a few percentages. This deviation was not previously noticed but since now we were examining

---
*apachalieva@lanl.gov
†alexander.wagner@ndsu.edu

small deviations from equilibrium, these differences between the predicted and measured equilibrium distributions have the same order of magnitude as the nonequilibrium contributions to the distribution function. Since the only ingredient in the analytical prediction of the MDLG equilibrium distribution is the distribution of particle displacements [9], we began to question the validity of the assumption that the distribution of the local displacements was truly given by a Maxwell-Boltzmann distribution, as expected.

This lead us to investigate the distribution of displacements for different finite time steps. For very short time steps Δt, the effect of particle interactions can be neglected and particles simply displace according to their current velocity. Therefore, the particle displacement can be expressed as a function of the velocity and given by $\delta x_j = v_j \Delta t$ for particle j. The Maxwell-Boltzmann distribution function $P_v(v_j)$ as given in Eq. (12) can be expressed in terms of the particle displacements for the limiting case of $\Delta t \to 0$ as

$$P(\delta x_j) = P_v \left(\frac{\delta x_j}{\Delta t} \right), \tag{1}$$

and it is given by a Gaussian distribution which is fully defined by its mean value and standard deviation. Without loss of generality, we set the net momentum of our simulations to zero which corresponds to zero mean value of the distribution function. The standard deviation can be obtained in two ways: measured directly from the MD simulation or approximated from the velocity autocorrelation function. By calculating the mean-squared displacement from an analytical approximation of the velocity autocorrelation function, we obtain a simple dependence including only one parameter. Details about the performed MD simulations, the derivation of the Gaussian distribution function and discussion of the results can be found in Secs. III and IV.

Regardless of the used method to obtain the mean-squared displacement, Fig. 1 shows that the resulting Gaussian functions—$P^{\text{G-T}}(X_i)$ and $P^{\text{G-M}}(X_i)$, do not agree with the measured MD probability distribution function $P^{\text{MD}}(X_i)$. As suspected from our studies of the deviation of nonequilibrium systems from equilibrium [11], the equilibrium distribution functions are close to a Gaussian distribution but they show noticeable deviations from the MD data. We emphasize that even though the disagreement between the two displacement functions is indeed small, it is of the same order of magnitude or larger than the deviation of a nonequilibrium distribution function.

In this paper, we investigate for which time steps the displacement distribution is given by a Maxwell-Boltzmann distribution and when a better description is needed. We found that the Maxwell-Boltzmann function is only valid in the extreme ballistic regime for very short Δt, and in the extreme diffusive regime for very large Δt. In an intermediate regime, the Maxwellian does not capture the distribution of the displacements and introduces an error to the collision operator. This is a practical issue that matters in many mesoscale methods such as Brownian dynamics [1], dissipative particle dynamics [2], stochastic rotation dynamics [3], and the lattice Boltzmann method [4,5].
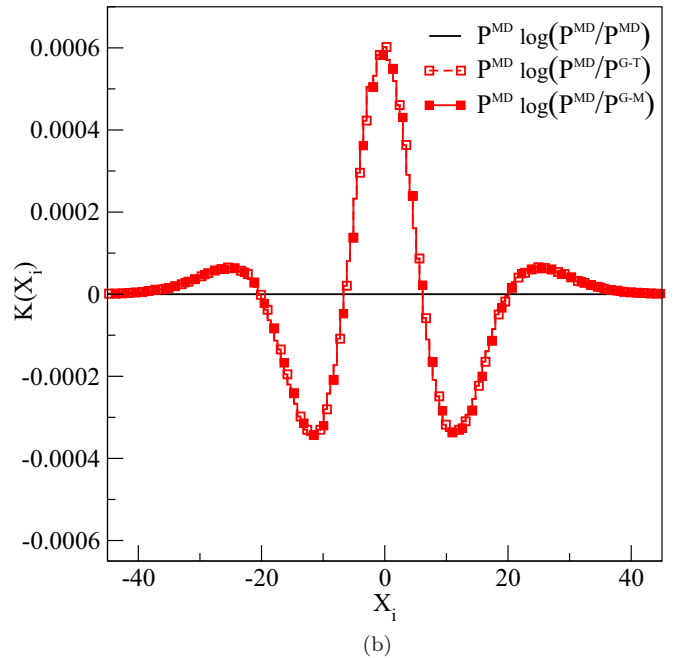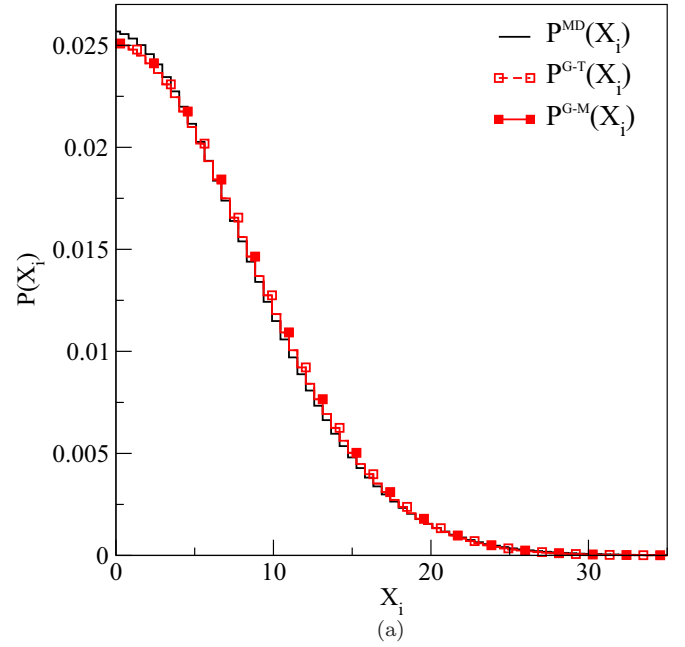

FIG. 1. (a) Displacements probability distribution functions. The solid line (black) depicts a PDF obtained from an MD simulation of LJ particles in equilibrium. The lines with empty or full squares (red) illustrate a Gaussian probability distribution function defined in Eq. (17) with mean-squared displacement obtained from the velocity autocorrelation function as given in Eq. (21) and with mean-squared displacement fitted directly to the MD data, respectively. Only the data for positive velocities have been depicted due to symmetry. Panel (b) shows the difference between the distributions per interval X_i as defined in Eq. (23). The presented data are for the standard parameters used in the paper and a coarse-grained time step $\Delta t = 3.2$.

III. SIMULATION SETUP

We are investigating a system of particles interacting with the standard 6-12 Lennard-Jones (LJ) intermolecular potential

defined as

$$V_{LJ}(r) = 4\varepsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right] \qquad (2)$$

with ε being the potential well depth, σ is the distance at which the inter-particle potential goes to zero, and r is the distance between two particles. We set the particle mass to $m = 1$ and the LJ particle diameter to $\sigma = 1$. All the MD simulations were executed using the open-source molecular dynamics software LAMMPS [12,13] that is developed by Sandia National Laboratories. We performed multiple MD simulations with $N = 99\,856$ particles in a two-dimensional (2D) square with length $L = 1000$ LJ units which corresponds to an area fraction of $\phi = 0.078387$. The area fraction ϕ for circular LJ particles with radius $a = \sigma/2$ is defined as the product of the particle surface area and the number of particles, divided by the square length L of the simulation box. The simulations were initialised with homogeneously distributed particles having kinetic energy that corresponds to a temperature of 20 in LJ units. The the kinetic theory considered in this paper (in particular the Poisson distribution of collision times) are expected to be correct only for dilute systems. Therefore, we focus our attention here on simulations of fairly dilute systems.

We have executed simulations of two-dimensional systems instead of three-dimensional ones to minimize computational cost. For a three-dimensional MD simulation to be computationally feasible, we need to reduce the domain size and adjust the number of particles to recover the same volume fraction as the 2D area fraction mentioned earlier. By reducing the domain size, we put a constraint on the coarse-grained time step Δt and therefore, on the maximum average particle displacement. Thus, it will not be possible to simulate extremely large time steps due to periodic image problems occurring when the particle displacements are larger than half of the simulation length L.

According to the definition of the LJ interaction potential in Eq. (2), we write the timescale as

$$\tau_{LJ} = \sqrt{\frac{m\sigma^2}{\varepsilon}}, \qquad (3)$$

which corresponds to the time in which a particle with kinetic energy of half the potential energy well ε traverses one diameter σ of a LJ particle. It is worth noting that there is a second timescale, i.e., the time it takes a particle with the kinetic energy of $1/2\, k_B T$ to transverse the diameter σ of a LJ particle, which is given by

$$\tau_{th} = \sqrt{\frac{m\sigma^2}{k_B T}} \qquad (4)$$

and we call this scale a thermal timescale. Note that for the temperature of 20 in LJ units, the thermal timescale is smaller than the LJ timescale τ_{LJ} by factor of $1/\sqrt{20} \approx 0.22$.

The simulation setup characterizes a standard semidilute gas in equilibrium with average velocity fixed to zero,

$$Nu_\alpha = \sum_{j=1}^{N} v_{j,\alpha} = 0, \qquad (5)$$

with N being the total number of MD particles.

TABLE I. LAMMPS simulation details.

Δt	MD step size (τ_{LJ})	MD output frequency	Output number	Total MD time steps	Total MD time (τ_{LJ})
0.01	0.0001	100	5000	500 000	50
0.1	0.0001	1 000	5000	5 000 000	500
0.2	0.0001	2 000	5000	10 000 000	1 000
0.4	0.0001	4 000	5000	20 000 000	2 000
0.8	0.0001	8 000	2000	16 000 000	1 600
1.6	0.0001	16 000	2000	32 000 000	3 200
3.2	0.0001	32 000	2000	64 000 000	6 400
6.4	0.0001	64 000	2000	128 000 000	12 800
12.8	0.0001	128 000	2000	256 000 000	25 600
25.6	0.0001	256 000	2000	512 000 000	51 200

The MD step size is set to $0.0001\,\tau_{LJ}$ with total MD simulation time varying from $50\,\tau_{LJ}$ to $51\,200\,\tau_{LJ}$ as shown in Table I. We chose a very small MD step size to ensure high accuracy of the MD simulation data. Our goal is to obtain results for MD simulations with wide regime range—from simulations with mean free time smaller than the time between collisions (ballistic regime) to simulations with much larger mean free path than the time step (diffusive regime). We define the dimensionless coarse-grained time step Δt as a product of the MD step size and the MD output frequency. The coarse-grained time step Δt varies from $0.01\,\tau_{LJ}$ to $25.6\,\tau_{LJ}$. To ensure the MD simulations have reached equilibrium state before we start collecting data, the initial $1\,200\,000$ MD iterations ($120\,\tau_{LJ}$) were discarded. The values depicted in Table I do not include the discarded iterations for clarity. The total number of saved MD iterations of per-particle data differs depending on the coarse-grained time step Δt. For simulations with smaller time step $\Delta t \in [0.01, 0.4]$, we saved 5000 coarse-grained iterations, while for simulations with $\Delta t \in [0.8, 25.6]$, the output number was reduced to 2000 coarse-grained iterations due to their high computational cost. This corresponds to 500 000 MD time steps for the MD simulation with the smallest executed coarse-grained time step $\Delta t = 0.01$, and 512 000 000 MD time steps for the simulation with the largest time step $\Delta t = 25.6$. Since we are simulating a semidilute gas in equilibrium, the total simulation number is irrelevant for the physical properties of the system, because they do not change once the gas has reached equilibrium state. However, we run the simulations for large number of iterations in order to produce large amounts of data which ensures sufficient averaging. An overview of the simulation parameters is given in Table I.

All the simulations were executed in parallel using 32 processors on the Darwin cluster at Los Alamos National Laboratory. The longest executed test case with $\Delta t = 25.6$ took about 120 h wall-clock-time. Depending on the number of coarse-grained iterations (2000 or 5000) the output data files took 20- or 50-GB memory space, respectively. The total memory space used for all LAMMPS simulations exceeded 350 GB.

We have performed a standard molecular dynamics simulation without the use of a thermostat. In the LAMMPS nomenclature this is called a NVE integration. The name is

related to a microcanonical ensemble NVE which is characterized by a constant number of particles (N), a constant volume (V), and a constant energy (E). By using the NVE thermostat, we sample from the microcanonical ensemble, thus we avoid any possible complications coming from the altered equations of motion a thermostat could introduce. However, to ensure the validity of the MD simulations, we have tested the canonical NVT thermostat which was used in earlier papers [9,14] and we have obtained equivalent results. For the canonical ensemble (NVT) the number of particle (N), volume (V) and temperature (T) are constant while the energy of the system fluctuates.

We analyze the collected MD data to recover the probability distribution function (PDF) $P(\delta x)$ of the displacements δx. To obtain an estimate for $P(\delta x)$, we define the particle displacement $\delta x_j(t)$ as

$$\delta x_{j,\alpha}(t + \Delta t) = x_{j,\alpha}(t) - x_{j,\alpha}(t + \Delta t), \qquad (6)$$

where $x_{j,\alpha}(t)$ is the position of particle j at time t, and α refers to the spatial coordinates $\alpha \in \{X, Y\}$ in 2D.

Two probability distribution functions can be compared in different ways: In principle, the PDF is defined as a function or it can be defined through an infinite set of moments. Given the experimental data set, we are of course limited in how well we can estimate the PDF. Therefore, here we use a combination of both approaches.

To obtain the full PDF description, we define a histogram $H(X_i)$ for the discrete displacement intervals X_i as follows:

$$H(X_i) = \frac{\sum_{t=0}^{T} \sum_{j=1}^{N} \Delta_{X_i}(\delta x_j(t))}{TN}, \qquad (7)$$

with number of MD particles N, number of the coarse-grained time steps T and with $\Delta_{X_i}(\delta x_j(t))$ being defined as

$$\Delta_{X_i}(\delta x_j(t)) = \begin{cases} 1, & \text{if } \delta x_j(t) \in X_i \\ 0, & \text{otherwise.} \end{cases} \qquad (8)$$

X_i is a histogram bin and corresponds to a range of $r_i \leqslant \delta x < r_{i+1}$ with i number of bins. In the current publication, we use $i = 200$ number of bins with equal bin width for a certain coarse-grained time step. The bin width depends on the particle displacements and varies for different time step Δt. The first and the last intervals are open at the edges to ensure that there are no empty bins in the histogram and that all possible displacements have been accounted for. This histogram has the following property:

$$\sum_i H(X_i) = 1. \qquad (9)$$

We can then estimate the probability

$$P(\delta x \in X_i) = \int_{\delta x \in X_i} P(\delta x)\, d\delta x \approx H(X_i). \qquad (10)$$

Even though the MD data are in discrete space and by using the collected MD displacements we are able to construct only a histogram as given in Eq. (7), we will further recall it as a probability distribution function. By collecting very large data sets for each coarse-grained time step Δt, we ensure that all histograms are very fine grained and thus agree very well with the underlying PDF as expressed in Eq. (10).

In our MD simulation setup, momentum is conserved. This means that we can also define the momentum through the displacements in addition to Eq. (5). We have

$$u_\alpha = \frac{\langle \delta x_{j,\alpha} \rangle}{\Delta t} = \frac{\sum_{j=1}^{N} \delta x_{j,\alpha}}{N \Delta t} = \frac{\sum_{j=1}^{N} v_{j,\alpha}}{N}, \qquad (11)$$

which are all equivalent. Even though, we have performed simulations with zero initial velocity we could obtain results for different mean velocities u_α by applying a Galilean transformation.

IV. GAUSSIAN DISTRIBUTION FUNCTION

The first theory for the probability distribution function of the displacements that we consider follows the assumption made by Parsa *et al.* [9]. For very short times the particle displacement is given by the velocity v_j of the particle j as $\delta x_j = v_j \Delta t$. Thus, we can write $\lim_{\Delta t \to 0} P(\delta x_j) = P_v(\delta x_j / \Delta t)$ using the probability distribution of the velocity given by

$$P_v(v_j) = \frac{1}{[2\pi k_B T]^{d/2}} \exp\left[\frac{(v_j - u_j)^2}{2k_B T} \right], \qquad (12)$$

where d is the number of dimensions and $k_B T$ is temperature of the system with k_B being the Boltzmann constant. Equation (12) is also known as the Maxwell-Boltzmann distribution which approximates the probability of particle moving in a certain direction. It holds for very short times Δt where the mean free time between two collisions is much shorter than the time step Δt. In this regime, particles undergo simple ballistic motion and the mean-squared displacement in one dimension is

$$\langle (\delta x_\alpha)^2 \rangle^{\text{ball}} = 2k_B T (\Delta t)^2. \qquad (13)$$

Then the probability for collisionless displacements is

$$P^{\text{ball}}(\delta x) = \frac{1}{[2\pi k_B T (\Delta t)^2]^{d/2}} \exp\left[-\frac{(\delta x - u\Delta t)^2}{2k_B T (\Delta t)^2} \right]. \qquad (14)$$

In a diffusive regime, the times are much longer than the mean free time and the particles undergo multiple collisions between time steps. Using the self-diffusion constant D, we write the mean-squared displacement in one dimension as

$$\langle (\delta x_\alpha)^2 \rangle^{\text{diff}} = 2dD(\Delta t). \qquad (15)$$

The probability distribution function of the displacements is given by

$$P^{\text{diff}}(\delta x) = \frac{1}{[4\pi dD(\Delta t)]^{d/2}} \exp\left[-\frac{(\delta x - u\Delta t)^2}{4dD(\Delta t)} \right]. \qquad (16)$$

Since both limiting cases are given by a Gaussian distribution function as shown in Eqs. (14) and (16), Parsa *et al.* [9] suggested that the intermediate probabilities can be well approximated by a single Gaussian distribution defined as

$$P^{\text{G}}(\delta x) = \frac{1}{[2\pi \langle (\delta x_\alpha)^2 \rangle]^{d/2}} \exp\left[-\frac{(\delta x - u\Delta t)^2}{2\langle (\delta x_\alpha)^2 \rangle} \right], \qquad (17)$$

with a mean-squared displacement $\langle (\delta x_\alpha)^2 \rangle$ which can be obtained theoretically or can be measured directly from an MD

simulation. The displacement of a particle is given by

$$\delta x = \int_0^{\Delta t} v(t)\, dt. \tag{18}$$

Now, for a simple semidilute gas system, we express the mean-squared displacement as a function of the velocity autocorrelation function,

$$
\begin{aligned}
\langle (\delta x_\alpha)^2 \rangle &= \left\langle \int dt \int dt'\, v(t) v(t') \right\rangle \\
&= \int dt \int dt'\, \langle v(t - t') v(0) \rangle \\
&= \int_{-\Delta t}^{\Delta t} (\Delta t - \delta t) \langle v(\delta t) v(0) \rangle \, d\delta t \\
&= 2 \int_0^{\Delta t} (\Delta t - \delta t) \langle v(\delta t) v(0) \rangle \, d\delta t.
\end{aligned}
\tag{19}
$$

For gases the velocity autocorrelation function is usually estimated by an exponential decay,

$$\langle v_\alpha(\delta t) v_\alpha(0) \rangle = k_B T \exp\left(-\frac{\Delta t}{\tau} \right), \tag{20}$$

where $k_B T$ is the temperature of the semidilute gas in LJ units and τ is an exponential decay constant which approximates the mean free time [15–19]. The velocity autocorrelation function for the simulated gas system is depicted in Fig. 2(a). We have approximated the mean free time to $\tau \approx 0.728$, which gives a good prediction of the velocity autocorrelation function for early times. As shown in Fig. 2(a), the velocity autocorrelation function has long range contributions for later times ($\Delta t > 4.0$) that is typical for two-dimensional systems [15–19]. The deviations resulting from the long-time tails are noticeable only for later times and larger displacements. In this work, we focus on results for $\Delta t = 3.2$, where the velocity autocorrelation function is well approximated by an exponential decay as defined in Eq. (20). For simplicity, we will therefore neglect the long-time tails shown in Fig. 2(a).

Now, the theoretical mean-squared displacement can be calculated according to Eq. (19) as

$$\langle (\delta x_\alpha)^2 \rangle = 2 k_B T \tau^2 \left[\exp\left(-\frac{\Delta t}{\tau} \right) + \frac{\Delta t}{\tau} - 1 \right]. \tag{21}$$

As shown in Fig. 2(b), this prediction recovers the mean-squared displacement very well. There are small deviations for later times which are not visible in log-log scale. These deviations are result of the long-time tails of the velocity autocorrelation function mentioned previously. This completes the definition of the Gaussian distribution function model using a mean-squared displacement obtained from Eq. (21). In general, $\langle (\delta x_\alpha)^2 \rangle$ can be also measured from the MD simulations. Later, we compare the Gaussian distribution functions obtained using these two approaches.

To estimate how good this PDF matches the MD data, we transform the formulation of $P(\delta x)$ from continuous to discrete using a histogram as defined in Eq. (10). This is realized by integrating the probability distribution function
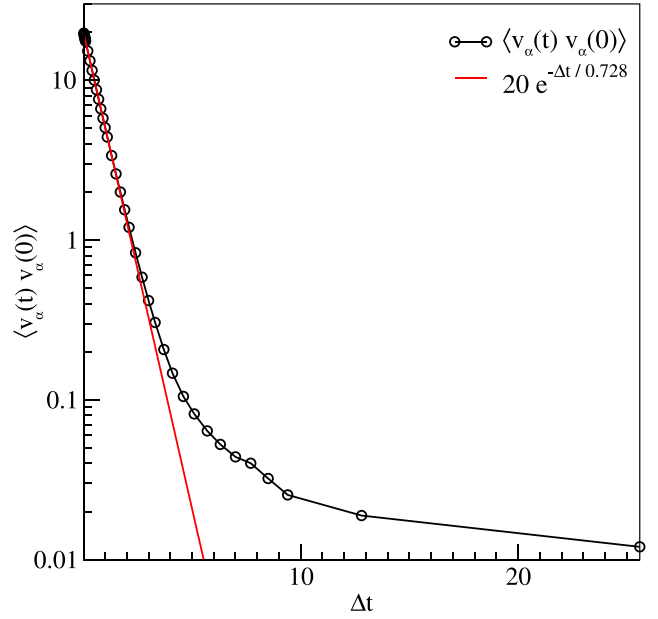
(a)
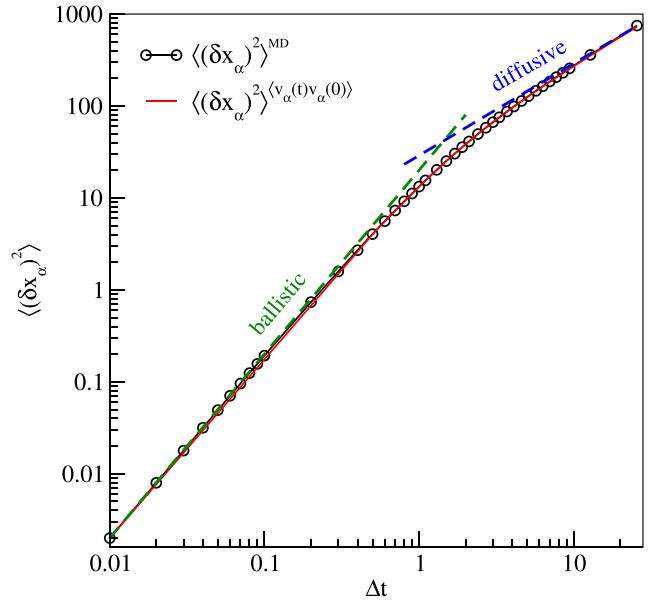
(b)

FIG. 2. (a) Velocity autocorrelation function measured from an MD simulation compared to an exponential decay with $\tau \approx 0.728$ as given in Eq. (20). The long-time tails are typical for two-dimensional systems [15–19]. (b) The mean-squared displacement directly measured from an MD simulation is compared to the theoretical value given in Eq. (21). Notice the two scaling regimes: $\langle (\delta x_\alpha)^2 \rangle \propto \Delta t$ for a ballistic regime with small times and $\langle (\delta x_\alpha)^2 \rangle \propto \Delta t^2$ for a diffusive regime with large times.

over predefined intervals X_i as

$$
\begin{aligned}
H(X_i) &= \int_{r_i}^{r_{i+1}} P(\delta x)\, d\delta x \\
&= \frac{1}{2} \left[\operatorname{erf}\left(\frac{r_i}{\sigma \sqrt{2}} \right) - \operatorname{erf}\left(\frac{r_{i+1}}{\sigma \sqrt{2}} \right) \right],
\end{aligned}
\tag{22}
$$

where X_i corresponds to a rage of $r_i \leqslant \delta x < r_{i+1}$ with number of bins $i = 200$. $\mathrm{erf}(r_i)$ is an error function encountered in integrating the normal distribution function with standard deviation σ and mean equal to zero. Using a histogram to compare two PDFs is a convenient method to analyze precisely where two or more distributing functions diverge.

To analyze how well the Gaussian distribution function fits the MD displacements in the transition regime, we consider a time step of $\Delta t = 3.2$. In Fig. 1(a), the MD displacements (black line) are plotted alongside a Gaussian distribution function $P^{\mathrm{G}\text{-}T}(X_i)$ with theoretical mean-squared displacement (red dashed line with empty squares) and a Gaussian distribution function $P^{\mathrm{G}\text{-}M}(X_i)$ with measured mean-squared displacement (red line with full squares). Both Gaussian distribution functions give an adequate prediction of the MD displacements distribution function; however, there are visible discrepancies at about 5%. Even though the deviations between the MD data and the proposed Gaussian distribution functions are small, they are of significant importance when examining nonequilibrium behavior and looking at small deviations from equilibrium.

Since the deviations between the Gaussian PDFs and the MD simulation data are relatively small, the following function is used to quantify more precisely the discrepancies:

$$K(X_i) = K(R \parallel Q) = R(X_i) \log \left[\frac{R(X_i)}{Q(X_i)} \right], \qquad (23)$$

where $R(X_i)$ and $Q(X_i)$ are probability distributions over an interval X_i. By performing a sum over all the bins X_i, we obtain the well-known Kullback-Leibler (KL) divergence [20] defined as

$$D_{\mathrm{KL}}(R \parallel Q) = \sum_i R(X_i) \log \left[\frac{R(X_i)}{Q(X_i)} \right]. \qquad (24)$$

The KL divergence measures the discrepancies of one probability distribution function to another. It is always nonnegative $D_{\mathrm{KL}}(R \parallel Q) \geqslant 0$ or equal to zero if and only if the probability distribution functions are identical $R(X_i) = Q(X_i)$ [20].

In Fig. 1(b), we show the discrepancies between the Gaussian probability distribution functions and the MD data per bin element X_i measured using Eq. (23). The solid line (black) depicts $K(P^{\mathrm{MD}} \parallel P^{\mathrm{MD}})$ which is zero by construction. The lines with full or empty symbols (red) display the divergence between the MD data and the Gaussian distribution functions with theoretical or measured mean-squared displacement, respectively. Note here that the $K(X_i)$ measure identifies both positive and negative deviations (which is necessary, since the integral of both probability distribution functions is 1) but as long as there is any deviation, the integral (or sum) in Eq. (24) always leads to a positive value. We can see a clear structure in the error of the MD data and the two Gaussian probability distribution functions. Thus, we conclude that a single Gaussian distribution function with the same standard deviation, being measured or theoretically obtained from the velocity autocorrelation function, differs significantly from the MD data in the intermediate regime.

The Kullback-Leibler divergence of the PDF models and the MD data are illustrated in Fig. 3. The divergence is calculated for a variety of time steps $\Delta t \in [0.01, 25.6]$. In
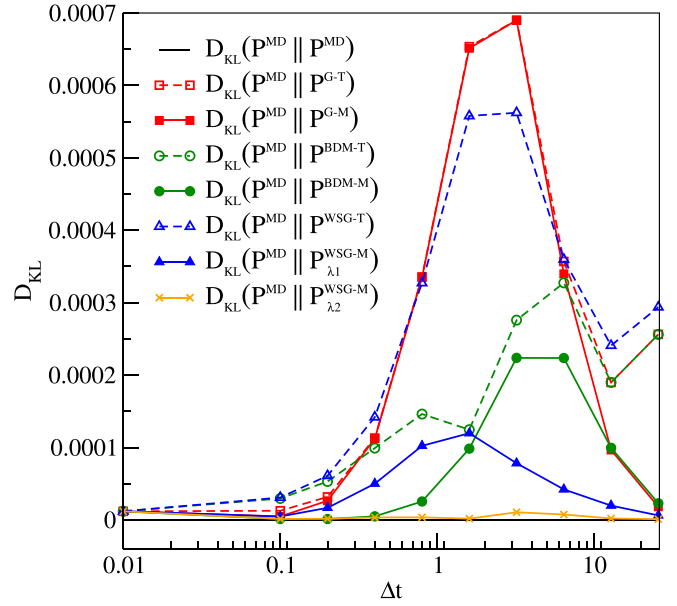

FIG. 3. Kullback-Leibler divergence results: Empty or full squares (red) for $D_{\mathrm{KL}}(P^{\mathrm{MD}} \parallel P^{\mathrm{G}\text{-}T})$ and $D_{\mathrm{KL}}(P^{\mathrm{MD}} \parallel P^{\mathrm{G}\text{-}M})$ discussed in Sec. IV; empty or full circles (green) for $D_{\mathrm{KL}}(P^{\mathrm{MD}} \parallel P^{\mathrm{BDM}\text{-}T})$ and $D_{\mathrm{KL}}(P^{\mathrm{MD}} \parallel P^{\mathrm{BDM}\text{-}M})$ discussed in Sec. V; empty or full triangles (blue) for $D_{\mathrm{KL}}(P^{\mathrm{MD}} \parallel P^{\mathrm{WSG}\text{-}T})$ and $D_{\mathrm{KL}}(P^{\mathrm{MD}} \parallel P^{\mathrm{WSG}\text{-}\lambda_1})$, and $\times$ symbols (yellow) for $D_{\mathrm{KL}}(P^{\mathrm{MD}} \parallel P^{\mathrm{WSG}\text{-}\lambda_2})$ discussed in Sec. VI. The $D_{\mathrm{KL}}(P^{\mathrm{MD}} \parallel P^{\mathrm{MD}})$ divergence (black line) is zero by definition and it is shown just as a comparison. All displacements PDFs show small error for very small Δt (ballistic regime) and for large Δt (diffusive regime). However, in the transition regime only the $P^{\mathrm{WSG}\text{-}\lambda_2}(X_i)$ distribution function with average number of collisions λ_2 gives a satisfactory description of the measured MD distribution function. The KL divergence is calculated for all time steps $\Delta t \in [0.01, 25.6]$ considered in this publication.

the current section, we focus on the two KL divergence measures $D_{\mathrm{KL}}(P^{\mathrm{MD}} \parallel P^{\mathrm{G}\text{-}T})$ and $D_{\mathrm{KL}}(P^{\mathrm{MD}} \parallel P^{\mathrm{G}\text{-}M})$ depicted by lines with full or empty squares (red), respectively. As expected, for purely ballistic test cases the constructed Gaussian distribution functions match very well the PDF obtained from MD data. In the transition regime, the estimated divergence increases rapidly and reaches a peak at $\Delta t = 3.2$, which indicates that the MD displacement function cannot be captured using a single Gaussian distribution function. For $\Delta t = 25.6$, the $K(P^{\mathrm{MD}} \parallel P^{\mathrm{G}\text{-}M})$ divergence is close to zero and we conclude that the simulation has reached diffusive regime. For some of the considered time steps, $P^{\mathrm{G}\text{-}M}(X_i)$ delivers slightly better results in comparison to $P^{\mathrm{G}\text{-}T}(X_i)$ but the improvement is not significant. For the particular case of $\Delta t = 3.2$, there is no visible difference between the two Gaussian distribution functions, which explains the complete overlap of the $K(P^{\mathrm{MD}} \parallel P^{\mathrm{G}\text{-}T})$ and $K(P^{\mathrm{MD}} \parallel P^{\mathrm{G}\text{-}M})$ results shown in Fig. 1(b).

To obtain a better theoretical formulation for the distribution of the equilibrium LJ displacements, we need to analyze rigorously the displacements' distribution function obtained from the MD data. One way to distinguish between two distribution functions is by looking at their moments. By estimating the PDF using the moments of the MD displacements,

we eliminate the small error introduced by the histogram in Eq. (10). From the MD simulation data, we calculate the kth moment as

$$\mu_k = \langle (\delta x)^k \rangle. \qquad (25)$$

Since we are looking at an ensemble average of particle displacements, the moments μ_k can be averaged in space and in time, leading to the following approximation

$$\mu_k = \frac{\sum_{t=1}^{T} \sum_{j=1}^{N} [x_j(t + \Delta t) - x_j(t)]^k}{TN} \qquad (26)$$

with N being the number of MD particles and T being the number of the coarse-grained time steps. The zeroth moment is given simply by the normalization as $\mu_0 = 1$. The first moment defines the average velocity u_α, which in our simulation setup is zero and leads to zero first- and third-order moments $\mu_1 = \mu_3 = 0$ due to symmetry. The second moment μ_2 is known in statistics as the variance or the mean-squared displacement and is given by $\mu_2 = \langle (\delta x)^2 \rangle$. The fourth moment $\mu_4 = \langle (\delta x)^4 \rangle$ is called kurtosis and it is a measure for the "tailedness" of a probability distribution function.

A probability distribution function is defined uniquely through an infinite set of moments. Generally, the better moments match, the better the distributions agree, and the higher order a moment is the less important it tends to be. It is therefore reasonable that we examine the agreements of the moments. The zeroth moment corresponds to normalization and always matches. The second moment should always match, but small errors can occur for theoretical distributions that use Eq. (21). The fourth-order moments at this point are unconstrained, and therefore the deviation of this moment from the experimental one should give a good estimate of the accuracy of the theoretical distribution. We therefore focus on the first two nontrivial moments—μ_2 and μ_4. The moments μ_0, μ_1, and μ_3 have been measured for completeness, but their value for LJ particles in equilibrium are expected to be $\mu_0 = 1$ and $\mu_1 = \mu_3 = 0$ for symmetry reasons.

As mentioned previously, the probability distribution function $P^G(\delta x)$ in Eq. (17) could be calculated using a theoretical or a measured $\langle (\delta x)^2 \rangle$. We measured the second and fourth moments of the Gaussian distribution functions and compared their deviation from the MD moments as shown in Fig. 4. The error is calculated in percentage.

The Gaussian distribution function with theoretical mean-squared displacement fails to reconstruct the second- and the fourth-order moments. The second-order moment error, depicted with a dashed line (black), is relatively small (below 3%). This error rapidly increases with larger time steps and reaches its highest point at $\Delta t = 25.6$. The $P^{G-T}(\delta x)$ fourth-order moment error is depicted in Fig. 4 as dashed line (red) with empty squares. The μ_4^{G-T} error is much larger than the μ_2^{G-T} error and increases very fast in the transition regime.

The second-order moment of $P^{G-M}(X_i)$ matches the MD second-order moment by construction. The fourth-order moment μ_4^{G-M}, however, differs from the measured fourth-order moment as shown in Fig. 4 (red line with full squares). For $\Delta t \in [0.8, 1.6]$, $P^{G-M}(\delta x)$ has a slightly larger fourth-order moment error than $P^{G-T}(\delta x)$. Unlike μ_4^{G-T}, which does not decrease with larger time steps, the μ_4^{G-M} error is large in
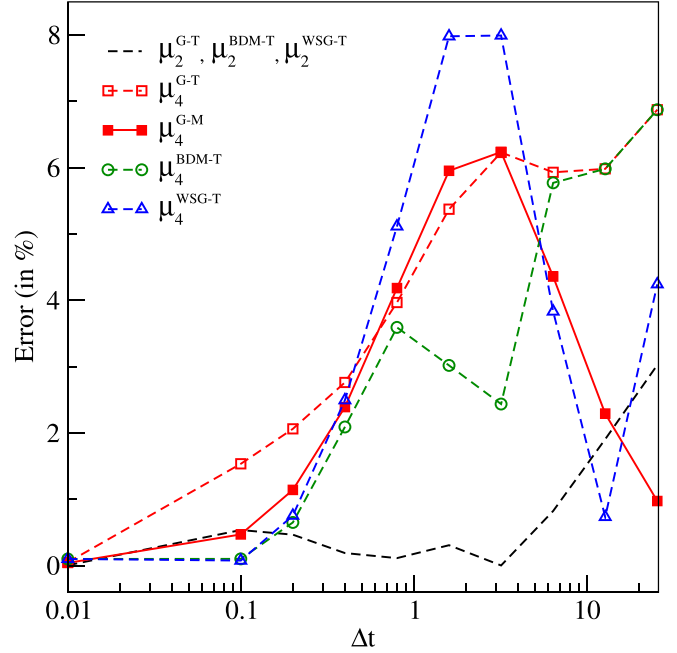

FIG. 4. Second- and fourth-order moment error calculated between the MD simulation data and the theoretical probability distribution functions. The second-order error is equivalent for all theoretical PDF models. The fourth-order error varies: The $P^{G-T}(\delta x)$ and $P^{G-M}(\delta x)$ errors are discussed in Sec. IV (red lines with empty or full squares); the $P^{BDM-T}(\delta x)$ errors discussed in Sec. V (green lines with circles); and the $P^{WSG-T}(\delta x)$ error is discussed in Sec. VI (blue lines with triangles). For some of the proposed distribution functions the second and the fourth-order moments have been fitted to the measured MD moments. These PDFs have zero second- and fourth-order error by construction and, therefore, they have not been depicted. The presented data are for the standard parameters used in the paper and a time step $\Delta t = 3.2$.

the transition regime and decreases to less than 1% for later times. We assume that the larger μ_4^{G-T} error is related to the larger second-order moment error of $P^{G-T}(\delta x)$. Figure 4 shows that the μ_4^{G-M} error is very small for early and late times which indicates that the Gaussian description with measured mean-squared displacement is valid for extreme ballistic and diffusive regimes.

Considering these results, we conclude that a single Gaussian distribution function cannot recover the MD displacements distribution function in a transition regime. In the following section, we construct a Gaussian mixture model which can be adjusted to capture better the MD simulation data.

V. BALLISTIC-DIFFUSIVE DISTRIBUTION FUNCTION

Going back to the assumption made by Parsa *et al.* [9], we take a slightly different approach by approximating the displacements PDF using a Gaussian mixture model with two components. The first component is a distribution function in a ballistic regime given by Eq. (14), while the second component is a distribution function in a diffusive regime defined in Eq. (16). We call this formulation ballistic-diffusive mixture

(BDM) model and define it as

$$P^{\text{BDM}}(\delta x) = \exp\left(-\frac{\Delta t}{\tau}\right) P^{\text{ball}}(\delta x)$$
$$+ \left[1 - \exp\left(-\frac{\Delta t}{\tau}\right)\right] P^{\text{diff}}(\delta x), \qquad (27)$$

where the ratio $\Delta t/\tau$ relates to the average number of collisions within a time interval Δt. The mean free time τ can be evaluated from the velocity autocorrelation function as given in Eq. (20). As shown in Fig. 2(a), the mean free time is estimated to $\tau \approx 0.728$ which agrees well with the measured velocity autocorrelation function for early times.

In a transition regime, the BDM model receives contributions from the ballistic and from the diffusive Gaussian distribution functions. The mixing coefficient $\exp(-\Delta t/\tau)$ depends on the time step and controls the ratio of the two probability distribution functions. For infinite small or infinite large time steps, $P^{\text{BDM}}(\delta x)$ is reduced to a single Gaussian distribution given by Eq. (14) or Eq. (16), respectively.

For the BDM model in Eq. (27), the ballistic contribution is fully defined by the simulation setup with standard deviation equal to $2k_B T (\Delta t)^2$ as given in Eq. (13). For the diffusive part $P^{\text{diff}}(\delta x)$, one could attempt to simply relate it to the self-diffusion constant D. This does not give the correct second-order moment though. Instead, we generalize the diffusive PDF from Eq. (16) as

$$P^{\text{diff}}(\delta x) = \frac{1}{\left[2\pi\sigma_{\text{diff}}^2\right]^{d/2}} \exp\left[-\frac{(\delta x - u\Delta t)^2}{2\sigma_{\text{diff}}^2}\right], \qquad (28)$$

where σ_{diff} is a free parameter and can be expressed as a function of the second-order moment μ_2 approximated by Eq. (21)

$$\mu_2 = \int_{-\infty}^{\infty} P^{\text{BDM}}(\delta x)(\delta x)^2\, d\delta x$$
$$= \int_{-\infty}^{\infty} \exp\left(-\frac{\Delta t}{\tau}\right) P^{\text{ball}}(\delta x)(\delta x)^2\, d\delta x$$
$$+ \int_{-\infty}^{\infty} \left[1 - \exp\left(-\frac{\Delta t}{\tau}\right)\right] P^{\text{diff}}(\delta x)(\delta x)^2\, d\delta x$$
$$= \exp\left(-\frac{\Delta t}{\tau}\right) k_B T (\Delta t)^2$$
$$+ \left[1 - \exp\left(-\frac{\Delta t}{\tau}\right)\right]\sigma_{\text{diff}}^2, \qquad (29)$$

with $\delta x \in X_i$. Now σ_{diff} given by

$$\sigma_{\text{diff}} = \sqrt{\frac{\mu_2 - \exp\left(-\frac{\Delta t}{\tau}\right) k_B T (\Delta t)^2}{\left[1 - \exp\left(-\frac{\Delta t}{\tau}\right)\right]}}. \qquad (30)$$

We examine the dependence of this diffusion constant on Δt in Fig. 6. Our original motivation would demand that $D = \sigma_{\text{diff}}^2/2\Delta t$ is a constant. However, this is not the case and we will see below that the BDM model only provides a modest improvement over the single Gaussian description. From now on, we will refer to this distribution function as theoretical BDM and denote it as $P^{\text{BDM-}T}(X_i)$, since the mean free time τ and the mean-squared displacement are estimated using the velocity autocorrelation function.
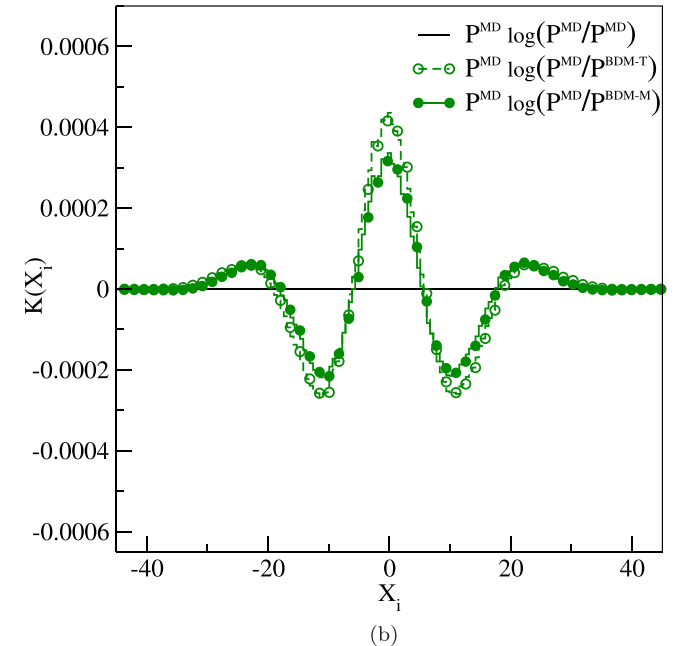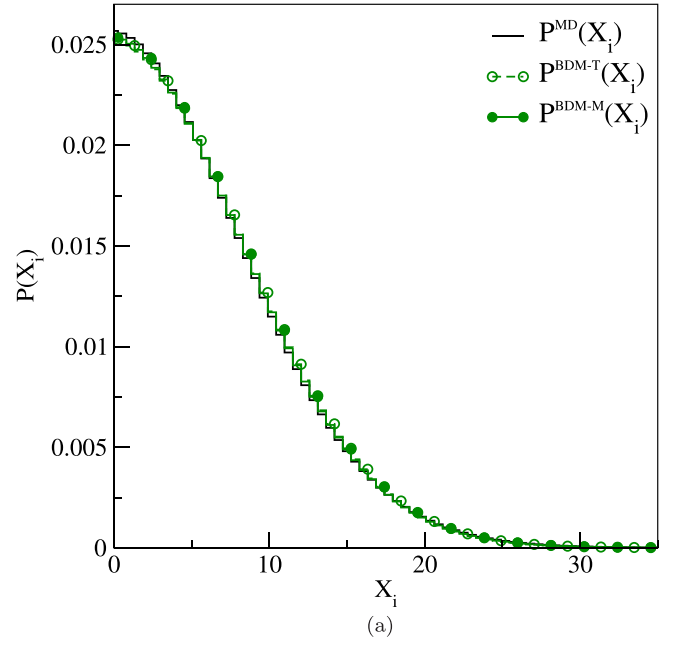

(a)

(b)

FIG. 5. (a) Displacements probability distribution functions. The solid line (black) depicts a PDF obtained from an MD simulation of LJ particles in equilibrium. The lines with empty or full circles illustrate the ballistic-diffusive distribution function defined in Eq. (27) with mean-squared displacement obtained from the velocity autocorrelation function as given in Eq. (21), and with mean-squared displacement fitted directly to the MD data, respectively. Only the data for positive velocities have been depicted due to symmetry. Panel (b) shows the difference between the distributions per interval X_i as defined in Eq. (23). The presented data are for the standard parameters used in the paper and a coarse-grained time step $\Delta t = 3.2$. The y axis has not been re-scaled for a better comparison with Fig. 1.

In Fig. 5(a), we show the resulting $P^{\text{BDM-}T}(X_i)$ distribution function, which resembles well the displacement distribution function obtained from the MD simulation. To assess the
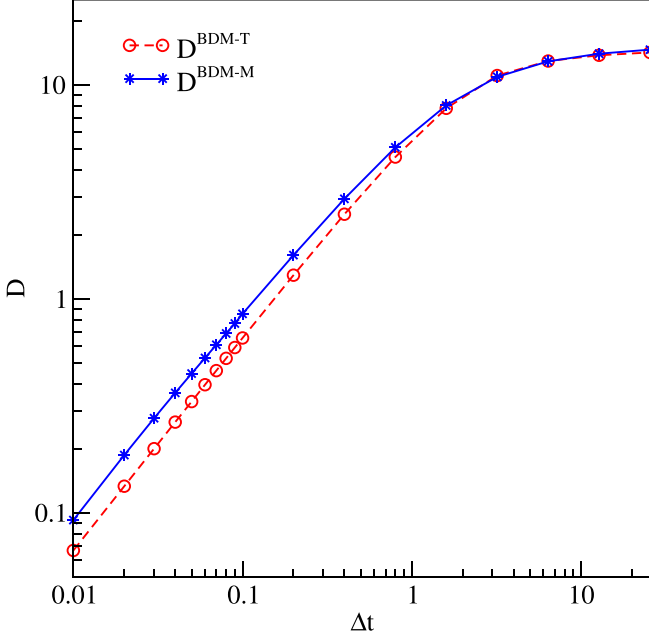
FIG. 6. Dependence of the self-diffusion constant to the time step Δt. The diffusion $D = \sigma_{\text{diff}}^2/2\Delta t$ converges to a constant for $P^{\text{BDM-}T}$ and $P^{\text{BDM-}M}$, however, for early times it is not fixed. This demonstrates that the BDM models do not capture the hydrodynamics properties.

discrepancies between the theoretical BDM and the MD distribution function, we calculate $K(P^{\text{MD}} \parallel P^{\text{BDM-}T})$ defined in Eq. (23). The results are displayed in Fig. 5(b). The theoretical ballistic-diffusive probability distribution function with $\Delta t = 3.2$ demonstrates a significant improvement in comparison to the single Gaussian distribution function shown in Fig. 1. However, there are noticeable deviations between the PDFs which we will investigate further.

The second- and fourth-order moment errors of the BDM distribution function are depicted in Fig. 4. The error is denoted as $\mu_2^{\text{BDM-}T}$ (back dashed line) and $\mu_4^{\text{BDM-}T}$ (green line with empty circles). The second-order moment error is equivalent to $\mu_2^{\text{G-}T}$ by construction. This error comes from the long tails of the velocity autocorrelation function shown in Fig. 2(a), which are not resolved in the theoretical approximation of the mean-squared displacement. Overall, the fourth-order moment error of the theoretical BDM model is smaller than the one calculated for the two Gaussian models discussed in Sec. IV. However, for later times this error increases and becomes as large as the theoretical Gaussian distribution function error.

In order to reduce the error, we construct a second version of the ballistic-diffusive mixture model where we fit the μ_2 and μ_4 moments directly to the MD data. This BDM model does not rely solely on the approximation of the average number of collisions ($\Delta t/\tau$), which cannot be measured precisely and depends on the approximation made for the velocity autocorrelation function.

In Sec. IV, we defined the mean-squared displacement in terms of the velocity autocorrelation function given by Eq. (19). Now, we define the fourth-order moment in a similar

way

$$
\begin{aligned}
\mu_4 &= \langle (\delta x_\alpha)^4 \rangle \\
&= \left\langle \int dt_1 v(t_1) \int dt_2 v(t_2) \int dt_3 v(t_3) \int dt_4 v(t_4) \right\rangle \\
&= \int dt_1 \int dt_2 \int dt_3 \int dt_4 \langle v(t_1)v(t_2)v(t_3)v(t_4) \rangle, \quad (31)
\end{aligned}
$$

where we need the four-point time correlators for the velocity, that are derived from the displacements given by Eq. (18). This integral, if feasible, would allow us to calculate theoretically the fourth-order moment and thus obtain a better approximation of the probability distribution function of displacements. However, we are unaware of a reliable way to derive this four-point velocity autocorrelation function and therefore, we measure the second- and the fourth-order moments directly from the MD simulation instead.

We have to make the following adjustments to the BDM probability distribution function, so that the second- and the fourth-order moments match the MD data: First, instead of calculating the mean-squared displacement from the velocity autocorrelation function, we use the measured mean-squared displacement for μ_2 in Eq. (30); second, instead of calculating the mean free path τ from the velocity autocorrelation function, we define it as a function of μ_2 and μ_4. Thus, the $P^{\text{BDM-}M}(X_i)$ distribution function has zero second- and fourth-order moments error by construction.

The fourth-order moment then has the form

$$
\begin{aligned}
\mu_4 &= \int_{-\infty}^{\infty} P^{\text{BDM}}(\delta x)(\delta x)^4 \, d\delta x \\
&= \int_{-\infty}^{\infty} \exp\left(-\frac{\Delta t}{\tau}\right) P^{\text{ball}}(\delta x)(\delta x)^4 \, d\delta x \\
&\quad + \int_{-\infty}^{\infty} \left[1 - \exp\left(-\frac{\Delta t}{\tau}\right)\right] P^{\text{diff}}(\delta x)(\delta x)^4 \, d\delta x \\
&= 3\left\{\sigma_{\text{diff}}^4 + \exp\left(-\frac{\Delta t}{\tau}\right)\left[(k_B T(\Delta t)^2)^2 - \sigma_{\text{diff}}^4\right]\right\} \quad (32)
\end{aligned}
$$

with σ_{diff} obtained using the measured second-order moment. Now, τ is not a constant anymore and is given by

$$
\tau = \frac{-\Delta t}{\ln\left\{\frac{\frac{\mu_4}{3} - \sigma_{\text{diff}}^4}{[k_B T(\Delta t)^2]^2 - \sigma_{\text{diff}}^4}\right\}}. \quad (33)
$$

Equations (30) and (33) define a system of linear equations with two unknowns. The system has a unique solution for σ_{diff} and τ as a function of μ_2, μ_4, and Δt. Thus, a second version of the ballistic-diffusive distribution function is derived and we refer to it as measured ballistic-diffusive distribution function $P^{\text{BDM-}M}(X_i)$ because it is fully defined by the MD moments. Details of the derivation are can be found in Appendix A.

As mentioned earlier, we demand $D = \sigma_{\text{diff}}^2/2\Delta t$ to be a constant, however, Fig. 6 illustrates that D converges to a constant for both BDM models but is not fixed for early time steps. This demonstrates that the BDM model does not capture the physical diffusion properties.

The $P^{\text{BDM-}M}(X_i)$ distribution function matches well the MD data as depicted in Fig. 5(a). The $K(P^{\text{MD}} \parallel P^{\text{BDM-}T})$ results

are illustrated in Fig. 5(b) and they show that the divergence between $P^{\text{BDM-}M}(X_i)$ and $P^{\text{MD}}(X_i)$ is smaller in comparison to the theoretical BDM distribution function. However, there is still error with well defined structure, which has to be accounted for.

To gain a better understanding of how the BDM model relates to the MD data and the Gaussian distribution functions, we calculate the KL divergence for $\Delta t \in [0.01, 25.6]$ as shown in Fig. 3. The dashed line with empty circles (green) corresponds to the KL divergence $D_{\text{KL}}(P^{\text{MD}} \parallel P^{\text{BDM-}T})$, while the solid line with full circles (green) illustrates the result of $D_{\text{KL}}(P^{\text{MD}} \parallel P^{\text{BDM-}M})$. The divergence is decreased by more than half compared to the KL divergence obtained from the Gaussian distribution functions. However, there is still clear error in the intermediate simulation regime.

Even though, we have fitted the second- and the fourth-order moments, we still have an unsatisfying approximation of the probability distribution function of the displacements. Thus, we conclude that a Gaussian mixture cannot capture the form of the distribution of the displacements for LJ particles in equilibrium. The remaining dependence of $D = \sigma_{\text{diff}}^2/2\Delta t$ on Δt suggests that it is not appropriate to assume that particles that have undergone just one collision will then follow a diffusive displacement. Instead, it might be useful to consider a range of distribution functions occurring after a number of collisions. We will follow up this idea in the next section.

VI. POISSON WEIGHTED SUM OF GAUSSIAN DISTRIBUTION FUNCTIONS

The number of collisions within a time interval plays an important role in the definition of the probability distribution function of displacements. We can prove this statement by a thought experiment: Consider a number of particles in a domain. When the particles undergo a collision their direction and velocity changes. This in turn means that the collisions also change the probability of certain displacements to occur.

In this section, we assume that the intermediate ballistic-diffusive regime could be described as a Poisson weighted sum of Gaussian distributions. One can consider that after a time step Δt the particles can be divided into groups depending on the number of collisions they have experienced. We model these particle collisions using the Poisson probability distribution function

$$P(\delta x) = \sum_{c=0}^{\infty} e^{-\lambda} \frac{\lambda^c}{c!}, \qquad (34)$$

where λ is effectively the average number of collisions given by

$$\lambda = \frac{\Delta t}{\tau}, \qquad (35)$$

where $\tau \approx 0.728$ is the mean free time obtained using Eq. (20). In this formulation the mean free time is considered to be an exponential decay constant. In principle the timing of the collisions should also be random (i.e., given by a Poisson process), but the resulting integrals over the collision times do not admit analytical solutions. Assuming that the collisions are evenly spaced may introduce a small error, but it makes the resulting displacements after c collisions again

Gaussian, which simplifies the application of our results. For details on arbitrary collision occurring at random time refer to Appendix B.

With this approximation the Poisson weighted sum of Gaussians (WSG) model is then given as

$$P^{\text{WSG}}(\delta x) = \sum_{c=0}^{\infty} e^{-\lambda} \frac{\lambda^c}{c!} \frac{\sqrt{(\lambda + 1)}}{\sqrt{2\pi(c+1)\langle(\delta x)^2\rangle}}$$
$$\times \exp\left[-\frac{(\lambda+1)(\delta x - u\Delta t)^2}{2(c+1)\langle(\delta x)^2\rangle}\right] \qquad (36)$$

for displacements δx in one dimension. In extreme regimes, being purely ballistic or purely diffusive, the probability distribution function $P^{\text{WSG}}(\delta x)$ is reduced to a single Gaussian distribution given by Eq. (14) or Eq. (16), respectively. However, in an intermediate regime, we will have contributions from multiple Gaussian distribution functions weighted by a Poisson distribution function.

By using the definition of the average number of collisions given in Eq. (35) and obtaining the mean free time and the mean-squared displacement based on the velocity autocorrelation function, we recover a fully defined theoretical version of the Poisson WSG model which we refer to as $P^{\text{WSG-}T}(X_i)$. This probability distribution function is illustrated in Fig. 7(a) by a dashed line with empty triangles (blue). $P^{\text{WSG-}T}(X_i)$ shows a good fit to the distribution function measured directly from the MD simulation but there are still visible discrepancies.

In Figs. 7(b)–7(d), the $K(P^{\text{MD}} \parallel P^{\text{WSG-}T})$ function is illustrated for three different time steps: $\Delta t = 0.01$, $\Delta t = 0.1$, and $\Delta t = 3.2$. The results for $\Delta t = 0.01$ show noise coming from the averaging procedure as one can see in Fig. 7(b). With increasing the time step, we start seeing some structure in the discrepancies between the theoretical weighted sum of Gaussians and the MD probability distribution function as shown in Fig. 7(c). For $\Delta t = 3.2$, one can see that the rate of discrepancies is as large as the one shown in Fig. 1(b) calculated for the single Gaussian distribution function but with an opposite sign. This observation suggests that the theoretical WSG does not capture well the distribution of the measured MD displacements.

To compare the overall performance of $P^{\text{WSG-}T}(X)$, we calculate its KL divergence as shown in Fig. 3 (blue dashed line with empty triangles). The $D_{\text{KL}}(P^{\text{MD}} \parallel P^{\text{WSG-}T})$ divergence is slightly smaller than the one measured for the Gaussian models presented in Sec. IV.

In order to find the source of the large KL divergence, we display the second- and fourth-order moments error in Fig. 4. The second-order moment error is equivalent to the error calculated for the other theoretical models ($\mu_2^{\text{G-}T} = \mu_2^{\text{BDM-}T} = \mu_2^{\text{WSG-}T}$). The $P^{\text{WSG-}T}(X_i)$ fourth-order moment error, however, is larger than the fourth-order moment error of the other two models. This is true especially for the intermediate regime and explains the poor results of the theoretical WSG model.

The average number of collisions λ plays an important role in the definition of the BDM and WSG models. Unfortunately, it is difficult to make a good approximation for λ based on the velocity autocorrelation function. Therefore, to
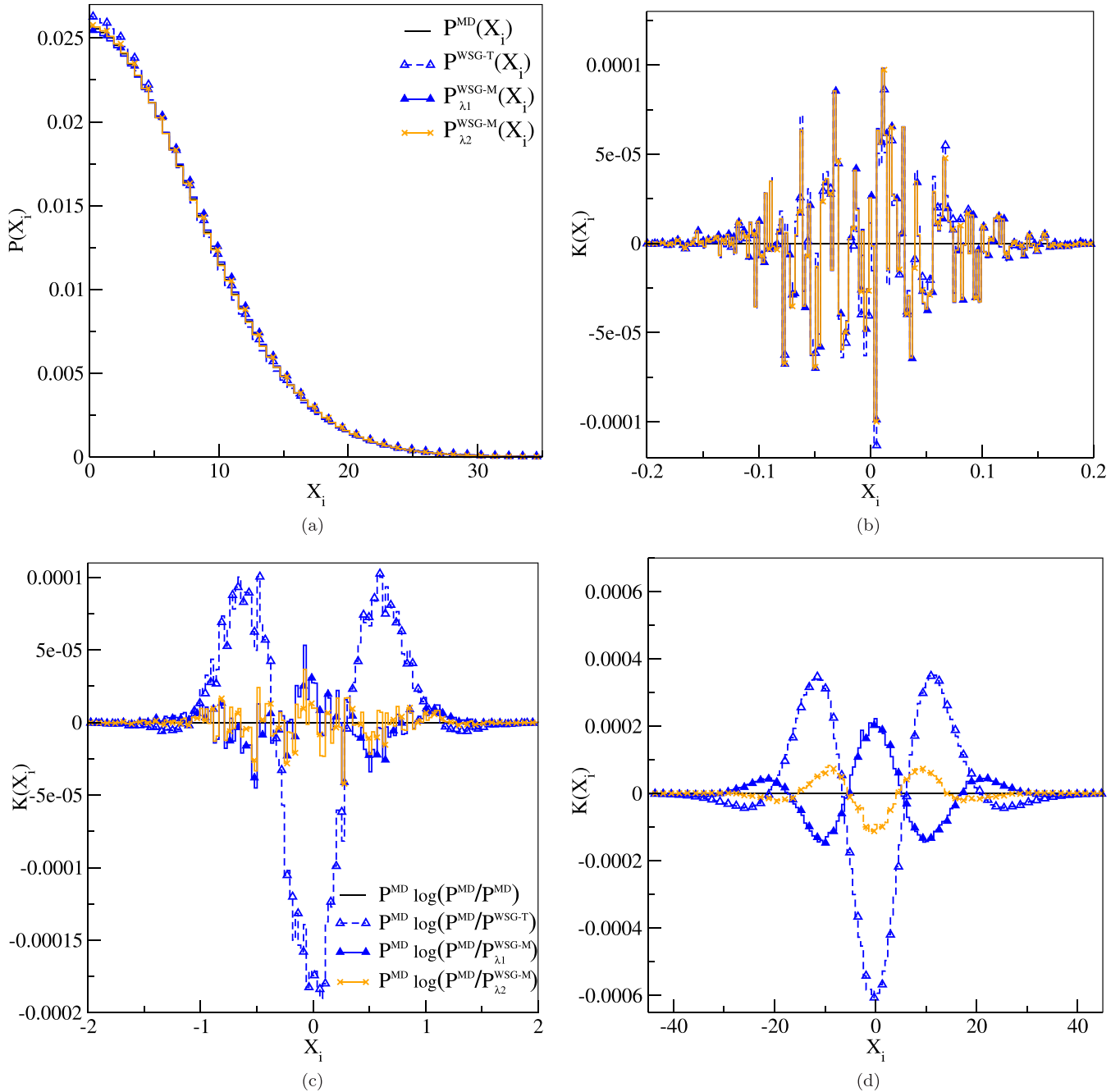
(a)

(b)

(c)

(d)

FIG. 7. (a) Displacements probability distribution functions. The solid line (black) depicts a PDF obtained from an MD simulation of LJ particles in equilibrium. The dashed line (blue) with empty triangles illustrates the PDF of the theoretical WSG defined in Eq. (36) with λ obtained using the theoretical velocity autocorrelation function from Eq. (20). The solid lines with full squares or x-symbols denote the Poisson WSG distribution function with average number of collisions λ_1 and λ_2, respectively. The time step is $\Delta t = 3.2$ and due to symmetry only the data for positive velocities has been depicted. Panels (b)–(d) show the difference between the distributions per interval X_i as defined in Eq. (23) for a variety of time steps: (b) $\Delta t = 0.01$, (c) $\Delta t = 0.1$, and (d) $\Delta t = 3.2$. The presented data are for the standard parameters used in the paper. The y axis of (a) and (d) have not been rescaled for a better comparison with Figs. 1 and 5.

reduce the error coming from the theoretical average number of collisions and to eliminate the second- and the fourth-order moment errors, we match these moments to the corresponding moments measured directly from the MD simulations. We derive the mean-squared displacement from the second-order Gaussian integral

$$\mu_2 = \int_{-\infty}^{\infty} P^{\mathrm{WSG}}(\delta x)(\delta x)^2 \, d\delta x = \int_{-\infty}^{\infty} \sum_{c=0}^{\infty} e^{-\lambda} \frac{\lambda^c}{c!} \frac{\sqrt{\lambda+1}}{\sqrt{2\pi(c+1)\langle(\delta x)^2\rangle}} \exp\left[-\frac{(\lambda+1)(\delta x - u\Delta t)^2}{2(c+1)\langle(\delta x)^2\rangle}\right](\delta x)^2 \, d\delta x \qquad (37)$$

and the fourth-order moment from the fourth-order Gaussian integral

$$\mu_4 = \int_{-\infty}^{\infty} P^{\text{WSG}}(\delta x)(\delta x)^4 \, d(\delta x)$$

$$= \int_{-\infty}^{\infty} \sum_{c=0}^{\infty} e^{-\lambda} \frac{\lambda^c}{c!} \frac{\sqrt{\lambda+1}}{\sqrt{2\pi(c+1)\langle(\delta x)^2\rangle}}$$

$$\times \exp\left[-\frac{(\lambda+1)(\delta x - u\Delta t)^2}{2(c+1)\langle(\delta x)^2\rangle}\right](\delta x)^4 \, d\delta x$$

$$= \frac{3\langle(\delta x)^2\rangle^2}{(\lambda+1)^2}[\lambda^2 + 3\lambda + 1]. \tag{38}$$

This ensures that the μ_2 and μ_4 moments are fully recovered from the WSG model. Now, we can express λ as a function of these parameters and solve the resulting quadratic equation

$$\frac{3\mu_2^2}{(\lambda+1)^2}[\lambda^2 + 3\lambda + 1] - \mu_4 = 0 \tag{39}$$

with $\mu_2 = \langle(\delta x)^2\rangle$ for brevity. The quadratic equation has the following solutions:

$$\lambda_{1,2} = \frac{-9\mu_2^2 \pm \sqrt{3[15\mu_2^4 - 4\mu_2^2\mu_4]} + 2\mu_4}{2[3\mu_2^2 - \mu_4]}. \tag{40}$$

Details of the derivation are omitted but they can be found in Appendix B.

Since the mean-squared displacement and the fourth-order moment depend wholly on the time step, we plot $\lambda_1(\Delta t)$, $\lambda_2(\Delta t)$, and $\lambda(\Delta t)$ from Eq. (35) as a function of Δt, which is depicted in Fig. 8. We see that the analytical expectation for λ from Eq. (35) is in better agreement with λ_2 for large Δt, while for small Δt it is in better agreement with λ_1. This is intriguing, and we do not fully understand the significance of this result. However, we should note here that both limits $\Delta t \rightarrow 0$ and $\Delta t \rightarrow \infty$ lead to a simple Gaussian distribution. For small Δt this is the case because there is only the $c = 0$ term in the Poisson distribution matters, and for large Δt because the Poisson distribution will be sharply peaked around $c = \lambda$, leading again to a simple Gaussian distribution function.

$P^{\text{WSG-}M}(X_i)$ has zero second- and fourth-order moment errors by construction, because these moments have been fitted to the MD simulation data.

The Kullback-Leibler divergence per element X_i for $P_{\lambda_1}^{\text{WSG-}M}(X_i)$ and $P_{\lambda_2}^{\text{WSG-}M}(X_i)$ is illustrated in Figs. 7(b)–7(d). In each figure, $K(P^{\text{MD}} \parallel P_{\lambda_1}^{\text{WSG-}M})$ and $K(P^{\text{MD}} \parallel P_{\lambda_2}^{\text{WSG-}M})$ are depicted for different time step. Figure 7(b) shows the error of $K(P^{\text{MD}} \parallel P_{\lambda_2}^{\text{WSG-}M})$ for $\Delta t = 0.01$ where the error is very small and is dominated by noise due to the averaging procedure. For the coarse-grained time step of $\Delta t = 0.1$, the error becomes larger and one sees small structures building; however, the noise is still dominant in the error contribution. In Fig. 7(d), we show the $K(P^{\text{MD}} \parallel P_{\lambda_2}^{\text{WSG-}M})$ results for $\Delta t = 3.2$. There is a clear structure of the error for both WSG-M probability distribution functions. In comparison to the Gaussian and the ballistic-diffusive mixture models, the WSG model with average number of collisions λ_1 and λ_2 shows much smaller error.
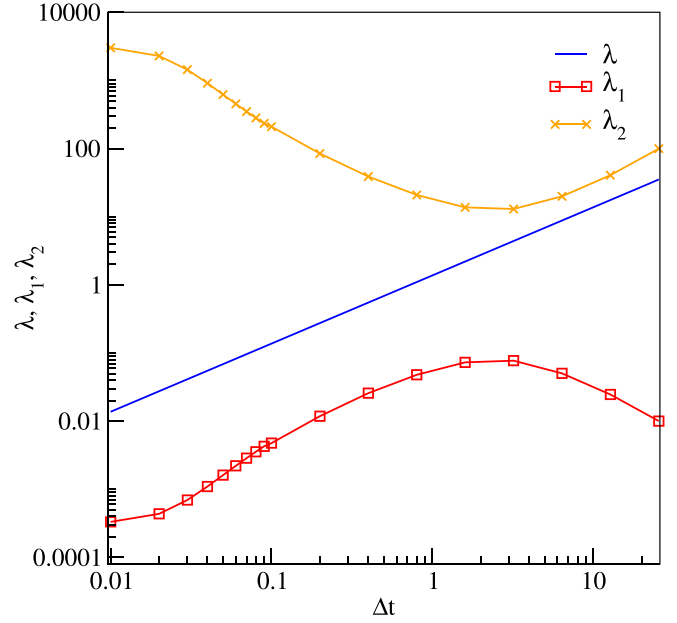

FIG. 8. Average number of collisions depending on the coarse-grained time step Δt. λ denotes the number of collisions obtained from the velocity autocorrelation theory given in Eq. (35), which is used for the calculation of the $P^{\text{WSG-}T}(X_i)$ distribution function. λ_1 and λ_2 are solutions of the quadratic equation given in Eq. (39). These values are used for the calculation of $P_{\lambda_1}^{\text{WSG-}M}(X_i)$ and $P_{\lambda_2}^{\text{WSG-}M}(X_i)$ distribution functions, respectively. λ_1 and λ_2 are obtained using the second and the fourth-order moments measured directly from the MD simulations.

For better comparison, we calculate the Kullback-Leibler divergence for $P_{\lambda_1}^{\text{WSG-}M}(X_i)$ and $P_{\lambda_2}^{\text{WSG-}M}(X_i)$ and display the results in Fig. 3. The KL divergence for λ_1 shows reduced error for the transition regime. The second solution of Eq. (39) λ_2, however, shows KL divergence close to zero for all time steps. This is a significant improvement comparing the results using a single Gaussian or a mixture of two Gaussian distribution functions.

The WSG probability distribution function strongly depends on the calculation of the average number of collisions. By using the theoretical average number of collisions obtained from the velocity autocorrelation function, the KL divergence $D_{\text{KL}}(P^{\text{MD}} \parallel P^{\text{WSG-}T})$ is almost as large as $D_{\text{KL}}(P^{\text{MD}} \parallel P^{\text{G}})$ for the Gaussian models. Even fitting the second- and the fourth-order moments is not sufficient to obtain a good estimation of the PDF obtained from the MD simulation. The KL divergence for the WSG model with λ_1 shows an improvement by about a factor of 6 but it still large in the transition regime. $P^{\text{WSG-}M}(X_i)$ with λ_2 gives a unique close to zero Kullback-Leibler divergence owing to the WSG model and the correct choice of the average number of collisions.

VII. CONCLUSIONS

In this article we have shown that displacement distributions are only of a Gaussian form for either very small times or for long times. The transition region, where a different distribution function is found roughly corresponds to the region where the motion of particles transitions from ballistic

to diffusive regime. One signal of the deviation is the fourth-order moment of the probability distribution function of displacements.

By allowing for the distribution to be a mixture of two distribution functions, one corresponding to the ballistic regime, and a second one to be selected to give the correct second- and fourth-order moments gives a PDF that agrees better with the MD distribution function, by about a factor of 3 measured by the Kullback-Leibler divergence.

Using the same amount of information, i.e., the second- and fourth-order moments, we found a different distribution function that gives a nearly perfect fit. This distribution was motivated by considering the distribution function as a mixture of Gaussian distributions that have undergone a number of collisions, which are given by a Poisson distribution.

We would like to add a small note of caution here, because both in internal discussion and in comments from a referee a pitfall in our thinking has emerged. It is tempting to think of this transition regime in terms of nonequilibrium phenomena, where non-Gaussian effects are commonly observed. One might therefore suspect that in this transition regime there might be a correlation in the velocities of the different particles that lead to the correlations in the displacements. However, the system considered in this paper is in equilibrium, and there are no correlations (at equal times) between the momenta of the different particles. Correlations build up only in time, through the collisions between the particles. So the distribution of the velocities remains Gaussian at all times, even when the distribution of the displacements becomes non-Gaussian in this regime.

This analytical description is very promising for the MDLG analysis of collision operators in nonequilibrium systems. It would be very helpful if a theoretical prediction of the fourth-order moment equivalent to the second-order moment derived from the velocity time correlation could be achieved, because then one could obtain the displacement distribution for all time steps through one measurement. The current approach still needs measurements of the fourth-order moment for each time step. Furthermore, the current study was done for a semidilute system. In future research, we anticipate to establish up to what density the distribution with Poisson weighted sum of Gaussians remains a valid description for the displacement distribution.

APPENDIX A: BALLISTIC-DIFFUSIVE DISTRIBUTION FUNCTION

The BDM probability distribution function is defined in Eq. (27). We derive the standard deviation σ_{diff} from the second-order Gaussian integral

$$
\begin{aligned}
\mu_2 &= \int_{-\infty}^{\infty} P^{\text{BDM}}(\delta x)(\delta x)^2 \, d\delta x \\
&= \int_{-\infty}^{\infty} \exp\left(-\frac{\Delta t}{\tau}\right) P^{\text{ball}}(\delta x)(\delta x)^2 \, d\delta x + \int_{-\infty}^{\infty} \left[1 - \exp\left(-\frac{\Delta t}{\tau}\right)\right] P^{\text{diff}}(\delta x)(\delta x)^2 \, d\delta x \\
&= \int_{-\infty}^{\infty} \frac{\exp\left(-\frac{\Delta t}{\tau}\right)}{[2\pi k_B T (\Delta t)^2]^{d/2}} \exp\left[-\frac{(\delta x)^2}{2 k_B T (\Delta t)^2}\right] (\delta x)^2 \, d\delta x + \int_{-\infty}^{\infty} \frac{\left[1 - \exp\left(-\frac{\Delta t}{\tau}\right)\right]}{\left[2\pi \sigma_{\text{diff}}^2\right]^{d/2}} \exp\left[-\frac{(\delta x)^2}{2\sigma_{\text{diff}}^2}\right] (\delta x)^2 \, d\delta x \\
&= \exp\left(-\frac{\Delta t}{\tau}\right) k_B T (\Delta t)^2 + \left[1 - \exp\left(-\frac{\Delta t}{\tau}\right)\right] \sigma_{\text{diff}}^2
\end{aligned}
\tag{A1}
$$

for one dimension ($d = 1$). Now, we express the standard deviation of $P^{\text{diff}}(\delta x)$ as

$$
\sigma_{\text{diff}} = \sqrt{\frac{\mu_2 - \exp\left(-\frac{\Delta t}{\tau}\right) k_B T (\Delta t)^2}{\left[1 - \exp\left(-\frac{\Delta t}{\tau}\right)\right]}}.
\tag{A2}
$$

This completes the definition of $P^{\text{BDM-T}}(X_i)$ using μ_2 and σ_{diff} recovered by Eqs. (21) and (20), respectively.

In the second version of the BDM model, we match the second- and the fourth-order moments measured directly from the MD simulations to the probability distribution function. The mean free time τ is not anymore a function of the velocity autocorrelation function but a free parameter. The derivation of σ_{diff} in Eqs. (A1) and (A2) is still valid. In addition, we fit the fourth-order moment μ_4 using the fourth-order Gaussian integral

$$
\begin{aligned}
\mu_4 &= \int_{-\infty}^{\infty} P^{\text{BDM}}(\delta x)(\delta x)^4 \, d\delta x \\
&= \int_{-\infty}^{\infty} \exp\left(-\frac{\Delta t}{\tau}\right) P^{\text{ball}}(\delta x)(\delta x)^4 \, d\delta x + \int_{-\infty}^{\infty} \left[1 - \exp\left(-\frac{\Delta t}{\tau}\right)\right] P^{\text{diff}}(\delta x)(\delta x)^4 \, d\delta x
\end{aligned}
$$

$$= \frac{3\sqrt{\pi}}{4}\left[\frac{\left(2k_BT(\Delta t)^2\right)^{5/2}}{\frac{\sqrt{2\pi k_BT(\Delta t)^2}}{\exp\left(-\frac{\Delta t}{\tau}\right)}} + \frac{\left(2\sigma_{\text{diff}}^2\right)^{5/2}}{\frac{\sqrt{2\pi\sigma_{\text{diff}}^2}}{1-\exp\left(-\frac{\Delta t}{\tau}\right)}}\right]$$

$$= \frac{3\exp\left(-\frac{\Delta t}{\tau}\right)(2k_BT(\Delta t)^2)^2\sqrt{2\pi k_BT(\Delta t)^2}}{4\sqrt{2\pi k_BT(\Delta t)^2}} + \frac{3\left[1-\exp\left(-\frac{\Delta t}{\tau}\right)\right]\left(2\sigma_{\text{diff}}^2\right)^2\sqrt{2\pi\sigma_{\text{diff}}^2}}{4\sqrt{2\pi\sigma_{\text{diff}}^2}}$$

$$= 3\left[\sigma_{\text{diff}}^4 + \exp\left(-\frac{\Delta t}{\tau}\right)\left((k_BT(\Delta t)^2)^2 - \sigma_{\text{diff}}^4\right)\right]. \tag{A3}$$

Now we derive the mean free time τ as a function of the time step Δt, and the second- and the fourth-order moments measured from the MD simulation

$$\exp\left(-\frac{\Delta t}{\tau}\right) = \frac{\frac{\mu_4}{3} - \sigma_{\text{diff}}^4}{(k_BT(\Delta t)^2)^2 - \sigma_{\text{diff}}^4} \qquad \tau = \frac{-\Delta t}{\ln\left[\frac{\frac{\mu_4}{3} - \sigma_{\text{diff}}^4}{(k_BT(\Delta t)^2)^2 - \sigma_{\text{diff}}^4}\right]}. \tag{A4}$$

Equations (A2) and (A4) define a system of linear equations with two unknowns. After substituting Eq. (A2) in Eq. (A4), we found a unique solution for τ given by

$$\exp\left(-\frac{\Delta t}{\tau}\right) = \frac{\mu_2^2 - \frac{\mu_4}{3}}{k_BT(\Delta t)^2[2\mu_2 - k_BT(\Delta t)^2] - \frac{\mu_4}{3}} \qquad \tau = \frac{-\Delta t}{\ln\left\{\frac{\mu_2^2 - \frac{\mu_4}{3}}{k_BT(\Delta t)^2[2\mu_2 - k_BT(\Delta t)^2] - \frac{\mu_4}{3}}\right\}}. \tag{A5}$$

The mean free time τ is a function of μ_2, μ_4, Δt and the temperature of the gas given in LJ units. The standard deviation σ_{diff} is recovered using Eq. (A2).

APPENDIX B: POISSON WEIGHTED SUM OF GAUSSIAN DISTRIBUTION FUNCTIONS

Without collisions particles will move with a constant velocity drawn from a Gaussian distribution function. In this case, the distribution of displacements is given by $P^{\text{ball}}(X_i)$ in Eq. (14). If we ought to calculate the distribution of particle displacements for particles that undergo a single collision at a random time $0 < t_c < \Delta t$, we would define a sum of two Gaussian distributed random numbers with a second moment given by

$$t_c^2 k_BT + (\Delta t + t_c)^2 k_BT = \left(\Delta t^2 + 2t_c^2 - 2t_c\Delta t\right)k_BT, \tag{B1}$$

which is less than the collisionless case except for $t_c = 0$ and $t_c = \Delta t$. The full distribution function in one dimension ($d = 1$) is then

$$P_{\delta x_c}(\delta x) = \int_{-\infty}^{\infty} \frac{1}{\left[2\pi k_BT t_c^2\right]^{d/2}}\exp\left[-\frac{(\delta x_c)^2}{2k_BT t_c^2}\right]\frac{1}{\left[2\pi k_BT\left(\Delta t^2 - 2\Delta t\,t_c + 2t_c^2\right)\right]^{d/2}}\exp\left[-\frac{(\delta x - \delta x_c)^2}{2k_BT\left(\Delta t^2 - 2\Delta t\,t_c + 2t_c^2\right)}\right]d\delta\delta x_c$$

$$= \frac{1}{\sqrt{2\pi k_BT\left(\Delta t^2 - 2\Delta t\,t_c + 2t_c^2\right)}}\exp\left[-\frac{(\delta x)^2}{2k_BT\left(\Delta t^2 - 2\Delta t\,t_c + 2t_c^2\right)}\right], \tag{B2}$$

where δx_c is the displacement for time 0 to t_c, and $(\delta x - \delta x_c)$ for time t_c to Δt. This results to a Gaussian distribution function with total displacement δx for a collision taking place at time t_c. To ensure that the time t_c is arbitrary and collisions at any time will be uniformly likely, we average over all possible collision times given by

$$P_{\delta t_c}(\delta x) = \frac{1}{\Delta t}\int_0^{\Delta t} P_{\delta x_c}(\delta x)\,dt_c$$

$$= \frac{1}{\Delta t}\int_0^{\Delta t}\frac{1}{\sqrt{2\pi k_BT\left(\Delta t^2 - 2\Delta t\,t_c + 2t_c^2\right)}}\exp\left[-\frac{(\delta x)^2}{2k_BT\left(\Delta t^2 - 2\Delta t\,t_c + 2t_c^2\right)}\right]dt_c. \tag{B3}$$

It is difficult to evaluate this integral analytically, but it can be solved numerically. However, this is the theory for only one collision occurring at a random time t_c. For the Poisson weighted sum of Gaussians in Sec. VI, we consider multiple collisions at multiple arbitrary times, which leads to high-dimensional integrals, whose solution is out of the scope of this publication. In addition to the numerical difficulty that multidimensional integrals pose, the resulting probability distribution functions are non-Gaussian. To avoid this, we assume that the collisions are evenly distributed which may introduce a small error.

The WSG probability distribution function is defined in Eq. (36) and recovers the second-order moment given by

$$
\begin{aligned}
\mu_2 &= \int_{-\infty}^{\infty} P^{\text{WSG}}(\delta x)(\delta x)^2 \, d\delta x \\
&= \int_{-\infty}^{\infty} \sum_{c=0}^{\infty} e^{-\lambda} \frac{\lambda^c}{c!} \frac{\sqrt{\lambda+1}}{\sqrt{2\pi(c+1)\langle(\delta x)^2\rangle}} \exp\left[-\frac{(\lambda+1)(\delta x)^2}{2(c+1)\langle(\delta x)^2\rangle}\right](\delta x)^2 \, d\delta x \\
&= \sum_{c=0}^{\infty} e^{-\lambda} \frac{\lambda^c}{c!} \frac{(c+1)\langle(\delta x)^2\rangle}{\lambda+1} \\
&= \langle(\delta x)^2\rangle \frac{e^{-\lambda}}{\lambda+1}\left(\sum_{c=0}^{\infty} \frac{c\lambda^c}{c!} + \sum_{c=0}^{\infty} \frac{\lambda^c}{c!}\right) \\
&= \langle(\delta x)^2\rangle e^{-\lambda} \frac{\lambda}{\lambda+1}\left(\sum_{c=0}^{\infty} \frac{\lambda^c}{c!} + \frac{e^{\lambda}}{\lambda}\right) \\
&= \langle(\delta x)^2\rangle.
\end{aligned}
\tag{B4}
$$

Analogously, one derives the fourth-order moment as

$$
\begin{aligned}
\mu_4 &= \int_{-\infty}^{\infty} P(\delta x)(\delta x)^4 \, d(\delta x) \\
&= \int_{-\infty}^{\infty} \sum_{c=0}^{\infty} e^{-\lambda} \frac{\lambda^c}{c!} \frac{\sqrt{\lambda+1}}{\sqrt{2\pi(c+1)\langle(\delta x)^2\rangle}} \exp\left[-\frac{(\lambda+1)(\delta x)^2}{2(c+1)\langle(\delta x)^2\rangle}\right](\delta x)^4 \, d\delta x \\
&= \sum_{c=0}^{\infty} e^{-\lambda} \frac{\lambda^c}{c!}\left\{\frac{3\sqrt{\pi}\left[\frac{2(c+1)\langle(\delta x)^2\rangle}{(\lambda+1)}\right]^{5/2}}{4\sqrt{\frac{2\pi(c+1)\langle(\delta x)^2\rangle}{(\lambda+1)}}}\right\} \\
&= \frac{3\langle(\delta x)^2\rangle^2}{(\lambda+1)^2}\left[e^{-\lambda}\sum_{c=0}^{\infty} \frac{\lambda^c(c^2+2c+1)}{c!}\right] \\
&= \frac{3\langle(\delta x)^2\rangle^2}{(\lambda+1)^2}\left[e^{-\lambda}\sum_{c=0}^{\infty} \frac{\lambda^c c^2}{c!} + 2e^{-\lambda}\sum_{c=0}^{\infty} \frac{\lambda^c c}{c!} + e^{-\lambda}\sum_{c=0}^{\infty} \frac{\lambda^c}{c!}\right] \\
&= \frac{3\langle(\delta x)^2\rangle^2}{(\lambda+1)^2}\left[e^{-\lambda}\sum_{c=1}^{\infty} \frac{\lambda^c c}{(c-1)!} + 2\lambda e^{-\lambda}\sum_{c=1}^{\infty} \frac{\lambda^{c-1}}{(c-1)!} + 1\right] \\
&= \frac{3\langle(\delta x)^2\rangle^2}{(\lambda+1)^2}\left[e^{-\lambda}\lambda\frac{d}{d\lambda}\sum_{c=1}^{\infty} \frac{\lambda^c}{(c-1)!} + 2\lambda + 1\right] \\
&= \frac{3\langle(\delta x)^2\rangle^2}{(\lambda+1)^2}\left[e^{-\lambda}\lambda\frac{d}{d\lambda}\lambda e^{\lambda} + \lambda + 1\right] \\
&= \frac{3\langle(\delta x)^2\rangle^2}{(\lambda+1)^2}\left[e^{-\lambda}\lambda(e^{\lambda} + \lambda e^{\lambda}) + 2\lambda + 1\right] \\
&= \frac{3\langle(\delta x)^2\rangle^2}{(\lambda+1)^2}[\lambda^2 + 3\lambda + 1].
\end{aligned}
\tag{B5}
$$

This description allows us to adjust λ, such that the fourth-order moment does converge to the measured MD value. We express λ as a function of Δt with mean-squared displacement and fourth-order moment measured directly from the MD simulation

$$
\frac{3\langle(\delta x)^2\rangle^2}{(\lambda+1)^2}[\lambda^2 + 3\lambda + 1] - \mu_4 = 0.
\tag{B6}
$$

After solving this quadratic equation, we obtain the following solutions:

$$
\lambda_{1,2} = \frac{-9\mu_2^2 \pm \sqrt{3[15\mu_2^4 - 4\mu_2^2\mu_4]} + 2\mu_4}{2[3\mu_2^2 - \mu_4]}.
\tag{B7}
$$

[1] D. L. Ermak and J. A. McCammon, Brownian dynamics with hydrodynamic interactions, J. Chem. Phys. **69**, 1352 (1978).

[2] P. J. Hoogerbrugge and J. M. V. A. Koelman, Simulating microscopic hydrodynamic phenomena with dissipative particle dynamics, Europhys. Lett. **19**, 155 (1992).

[3] T. Ihle and D. M. Kroll, Stochastic rotation dynamics: A Galilean-invariant mesoscopic model for fluid flow, Phys. Rev. E **63**, 020201(R) (2001).

[4] Y. H. Qian, D. D'Humières, and P. Lallemand, Lattice BGK models for Navier-Stokes equation, Europhys. Lett. **17**, 479 (1992).

[5] X. He and L.-S. Luo, Theory of the lattice Boltzmann method: From the Boltzmann equation to the lattice Boltzmann equation, Phys. Rev. E **56**, 6811 (1997).

[6] J. Masoliver, Three-dimensional telegrapher's equation and its fractional generalization, Phys. Rev. E **96**, 022101 (2017).

[7] J. Masoliver and K. Lindenberg, Two-dimensional telegraphic processes and their fractional generalizations, Phys. Rev. E **101**, 012137 (2020).

[8] P. L. Bhatnagar, E. P. Gross, and M. Krook, Model for collision processes in gases. I. Small amplitude processes in charged and neutral one-component systems, Phys. Rev. **94**, 511 (1954).

[9] M. R. Parsa and A. J. Wagner, Lattice gas with molecular dynamics collision operator, Phys. Rev. E **96**, 013314 (2017).

[10] M. R. Parsa and A. J. Wagner, Large Fluctuations in Nonideal Coarse-Grained Systems, Phys. Rev. Lett. **124**, 234501 (2020).

[11] A. Pachalieva and A. Wagner (unpublished).

[12] S. Plimpton, Fast parallel algorithms for short-range molecular dynamics, J. Comput. Phys. **117**, 1 (1995).

[13] LAMMPS Official Website, http://lammps.sandia.gov.

[14] M. R. Parsa, A. Pachalieva, and A. J. Wagner, Validity of the molecular-dynamics-lattice-gas global equilibrium distribution function, Int. J. Mod. Phys. C **30**, 1941007 (2019).

[15] G. E. Uhlenbeck and L. S. Ornstein, On the theory of the Brownian motion, Phys. Rev. **36**, 823 (1930).

[16] M. S. Green, Markoff random processes and the statistical mechanics of timedependent phenomena, J. Chem. Phys. **20**, 1281 (1952).

[17] M. S. Green, Markoff random processes and the statistical mechanics of timedependent phenomena. II. Irreversible processes in fluids, J. Chem. Phys. **22**, 398 (1954).

[18] R. Kubo, Statistical-mechanical theory of irreversible processes. I. General theory and simple applications to magnetic and conduction problems, J. Phys. Soc. Jpn. **12**, 570 (1957).

[19] D. A. Weitz, D. J. Pine, P. N. Pusey, and R. J. A. Tough, Nondiffusive Brownian motion studied by diffusing-wave spectroscopy, Phys. Rev. Lett. **63**, 1747 (1989).

[20] S. Kullback and R. A. Leibler, Annals of mathematical statistics, Ann. Math. Stat. **22**, 79 (1951).