TUM School of Engineering and Design
Technische Universität München

Molecular Dynamics Lattice Gas Analysis Tool

Aleksandra A. Pachalieva

Vollständiger Abdruck der von der TUM School of Engineering and Design der Technischen Universität München zur Erlangung des akademischen Grades einer

Doktorin der Ingenieurwissenschaften (Dr.-Ing.)

genehmigten Dissertation.

Vorsitz:

Prof. Dr. Julija Zavadlav

Prüfende der Dissertation:

1. Prof. Dr.-Ing. Nikolaus A. Adams
2. Prof. Dr. Alexander J. Wagner

Die Dissertation wurde am 20.12.2021 bei der Technischen Universität München eingereicht und durch die TUM School of Engineering and Design am 17.05.2022 angenommen.

Preface

This work has evolved during my time as a research associate at the Chair of Aerodynamics and Fluid Mechanics, Technical University of Munich from 2016 to 2021. It is a product of a collaboration between Technical University of Munich, North Dakota State University and Los Alamos National Laboratory.

First, I want to express my sincere gratitude to *Prof. Dr.–Ing. Nikolaus A. Adams* and *PD Dr.-Ing. habil. Thomas Indinger* for their continuous support and encouragement. They provided me with guidance and advice, when I mostly needed it, and with many liberties to find, and follow my own path.

My appreciation goes to *Prof. Alexander J. Wagner* for the excellent supervision during the last four years of my Ph.D. Thank you for the countless ways you have supported me and my work.

I would also like to thank my colleagues and friends from our industrial collaborator Altair Engineering Inc. – *Christoph Niedermeier*, *Hanna Ketterle*, *Shezad Afzal*, *Andrea Pascuali* and *Amin Safi*. Special thanks to *Christoph* and *Hanna*, who were there every step of the way with inspiring ideas, moral support and true friendship.

This work would have been impossible without my colleagues *Marco Kiewat* and *Daiki Matsumoto* from the research group for Automotive Aerodynamics at the Chair of Aerodynamics and Fluid Mechanics of the Technical University of Munich. Your good humor and positive attitude made the long office hours much more fun. I truly enjoyed working with you.

I would like to thank also my current employer Los Alamos National Laboratory and my supervisors and collaborators *Kipton Barros*, *Hari Viswanathan*, *Nicholas Lubbers*, *Dan O'Malley*, and *Jeffrey De'Haven Hyman* for giving me the freedom to work on my dissertation while exploring new possibilities.

This work would have been impossible without the unconditional love and support of my family and friends. A special shout-out goes to my parents, *Mariana* and *Atanas Pachalievi*, who taught me to dream big and never give up, and to my sister, *Radost Miteva*, who was always there to give me the needed push in the right direction.

Abstract

The aim of this thesis is to study the derivation of the lattice Boltzmann method from a coarse-graining of an underlying Molecular Dynamics simulation.

In the rapid development of the lattice Boltzmann method, the link to its predecessor Lattice-Gas Cellular Automata has been compromised. To recover this link, we introduce the Molecular Dynamics Lattice Gas analysis tool, which coarse-grains the trajectories of an underlying Molecular Dynamics simulation to reproduce the formalism of the lattice Boltzmann method. The Molecular Dynamics Lattice Gas approach establishes a direct link between a Lattice Gas method and a coarse-graining of a Molecular Dynamics simulation. As a result, the Molecular Dynamics Lattice Gas method allows us to connect the lattice Boltzmann method to physical reality.

We use this analysis tool to derive more precisely key elements of the lattice Boltzmann method such as its equilibrium distribution function, relaxation parameters, and collision rules. Furthermore, the MDLG method can be extended to derive the fluctuating properties, forcing fields, and boundary conditions of the lattice Boltzmann method; however, this is beyond the scope of this work.

Kurzfassung

Das Ziel dieser Arbeit ist die Ableitung der Lattice-Boltzmann-Methode aus der Vergröberung einer zugrunde liegenden Molekulardynamik-Simulation.

Bei der schnellen Entwicklung der Lattice-Boltzmann-Methode wurde die Verbindung zu ihrem Vorgänger Gitter-Gas-Zellular-Automaten vernachlässigt. Um diese Verbindung zu rekonstruieren, stellen wir das Analysewerkzeug Molekulardynamik-Gitter-Gas vor, das die Trajektorien einer zugrunde liegenden Molekulardynamik-Simulation vergröbert, um die Lattice-Boltzmann-Methode abzuleiten. Der Molekulardynamik-Gitter-Gas-Ansatz stellt eine direkte Verbindung zwischen einer Lattice-Gas-Methode und einer Molekulardynamik-Simulation her. Als Ergebnis ermöglicht uns die Molekulardynamik-Gitter-Gas-Methode, die Lattice-Boltzmann-Methode mit einer physikalischen Realität zu verbinden.

Wir verwenden dieses Analysewerkzeug, um Schlüsselelemente der Lattice-Boltzmann-Methode wie ihre Gleichgewichtsverteilungsfunktion, Relaxationsparameter und Kollisionsregeln genauer abzuleiten. Darüber hinaus lassen sich mit dem MDLG-Verfahren auch die fluktuierende Eigenschaften, Kraftfelder und Randbedingungen des Lattice-Boltzmann-Verfahrens ableiten, dies ist jedoch nicht Gegenstand dieser Arbeit.

Contents

List of Figures

List of Tables

"An intelligent being knowing, at a given instant of time, all forces acting in nature, as well as the momentary positions of all things of which the universe consists, would be able to comprehend the motions of the largest bodies of the world and those of the smallest atoms in one single formula, provided it were sufficiently powerful to all the data to analysis; to it, nothing would be uncertain, both future and past would be present before its eyes."

— Pierre Simon Laplace,

Theorie analytique des probabilites
(Paris: V. Courcier, 1820)

CHAPTER 1

Introduction

Over the last three decades, the LBM has experienced a meteoric rise. It has become the method of choice in a number of areas such as automotive and, to a lesser degree, aerospace industries [2, 3]. With its intrinsic parallelism and ability to cope with complex geometries, the LBM has been developed to solve problems in turbulence [4, 5, 6], multiphase and multicomponent flows [7, 8, 9, 10], and thermal and reactive flows [11, 12], to name just a few. The LBM has also proven useful beyond Newtonian mechanics in disciplines such as quantum and relativistic fluids [13, 14, 15, 16].

To understand the lattice Boltzmann method, we first need to show how it relates to other Computational Fluid Dynamics (CFD) methods [17, 18, 19, 20, 21]. Fluid flows can be defined at different length scales as shown in Fig. 1.1. We have the following regimes from small to large length scale: *microscale*, *mesoscale*, and *macroscale* [22].

First, in the *microscopic* description, the length scale l_{a} is the size of the fluid atom or molecule, with a time scale $t_{\mathrm{c}} \to 0$, being the time needed for a collision. This regime is often referred to as a molecular description governed by Newtonian dynamics [23, 24, 25]. In Molecular Dynamics (MD), the collisions are resolved, thus the method needs to execute a large number of iterations to simulate a single physical second. This makes the MD extremely expensive and therefore, inefficient for large simulations.

Second in the hierarchy is the *mesoscopic* scale, where we do not track the evolution of individual particles, but rather a distribution of particles. This regime is described by the kinetic theory of gases [26, 27]. The *mesoscopic* length and time scales are given by the distance or time between two consecutive collisions, also known as mean free path l_{mfp}, and mean free time t_{mft}, respectively.

Third, the *macroscopic* regime considers a continuum description with length scale l and time scale t much larger than the one of the previously described regimes ($l_{\mathrm{a}} \ll l_{\mathrm{mfp}} \ll l$, and $t_{\mathrm{c}} \ll t_{\mathrm{mft}} \ll t$). The governing equation in *macroscopic* regime is the

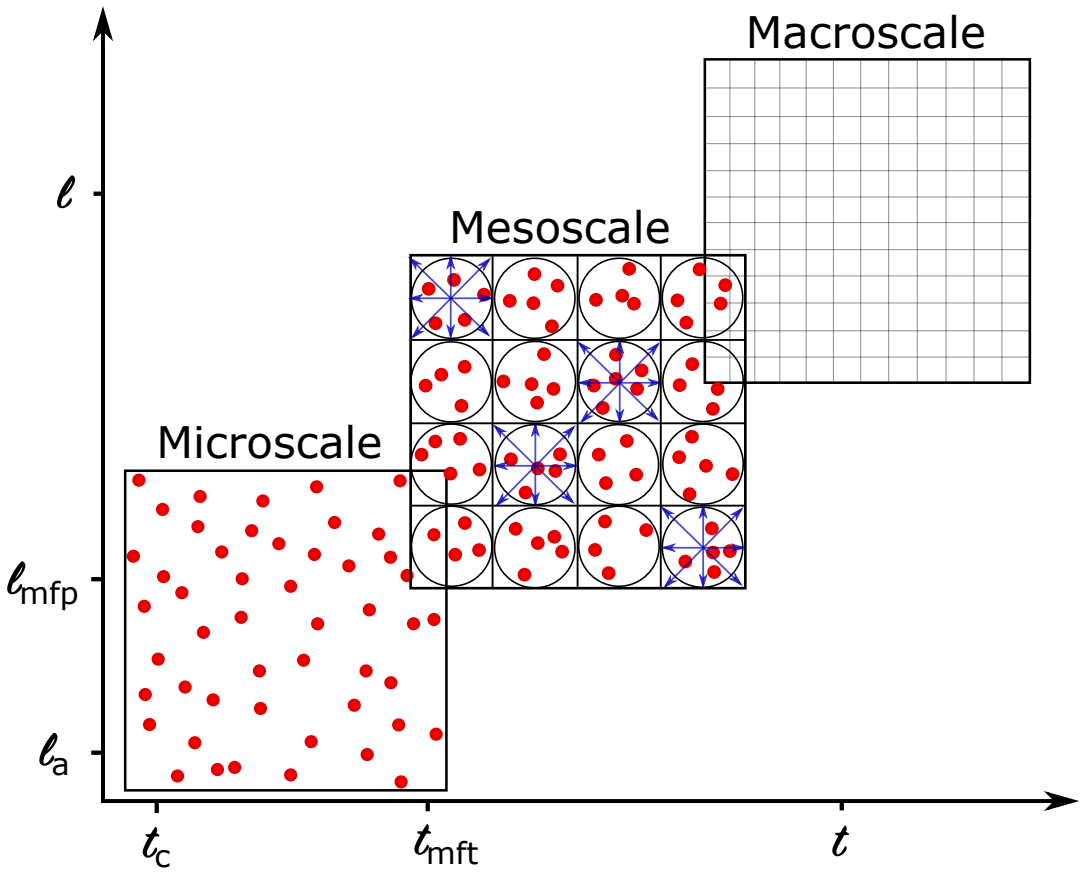


Figure 1.1: The hierarchy of length and time scales for fluid dynamics simulations. They obey the following rules $l_a \ll l_{mfp} \ll l$, and $t_c \ll t_{mft} \ll t$.

Navier-Stokes Equation (NSE) [28], and the observables are quantities such as velocity and density.

The LBM is a *mesoscale* method that tracks the distribution of particles [1]. Initially, it was developed as a postprocessing algorithm to eliminate the statistical noise of the Lattice-Gas Cellular Automata (LGCA) method and to obtain useful macroscopic quantities such as velocity and density [27, 29, 30, 31]. Later, it was realized that LBM can solve some of the shortcomings of the lattice gases, and thus, can be used as a standalone CFD method. In the last three decades, scientists have developed the LBM and in this process, they have found three independent ways to derive the LBM as shown schematically in Fig. 1.2.

The *bottom-up* approach corresponds to the already described link between LGCA and LBM methods. The first LBM had a striking similarity to its predecessors by deriving LBM from its LGCA foundation. In his book, Sauro Succi calls this derivation method the "statistical physics approach" [1]. Later, when significant improvements were suggested to the LBM definition, this link was broken.

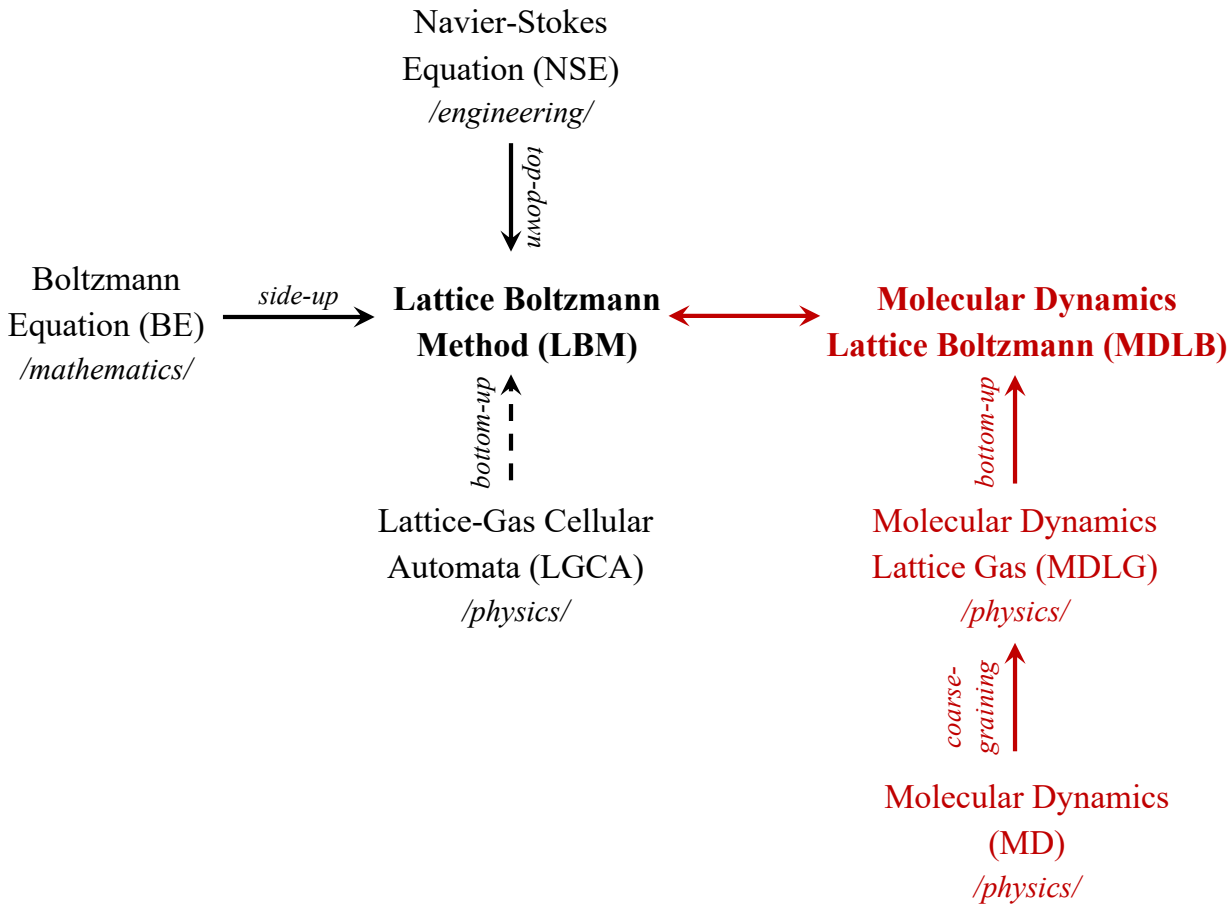


Figure 1.2: Ways to derive LBM. The LBM was first derived from the LGCA using a bottom-up approach based on statistical physics. Later, this link was compromised after realizing that the LBM can be significantly improved by deriving it from Navier-Stokes and Boltzmann equations. We introduce a novel approach to derive the LBM through a coarse-graining of MD simulation. In this way, we restore LBM's lost connection to a physical reality. This figure has been adapted from Sauro Succi's "The Lattice Boltzmann Equation: For Complex States of Flowing Matter" [1] book, and it has been extended to show how the novel MDLB approach relates to the LBM derivation methods.

The *top-down* approach is also called "reverse engineering", where the LBM can be used to recover the NSE. This derivation method ensures that the LBM can be used to simulate the behavior of fluids. Many of the existing LBM solvers were derived using the top-down approach.

The derivation of LBM from the Boltzmann equation provides a "mathematical" perspective to the method. It shows that LBM is a discrete representation of the continuum Boltzmann equation [32]. However, due to the limited number of discrete velocity directions, LBM is, in general, a poor solver of the Boltzmann equation. Nevertheless, this derivation contributed to significant improvements to the LBM such as usage of Maxwellian equilibrium distribution function and an Bhatnagar–Gross–Krook (BGK) collision rule. This link between kinetic theory and LBM allowed for the development of methods for higher-order thermal problems, relativistic schemes as well as entropic and non-ideal LBMs. Furthermore, it proves that as a *mesoscopic* method LBM can solve problems beyond the hydrodynamic limit [22].

Each of these derivation methods brings a unique perspective and inspires further development of the LBM. However, suggesting an improvement from one of the derivation methods could compromise others. Examples for this are (1) the introduction of the Maxwellian instead of previously used Fermi-Dirac equilibrium distribution function, and (2) the BGK collision rule, both coming from kinetic theory of gases. In the Fermi-Dirac equilibrium distribution function the velocity is viscosity dependent, and in the advection term, it has a density dependent prefactor that requires additional rescaling to recover the NSE [33]. This drawback is not present when using the Maxwellian equilibrium function. The BGK collision rule made LBM more efficient by not relying on complex LGCA collision rules as discussed in Section 2.4. The resulting LBM method with Maxwellian equilibrium distribution function and a BGK collision operator cannot be recovered from the LGCA. This means that the physical underpinning of the resulting LBM method is lost. One could say that as long the method performs well, there is no reason for concern, but we believe that this can threaten the very foundations that made the LBM such a surprisingly strong competitor to standard CFD methods. LBM is not simply another method for discretizing the NSE, and the fact that it outperforms classical CFD in modeling of automotive hydrodynamics and it makes significant inroads in the aerospace industry is proof [3, 2]. The success of the lattice Boltzmann method must be firmly attributed to its grounding in some physical reality, and the inability to do this is a worrying shortcoming.

One example of the missing link between the LBM and physical reality is the ability of the former to experience over-relaxation, which is considered to be a numerical trick without a physical underpinning. LBM with over-relaxation can no longer be related to LGCA (bottom-up derivation) methods by a statistical average, since the local collisions can only achieve equilibrium, but never over-relax. Deriving over-relaxation as a discretization of the Boltzmann equation (side-up derivation) equally fails to justify

over-relaxation as shown by Bösch and Karlin [34]. Over-relaxation is also not a concept used in continuum mechanics and the Navier-Stokes equation (top-down derivation). Then why is over-relaxation such an intrinsic part of the LBM?

We anticipate to answer this and other questions by deriving the lattice Boltzmann method in a novel way using a direct mapping from an average over MD simulations onto LBM. This approach has its roots in the Molecular Dynamics Lattice Gas (MDLG) method pioneered by Parsa *et al.* [35].

The MDLG method restores the broken connection between the LBM and the physical reality. It provides a new perspective for the justification of the LBM, which can be used to analyze the building blocks of the LBM. Within this work, we derive the LBM equilibrium distribution function and the collision operator for a simple shear show. For the latter, we observe a natural transition from under-relaxation to over-relaxation depending on the regime (ballistic to diffusive) of the underlying MD simulation.

Future work is needed to obtain a generic MDLG description of the collision operator. The question of the form of the collision operator has been explored by a number of research groups, which is the reason why nowadays there is a "zoo" of collision rules [36] such as Single-Relaxation Time (SRT) LBM [37], Multi-Relaxation Time (MRT) LBM [38, 39, 40, 41], Two-Relaxation Time (TRT) LBM [42, 43], regularized LBM [44, 45, 46], entropic LBM [47, 48, 49, 50], cascaded LBM [51, 52], and cumulant LBM [53, 54, 55]. Another area of interest is bringing back fluctuations to the noise-free LBM, which are much needed when simulating thermodynamic systems [56, 57, 58, 59]. Furthermore, one can consider improving the following LBM properties: incorporating force fields [60, 61, 62], handling boundary conditions for complex flows [63, 64, 65, 66], or estimating coefficients having crucial influence on the flow for multiphase and multicomponent systems [7, 8, 9, 10].

The rest of the thesis is structured as follows: In Chapter 2, we briefly discuss the foundations of this work and introduce the relevant CFD methods. In Chapter 3 we summarize the main contributions of this thesis. In Chapter 4, we conclude our work and give suggestions for further application areas and improvements of the MDLG analysis tool.

The list of peer-reviewed publications and selected talks can be found in Appendix A. The full texts of the peer-reviewed publications can be found as follows: In Appendix B, we investigate the validity of the MDLG equilibrium distribution function. In Appendices C and D, we propose a more accurate MDLG equilibrium distribution function and show that it can recover the measured MDLG equilibrium function. In Appendix E, we investigate a non-equilibrium MD system and measure over-relaxation using the MDLG analysis tool. This confirms that the MDLG method is a powerful novel way to derive the LBM and restore the link between the LGCA and LBM.

CHAPTER 2

Methods

In this chapter, we provide a brief introduction to the numerical methods, which serve as a foundation for this work.

2.1 Molecular dynamics

The MD simulation is a microscopic method [67, 22], tracking the individual position and the velocity of particles for a given domain. The MD simulations are often compared to actual experiments due to the similarities of the procedure [24]: prepare a sample of the system of interest consisting of N particles; solve Newton's equation of motion until the system equilibrates, and then collect sufficient data until the statistical noise is reduced to a minimum to ensure accuracy of the measurement.

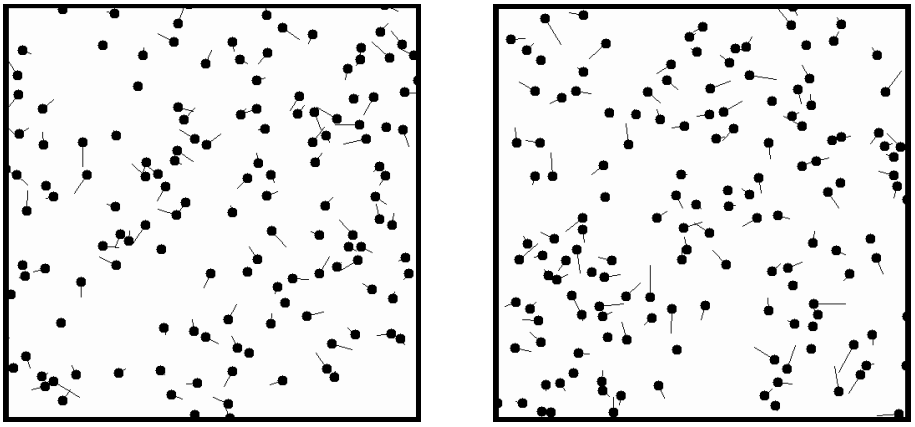
MD is a fundamental approach that models the actual physical forces using inter-molecular interactions. Once we know the total force applied to each particle, we can write Newton's second law

$$\frac{d^2\boldsymbol{x}_i}{dt^2} = \frac{\boldsymbol{f}_i}{m_i} = \frac{1}{m_i}\sum_{j\neq i}\boldsymbol{f}_{ij} \tag{2.1}$$

where $\boldsymbol{x}_i$ is the particle position, $\boldsymbol{f}_i(t)$ is the total force on the i-th particle from all other particles. Having this information, we can calculate the acceleration of each particle. Then, we obtain the updated particle position by integrating the Newton's second law. The most commonly used integrator is the velocity Verlet algorithm [24] given as

$$\begin{aligned}
\boldsymbol{x}_i(t+\Delta t) &= \boldsymbol{x}_i(t) + \boldsymbol{v}_i(t)\Delta t + \frac{1}{2}\frac{\boldsymbol{f}_i(t)}{m_i}\Delta t^2, \\
\boldsymbol{v}_i(t+\Delta t) &= \boldsymbol{v}_i(t) + \frac{1}{2}\left(\frac{\boldsymbol{f}_i(t)}{m_i} + \frac{\boldsymbol{f}_i(t+\Delta t)}{m_i}\right),
\end{aligned} \tag{2.2}$$

where $\boldsymbol{v}_i$ is the particle velocity. In this algorithm, the velocity and position are calculated at the same value of the time variable. A schematic representation of a two-dimensional MD simulation is given in Fig. 2.1.



(a) Time step: t **(b)** Time step: $t + \Delta t$

Figure 2.1: Schematic representation of a two-dimensional MD simulation. The direction of the lines represent the velocity direction of each particle, while the lines length represent the velocity magnitude.

MD is a useful method for simulating microscale phenomena such as chemical reactions, protein folding, nanoconfinement effects, and phase change. Even though this approach is very accurate, it is not typically used for macroscopic phenomena due to its high computational cost even when using the current HPC technologies. This happens because MD tracks individual positions of atoms, which is computationally expensive and effectively limits the simulation time and domain size. This makes MD highly impractical as a Navier-Stokes solver, and one should use a more suitable numerical method for such applications.

Our MDLG analysis tool uses particle positions obtained from an underlying MD simulation to calculate the particle displacements for a specific coarse-grained time step. This procedure allows for first principles investigation of the LBM.

2.2 Continuity and Navier-Stokes equations

The equations of interest for conventional CFD methods are the continuity and the Navier-Stokes equations. These equations are valid as long as the representative physical length scale of the system is much larger than the mean free path of the molecules representing the fluid. Such fluids are referred to as a continuum. The ratio between the

mean free path, λ, and the representative length scale, L, is called the Knudsen number

$$Kn = \frac{\lambda}{L} \tag{2.3}$$

The NSE is valid for $Kn < 0.01$. For larger Knudsen numbers $0.01 < Kn < 0.1$, it can still be used; however, such flows require special boundary conditions.

The two-dimensional continuity and Navier-Stokes equations for steady, incompressible flow with constant properties are

continuity

$$\frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y} = 0 \tag{2.4}$$

x-momentum

$$\frac{\partial u_x}{\partial t} + u_x\frac{\partial u_x}{\partial x} + u_y\frac{\partial u_x}{\partial y} = -\frac{1}{\rho}\frac{\partial P}{\partial x} + \nu\left(\frac{\partial^2 u_x}{\partial x^2} + \frac{\partial^2 u_x}{\partial y^2}\right) + f_x \tag{2.5}$$

y-momentum

$$\frac{\partial u_y}{\partial t} + u_x\frac{\partial u_y}{\partial x} + u_y\frac{\partial u_y}{\partial y} = -\frac{1}{\rho}\frac{\partial P}{\partial y} + \nu\left(\frac{\partial^2 u_y}{\partial x^2} + \frac{\partial^2 u_y}{\partial y^2}\right) + f_y \tag{2.6}$$

where u_x and u_y are the velocities in x- and y- directions, respectively. P is the pressure, ρ is the density of the fluid, ν is the viscosity and f_x and f_y are the body forces for a two-dimensional flow. The continuity equation represent the conservation of mass, while the NSE handles the conservation of momentum for Newtonian fluids. Due to the nonlinear character of the full compressible NSE a generic analytical solution does not exist. In order to obtain analytical, exact and approximate solutions to the NSE, it is necessary to simplify the equations by making assumptions about the flow, the fluid or the geometry of the problem. Some of the commonly used assumptions are: the flow is laminar, steady, two-dimenstional; the fluid is incompressible with constant properties; and the geometry contains parallel plates. Examples of analytic solutions are the parallel flow through a straight channel, Hagen-Poiseuille flow and the Couette flow [68].

When an analytical solution is not possible, the NSE need to be discretized in space and time. In this way, one can approximate their solution using a computer. Commonly used methods to solve the continuity and Navier-Stokes equation are the finite difference, finite volume and finite element method [69, 70]. Details about these conventional Navier-Stokes solvers is out of the scope of this dissertation.

The standard LBM can recover the continuity and Navier-Stokes equations and it is often seen as a second-order accurate solver for the weakly compressible NSE.

2.3 Boltzmann equation

The kinetic theory of gases is considered a bridge between the microscale methods, where we track the motion of individual particles, and the macrosocale where we describe the fluid flow using quantities such as density and velocity. As shown in Fig. 1.1, the mesoscopic kinetic theory describes the distribution of particles in a gas, where the length and time scales are around the mean free path and the mean free time (l_{mfp} and t_{mft}), respectively.

The Boltzmann Equation is given by

$$\frac{\partial f}{\partial t} + \xi_\beta \frac{\partial f}{\partial x_\beta} + \frac{F_\beta}{\rho} \frac{\partial f}{\partial \xi_\beta} = \Omega(f) \tag{2.7}$$

where f is the probability density function which represents the probability of finding a particle at position x at time t. $\Omega(f) = df/dt$ is the effect of collisions on f. The parameter ξ describes the particle velocity with the first two terms representing the advection of the distribution function with velocity ξ. The third term describes the forces applied on the velocity. On the right hand side, we have a source term which redistributes f according to collisions and it is called collision rule.

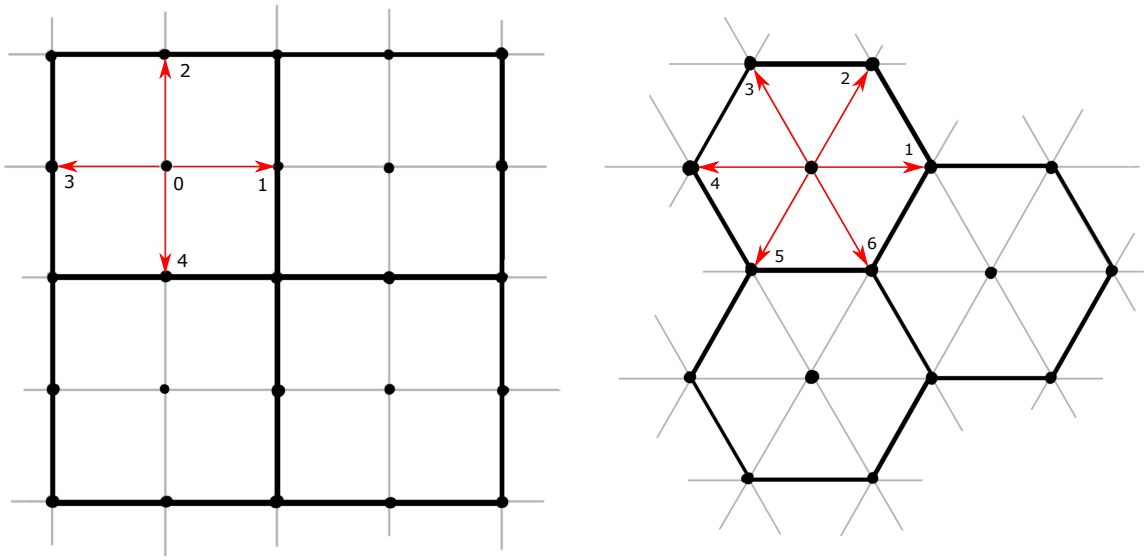
The collision rule in kinetic theory is mathematically represented by a double integral over the velocity space. It models all outcomes of two-particle collisions considering the intermolecular forces.

The LBM is considered a numerical solver for the Boltzmann equation [9]. However, for regimes where a good solution for the Boltzmann equation is needed, the LBM is not applicable due to the small number of discrete velocities. This limitation of the LBM is also its main advantage; it allows for much simpler collision rule. This link to kinetic theory enables LBM to simulate phenomena beyond the hydrodynamic limit.

2.4 Lattice-gas cellular automata

LGCA emerged as a Boolean analogue of MD and provided a much needed trade-off between computational efficiency and physical accuracy. Later on in 1986, Frisch, Hasslacher, and Pomeau [71] published an LGCA method capable of simulating fluid flows. Compared to previously suggested LGCA models such as the Hardy-Pomeau-de Pazzis (HPP) [72], the Frisch-Hasslacher-Pomeau (FHP) model uses hexagon lattice cells instead of the square domain used by the HPP model as shown in Fig. 2.2. The HPP model obeys the conservation laws; however, it did not have enough isotropy to solve the Navier-Stokes equations.

In general, the LGCA methods adhere to the following rules:



(a) **(b)**

Figure 2.2: (a) Square lattice used in the HPP model with only four prescribed directions; (b) Hexagon lattice used in the FHP model with six prescribed directions.

- Particles can move only along the discrete directions allowed for the specific LGCA model, *e.g.,* in the FHP model only six discrete directions $\boldsymbol{c}_i$ are allowed. This leads to a Fermi-Dirac equilibrium distribution function;
- Particles can move only to the nearest neighbor;
- Exclusion principle, which says that only one particle can be present in a node at any time.

The absence (value of 0) or the presence (value of 1) of a particle at a site $\boldsymbol{x}$ and time t is defined by a Boolean occupation number $n_i(\boldsymbol{x}, t)$. The mass and momentum can be recovered from the occupation numbers [27, 73] as

$$\rho(\boldsymbol{x}, t) = \frac{m}{v_0} \sum_i n_i(\boldsymbol{x}, t), \tag{2.8}$$

$$\rho u(\boldsymbol{x}, t) = \frac{m}{v_0} \sum_i \boldsymbol{c}_i n_i(\boldsymbol{x}, t), \tag{2.9}$$

where m is the particle mass, and v_0 is the volume of the node. The LGCA evolution equation can be divided into two simple steps – *collision* and *streaming*. The collisions are local operations that redistribute particles according to collision rules

$$n_i^*(\boldsymbol{x}, t) = n_i(\boldsymbol{x}, t) + \Xi_i(\boldsymbol{x}, t), \tag{2.10}$$

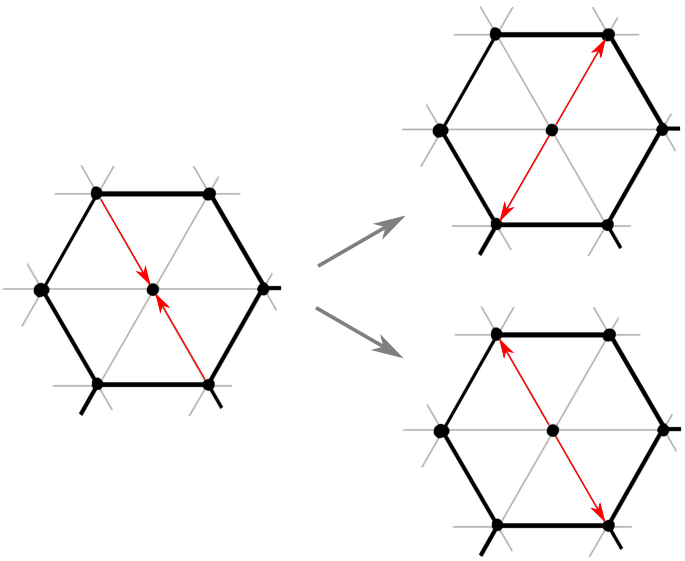


Figure 2.3: FHP collision with two possible collision outcomes.

that conserve particle number and local momentum

$$\sum_i \Xi_i(\boldsymbol{x}, t) = 0, \qquad \text{and} \qquad \sum_i \boldsymbol{c}_i \Xi_i(\boldsymbol{x}, t) = 0, \tag{2.11}$$

as shown in Fig. 2.3. The effectiveness of collisions can be tuned [33, 74], and as a result, the viscosity can be reduced up to a certain degree. However, depending on the LGCA model and the number of discrete directions $\boldsymbol{c}_i$, it becomes cumbersome to express all the possible collision rules mathematically [27, 29].

After the collision step, particles propagate from one node to a neighboring node following the prescribed velocity direction $\boldsymbol{c}_i$ according to the following equation

$$n_i(\boldsymbol{x} + \boldsymbol{c}_i \Delta t, t + \Delta t) = n_i^\star(\boldsymbol{x}, t). \tag{2.12}$$

The complete LGCA evolution equation is given by

$$n_i(\boldsymbol{x} + \boldsymbol{c}_i \Delta t, t + \Delta t) = n_i(\boldsymbol{x}, t) + \Xi_i(\boldsymbol{x}, t). \tag{2.13}$$

The main advantage of the LGCA model is its perfect collisions due to the Boolean occupation numbers, which ensure that there is no round-off error, commonly observed for floating-point operations. In addition, the collision is a local operation, where each site is calculated independently of the others, and thus, no communication between processors is needed making the model intrinsically parallel.

At this stage of the development of the LGCA, their main disadvantage is the large amount of statistical noise, which could be helpful in certain applications such as measuring thermal fluctuations of a real gas [27]. However, when analyzing the properties of a macroscopic simulation, the fluctuations are rather a source of concern. LGCA methods also suffer from Galilean invariance and velocity-dependent fluid pressure caused by the exclusion principle. In addition, they are not capable of simulating high Reynolds numbers due to high viscosity. Even though the collision rule is perfectly parallelizable, in higher dimensions, the complexity of the collision rules grows exponentially, *e.g.,* in three-dimensions with 24 velocities, the model has 2^{24} collision rules in a node [22]. In such cases, lookup tables are often used to assign a collision rule [1].

Later on, Boghosian *et al.* [75] introduced the integer lattice gases, where one can control the level of fluctuations while maintaining the exact conservation laws and having unconditional stability. In the pursuit of understanding better the fluctuating systems, Blommel *et al.* [76] constructed a new integer lattice gas with Monte Carlo collision operator.

In the late 80s, the LBM was proposed as a remedy to most of the shortcomings of the LGCA [77, 78, 79]. The original LBM was derived from lattice gases, and its link to kinetic theory was not fully understood until the mid-90s [9].

2.5 Lattice Boltzmann method

In 1988, in order to reduce the statistical noise of the LGCA method, McNamara and Zanetti [77] replaced the Boolean occupation number n_i with its expectation value given by

$$f_i = \langle n_i \rangle \tag{2.14}$$

resulting in an Lattice Boltzmann Equation (LBE). This increased the computational efficiency of the method, since additional averaging over lattice gas results is no longer required. This counteracts the higher computational cost arising from transitioning from a Boolean lattice gas to an LBM requiring real numbers. Since, the LBM was derived directly from the revolutionary FHP lattice gas method [71], some limitations inherited from its predecessor remained. The most significant shortcoming of the novel LBM are low Reynolds flows due to low collisionality of the underlying LGCA rules, and mildly Galilean invariance violating terms, caused by the usage of a Fermi-Dirac distribution function [33].

Later, the LBM was connected to the kinetic theory of gases by introducing the BGK [37] collision operator with Maxwell-Boltzmann equilibrium distribution function, which removed the velocity dependent terms in the NSE [80]. These two modifications allowed for the LBM method to become one of the leading CFD methods worldwide.

The BGK-LBM evolution equation is defined as

$$f_i(\boldsymbol{x} + \boldsymbol{c}_i\Delta t, t + \Delta t) = f_i(\boldsymbol{x}, t) + \Omega_i(\boldsymbol{x}, t), \tag{2.15}$$

and similarly to the LGCA can be divided into two simple steps:

$$\textit{collision step:} \quad f_i^\star(\boldsymbol{x}, t) = f_i(\boldsymbol{x}, t) + \Omega_i(\boldsymbol{x}, t), \tag{2.16}$$

$$\textit{streaming step:} \quad f_i(\boldsymbol{x} + \boldsymbol{c}_i\Delta t, t + \Delta t) = f_i^\star(\boldsymbol{x}, t). \tag{2.17}$$

The *collision step* redistributes the populations $f_i(\boldsymbol{x}, t)$ and relaxes them towards equilibrium at rates corresponding to the relaxation matrix Λ_{ij}. The MRT-BGK collision operator is given by

$$\Omega_i(\boldsymbol{x}, t) = \sum_j \Lambda_{ij} \left(f_j^{\mathrm{eq}}(\boldsymbol{x}, t) - f_j(\boldsymbol{x}, t) \right). \tag{2.18}$$

A simpler version of the BGK collision operator is the SRT-BGK

$$\Omega_i(\boldsymbol{x}, t) = \frac{\Delta t}{\tau} \left(f_i^{\mathrm{eq}}(\boldsymbol{x}, t) - f_i(\boldsymbol{x}, t) \right), \tag{2.19}$$

where all the moments relax to the same relaxation time τ. The LBM collision operator is constructed such that total density and momentum are conserved

$$\sum_i \Omega_i(\boldsymbol{x}, t) = 0, \tag{2.20}$$

$$\sum_i \boldsymbol{c}_i \Omega_i(\boldsymbol{x}, t) = 0. \tag{2.21}$$

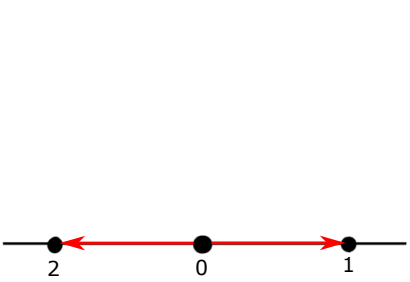
Often used equilibrium distribution function in LBM is the second-order truncation of the Maxwell-Boltzmann distribution function given by

$$f_i^{\mathrm{eq}}(\boldsymbol{x}, t) = \rho w_i \left(1 + \frac{\boldsymbol{u} \cdot \boldsymbol{c}_i}{c_s^2} + \frac{(\boldsymbol{u} \cdot \boldsymbol{c}_i)^2}{2c_s^4} - \frac{\boldsymbol{u} \cdot \boldsymbol{u}}{2c_s^2} \right), \tag{2.22}$$

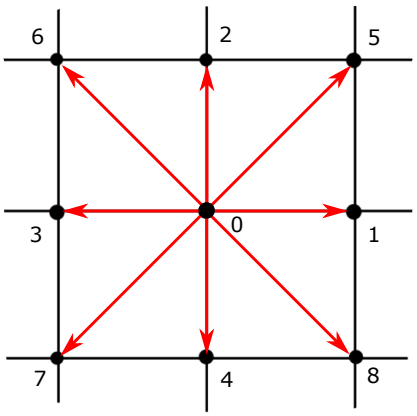
with a velocity set defined as a combination of the discrete velocities $\boldsymbol{c}_i$ and their corresponding weights w_i given by $\{\boldsymbol{c}_i, w_i\}$. The velocity set is often denoted as DdQq with d being the number of spatial dimensions, and q the number of discrete velocities. Commonly used velocity sets are D1Q3, D2Q9 and D3Q27 as shown in Fig. 2.4. Their corresponding weights w_i are given in Table 2.1. The parameter c_s describes the relationship between the pressure and the density for an isothermal model given by

$$p = c_s^2 \rho, \tag{2.23}$$

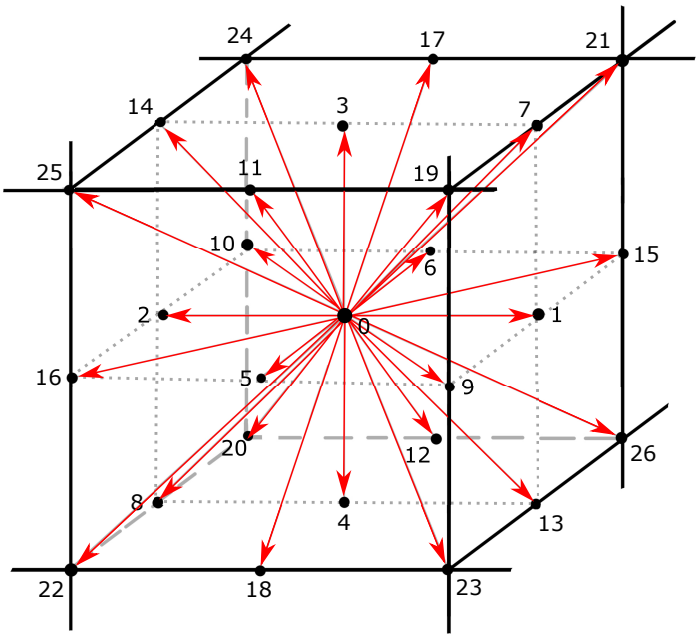
which corresponds to the speed of sound. For all the lattices presented in this overview, the speed of sound is $c_s^2 = (1/3)\Delta x^2/\Delta t^2$, where Δx and Δt are the space and time resolution, respectively. In LBM, these parameters are typically set to unity ($\Delta x = \Delta t = 1$).



(a) D1Q3 velocity set



(b) D2Q9 velocity set



(c) D3Q27 velocity set

Figure 2.4: Commonly used velocity sets: (a) D1Q3 velocity set; (b) D2Q9 velocity set; (c) D3Q19 velocity set. The weights for each velocity length are given in Table 2.1.

DdQq	**Velocities** $\boldsymbol{c}_i$	**Length** $\lvert\boldsymbol{e}_i\rvert$	**Weights** w_i
D1Q3	(0)	0	$2/3$
	(± 1)	1	$1/3$
D2Q9	$(0,0)$	0	$4/9$
	$(\pm 1,0),(0,\pm 1)$	1	$1/9$
	$(\pm 1,\pm 1),(\pm 1,\pm 1)$	$\sqrt{2}$	$1/36$
D3Q27	$(0,0,0)$	0	$8/27$
	$(\pm 1,0,0),(0,\pm 1,0),(0,0,\pm 1)$	1	$2/27$
	$(\pm 1,\pm 1,0),(\pm 1,0,\pm 1),(0,\pm 1,\pm 1)$	$\sqrt{2}$	$1/54$
	$(\pm 1,\pm 1,\pm 1)$	$\sqrt{3}$	$1/216$

Table 2.1: Properties of the most popular DdQq lattices used to solve the NSE. The speed of sound for all these velocity sets is $c_s = 1/\sqrt{3}$.

The discrete-velocity distribution function $f_i(\boldsymbol{x},t)$ and the equilibrium distribution function $f_i^{\mathrm{eq}}(\boldsymbol{x},t)$ conserve density and momentum

$$\rho = \sum_i f_i(\boldsymbol{x},t) = \sum_i f_i^{\mathrm{eq}}(\boldsymbol{x},t), \tag{2.24}$$

$$\rho\boldsymbol{u} = \sum_i \boldsymbol{c_i} f_i(\boldsymbol{x},t) = \sum_i \boldsymbol{c_i} f_i^{\mathrm{eq}}(\boldsymbol{x},t). \tag{2.25}$$

The *streaming step* propagates the collided distribution function $f_i^{\star}(\boldsymbol{x},t)$ to the neighboring cells as given in Eq. (2.17).

The NSE can be recovered from the LBE through the Chapman-Enskog expansion [81], or the Taylor expansion [82]. According to the derivation, the kinematic shear viscosity in LBM is given by the relaxation time

$$\nu = c_s^2 \left(\tau - \frac{\Delta t}{2}\right) \tag{2.26}$$

and the kinematic bulk viscosity is then $\nu_B = 2\nu/3$. Depending on the value of $\tau/\Delta t$ as given in Eq. (2.19), we define three different regimes of the collision operator [22]:

- *under-relaxation*: with $\tau/\Delta t > 1$, for which $f_i(\boldsymbol{x},t)$ decays exponentially towards equilibrium.

- *full-relaxation*: with $\tau/\Delta t = 1$, for which $f_i(\boldsymbol{x},t)$ decays directly to equilibrium.

- *over-relaxation*: with $1/2 < \tau/\Delta t < 1$, for which $f_i(\boldsymbol{x}, t)$ oscillates around equilibrium with a decreasing amplitude. $\tau/\Delta t \geq 1/2$ is a necessary condition for stability.

Over-relaxation is commonly used to reduce the effective LBM viscosity, which allows for simulating high Reynolds number flows.

In the last three decades, LBM has been constantly improved, and it has grown in popularity. Nowadays, it is a model of choice for automotive hydrodynamics and the aerospace industry due to its local operations, close to perfect parallelization, and simple meshing. However, LBM is often seen as a black box, and even commonly used properties such as over-relaxation, equilibrium distribution function, fluctuations, and collision rules, are not always grounded in physical reality. It is concerning that over the years the LBM has lost its foundation in LGCA, and thus, in its current form, cannot be derived from it. We developed a novel approach to derive the LBM from an underlying MD simulation, which restores the connection between lattice gases and LBM.

2.6 Molecular dynamics lattice gas method

We developed a novel method to derive the LBM using a direct mapping approach from an average over MD simulations onto LBM. This coarse-graining approach has its roots in the MDLG method, pioneered by Parsa *et al.* [35].

The MDLG consists of overlaying a square lattice with lattice spacing Δx onto an MD simulation. Then, we track the migration of the particles of the MD simulation to imposed lattice positions with displacement $\boldsymbol{c}_i$ after a time step Δt, as shown in Fig. 2.5a. The integer lattice gas occupation number $n_i(\boldsymbol{x}, t)$ is given by

$$n_i(\boldsymbol{x}, t) = \sum_j \Delta_{\boldsymbol{x}}[x_j(t)] \Delta_{\boldsymbol{x}-\boldsymbol{c}_i}[x_j(t - \Delta t)], \tag{2.27}$$

where $\Delta_{\boldsymbol{x}}[x_j(t)] = 1$, if particle j is in lattice cell $\boldsymbol{x}$ at time t, and $\Delta_{\boldsymbol{x}}[x_j(t)] = 0$, otherwise. Fig. 2.5b shows a schematic representation of the lattice velocities for an D2Q49 lattice. The MDLG evolution equation describes essentially a lattice gas defined as

$$n_i(\boldsymbol{x} + \boldsymbol{c}_i \Delta t, t + \Delta t) = n_i(\boldsymbol{x}, t) + \Xi_i(\boldsymbol{x}, t), \tag{2.28}$$

where the collision operator $\Xi_i(\boldsymbol{x}, t)$ is defined as

$$\Xi_i(\boldsymbol{x}, t) = n_i(\boldsymbol{x} + \boldsymbol{c}_i \Delta t, t + \Delta t) - n_i(\boldsymbol{x}, t). \tag{2.29}$$

This lattice gas rigorously conserves the hydrodynamic properties of the system up to the coarse-graining approximation because it is obtained from a coarse-graining of an underlying MD simulation.

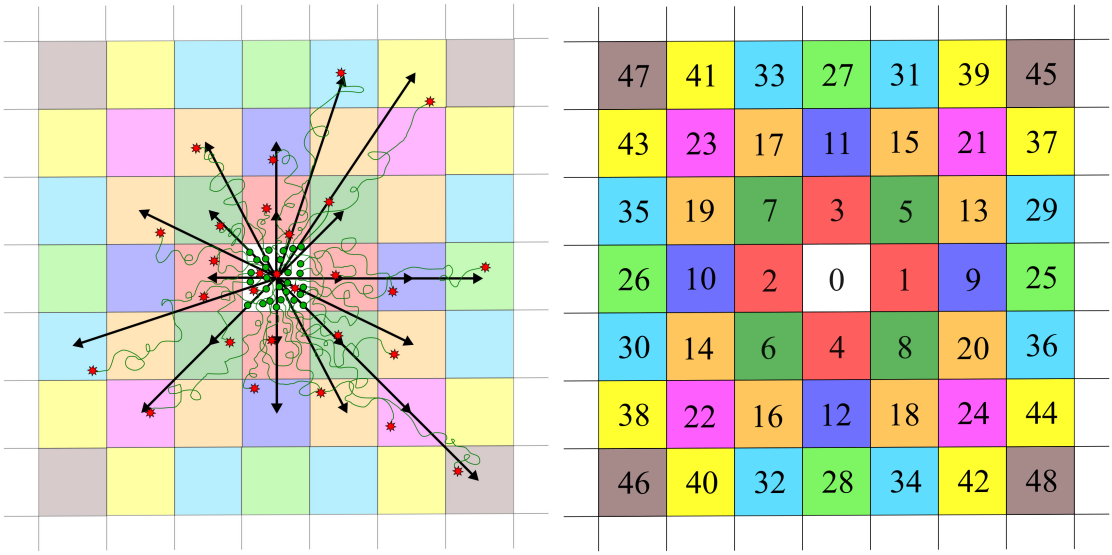


Figure 2.5: (a) Sketch of the MDLG analysis. A lattice is superimposed onto the MD simulation domain. The movement of the particles is tracked from the central node using their MD trajectories. The green circles represent the position of the particles at time $t - \Delta t$ and the red circles are their respective positions at time t. Using the particle trajectories and the imposed lattice, the occupation number n_i is defined as given in Eq. (2.27). The black arrows are the lattice velocities. Only the lattice velocities which have at least one particle within their area (i.e. non-zero occupation number) are shown. (b) Schematic representation of the D2Q49 lattice with the numbering convention for the lattice velocities in two dimensions. The central point 0 corresponds to the zeroth-velocity $v_0 = (0, 0)$ and the rest of the velocities are given as a vector connecting the central point and the lattice point in question as shown in (a). The velocities are color coded depending on their length.

2.7 Molecular dynamics lattice Boltzmann

To obtain a lattice Boltzmann representation of the MDLG method, we average over an ensemble of MD simulations of the same macroscopic state. Thus, we obtain the particle probability distribution function

$$f_i = \langle n_i \rangle_{\text{neq}}. \tag{2.30}$$

We call this method MDLB. The MDLB evolution equation is given as a non-equilibrium ensemble average of Eq. (2.28) equivalent to

$$f_i(\boldsymbol{x} + \boldsymbol{c}_i \Delta t, t + \Delta t) = f_i(\boldsymbol{x}, t) + \Omega_i(\boldsymbol{x}, t), \quad \text{with} \quad \Omega_i(\boldsymbol{x}, t) = \langle \Xi_i(\boldsymbol{x}, t) \rangle_{\text{neq}}, \tag{2.31}$$

where $\Omega_i(\boldsymbol{x}, t)$ is the MDLB collision operator. In the MDLB description the probability distribution function can be derived by tracking the trajectories of a non-equilibrium MD simulation given as

$$\begin{aligned} f_i &= \langle n_i \rangle^{\text{neq}} \\ &= \left\langle \sum_j \Delta_{\boldsymbol{x}}[x_j(t)] \Delta_{\boldsymbol{x}-\boldsymbol{c}_i}[x_j(t-\Delta t)] \right\rangle \\ &= N \int dx_1 \int d\delta x_1 \; P^{(1)}(x_1, \delta x_1) \Delta_{\boldsymbol{x}}[x_1] \Delta_{\boldsymbol{x}-\boldsymbol{c}_i}[x_1 - \delta x_1], \end{aligned} \tag{2.32}$$

where N is the total number of particles and $P^{(1)}$ is the one-particle displacements Probability Distribution Function (PDF). When considering an equilibrium MD simulation, we can derive the equilibrium distribution function f_i^{eq}.

The equilibrium distribution function f_i^{eq} is a key element in the LBM context, and the MDLB formulation allows us to derive f_i^{eq} analytically from the one-particle displacements PDF. In the publications [83, 84], we study the optimal form of the one-particle displacement distribution function. The simplest one-particle displacement distribution function, which can be used to derive an equilibrium distribution function is the single Gaussian. We discuss the validity of this equilibrium distribution function depending on the area fraction of the MD simulation in Parsa *et al.* [85]. Under further investigation, we observe small discrepancies (approx. 5%) between the measured f_i^{eq} and the one obtained from a simple Gaussian distribution function. Therefore, in the publications [83, 84], we propose more complex one-particle distribution function – a Poisson weighted sum of Gaussians, which significantly improves the accuracy of the MDLB equilibrium distribution function.

Another key component of the LBM is the collision operator Ω_i, which in the MDLB description is derived as

$$\begin{aligned} \Omega_i &= \langle \Xi_i \rangle \\ &= \int dx_1 \int d\delta_1 \cdots \int dx_N \int d\delta_N P^N(x_1, \delta_1, \ldots, x_N, \delta_N, t + \Delta t) \Delta_{\boldsymbol{x}+\boldsymbol{c}_i}(x_j) \Delta_{\boldsymbol{x}}(x_j - \delta_j) \\ &\quad - \int dx_1 \int d\delta_1 \cdots \int dx_N \int d\delta_N P^N(x_1, \delta_1, \ldots, x_N, \delta_N, t) \Delta_{\boldsymbol{x}}(x_j) \Delta_{\boldsymbol{x}-\boldsymbol{c}_i}(x_j - \delta_j) \\ &= N \int dx_1 \int d\delta_1 P^1(x_1, \delta_1, t + \Delta t) \Delta_{\boldsymbol{x}+\boldsymbol{c}_i}(x_1) \Delta_{\boldsymbol{x}}(x_1 - \delta_1) \\ &\quad - N \int dx_1 \int d\delta_1 P^1(x_1, \delta_1, t) \Delta_{\boldsymbol{x}}(x_1) \Delta_{\boldsymbol{x}-\boldsymbol{c}_i}(x_1 - \delta_1) \end{aligned} \tag{2.33}$$

For stationary problems the one-particle distribution function does not depend on time. In such cases, we obtain

$$\begin{aligned}\Omega_i &= \langle \Xi_i \rangle \\ &= N \int dx_1 \int d\delta_1 P^1(x_1, \delta_1) \Delta_{\boldsymbol{x}+\boldsymbol{c}_i}(x_1) \Delta_{\boldsymbol{x}}(x_1 - \delta_1) \\ &\quad - N \int dy_1 \int d\gamma_1 P^1(y_1, \gamma_1) \Delta_{\boldsymbol{x}}(y_1) \Delta_{\boldsymbol{x}-\boldsymbol{c}_i}(y_1 - \gamma_1).\end{aligned} \tag{2.34}$$

The MDLB formulation allows us to express $f_i(\boldsymbol{x}, t)$ and Ω_i as quantities which depend solely on the one-particle displacements PDF. We use this description to provide crucial information about the collision rule. In Section E, we show the derivation of the MDLB collision operator, and its ability to exhibit over-relaxation. This is a fundamental result, since until now, the over-relaxation of the LBM has been considered a numerical trick to improve the stability and reduce the viscosity of the simulations [34].

As mentioned in Section 1, there is a variety of LBM collision rules and currently there is no way to distinguish which one is more accurate in a specific application. The MDLB method allows for measuring the true collision rule from an underlying non-equilibrium MD simulation after a sufficient averaging, and thus, it can rank the applicability of the existing collision operators. Furthermore, using the MDLG description, we could reintroduce fluctuations to the LBM and derive better boundary conditions and forcing LBM schemes. We anticipate to answer these questions using the MDLG analysis; however, they are out of scope of this work.

CHAPTER 3

Contributions

The main contribution of this work is the development of a novel derivation method for the lattice Boltzmann method from an underlying Molecular Dynamics simulation. This research grew out of the need to recover the missing connection between the LBM and its predecessor LGCA. As mentioned in Chapter 1, improvements of the LBM from the top-down and the side-up derivation methods could compromise other derivation procedures as shown in Fig. 1.2. Example for this is the failure to recover the LBM from LGCA after introducing the Maxwellian equilibrium distribution function and the BGK collision rule.

The MDLG analysis tool is pioneered by Parsa *et al.* [35] and aims to reestablish the missing link by coarse-graining of an MD simulation.

This chapter summarizes the main contributions for each of the publications, part of this thesis. The author's publications support this statement and can be found in Appendices B-E.

3.1 Publication 1: Validity of the MDLG global equilibrium distribution function

The publication *M. R. Parsa, A. Pachalieva and A. J. Wagner, "Validity of the Molecular-Dynamics-Lattice-Gas Global Equilibrium Distribution Function", International Journal of Modern Physics C, Vol. 30, No. 10, 1941007 (2019)* [85] discusses the validity range of the MDLG equilibrium distribution function for a variety of area fractions.

This work helps to understand the properties and limitations of the MDLG equilibrium probability distribution function in respect to different volume fractions. In general, one can calculate the equilibrium distribution function using a single parameter λ, defined as a ratio between the velocity correlation function and the mean-squared displacement of the underlying MD simulation. However, for higher volume fractions this relation breaks

because the velocity auto-correlation function cannot reproduce the correct behavior of the mean-squared displacement anymore. This leads to wrong estimation of the equilibrium distribution function compared to the measured one from the MD simulation. In order to recover the correct analytical MDLG f_j^{eq} for high volume fractions, we need to use the measured mean-squared displacement instead of the theoretical one. Thus, the MDLG equilibrium distribution function appears to be universally defined as a function of the mean-squared displacement and the lattice size of the underlying MD system.

Exact knowledge of the equilibrium distribution is important when studying a non-equilibrium system. In a future publication, we intend to investigate whether the MDLB collision operator can be expressed in terms of an BGK collision operator $\Omega_i(\boldsymbol{x}, t) = \sum_j \Lambda_{ij}[f_j^{\mathrm{eq}} - f_j(\boldsymbol{x}, t)]$, with Λ_{ij} being the relaxation matrix. Here, the knowledge of an exact local equilibrium distribution is of the essence. Preliminary studies show that for non-equilibrium cases, we observe a small difference (approximately 0.5%) between the measured and the analytical f_j^{eq}. The reason for this discrepancy might be due to the assumption made by Parsa *et al.* [35] that the displacements probability distribution for all scales (from ballistic to diffusive regime) is given by a single Gaussian distribution function. This assumption was made because a Gaussian distribution can describe the probability distribution for very short or very long time scales. However, intermediate probabilities between the ballistic and the diffusive regime might have different behavior. Those intermediate regimes, due to their nature, might be better approximated, for example, by a combination of two Gaussian distributions. These considerations are a subject of our next publication [83].

Overall, the exact knowledge of the MDLG equilibrium distribution function is essential when looking at the deviations from local equilibrium, such as the investigation of collision operators, which we discuss in Pachalieva *et al.* [86]. Another area of interest is when analyzing the fluctuating behavior of the MDLG systems, which will elucidate the correct form of fluctuations for non-ideal systems.

3.2 Publication 2: Non-Gaussian distribution of displacements for LJ particles in equilibrium

In the publication A. Pachalieva *and A. J. Wagner, "Non-Gaussian distribution of displacements for Lennard-Jones particles in equilibrium", Physical Review E, 102(5), 053310 (2020)* [83], we show that the displacements equilibrium distribution for large range of coarse-graining scales cannot be approximated using a Gaussian distribution, as assumed previously.

In statistical mechanics, the velocity distribution function of particles is often given by a Gaussian distribution function referred to as the Maxwell-Boltzmann distribution. Here, we focus on the distribution of particle displacements as these displacements can

be seen as time averages of the velocities. Therefore, the displacement distribution function is typically assumed to be a Gaussian distribution function as well. However, we investigate the displacements distribution for different finite time steps and found that the Maxwell-Boltzmann function is only valid in the extreme ballistic (very short Δt) and diffusive (very large Δt) regimes. In the transition regime, the Maxwellian does not capture the correct distribution of displacements and introduces an error to the collision operator.

We propose a novel one-particle PDF using a Poisson Weighted Sum of Gaussians (WSG) for the intermediate regime. The Poisson WSG shows a better agreement with the measured displacements from the MD simulation. 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. A drawback of this description is that no theoretical solution is available for the fourth-order moment, which needs to be measured for each time step apriori.

It is important to mention that the system considered in this publication is in equilibrium, and there are no correlations in the velocities of different particles. This means that the distribution of the velocities remains Gaussian at all times, even when the distribution of the displacements in the transition regime becomes non-Gaussian.

Capturing the correct behavior of the displacement distribution function would allow us to recover a valid equilibrium distribution function. This is of practical relevance for mesoscale simulation methods like Brownian Dynamics [87], Dissipative Particle Dynamics [88], Stochastic Rotation Dynamics [89], and the lattice Boltzmann method [80, 9], to name but a few. However, a question remains if the one-particle PDF can reproduce more accurately the equilibrium distribution function, which is the subject of our next publication [84].

3.3 Publication 3: MDLG equilibrium distribution function for LJ particles

In the publication *"A. Pachalieva and A. J. Wagner, "Molecular dynamics lattice gas equilibrium distribution function for Lennard-Jones particles", Philosophical Transactions of the Royal Society A, (2021)"* [84] we investigate the ability of the newly proposed Poisson WSG one-particle probability distribution function to approximate the measured MD equilibrium distribution function.

The results show that the newly proposed Poisson WSG model captures much better the MD equilibrium distribution function compared to the previously used simple Gaussian distribution function. This remains true for purely ballistic and purely diffusive regimes where the Poison WSG distribution function is reduced to a Gaussian. Even though the deviations between the Poison WSG and the single Gaussian distribution

function are relatively small, they are of significant importance when a non-equilibrium system is analyzed. Typically, in hydrodynamics, the system of interest is not too far from equilibrium, and thus even small deviations play a crucial role in estimating the correct equilibrium distribution function. A key interest of having a good approximation of the equilibrium distribution function is to analyze non-equilibrium predictions of the MDLG mapping. In future work, we aim to investigate the MDLG predictions for LBM collision operators [86]. This task would be nearly impossible without an in-depth understanding of the equilibrium behavior of the MDLG mapping.

3.4 Publication 4: Connecting LBM to physical reality by coarse-graining MD simulation

The publication *"A. Pachalieva and A. J. Wagner, "Connecting lattice Boltzmann methods to physical reality by coarse-graining Molecular Dynamics simulation", submitted to Physical Review Letters, (2021)"* [86] discusses the frequently used over-relaxation, which has been seen as lacking physical motivation.

Original LBMs are obtained from Boltzmann averages of lattice gases, where local collisions conserve particle number and local momentum. In LGCA, the local collisions can only bring the simulations to a state of local equilibrium. Later on, to solve the problem of Galilean invariance, the Fermi-Dirac equilibrium distribution function was replaced by a Maxwell-Boltzmann equilibrium distribution function and an BGK collision term defined as

$$f_i(\boldsymbol{x} + \boldsymbol{c}_i\Delta t, t + \Delta t) = f_i(\boldsymbol{x}, t) + \Omega_i, \tag{3.1}$$

with an BGK collision operator

$$\Omega_i = \sum_j \Lambda_{ij}[f_j^{\mathrm{eq}} - f_j(\boldsymbol{x}, t)], \tag{3.2}$$

where the f_i are continuous densities associated with a lattice velocity $\boldsymbol{c}_i$ that represent an expectation value for the number of particles moving from lattice site $\boldsymbol{x} - \boldsymbol{c}_i\Delta t$ to lattice site $\boldsymbol{x}$ at time t. The BGK collision term redistributes those densities and relaxes them towards an imposed local equilibrium distribution f_i^{eq}. In the simplest case, the relaxation matrix Λ_{ij} has a single relaxation time $\Lambda_{ij} = (1/\tau)\delta_{ij}$, where $\tau = 1$ implies that local equilibrium is reached in one time step.

Such lattice Boltzmann methods cannot be justified by averages over Boolean lattice gases and require a different derivation method. Deriving LBM from the Boltzmann equation is equally unable to recover over-relaxation as shown by Bösch and Karlin [34].

In this publication, we show that LBM can be derived using the MDLG analysis tool. The MDLB collision operator is directly calculated using a one-particle displacement

probability distribution, which can be obtained from an MD simulation. Furthermore, this approach shows that such lattice Boltzmann collision operators naturally transition from under- to over-relaxation. Thus, the over-relaxation in LBM can be derived from first principles and is a consequence of the coarse-grained representation of the LBM.

CHAPTER 4

Concluding Discussion

Within this work, we have shown that the LBM can be derived directly from an explicit coarse-graining procedure using microscopic information such as MD simulations, which we call Molecular Dynamics Lattice Gas analysis tool. To conduct the analysis, we track the migration of particles between coarse-grained time steps, which is sufficient to derive the LBM.

The MDLG analysis tool is a novel LBM derivation method, that restores the link between LBM and its microscopic nature. The MDLG tool has already proved to be useful in understanding the key elements of the LBM such as equilibrium distribution function and the collision operator.

We derived an improved MDLB equilibrium distribution function, which is of key interest when looking at systems with small deviations from local equilibrium [83, 84]. In such cases, even small deviations of the equilibrium distribution function can compromise the derivation of a collision operator. Our study showed that a single Gaussian could not approximate the one-particle displacement distribution function used for deriving the equilibrium distribution function in the transition region between ballistic and diffusive regimes. In order to derive a better equilibrium distribution function, we propose a novel one-particle distribution function, that uses the second- and fourth-order moments measured from an MD simulation. This distribution is defined as a mixture of Gaussian distributions that have undergone a number of collisions, given by a Poisson distribution. Using this distribution function, we obtain nearly perfect approximation of the measured displacement distribution function.

Next, we studied the MDLG collision operator obtained from a Boltzmann average from an underlying MD data. We derive the LBM collision operator for a simple shear flow using the MDLG description. More importantly, in our study we show that the most frequently used property of the LBM, over-relaxation, arises naturally from physical lattice gases. Thus, we confirm the firm foundation of the LBM in physical reality.

In the future, we anticipate to use the MDLG analysis tool to derive a generic LBM collision operator. At this point a critical question remains of which collision operator captures the best the underlying physical system. Answering this question will resolve a debate, that has entertained the LBM community for over three decades.

Another key element of the LBM is how to reintroduce fluctuations relevant when considering a thermodynamic non-ideal system. Parsa *et al.* has made significant progress in this field measuring two-particle correlation functions and observing large fluctuations in a non-ideal gas system [90]. More research is needed to identify the true form of the thermal fluctuations and reintroduce them to the noise-free LBM.

Furthermore, one should investigate how to reproduce microscopic force fields and boundary conditions at the mesoscopic level. Boundary conditions and external forces play an important role in many hydrodynamic problems and can benefit from more precise representation. This can be achieved using the MDLG analysis tool.

In conclusion, the MDLG method is a powerful tool that restores the lost connection between LBM and the microscale. It allows us to look at the LBM once again as a ensemble average of an LGCA. The MDLG tool has the potential to recover the LBM from an underlying physical system and thereby re-affirm its foundation in physical reality.

APPENDIX A

List of publications

A.1 Peer-reviewed journal publications

- **A. Pachalieva** and A.J. Wagner, "Over-relaxation as a physical effect", *Physical Review Letters* (submitted).
- **A. Pachalieva** and A. J. Wagner, "Molecular dynamics lattice gas equilibrium distribution function for Lennard-Jones particles", *Philosophical Transactions of the Royal Society A*, (2021) (published).
- **A. Pachalieva** and A.J. Wagner, "Non-Gaussian distribution of displacements for Lennard-Jones particles in equilibrium", *Physical Review E*, 102(5), 053310, (2020) (published).
- M.R. Parsa, **A. Pachalieva** and A.J. Wagner, "Validity of the Molecular-Dynamics-Lattice-Gas global equilibrium distribution function", *International Journal of Modern Physics C*, Vol. 30, No. 10, 1941007, (2019) (published).
- **A. Pachalieva**, C.A. Niedermeier and T. Indinger, "Massively parallel GPU implementation of the Lattice-Boltzmann Method for the simulation of coupled aero-thermodynamic systems", *International Journal of Automotive Engineering*, Vol.10, Issue.1, pp.125-132, (2019) (published).

A.2 Conference proceedings

- **A. Pachalieva**, C.A. Niedermeier and T. Indinger, "Numerical simulation of coupled aero-thermodynamic systems based on the Lattice-Boltzmann Method", *2018 JSAE Annual Congress Proceedings (Spring)*, No.78-18, (2018).

A.3 Invited talks

- **A. Pachalieva** and A.J. Wagner, Deriving LBM Collision Operator using the Coarse-Graining MDLG Approach, *North Dakota State University*, January 27, 2020, Fargo, ND, USA,
- **A. Pachalieva** and A.J. Wagner, Is Over-Relaxation a Physical Effect? - Deriving LBM Collision Operator Using the MDLG Approach, *City College of New York*, December 5, 2018, New York, NY, USA.
- **A. Pachalieva** and A.J. Wagner, Is Over-Relaxation a Physical Effect? - Deriving LBM Collision Operator Using the MDLG Approach, *Worcester Polytechnic Institute*, November 28, 2018, Worcester, MA, USA.
- **A. Pachalieva** and C.A. Niedermeier and T. Indinger, Numerical Simulations of Coupled Aero-Thermodynamic Systems Based on the Lattice-Boltzmann Method, *North Dakota State University*, August 27, 2018, Fargo, ND, USA.
- **A. Pachalieva** and C.A. Niedermeier and T. Indinger, DNS of Passive Scalar Transport in Turbulent Channel Flow using the Lattice Boltzmann Method on Massively Parallel GPU Architectures, *Tokyo Institute of Technology*, May 29, 2018, Tokyo, Japan.

A.4 Contributed talks and posters

- **A. Pachalieva**, N. Lubbers, M. Mehana, Y. Chen, Q. Kang and H. Viswanathan, Incorporating Nanoconfinement Effects into Lattice Boltzmann Method using Machine Learning and Molecular Dynamics, *Mechanistic Machine Learning and Digital Twins for Computational Science, Engineering & Technology (MMLDT-CSET) 2021*, September 26-29, 2021, San Diego, CA, USA.
- **A. Pachalieva** and A.J. Wagner, Deriving Lattice Boltzmann from Molecular Dynamics, *30th International Conference on Discrete Simulation of Fluid Dynamics (DSFD)*, September 13-17, 2021, Viterbo, Italy.
- **A. Pachalieva** and A.J. Wagner, Deriving LBM Collision Operator using the Coarse-Graining MDLG Approach, *APS March Meeting 2021*, March 15-19, 2021, Chicago, IL, USA.
- **A. Pachalieva** and A.J. Wagner, Molecular Dynamics Lattice Gas Equilibrium Distribution Function for Lennard-Jones Particles, *29th International Conference on Discrete Simulation of Fluid Dynamics (DSFD)*, July 13-17, 2020, Viterbo, Italy.

- **A. Pachalieva**, C.A. Niedermeier, T. Indinger and N. Adams, DNS of Passive Scalar Transport in Turbulent Channel Flow using the Lattice Boltzmann Method on Massively Parallel GPU Architectures, *15th International Conference for Mesoscopic Methods in Engineering and Science (ICMMES)*, July 9-13, 2018, Newark, DE, USA.

- **A. Pachalieva** and A.J. Wagner, Poster Award for "Molecular Dynamics Lattice Gas Analysis Tool", *27th International Conference on Discrete Simulation of Fluid Dynamics (DSFD)*, June 25-29, 2018, Worcester, MA, USA.

- **A. Pachalieva**, C.A. Niedermeier and T. Indinger, Numerical Simulation of Coupled Aero-Thermodynamic Systems based on the Lattice-Boltzmann Method, *Society of Automotive Engineers of Japan (JSAE) Annual Congress (Spring)*, May 23-25, 2018, Pacifico Yokohama, Japan.

APPENDIX B

Publication 1: Validity of the molecular dynamics lattice gas global equilibrium distribution function

The MDLG method establishes a direct link between a lattice-gas method and the coarse-graining of a molecular dynamics (MD) approach. Due to its connection to MD, the MDLG rigorously recovers the hydrodynamics and allows to validate the behavior of the lattice-gas or lattice-Boltzmann methods directly without using the standard kinetic theory approach. In this paper, we show that the analytical definition of the equilibrium distribution function remains valid even for very high volume fractions.

Preprint of an article published in

International Journal of Modern Physics C, Vol. 30, No. 10, 1941007 (2019) [85]
https://doi.org/10.1142/S0129183119410079

[1]Source: https://www.worldscientific.com/page/authors/author-rights (visited on July 7, 2022)

Alexander J. Wagner
Department of Physics
North Dakota State University
Fargo, 58108 ND, USA
alexander.wagner@ndsu.edu

Aleksandra A. Pachalieva
Chair of Aerodynamics and Fluid Mechanics
TUM School of Engineering and Design
Technical University of Munich
85748 Garching, Germany
aleksandra.pachalieva@tum.de

Letter of confirmation concerning publications included in Aleksandra A. Pachalieva's dissertation

To Whom It My Concern,

The authors of the publication "***M. R. Parsa, A. Pachalieva and A. J. Wagner, "Validity of the Molecular-Dynamics-Lattice-Gas Global Equilibrium Distribution Function", International Journal of Modern Physics C, Vol. 30, No. 10, 1941007 (2019)***" confirm that Aleksandra A. Pachalieva can use the publication as part of her publication-based dissertation.

Sincerely,

Alexander J. Wagner

International Journal of Modern Physics C
Vol. 30, No. 10 (2019) 1941007 (9 pages)

DOI: 10.1142/S0129183119410079

Validity of the molecular-dynamics-lattice-gas global equilibrium distribution function

M. Reza Parsa

Program in Materials and Nanotechnology
North Dakota State University, Fargo, ND 58108, USA

Department of Applied Mathematics
University of California, Merced, CA, 95343, USA
mparsa@ucmerced.edu

Aleksandra Pachalieva

Department of Mechanical Engineering
Technical University of Munich, 85748 Garching, Germany
aleksandra.pachalieva@tum.de

Alexander J. Wagner*

Department of Physics, North Dakota State University
Fargo, 58108 ND, USA
alexander.wagner@ndsu.edu

Received 18 November 2018
Accepted 26 February 2019
Published 22 August 2019

The molecular-dynamics-lattice-gas (MDLG) method establishes a direct link between a lattice-gas method and the coarse-graining of a molecular dynamics (MD) approach. Due to its connection to MD, the MDLG rigorously recovers the hydrodynamics and allows to validate the behavior of the lattice-gas or lattice-Boltzmann methods directly without using the standard kinetic theory approach. In this paper, we show that the analytical definition of the equilibrium distribution function remains valid even for very high volume fractions.

Keywords: Molecular dynamics; lattice gas; lattice Boltzmann method.

PACS Nos.: 11.25.Hf, 123.1K.

1. Introduction

The lattice gas automata were introduced by Frisch *et al.*[1] in 1986. These methods describe the presence of a particle using Boolean states and thus exhibit perfect collisions. However, the lattice gas methods suffer from statistical noise and their

*Corresponding author.

collision rules can have very complex mathematical representation.[2] Later on, Boghosian *et al.*[3] introduced the integer lattice gases, where one can control the level of fluctuations, while maintaining the exact conservation laws and having unconditional stability. In the pursuit of understanding better the fluctuating systems, Blommel *et al.*[4] constructed a new integer lattice gas with Monte Carlo collision operator.

The Molecular-Dynamics-Lattice-Gas (MDLG) method[5,6] establishes a direct link between a lattice gas method and the coarse-graining of a molecular dynamics (MD) simulation. After comparing the equilibrium properties of the MDLG method to the lattice Boltzmann equilibrium, Parsa *et al.*[5] found that for any dilute gases for coarse-graining lattice spacing Δx exists a coarse-grained time step Δt such that the MDLG equilibria resembles the lattice Boltzmann method. However, an open question remained about the range of validity for the predicted analytical solution of the equilibrium distribution function.

The novelty of the current publication is to investigate the behavior of the MDLG method for varying volume fractions of the underlying MD simulation. Such a test was not performed in the initial publication by Parsa *et al.*[5] This analysis shows that the analytical MDLG equilibrium function remains valid even for very high volume fractions.

The rest of the paper is summarized as follows: In Sec. 2, we elaborate upon the main components of the MDLG method. Our findings on how the equilibrium distribution function behaves with varying nominal volume fraction are presented in Sec. 3. Finally, some concluding remarks and future discussions are mentioned in Sec. 4.

2. Molecular-Dynamics-Lattice-Gas Method

The MDLG method was originally introduced by Parsa *et al.*[5] It is based on an underlying MD simulation, where we track the migration of the particles to imposed lattice positions with displacement v_i after a time step Δt, as shown in Fig. 1.

This gives the integer lattice gas occupation number

$$n_i(x,t) = \sum_j \Delta_x[x_j(t)]\Delta_{x-v_i}[x_j(t-\Delta t)], \tag{1}$$

where $\Delta_x[x_j(t)] = 1$, if particle x is in lattice cell x at time t, and $\Delta_x[x_j(t)] = 0$, otherwise.

The MDLG evolution equation then takes the form of a lattice gas

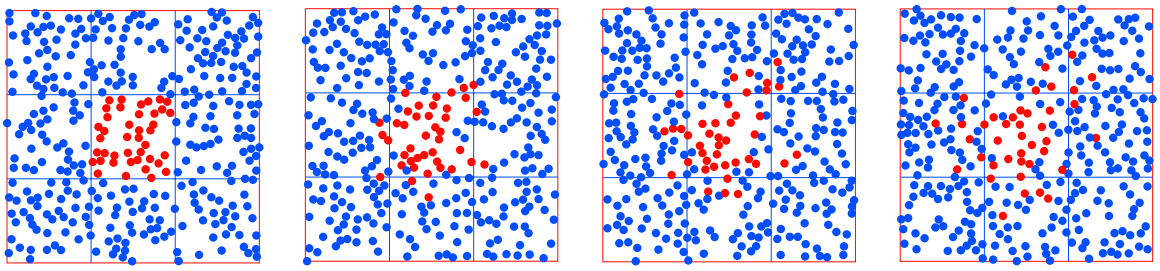
$$n_i(x + v_i, t + \Delta t) = n_i(x,t) + \Xi_i, \tag{2}$$

where the collision operator Ξ_i is defined as

$$\Xi_i = n_i(x + v_i, t + \Delta t) - n_i(x,t). \tag{3}$$

Since this analysis is based on a MD simulation, the resulting lattice gas model rigorously conserves the hydrodynamic properties of the system, up to the coarse-graining approximation. We define a Boltzmann average of the MDLG as

$$f_i = \langle n_i \rangle_{\text{neq}}, \tag{4}$$



(a) $a^2 = 0$ (b) $a^2 = 0.003$ (c) $a^2 = 0.010$ (d) $a^2 = 0.019$

Fig. 1. (Color online) Representation of the MDLG algorithm: a lattice (blue line) is overlaid onto the MD simulation. The position of each particle is tracked depending on an *a priori* chosen time (Δt) and space (Δx) discretization for the MDLG algorithm. The particles in the central lattice are colored in red to allow the reader to track their movement, however, each particle in the MDLG method is tracked and its occupation has been considered at each time step. The results are obtained from a MD simulation with volume fraction of $\phi = 0.0785$, $\kappa_B T = 50$ in Lennard–Jones (LJ) units, $\lambda = 0.5$, $\Delta x = 50$ and $\Delta t = 0.0001$. The shown time frames correspond to (a) $t = 0$, (b) $t = 0.3$, (c) $t = 0.6$ and (d) $t = 0.9$ in LJ units.

and by taking a nonequilibrium ensemble average of Eq. (2), we obtain the MDLB evolution equation:

$$f_i(x + v_i, t + \Delta t) = f_i(x, t) + \Omega_i \quad \text{with } \Omega_i = \langle \Xi_i \rangle_{\text{neq}}. \tag{5}$$

The global equilibrium distribution function can be numerically approximated by averaging the lattice gas densities n_i over the whole lattice and all iterations for an equilibrium system:

$$f_i^{\text{eq}} = \langle n_i \rangle_{\text{eq}} = \sum \langle \Delta_x[x_j(t)] \Delta_{x - v_i}[x_j(t - \Delta t)] \rangle. \tag{6}$$

Under the assumption that for an ideal gas the displacements δx of the particles are independent and that their probability distribution is a Gaussian with variance given by the mean-squared displacement $\langle (\delta x)^2 \rangle$, the equilibrium function can be analytically predicted to be

$$\frac{f_i^{\text{eq}}}{\rho^{\text{eq}}} = \prod_{\alpha=1}^{d} f_{i,\alpha}^{\text{eq}}, \tag{7}$$

where

$$\begin{aligned} f_{i,\alpha}^{\text{eq}} = N \Bigg(& e^{-\frac{(u_{i,\alpha}-1)^2}{2a^2}} - 2e^{-\frac{u_{i,\alpha}^2}{2a^2}} + e^{-\frac{(u_{i,\alpha}+1)^2}{2a^2}} \Bigg) \\ & + \frac{u_{i,\alpha} - 1}{2} \left[\text{erf}\left(\frac{u_{i,\alpha} - 1}{\sqrt{2}a} \right) - \text{erf}\left(\frac{u_{i,\alpha}}{\sqrt{2}a} \right) \right] \\ & + \frac{u_{i,\alpha} + 1}{2} \left[\text{erf}\left(\frac{u_{i,\alpha} + 1}{\sqrt{2}a} \right) - \text{erf}\left(\frac{u_{i,\alpha}}{\sqrt{2}a} \right) \right], \end{aligned} \tag{8}$$

and

$$a^2 = \frac{\langle(\delta x_\alpha)^2\rangle}{(\Delta x)^2}, \quad N = \frac{a}{\sqrt{2\pi}}, \quad u_{i,\alpha} = v_{i\alpha} - u_\alpha. \tag{9}$$

where Δx is the lattice size. This is the main result from the publication by Parsa *et al.*[5] and the derivation can be found there. The weights obtained by the MDLB method resemble to high extent the standard D2Q9 weights ($w_0 = 4/9, w_{1-4} = 1/9, w_{5-8} = 1/36$) for a specific time and space discretization with $a^2 = 1/6$. The weights of the MDLB method for $a^2 = 1/6$ obtained from Eq. (7) are

$$w_0 = 0.45721, \tag{10}$$

$$w_{1-4} = 0.10883, \tag{11}$$

$$w_{5-8} = 0.025907, \tag{12}$$

where sum over all the D2Q9 velocity weights is slightly lower than 1.0 (≈ 0.996158) because the MDLB does not impose restrictions on the number of velocities and thus, higher velocities also have small contributions to the total sum. We can make predictions for the form of higher-order lattice velocity sets like the one published in Refs. 7–9. However, this is out of the scope of the current publication. Due to the lack of a velocity set restriction, the MDLB models also do not have a lattice velocity restriction typical for the LBM methods. The equilibrium distribution function given in Eq. (7) does not resemble any of the already published formulations of the equilibrium distribution function because it is not explicitly restricted to a specific velocity set. As a comparison, please, refer to the MCLG model published by Blommel *et al.*,[4] where an *apriori* restriction of the number of velocities is made and the equilibrium distribution function obtained by the authors recovers the entropic formulation of the equilibrium distribution function given by Ansumali *et al.*[10]

For ideal gas systems, the mean-square displacement is given in relation to the velocity correlation function

$$\langle(\delta x)^2\rangle = 2\int_0^t dt'(t-t')\langle v_\alpha(t')v_\alpha(0)\rangle, \tag{13}$$

which is well approximated by an exponential decay given by

$$\langle v_\alpha(t')v_\alpha(0)\rangle = k_B T \exp\left(-\frac{t}{\lambda}\right), \tag{14}$$

with k_B being the Boltzmann constant, T the temperature in Lennard–Jones (LJ) units, and λ the exponential decay constant.[11–13] Now, we can express the mean-square displacement as a function of a single free parameter λ:

$$\langle(\delta x)^2\rangle = 2\kappa_B T\lambda^2\left(\exp\left(-\frac{t}{\lambda}\right) + \frac{t}{\lambda} - 1\right). \tag{15}$$

that is depicted in Fig. 3.

The assumption of uncorrelated displacements, which was used for the prediction of the equilibrium distribution, is likely only valid when the system can be approximated as an ideal gas. Our aim in this paper is, therefore, to find the validity range of this assumption by comparing the measured equilibrium distribution and the theoretical prediction for different nominal volume fractions.

3. MDLG Equilibrium Distribution Function for Different Volume Fractions

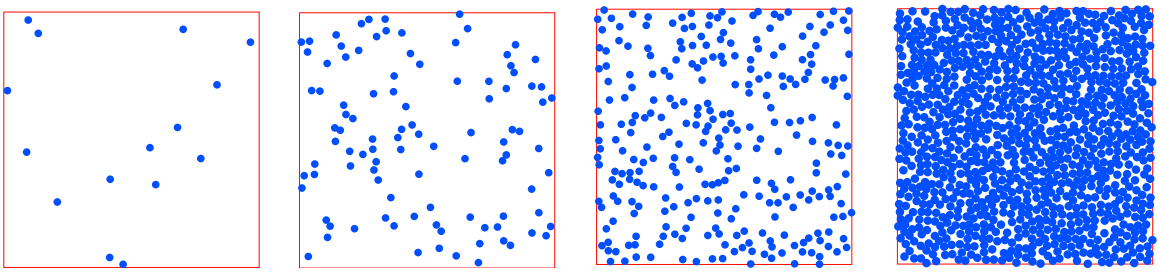
We have chosen four different setups with varying volume fractions for our underlying MD simulations with standard LJ interaction potential defined as

$$V_{\mathrm{LJ}} = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right], \tag{16}$$

where ε is the depth of the potential well, σ is the distance where the LJ potential is zero and r is the distance between particles. The MD simulations were performed using the open-source LAMMPS framework. We vary the volume fraction ϕ of the MD simulations from $\phi = 0.0078$ to $\phi = 0.8722$ (where the low value corresponds to the density employed in Ref. 5). The velocity has been fixed to $u = 0$, to allow for a deliberative analysis of the MDLG method in respect only to the volume fraction. For details, how the system behaves for different velocities, please, refer to Ref. 5. The volume fraction ϕ is defined for circular LJ particles with radius $r = \sigma$ with σ being the distance at which the inter particle potential goes to zero. Even though, the simulated volume fraction is above the maximum package density for hard spheres ($\phi = 0.7405$), we still observe diffusion due to the high temperature of the system (50 LJ units). When we increase the volume fraction even more, the system goes into a solid state and the dynamics slows down considerably. A visual representation of the variation of the used volume fractions can be seen in Figs. 2(a)–2(d).

All the simulations were initialized with homogeneously distributed particles. The total number of measured iterations for each simulation setup is 2 000 000 with a time step of $0.0001\,\tau$, which corresponds to a timescale of $\tau = 200$. As in the previous study,[5] a sufficient number of initial iterations, ($10^6, 10^6, 10^6, 10^7$, respectively), were discarded from the sampling process to ensure that the system has reached an equilibrium state before the probing.

After fitting the exponential constant λ of the velocity correlation function in Eq. (14) from the MD data, we can use the relation between the velocity correlation function and the mean-square displacement to obtain an analytical representation of the mean-square displacement as given in Eq. (13) for each of the nominal volume fractions (from $\phi = 0.0078$ to $\phi = 0.8722$). In Fig. 3, we compare the predicted mean-square displacement from Eq. (13) and the measured value from the underlying MD simulation data. There is an excellent agreement between the theoretical mean-square displacement and the measured one based on the velocity correlation exponential fit for volume fractions from $\phi = 0.0078$ to $\phi = 0.1962$. Those volume



(a) $\phi = 0.0078$ (b) $\phi = 0.0785$ (c) $\phi = 0.1962$ (d) $\phi = 0.8722$

Fig. 2. (Color online) Visual representation and comparison between the volume fractions used for the underlying MD simulations. The LJ particles are represented by circles with radius $r = \sigma$ with σ being the distance at which the inter particle potential is zero.

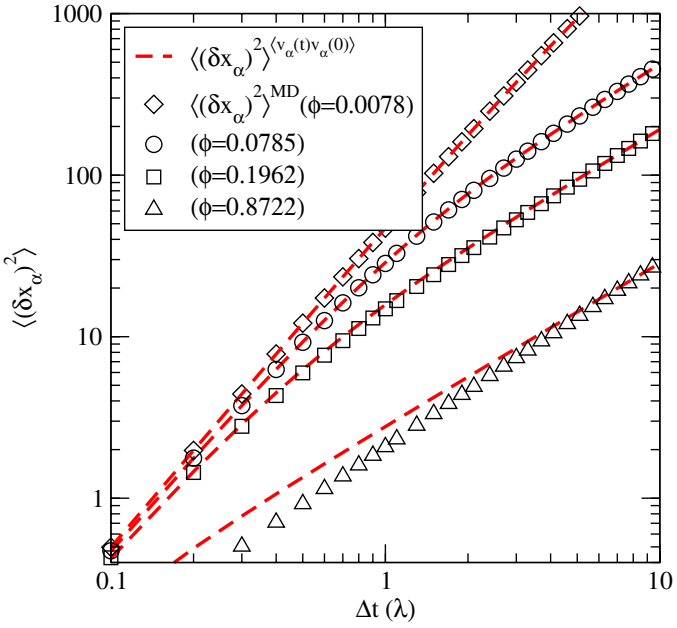
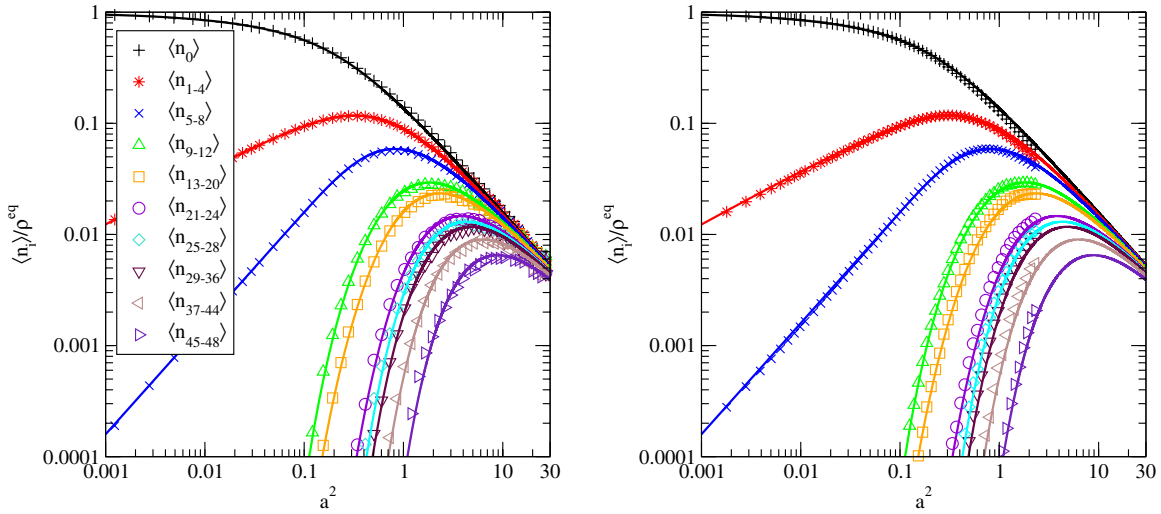


Fig. 3. (Color online) Measured mean-square displacement from the MD simulation data (symbols) for different volume fractions compared to the corresponding predicted values (dashed line).

fractions describe, in general, more gassy systems. For denser systems with $\phi = 0.8722$, the measured mean-square displacement deviate from the predicted one. As shown in Fig. 2(d), a system with $\phi = 0.8722$ is rather in a state close to a melted solid under high pressure.

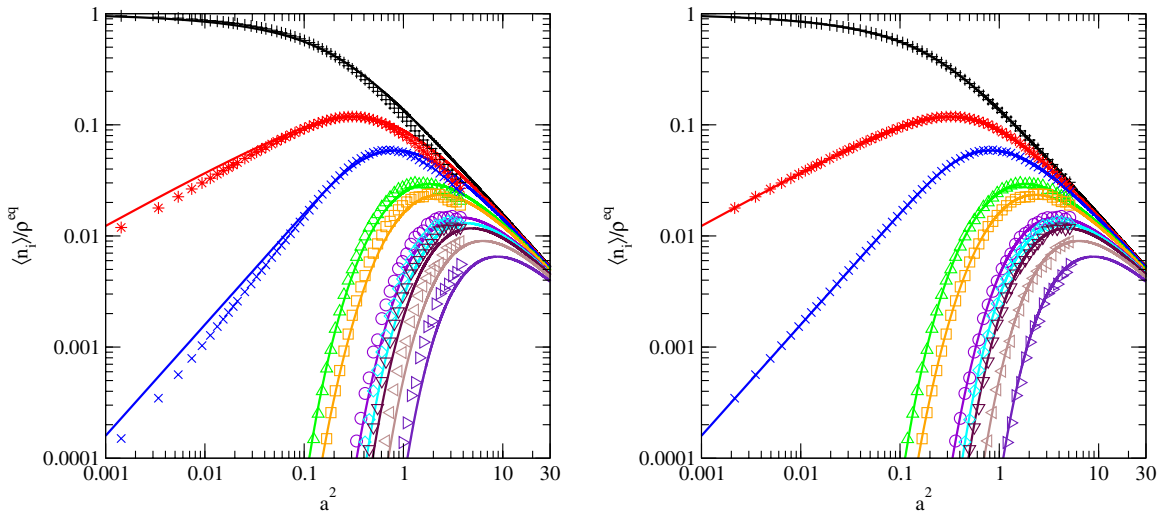
To validate the analytical solution of the MDLG global equilibrium distribution function f_i^{eq} in Eq. (7), we compare it to the measured f_i^{eq} from the underlying MD simulation as a function of a^2 from Eq. (9). In Figs. 4 and 5, we show that the theoretical prediction of the MDLG global equilibrium distribution function f_i^{eq} agrees well with the measured one from them MD simulations as long as the mean-square displacement can be approximated from the theory (see Eq. (13) as depicted in Fig. 3). However, once this relation breaks, the measured global equilibrium

distribution function also starts to deviate from the theoretically predicted values as shown in Fig. 5. For systems with volume fraction around and above $\phi = 0.8722$, the MDLG global equilibrium distribution function is not well approximated from the theoretical mean-square displacement given in Eq. (15).



(a) $\phi = 0.0078$ (b) $\phi = 0.1962$

Fig. 4. (Color online) Comparison between the measured equilibrium distribution function f_i^{eq} from an underlying MD simulation and the analytical solution given in Eq. (7) for different volume fractions from $\phi = 0.0078$ to $\phi = 0.1962$. The results for volume fraction $\phi = 0.0785$ has been already published in Parsa *et al.*,[5] where we observed the same behavior.



(a) Theoretical $\langle(\delta x)^2\rangle$, $\phi = 0.8722$ (b) Measured $\langle(\delta x)^2\rangle$, $\phi = 0.8722$

Fig. 5. (Color online) Global equilibrium distribution functions f_i^{eq} for volume fraction $\phi = 0.8722$: (a) The equilibrium distribution function is obtained from the fitted mean-square displacement and Eq. (15). There is a strong deviation between the measured and the theoretical data; (b) The equilibrium distribution function is calculated using the measured mean-square displacement from the MD simulation. The legend can be seen in Fig. 4 (omitted for clarity).

As a remedy, instead of calculating the theoretical value of $\langle(\delta x)^2\rangle$, we can measure the mean-square displacement as a function of the time step Δt directly and use this numerical value to obtain a^2. There is a noticeable difference between the MDLG global equilibrium distribution function obtained from the predicted and from the measured mean-square displacement as depicted in Figs. 5(a) and 5(b), respectively. Using the measured mean-square displacement, however, we recover the correct MDLG global equilibrium distribution function from theory even for high volume fractions.

We find, therefore, that the analytical prediction of the equilibrium distribution function is correct for a large range of volume fractions, even after the relation between the velocity correlation function and the mean-square displacement breaks and correlations become important. In this case, we need to use the actual mean-square displacement measured from the MD simulation instead of using the theoretical solution in Eq. (14) to obtain a^2.

4. Conclusions

In this work, we have investigated the behavior of the MDLG equilibrium function in respect to different volume fractions. We have shown that for a very large range of volume fractions, the equilibrium distribution function can be obtained from a single parameter λ based on the relation between the velocity correlation function and the mean-square displacement of the underlying MD simulation. However, for higher volume fractions, this correlation does not apply, and the exponential fit from the velocity correlation function does not reproduce the right behavior of the mean-square displacement anymore, which leads also to wrong estimation of the equilibrium distribution function in respect to the measured one from the MD simulation. We found that by using the measured mean-square displacement instead of the theoretical, we can still recover the correct analytical MDLG equilibrium distribution function. We conclude that the MDLG equilibrium distribution function appears to be universally defined as a function of the mean-square displacement and the lattice size of the underlying MD system.

An exact knowledge of the equilibrium distribution is important, if one wishes to examine the deviation of a system from local equilibrium. In a future publication, we intend to investigate whether the MDLG collision operator can be expressed in terms of a BGK collision operator $\Omega_i = \sum_i \Lambda_{ij}(f_j^{\mathrm{eq}} - f_j)$ with Λ_{ij} being the relaxation matrix. Here, the knowledge of an exact local equilibrium distribution is of the essence. Preliminary studies show that for nonequilibrium cases, we observe a small difference (approximately, 0.5%) between the measured and the analytical global equilibrium distribution function. The reason for this discrepancy might be due to the assumption made in Ref 5, that the probability distribution for the displacement is described as a Gaussian for all scales (from ballistic to diffusive regime). This assumption was made, because the probability distribution for short scales as well as the one for long scales can be described by a Gaussian distribution. However,

intermediate probabilities between the ballistic and the diffusive regime could have different behavior, as for example the well-known *cage effect* described by Chong *et al.*[14] Those intermediate regimes due to their nature, might be better approximated, for example, by a combination of two Gaussian distributions. These considerations are outside the scope of the current paper and will be a subject of future research.

Additionally, the knowledge of the equilibrium distributions is essential when analyzing the fluctuation behavior of MDLG systems, which will elucidate the correct form of fluctuations for nonideal systems. This is the subject of a following publication.[15]

Acknowledgments

AP acknowledges support from the German Federal Ministry of Education and Research (BMBF) in the scope of the project Aerotherm (reference numbers: 01IS16016A-B).

References

1. U. Frisch, B. Hasslacher and Y. Pomeau, *Phys. Rev. Lett.* **56**, 1505 (1986).
2. D. A. Wolf-Gladrow, *Lattice-Gas Cellular Automata and Lattice Boltzmann Models* (Springer, New York, 2000).
3. B. M. Boghosian, J. Yepez, F. J. Alexander and N. H. Margolus, *Phys. Rev. E* **55**, 4137 (1997).
4. T. Blommel and A. J. Wagner, *Phys. Rev. E* **97**, 023310 (2018).
5. M. R. Parsa and A. J. Wagner, *Phys. Rev. E* **96**, 013314 (2017).
6. M. R. Parsa, Lattice gases with molecular dynamics collision operator, PhD Dissertation, North Dakota State University (2018).
7. X. Shan, *Phys. Rev. E* **81**, 036702 (2010).
8. S. Chikatamarla and I. V. Karlin, *Phys. Rev. E* **79**, 046701 (2009).
9. M. Atif, M. Namburi and S. Ansumali, *Phys. Rev. E* **98**, 053311 (2018).
10. S. Ansumali, I. V. Karlin and H. C. Öttinger, *Europhys. Lett.* **63**(6), 798 (2003).
11. M. S. Green, *J. Chem. Phys.* **20**, 1281 (1952).
12. R. Kubo, *J. Phys. Soc. Jpn.* **12**, 570 (1957)
13. D. A. Weitz, D. J. Pine, P. N. Pusey and R. J. A. Tough, *Phys. Rev. Lett.* **63**, 1747 (1989).
14. S.-H. Chong and M. Fuchs, *Phys. Rev. Lett.* **88**, 185702 (2002).
15. M. R. Parsa and A. J. Wagner, Large fluctuations in non-ideal coarse-grained systems, arXiv:1905.07611 [physics.comp-ph] (2019).

APPENDIX C

Publication 2: Non-Gaussian distribution of displacements for Lennard-Jones particles in equilibrium

---

Most mesoscale simulation methods assume Gaussian distributions of velocity-like quantities. These quantities are not true velocities, 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.

Preprint of an article published in

Physical Review E, 102(5), 053310 (2020) [83]
https://doi.org/10.1103/PhysRevE.102.053310

https://journals.aps.org/pre/abstract/10.1103/PhysRevE.102.053310

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*

(Received 9 June 2020; accepted 19 October 2020; published 19 November 2020)

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.

DOI: 10.1103/PhysRevE.102.053310

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].

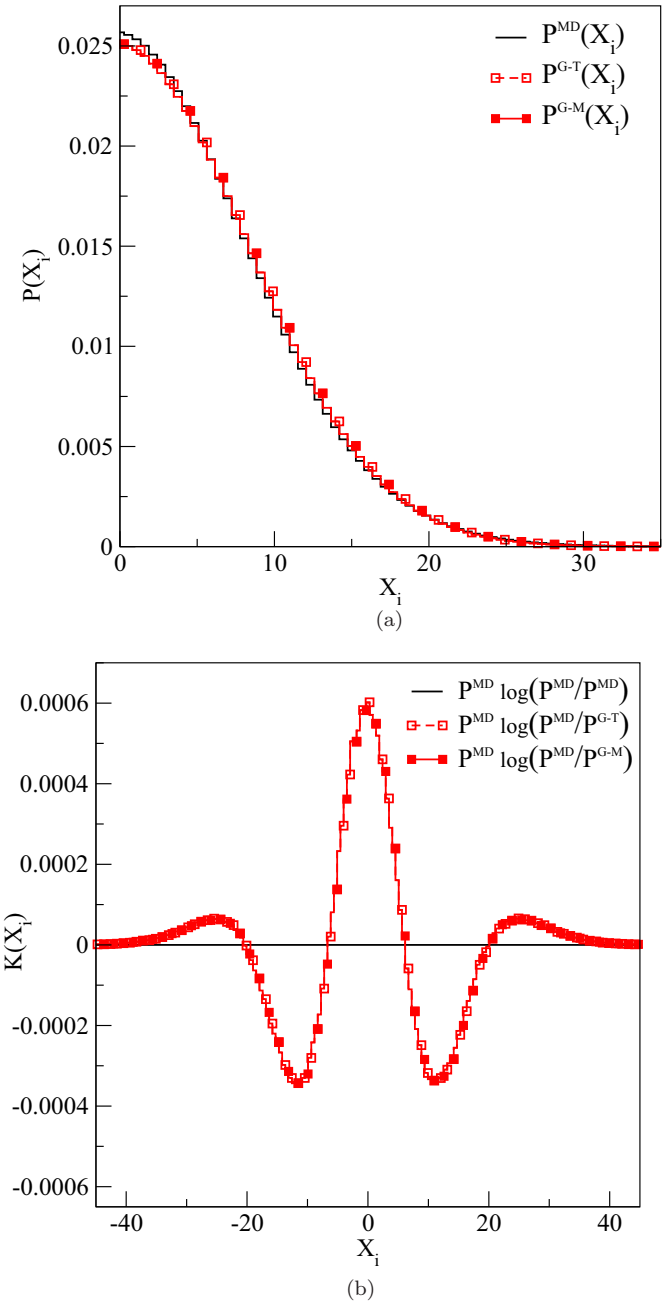


FIG. 1. (a) Displacements probability distribution functions. The solid line (black) depicts a PDF obtained from an MD simulation of LJ particles in equilibrium. The lines with empty or full squares (red) illustrate a Gaussian probability distribution function defined in Eq. (17) with mean-squared displacement 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

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

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

$$\tau_{\mathrm{LJ}} = \sqrt{\frac{m\sigma^2}{\varepsilon}}, \tag{3}$$

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 $1/2\ k_BT$ to transverse the diameter σ of a LJ particle, which is given by

$$\tau_{\mathrm{th}} = \sqrt{\frac{m\sigma^2}{k_BT}} \tag{4}$$

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,

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

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 $0.0001\ \tau_{\mathrm{LJ}}$ with total MD simulation time varying from $50\ \tau_{\mathrm{LJ}}$ to $51\,200\ \tau_{\mathrm{LJ}}$ 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 $0.01\ \tau_{\mathrm{LJ}}$ to $25.6\ \tau_{\mathrm{LJ}}$. To ensure the MD simulations have reached equilibrium state before we start collecting data, the initial 1 200 000 MD iterations ($120\ \tau_{\mathrm{LJ}}$) 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 ensamble 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 ensamble, 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 ensamble (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

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

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:

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

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

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

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:

$$\sum_i H(X_i) = 1. \tag{9}$$

We can then estimate the probability

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

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

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

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

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

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

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

Then the probability for collisionless displacements is

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

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

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

The probability distribution function of the displacements is given by

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

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

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

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

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

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

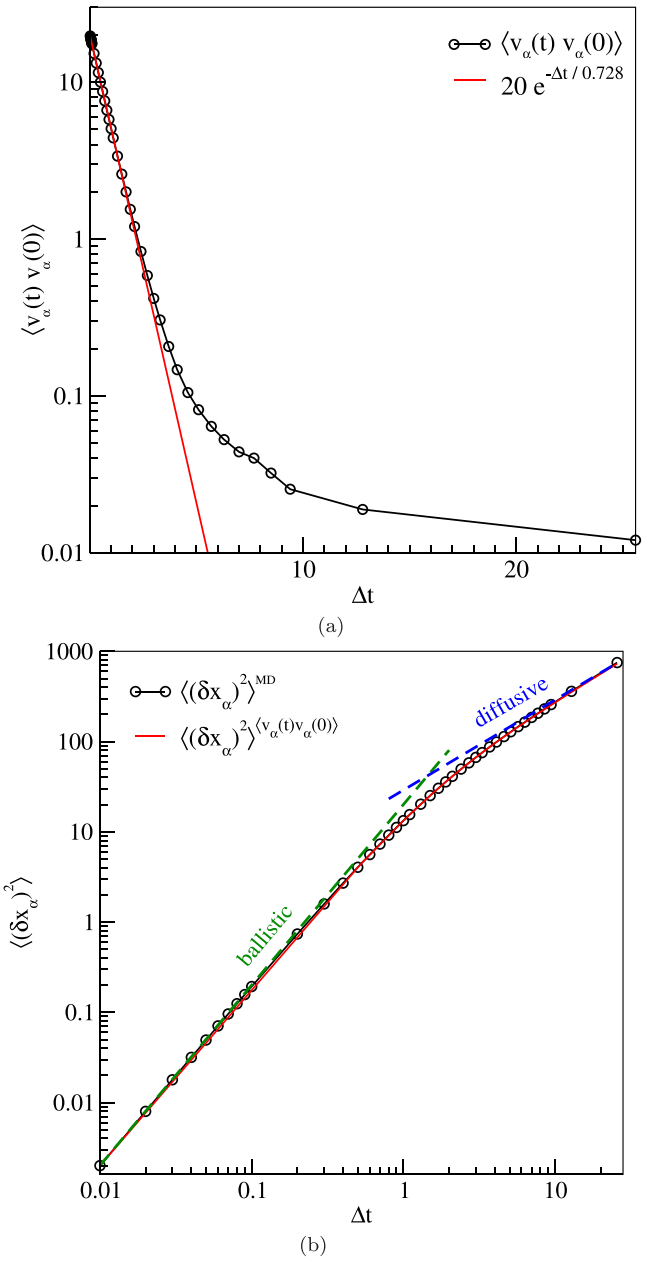


FIG. 2. (a) Velocity autocorrelation function measured from an MD simulation compared to an exponential decay with $\tau \approx 0.728$ as given in Eq. (20). The long-time tails are typical for two-dimensional systems [15–19]. (b) The mean-squared displacement directly measured from an MD simulation 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

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

where X_i corresponds to a rage of $r_i \leqslant \delta x < r_{i+1}$ with number of bins $i = 200$. 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 $P^{\text{G-}T}(X_i)$ with theoretical mean-squared displacement (red dashed line with empty squares) and a Gaussian distribution function $P^{\text{G-}M}(X_i)$ 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:

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

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

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

The KL divergence measures the discrepancies of one probability distribution function to another. It is always nonnegative $D_{\text{KL}}(R \parallel Q) \geqslant 0$ 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 $K(P^{\text{MD}} \parallel P^{\text{MD}})$ 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 the current section, we focus on the two KL divergence measures $D_{\text{KL}}(P^{\text{MD}} \parallel P^{\text{G-}T})$ and $D_{\text{KL}}(P^{\text{MD}} \parallel P^{\text{G-}M})$ 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 $K(P^{\text{MD}} \parallel P^{\text{G-}M})$ divergence is close to zero and we conclude that the simulation has reached diffusive regime. For some of the considered time steps, $P^{\text{G-}M}(X_i)$ delivers slightly better results in comparison to $P^{\text{G-}T}(X_i)$ 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 $K(P^{\text{MD}} \parallel P^{\text{G-}T})$ and $K(P^{\text{MD}} \parallel P^{\text{G-}M})$ 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,

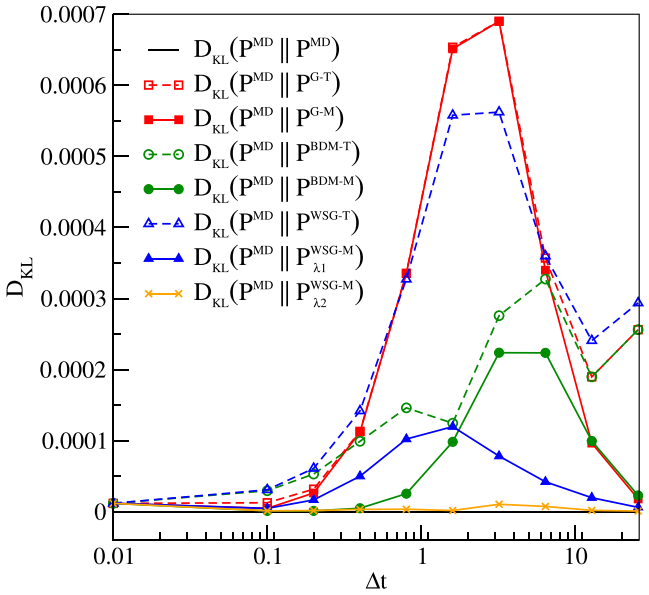


FIG. 3. Kullback-Leibler divergence results: Empty or full squares (red) for $D_{\text{KL}}(P^{\text{MD}} \parallel P^{\text{G-}T})$ and $D_{\text{KL}}(P^{\text{MD}} \parallel P^{\text{G-}M})$ discussed in Sec. IV; empty or full circles (green) for $D_{\text{KL}}(P^{\text{MD}} \parallel P^{\text{BDM-}T})$ and $D_{\text{KL}}(P^{\text{MD}} \parallel P^{\text{BDM-}M})$ discussed in Sec. V; empty or full triangles (blue) for $D_{\text{KL}}(P^{\text{MD}} \parallel P^{\text{WSG-}T})$ and $D_{\text{KL}}(P^{\text{MD}} \parallel P^{\text{WSG-}\lambda_1})$, and $\times$ symbols (yellow) for $D_{\text{KL}}(P^{\text{MD}} \parallel P^{\text{WSG-}\lambda_2})$ discussed in Sec. VI. The $D_{\text{KL}}(P^{\text{MD}} \parallel P^{\text{MD}})$ 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 $P^{\text{WSG-}\lambda_2}(X_i)$ 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.

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

$$\mu_k = \langle (\delta x)^k \rangle. \tag{25}$$

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

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

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 $P^{\mathrm{G}}(\delta x)$ 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 $P^{\mathrm{G}\text{-}T}(\delta x)$ fourth-order moment error is depicted in Fig. 4 as dashed line (red) with empty squares. The $\mu_4^{\mathrm{G}\text{-}T}$ error is much larger than the $\mu_2^{\mathrm{G}\text{-}T}$ error and increases very fast in the transition regime.

The second-order moment of $P^{\mathrm{G-M}}(X_i)$ matches the MD second-order moment by construction. The fourth-order moment $\mu_4^{\mathrm{G}\text{-}M}$, however, differs from the measured fourth-order moment as shown in Fig. 4 (red line with full squares). For $\Delta t \in [0.8, 1.6]$, $P^{\mathrm{G}\text{-}M}(\delta x)$ has a slightly larger fourth-order moment error than $P^{\mathrm{G}\text{-}T}(\delta x)$. Unlike $\mu_4^{\mathrm{G}\text{-}T}$, which does not decrease with larger time steps, the $\mu_4^{\mathrm{G}\text{-}M}$ error is large in the transition regime and decreases to less than 1% for later times. We assume that the larger $\mu_4^{\mathrm{G}\text{-}T}$ error is related to the larger second-order moment error of $P^{\mathrm{G}\text{-}T}(\delta x)$. Figure 4 shows that the $\mu_4^{\mathrm{G}\text{-}M}$ 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.

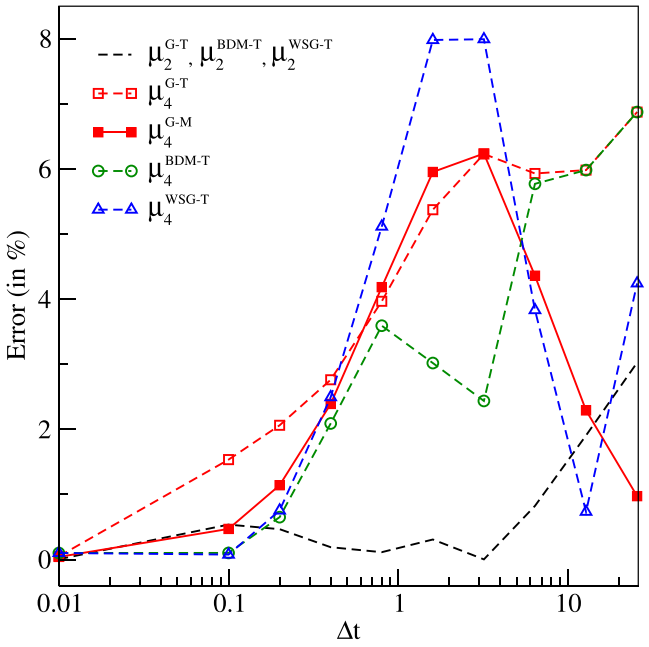


FIG. 4. Second- and fourth-order moment error calculated between the MD simulation data and the theoretical probability distribution functions. The second-order error is equivalent for all theoretical PDF models. The fourth-order error varies: The $P^{\mathrm{G}\text{-}T}(\delta x)$ and $P^{\mathrm{G}\text{-}M}(\delta x)$ errors are discussed in Sec. IV (red lines with empty or full squares); the $P^{\mathrm{BDM}\text{-}T}(\delta x)$ errors discussed in Sec. V (green lines with circles); and the $P^{\mathrm{WSG}\text{-}T}(\delta x)$ 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 $\Delta t = 3.2$.

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

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

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, $P^{\mathrm{BDM}}(\delta x)$ 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 $2k_BT(\Delta t)^2$ as given in Eq. (13). For the diffusive part $P^{\mathrm{diff}}(\delta x)$, 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

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

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)

$$\begin{aligned}\mu_2 &= \int_{-\infty}^{\infty} P^{\mathrm{BDM}}(\delta x)(\delta x)^2\, d\delta x \\ &= \int_{-\infty}^{\infty} \exp\left(-\frac{\Delta t}{\tau}\right) P^{\mathrm{ball}}(\delta x)(\delta x)^2\, d\delta x \\ &\quad + \int_{-\infty}^{\infty} \left[1 - \exp\left(-\frac{\Delta t}{\tau}\right)\right] P^{\mathrm{diff}}(\delta x)(\delta x)^2\, d\delta x \\ &= \exp\left(-\frac{\Delta t}{\tau}\right) k_BT(\Delta t)^2 \\ &\quad + \left[1 - \exp\left(-\frac{\Delta t}{\tau}\right)\right]\sigma^2_{\mathrm{diff}}, \end{aligned} \quad (29)$$

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

$$\sigma_{\mathrm{diff}} = \sqrt{\frac{\mu_2 - \exp\left(-\frac{\Delta t}{\tau}\right) k_BT(\Delta t)^2}{\left[1 - \exp\left(-\frac{\Delta t}{\tau}\right)\right]}}. \quad (30)$$

We examine the dependence of this diffusion constant on Δt in Fig. 6. Our original motivation would demand that $D = \sigma^2_{\mathrm{diff}}/2\Delta t$ 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 $P^{\mathrm{BDM\text{-}T}}(X_i)$, since the mean free time τ and the mean-squared displacement are estimated using the velocity autocorrelation function.

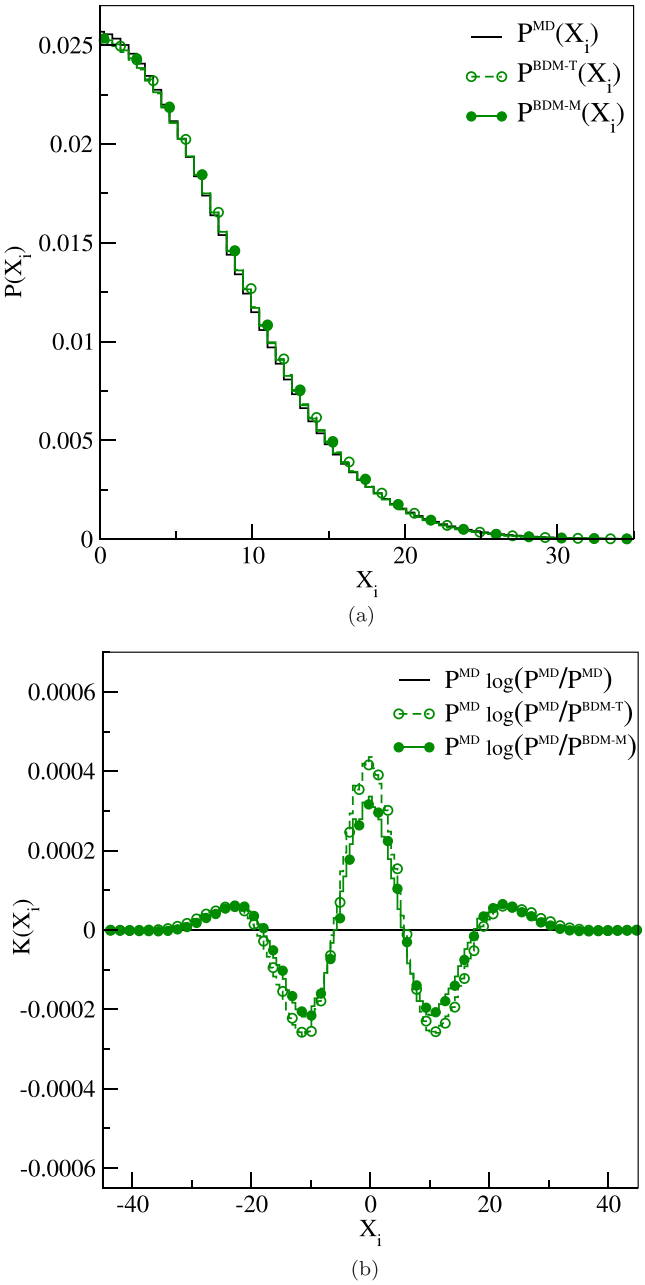


FIG. 5. (a) Displacements probability distribution functions. The solid line (black) depicts a PDF obtained from an MD simulation of LJ particles in equilibrium. The lines with empty or full circles illustrate the ballistic-diffusive distribution function defined in Eq. (27) with mean-squared displacement 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 $P^{\mathrm{BDM\text{-}T}}(X_i)$ distribution function, which resembles well the displacement distribution function obtained from the MD simulation. To assess the

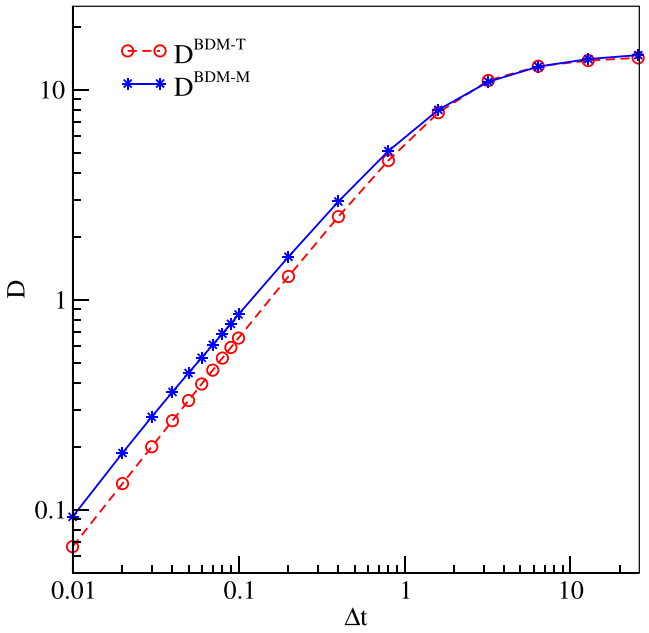


FIG. 6. Dependence of the self-diffusion constant to the time step Δt. The diffusion $D = \sigma^2_{\rm diff}/2\Delta t$ converges to a constant for $P^{\text{BDM-}T}$ and $P^{\text{BDM-}M}$, however, for early times it is not fixed. This demonstrates that the BDM models do not capture the hydrodynamics properties.

discrepancies between the theoretical BDM and the MD distribution function, we calculate $K(P^{\rm MD} \parallel P^{\text{BDM-}T})$ 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

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

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

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

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

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

Equations (30) and (33) define a system of linear equations with two unknowns. The system has a unique solution for $\sigma_{\rm diff}$ 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 $D = \sigma^2_{\rm diff}/2\Delta t$ 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 $K(P^{\rm MD} \parallel P^{\text{BDM-}T})$ 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 $D = \sigma^2_{\text{diff}}/2\Delta t$ 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

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

where λ is effectively the average number of collisions given by

$$\lambda = \frac{\Delta t}{\tau}, \tag{35}$$

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

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

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

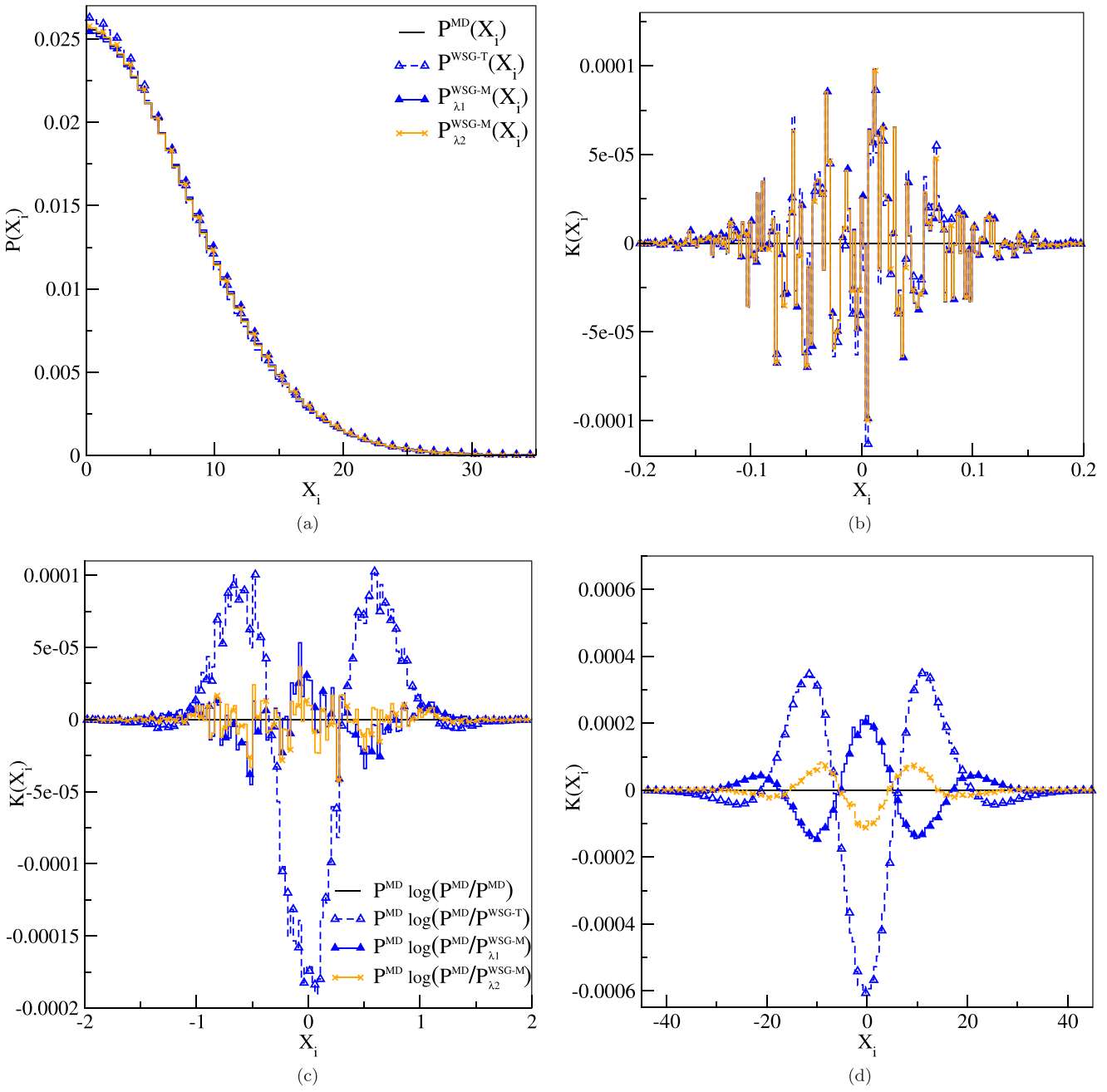


FIG. 7. (a) Displacements probability distribution functions. The solid line (black) depicts a PDF obtained from an MD simulation of LJ particles in equilibrium. The dashed line (blue) with empty triangles illustrates the PDF of the theoretical WSG defined in Eq. (36) with λ obtained using the theoretical 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

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

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

$$\begin{aligned}\mu_4 &= \int_{-\infty}^{\infty} P^{\mathrm{WSG}}(\delta x)(\delta x)^4\, d(\delta x)\\ &= \int_{-\infty}^{\infty}\sum_{c=0}^{\infty} e^{-\lambda}\frac{\lambda^c}{c!}\frac{\sqrt{\lambda+1}}{\sqrt{2\pi(c+1)\langle(\delta x)^2\rangle}}\\ &\quad\times\exp\left[-\frac{(\lambda+1)(\delta x-u\Delta t)^2}{2(c+1)\langle(\delta x)^2\rangle}\right](\delta x)^4\, d\delta x\\ &= \frac{3\langle(\delta x)^2\rangle^2}{(\lambda+1)^2}[\lambda^2+3\lambda+1]. \end{aligned} \tag{38}$$

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:

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

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 $\Delta t\to 0$ and $\Delta t\to\infty$ 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 $P^{\text{WSG-}M}_{\lambda_1}(X_i)$ and $P^{\text{WSG-}M}_{\lambda_2}(X_i)$ is illustrated in Figs. 7(b)–7(d). In each figure, $K(P^{\mathrm{MD}}\parallel P^{\text{WSG-}M}_{\lambda_1})$ and $K(P^{\mathrm{MD}}\parallel P^{\text{WSG-}M}_{\lambda_2})$ are depicted for different time step. Figure 7(b) shows the error of $K(P^{\mathrm{MD}}\parallel P^{\text{WSG-}M}_{\lambda_2})$ 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 $K(P^{\mathrm{MD}}\parallel P^{\text{WSG-}M}_{\lambda_2})$ 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.

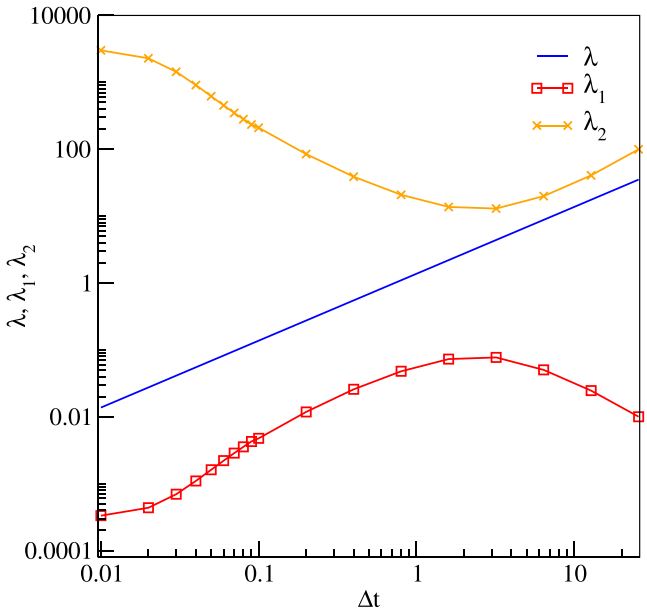


FIG. 8. Average number of collisions depending on the coarse-grained time step Δt. λ denotes the number of collisions obtained from the velocity autocorrelation theory given in Eq. (35), which is used for the calculation of the $P^{\text{WSG-}T}(X_i)$ distribution function. λ_1 and λ_2 are solutions of the quadratic equation given in Eq. (39). These values are used for the calculation of $P^{\text{WSG-}M}_{\lambda_1}(X_i)$ and $P^{\text{WSG-}M}_{\lambda_2}(X_i)$ 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 $P^{\text{WSG-}M}_{\lambda_1}(X_i)$ and $P^{\text{WSG-}M}_{\lambda_2}(X_i)$ 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 $D_{\mathrm{KL}}(P^{\mathrm{MD}}\parallel P^{\text{WSG-}T})$ is almost as large as $D_{\mathrm{KL}}(P^{\mathrm{MD}}\parallel P^{\mathrm{G}})$ 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.

ACKNOWLEDGMENTS

A.W. and A.P. thank M. Reza Parsa and Michael E. Cates for the helpful discussions. A.P. was partially supported by the Center for Nonlinear Studies and the Laboratory Directed Research and Development program at Los Alamos National Laboratory, and the German Federal Ministry of Education and Research in the scope of the project Aerotherm (Grant No. 01IS16016A-B).

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

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

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

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

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

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

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

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

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

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

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

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

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

Analogously, one derives the fourth-order moment as

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

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

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

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

$$\lambda_{1,2}=\frac{-9\mu_2^2\pm\sqrt{3\left[15\mu_2^4-4\mu_2^2\mu_4\right]}+2\mu_4}{2\left[3\mu_2^2-\mu_4\right]}. \qquad \text{(B7)}$$

[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).

APPENDIX D

Publication 3: Molecular dynamics lattice gas equilibrium distribution function for Lennard-Jones particles

The molecular dynamics lattice gas method maps a molecular dynamics simulation onto a lattice gas using a coarse-graining procedure. This is a novel fundamental approach to derive the lattice Boltzmann method by taking a Boltzmann average over the molecular dynamics lattice gas. A key property of the lattice Boltzmann method is the equilibrium distribution function, which was originally derived by assuming that the particle displacements in the molecular dynamics simulation are Boltzmann distributed. However, we recently discovered that a single Gaussian distribution function is not sufficient to describe the particle displacements in a broad transition regime between free particles and particles undergoing many collisions in one time step. In a recent publication, we proposed a Poisson weighted sum of Gaussians which shows better agreement with the molecular dynamics data. We derive a lattice Boltzmann equilibrium distribution function from the Poisson weighted sum of Gaussians model and compare it to a measured equilibrium distribution function from molecular dynamics data and to an analytical approximation of the equilibrium distribution function from a single Gaussian probability distribution function.

Preprint of an article published in
Philosophical Transactions of the Royal Society A **379**: 20200404. 20200404 [84]
https://doi.org/10.1098/rsta.2020.0404

[3]Source: https://royalsociety.org/journals/permissions/ (visited on July 7, 2022)

PHILOSOPHICAL TRANSACTIONS A

royalsocietypublishing.org/journal/rsta

Research

Cite this article: Pachalieva A, Wagner AJ. 2021 Molecular dynamics lattice gas equilibrium distribution function for Lennard–Jones particles. *Phil. Trans. R. Soc. A* **379**: 20200404.
https://doi.org/10.1098/rsta.2020.0404

Accepted: 19 March 2021

One contribution of 15 to a theme issue 'Progress in mesoscale methods for fluid dynamics simulation'.

Subject Areas:
computational physics, fluid mechanics

Keywords:
molecular dynamics, lattice gas method, lattice Boltzmann method, coarse-graining

Author for correspondence:
Alexander J. Wagner
e-mail: alexander.wagner@ndsu.edu

Molecular dynamics lattice gas equilibrium distribution function for Lennard–Jones particles

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

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

AP, 0000-0003-1246-0410; AJW, 0000-0001-7248-8234

The molecular dynamics lattice gas (MDLG) method maps a molecular dynamics (MD) simulation onto a lattice gas using a coarse-graining procedure. This is a novel fundamental approach to derive the lattice Boltzmann method (LBM) by taking a Boltzmann average over the MDLG. A key property of the LBM is the equilibrium distribution function, which was originally derived by assuming that the particle displacements in the MD simulation are Boltzmann distributed. However, we recently discovered that a single Gaussian distribution function is not sufficient to describe the particle displacements in a broad transition regime between free particles and particles undergoing many collisions in one time step. In a recent publication, we proposed a Poisson weighted sum of Gaussians which shows better agreement with the MD data. We derive a lattice Boltzmann equilibrium distribution function from the Poisson weighted sum of Gaussians model and compare it to a measured equilibrium distribution function from MD data and to an analytical approximation of the equilibrium distribution function from a single Gaussian probability distribution function.

This article is part of the theme issue 'Progress in mesoscale methods for fluid dynamics simulation'.

THE ROYAL SOCIETY PUBLISHING

1. Introduction

The molecular dynamics lattice gas (MDLG) method [1,2] uses a coarse-graining procedure to establish a direct link between microscopic methods—in particular, molecular dynamics (MD) simulation, and mesoscale methods such as lattice gas (LG) [3,4] and lattice Boltzmann methods (LBM) [5,6]. The MDLG fully relies on MD data and as such it rigorously recovers the hydrodynamics of the underlying physical system, and can be used to verify the behaviour and examine the properties of the LG or the LBM methods directly without using the standard kinetic theory approach. Aspects that can be examined include fluctuating [7–10], thermal [11,12], multi-phase and multi component systems [5,13–15].

A key feature in the LBM is the equilibrium distribution function. The LBM equilibrium distribution was originally derived by analogy to the continuous Boltzmann equation, where the equilibrium distribution for the velocities is a Maxwell Boltzmann distribution. Similarly, the LBM moments of the discrete velocity distribution were matched, to the degree possible, with the velocity moments of the Maxwell Boltzmann distribution. In the alternative derivation of the LBM from MD, it was shown that these previously postulated equilibrium distributions are indeed, at least approximately, consistent with the MDLG approach for specific discretization combinations for lattice and time spacing.

In the original MDLG calculation of the equilibrium distribution by Parsa *et al.* [1], it was assumed that the particle displacements in the MD simulation are also Boltzmann distributed. This assumption gave an adequate prediction of the global equilibrium distribution function of the LBM. However, later on by examining more carefully the equilibrium system, we noticed small deviations (up to 5%) between the analytically predicted and the measured equilibrium distribution functions. These deviations were traced back to the prediction of the one-particle displacement distribution function. In Pachalieva *et al.* [16], we proposed a correction of the displacement distribution function, which shows that a dilute gas with area fraction of $\phi = 0.0784$ and temperature of 20 LJ is better approximated by a Poisson weighted sum of Gaussians (WSG) probability distribution function. This probability distribution function takes into account that after a time step Δt the particles can be divided into groups depending on the number of collisions they have experienced. In principle, the timing of the collisions should be random (given by a Poisson process), however, the resulting integrals over the collision times do not allow for an analytical solution. Thus, we assume that the particle collisions are evenly spaced, which may introduce a small error but it makes the resulting displacements again Gaussian distributed. For details, please refer to [16]. The Poisson weighted sum of Gaussians probability distribution function also delivers better results for a purely ballistic and purely diffusive regimes (for very small or very large time steps, respectively), where the Poisson WSG formulation is reduced to a single Gaussian. In the current publication, we show that the original premise of the paper [16] does indeed hold. We derive the MDLG equilibrium distribution function from the Poisson WSG one-particle displacement function and show that it compares favourably to a measured equilibrium distribution function from MD simulation, whereas the single Gaussian equilibrium distribution function is a much poorer prediction. Our findings show that the Poisson WSG approximates the measured equilibrium distribution function significantly better.

The rest of the paper is summarized as follows: we briefly describe the MDLG analysis method in §2. In §3, we derive the equilibrium distribution function from one-particle displacement function. In §3a, we show how to derive the equilibrium distribution function when the distribution is given by a single Gaussian and in §3b when the displacements are instead distributed according to a Poisson WSG one-particle displacement function. In §4, we give a detailed description of the MD simulation set-up used to obtain the MD data. The MD trajectories are later used to validate the theoretical solutions of the equilibrium distribution function. In §5, we compare the equilibrium distribution function obtained on one hand from theory, using either a single Gaussian or the novel Poisson weighted sum of Gaussians probability distribution function, and on the other hand, measured from MD data. Our analysis shows significant improvement of the equilibrium distribution function analytical prediction when

the Poisson WSG model is used. Finally, in §6, we give a brief conclusion and suggestions for future work.

2. Molecular dynamics lattice gas method

In the MDLG analysis, we impose a lattice onto an MD simulation of Lennard–Jones (LJ) particles and track the migration of the particles from one lattice position to another with displacement v_i after a time step Δt as shown in figure 1*a*. A schematic of the lattice is given in figure 1*b* where the numbers 0 to 49 represent the i index of the occupation number of a D2Q49 velocity set. We run MD simulations and analyse the particles' trajectories to obtain MDLG occupation numbers defined as

$$n_i(x,t) = \sum_j \Delta_x[x_j(t)]\Delta_{x-v_i}[x_j(t-\Delta t)], \tag{2.1}$$

with the delta function $\Delta_x[x_j(t)] = 1$, if particle x is in the lattice cell at time t, and $\Delta_x[x_j(t)] = 0$, otherwise. Here, the $x_j(t)$ is the position of the j-th particle at time t and v_i is the particle displacement, which in the MDLG description is strongly correlated to the lattice velocities. We can now cast the evolution of the occupation numbers n_i in the form of an LG evolution equation as

$$n_i(x+v_i, t+\Delta t) = n_i(x,t) + \Xi_i, \tag{2.2}$$

by defining the LG collision operator Ξ_i in terms of the occupation numbers as

$$\Xi_i = n_i(x+v_i, t+\Delta t) - n_i(x,t). \tag{2.3}$$

The molecular dynamics lattice Boltzmann (MDLB) distribution function is defined as a Boltzmann ensemble average of the MDLG occupation numbers n_i and it is given by

$$f_i = \langle n_i \rangle_{\text{neq}}. \tag{2.4}$$

By taking the non-equilibrium ensemble average of equation (2.2), we obtain the MDLB evolution equation

$$f_i(x+v_i, t+\Delta t) = f_i(x,t) + \Omega_i, \quad \text{with } \Omega_i = \langle \Xi_i \rangle_{\text{neq}}, \tag{2.5}$$

where Ω_i is the MDLB collision operator. A key element of the LBM is the global equilibrium distribution function, which in the MDLB context is defined as an average of the LG densities n_i over the whole MD domain and all iterations of an equilibrium MD simulation. The MDLB equilibrium distribution function is given by

$$\begin{aligned} f_i^{\text{eq}} &= \langle n_i \rangle_{\text{eq}} \\ &= \left\langle \sum_j \Delta_x[x_j(t)]\Delta_{x-v_i}[x_j(t-\Delta t)] \right\rangle_{\text{eq}} \\ &= M \int dx_1 \int d\delta x_1 \, P^{(1),\text{eq}}(x_1, \delta x_1)\Delta_x[x_1]\Delta_{x-v_i}[x_1 - \delta x_1], \end{aligned} \tag{2.6}$$

where M is the total number of particles and $P^{(1),\text{eq}}$ is the one-particle displacement distribution function in equilibrium. This allows us to obtain the equilibrium distribution function f_i^{eq} analytically from the one-particle displacements probability distribution function (PDF).

3. Derivation of the molecular dynamics lattice Boltzmann equilibrium distribution function

In the MDLB formulation, the equilibrium distribution function depends solely on the one-particle displacement distribution function. Thus, knowing $P^{(1),\text{eq}}$ is crucial for predicting the equilibrium distribution function. In the following subsections, we derive the equilibrium

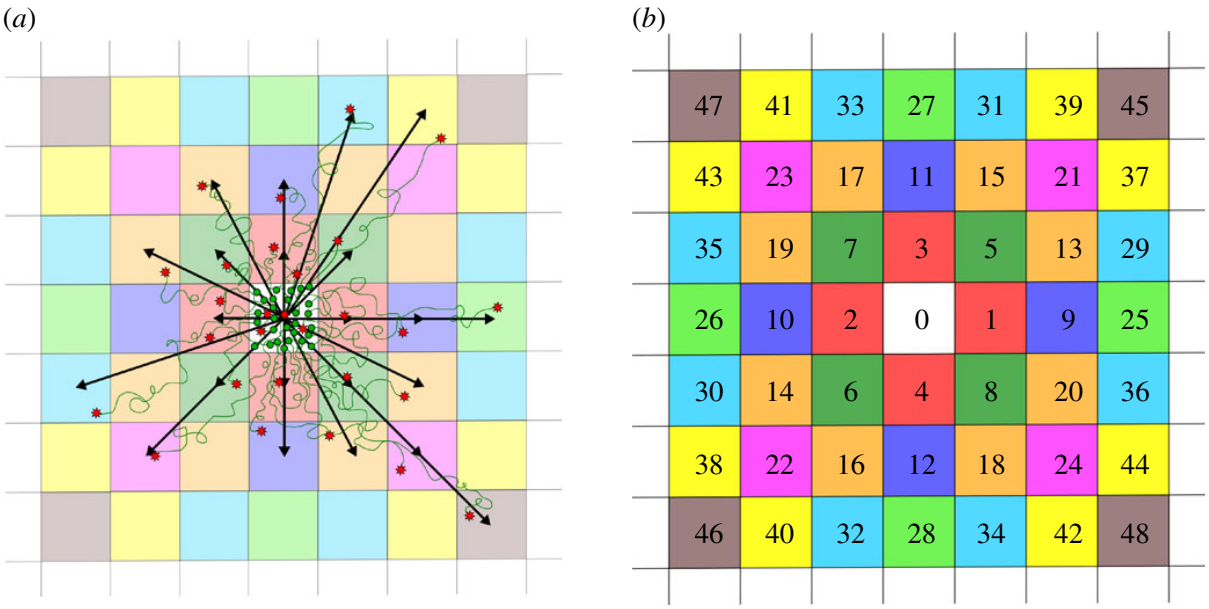


Figure 1. (*a*) Sketch of the MDLG analysis. A lattice is superimposed onto the MD simulation domain. The movement of the particles is tracked from the central node using their MD trajectories. The circles (green) represent the position of the particles at time $t - \Delta t$ and the stars (red) are their respective positions at time t. Using the particle trajectories and the imposed lattice, the occupation number n_i is defined as given in equation (2.1). The black arrows are the lattice velocities. Only the lattice velocities which have at least one particle within their area (i.e. non-zero occupation number) are shown. (*b*) Schematic of the D2Q49 lattice with the numbering convention for the lattice velocities in two dimensions. The central point 0 corresponds to the zeroth-velocity $v_0 = (0, 0)$ and the rest of the velocities are given as a vector connecting the central point and the lattice point in question as shown in (*a*). The velocities are colour-coded depending on their length. (Online version in colour.)

distribution function from (a) a single Gaussian probability distribution function and (b) from a Poisson weighted sum of Gaussians probability distribution function.

(a) Single Gaussian distribution model

In Parsa *et al.* [1], a good approximation of the MDLB equilibrium distribution function is given by a single Gussian in one dimension ($d = 1$)

$$P_\alpha^{\mathrm{G}}(\delta x) = \frac{1}{[2\pi\langle(\delta x_\alpha)^2\rangle]^{d/2}} \exp\left[-\frac{(\delta x_\alpha - u_\alpha \Delta t)^2}{2\langle(\delta x_\alpha)^2\rangle}\right], \tag{3.1}$$

with displacements δx_α, second-order moment $\langle(\delta x_\alpha)^2\rangle$ and mean velocity u_α. The solution factorizes for higher dimensions and it is given by

$$P^{\mathrm{G}}(\delta x) = \prod_{\alpha=1}^{d} P_\alpha^{\mathrm{G}}(\delta x). \tag{3.2}$$

Following equation (2.6), the equilibrium distribution function can be expressed as

$$\frac{f_i^{\mathrm{eq,G}}}{\rho^{\mathrm{eq}}} = \prod_{\alpha=1}^{d} f_{i,\alpha}^{\mathrm{eq,G}}, \tag{3.3}$$

with ρ^{eq} being the mass density. The equilibrium distribution function $f_{i,\alpha}^{\mathrm{eq,G}}$ in one dimension is given by

$$\begin{aligned} f_{i,\alpha}^{\mathrm{eq,G}} = {} & N\left(\mathrm{e}^{-((u_{i,\alpha}-1)^2/2a^2)} - 2\,\mathrm{e}^{-(u_{i,\alpha}^2/2a^2)} + \mathrm{e}^{-((u_{i,\alpha}+1)^2/2a^2)}\right) + \frac{u_{i,\alpha}-1}{2} \\ & \times\left[\mathrm{erf}\left(\frac{u_{i,\alpha}-1}{a\sqrt{2}}\right) - \mathrm{erf}\left(\frac{u_{i,\alpha}}{a\sqrt{2}}\right)\right] + \frac{u_{i,\alpha}+1}{2}\left[\mathrm{erf}\left(\frac{u_{i,\alpha}+1}{a\sqrt{2}}\right) - \mathrm{erf}\left(\frac{u_{i,\alpha}}{a\sqrt{2}}\right)\right], \end{aligned} \tag{3.4}$$

with

$$a^2 = \frac{\langle(\delta x_\alpha)^2\rangle}{(\Delta x)^2} \quad \text{and} \quad N = \frac{a}{\sqrt{2\pi}}, \qquad u_{i,\alpha} = v_{i,\alpha} - u_\alpha, \tag{3.5}$$

where $\langle(\delta x_\alpha)^2\rangle$ is the mean-squared displacement, Δx is the lattice size, and u_α is the mean velocity. We have performed MD simulations with mean velocity set to zero, however, we could obtain results for different mean velocities u_α by applying a Galilean transformation. We have set the value of a^2 to approximately 1/6 for which the MDLG results agree with the values of the D2Q9 lattice Boltzmann weights. For details regarding the derivation of the Gaussian equilibrium distribution function, please refer to [1].

Even though this formulation shows very good agreement with the measured equilibrium distribution function from MD simulations, under more careful investigation we found that there are discrepancies of up to about 5% for certain parameter regimes. This means that the displacement distribution function cannot be fully captured by a single Gaussian and a more complex distribution function has to be applied.

(b) Poisson weighted sum of Gaussians model

In Pachalieva *et al.* [16], we have introduced a correction of the displacements PDF proposed by Parsa *et al.* [1] using a Poisson weighted sum of Gaussians (WSG) instead of a single Gaussian distribution function. The Poisson WSG is given by

$$P^{\mathrm{WSG}}(\delta x) = \sum_{c=0}^{\infty} \mathrm{e}^{-\lambda} \frac{\lambda^c}{c!} P^c(\delta x), \tag{3.6}$$

where the $P^c(\delta x)$ probability distribution function also factorizes for higher dimensions equivalently to the single Gaussian distribution function as given in equation (3.2). The one-dimensional Poisson weighted sum of Gaussians probability distribution function $P^c_\alpha(\delta x)$ is then given by

$$P^c_\alpha(\delta x) = \left[\frac{(\lambda+1)}{2\pi(c+1)\langle(\delta x_\alpha)^2\rangle}\right]^{d/2} \exp\left[-\frac{(\lambda+1)(\delta x_\alpha - u_\alpha \Delta t)^2}{2(c+1)\langle(\delta x_\alpha)^2\rangle}\right], \tag{3.7}$$

where δx_α is the displacement in one dimension, $\langle(\delta x_\alpha)^2\rangle$ is the second-order moment, u_α is the mean velocity, c is the number of occurrences and λ is the average number of collisions. The fact that the new displacement distribution function is just a sum of Gaussians makes the calculation of the new MDLG equilibrium functions surprisingly simple. Thus, we obtain

$$f_i^{\mathrm{eq}} = \sum_{c=0}^{\infty} \mathrm{e}^{-\lambda} \frac{\lambda^c}{c!} f_i^{c,\mathrm{eq}}. \tag{3.8}$$

The $f_i^{c,\mathrm{eq}}$, similar to equation (3.3), is given by

$$\frac{f_i^{c,\mathrm{eq}}}{\rho^{\mathrm{eq}}} = \prod_{\alpha=1}^{d} f_{i,\alpha}^{c,\mathrm{eq}}, \tag{3.9}$$

where ρ^{eq} is the mass density and $f_{i,\alpha}^{c,\mathrm{eq}}$ in one dimension is given by

$$f_{i,\alpha}^{c,\mathrm{eq}} = \left\{\frac{N_c}{2\sqrt{\pi}}\left(\mathrm{e}^{-((u_{i,\alpha}-1)^2/N_c^2)} - 2\,\mathrm{e}^{-(u_{i,\alpha}^2/N_c^2)} + \mathrm{e}^{-((u_{i,\alpha}+1)^2/N_c^2)}\right)\right.$$
$$\left. + \frac{(u_{i,\alpha}-1)}{2}\left[\mathrm{erf}\left(\frac{(u_{i,\alpha}-1)}{N_c}\right) - \mathrm{erf}\left(\frac{u_{i,\alpha}}{N_c}\right)\right] + \frac{(u_{i,\alpha}+1)}{2}\left[\mathrm{erf}\left(\frac{(u_{i,\alpha}+1)}{N_c}\right) - \mathrm{erf}\left(\frac{u_{i,\alpha}}{N_c}\right)\right]\right\}, \tag{3.10}$$

with

$$N_c = \sqrt{\frac{2a^2(c+1)}{\lambda+1}} \tag{3.11}$$

where a^2 and $u_{i,\alpha}$ are defined in equation (3.5). The one-dimensional equilibrium distribution function given in equation (3.10) is similar to the single Gaussian equilibrium distribution function in equation (3.4), however, their weighting factors are not the same. The equilibrium distribution function in equation (3.10) also takes into account the average number of collisions λ, which needs to be defined.

One way to approximate the average number of collisions λ is by using the velocity auto-correlation function. However, the auto-correlation function is just a theoretical approximation and is not exact. To eliminate the second-order and the fourth-order moment errors, we match these moments to the corresponding ones measured directly from the MD simulations. The second-order moment of the Poisson WSG one-particle distribution function can be derived from the second-order Gaussian integral

$$\begin{aligned}\mu_2 &= \int_{-\infty}^{\infty} P^{\mathrm{WSG}}(\delta x)(\delta x)^2 \,\mathrm{d}\delta x \\ &= \int_{-\infty}^{\infty} \sum_{c=0}^{\infty} \mathrm{e}^{-\lambda} \frac{\lambda^c}{c!} \frac{\sqrt{\lambda+1}}{\sqrt{2\pi(c+1)\langle(\delta x)^2\rangle}} \exp\left(-\frac{(\lambda+1)(\delta x - u\Delta t)^2}{2(c+1)\langle(\delta x)^2\rangle}\right)(\delta x)^2 \,\mathrm{d}\delta x \\ &= \langle(\delta x)^2\rangle. \end{aligned} \tag{3.12}$$

Analogously, we obtain the fourth-order moment from the fourth-order Gaussian integral

$$\begin{aligned}\mu_4 &= \int_{-\infty}^{\infty} P^{\mathrm{WSG}}(\delta x)(\delta x)^4 \,\mathrm{d}(\delta x) \\ &= \int_{-\infty}^{\infty} \sum_{c=0}^{\infty} \mathrm{e}^{-\lambda} \frac{\lambda^c}{c!} \frac{\sqrt{\lambda+1}}{\sqrt{2\pi(c+1)\langle(\delta x)^2\rangle}} \exp\left(-\frac{(\lambda+1)(\delta x - u\Delta t)^2}{2(c+1)\langle(\delta x)^2\rangle}\right)(\delta x)^4 \,\mathrm{d}\delta x \\ &= \frac{3\langle(\delta x)^2\rangle^2}{(\lambda+1)^2}\left[\lambda^2 + 3\lambda + 1\right]. \end{aligned} \tag{3.13}$$

By solving the quadratic equation for λ

$$\frac{3\mu_2^2}{(\lambda+1)^2}\left[\lambda^2 + 3\lambda + 1\right] - \mu_4 = 0 \tag{3.14}$$

we find the following solutions:

$$\lambda_{1,2} = \frac{-9\mu_2^2 \pm \sqrt{3[15\mu_2^4 - 4\mu_2^2\mu_4]} + 2\mu_4}{2[3\mu_2^2 - \mu_4]}. \tag{3.15}$$

where $\mu_2 = \langle(\delta x)^2\rangle$ and μ_4 are the second- and fourth-order displacement moments, respectively. We use the moments measured from MD simulations, which ensures that the Poisson weighted sum of Gaussians model has the same μ_2 and μ_4 moments. In Pachalieva *et al.* [16], we show that λ_2 provides an optimal solution, which we use to derive the Poisson WSG equilibrium distribution function. For detailed derivation and discussion of the Poisson WSG displacement distribution function, please refer to Pachalieva *et al.* [16].

Meaningfully comparing two probability distribution functions is a non-trivial task since often there are significant deviations in the tails of the distribution that would show up in a simpler measure like dividing the distributions. However, since the tails carry little weight, these deviations are not relevant for the system. In Pachalieva *et al.* [16], we used the Kullback–Leibler (KL) divergence, a tool commonly used in machine learning. The element-wise definition of this function is given by

$$K(X_i) = K(R \parallel Q) = R(X_i)\log\left(\frac{R(X_i)}{Q(X_i)}\right), \tag{3.16}$$

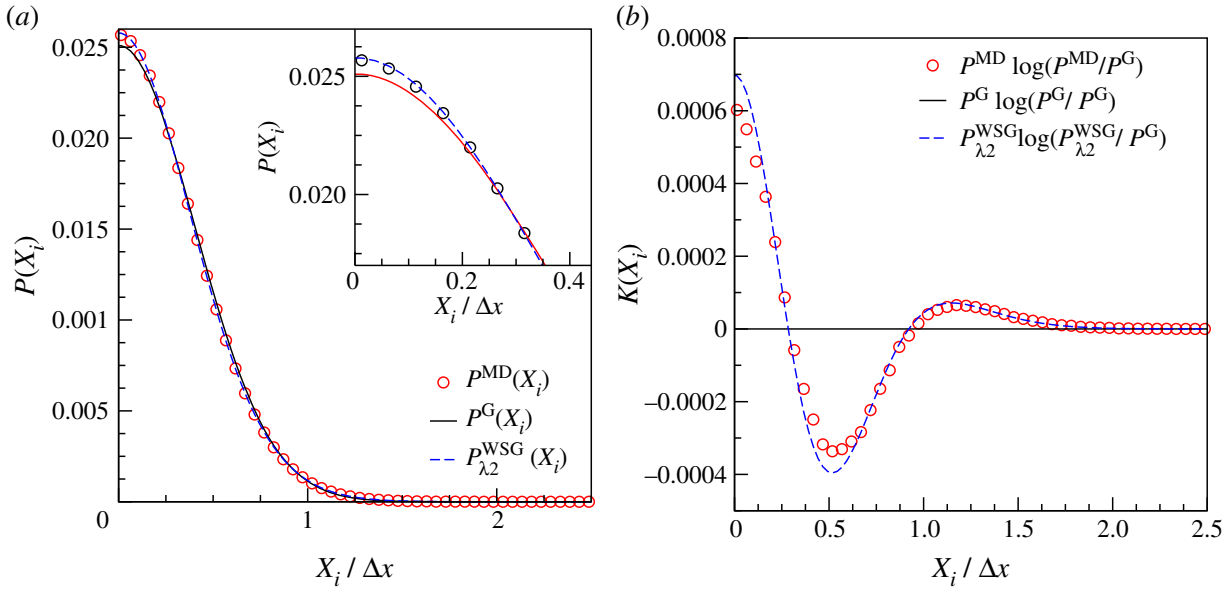


Figure 2. (*a*) Displacements probability distribution functions. The symbols (red) depict a PDF obtained from an MD simulation of LJ particles in equilibrium. The line (black) illustrates a Gaussian probability distribution function defined in equation (3.1) with mean-squared displacement fitted directly to the MD data. The dashed line (blue) represents the Poisson WSG obtained from equation (3.6). Only the data for positive velocities has been depicted due to symmetry. (*b*) The difference between the distributions per interval X_i as defined in equation (3.16). The presented data are for the standard parameters used in the paper and a coarse-grained time step $\Delta t = 3.2$. (Online version in colour.)

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 KL divergence [17] defined as

$$D_{\mathrm{KL}}(R \parallel Q) = \sum_i R(X_i) \log\left(\frac{R(X_i)}{Q(X_i)}\right). \tag{3.17}$$

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

In figure 2*a*, we see the true probability distribution function obtained from the MD data $P^{\mathrm{MD}}(X_i)$, the Gaussian probability distribution function $P^{\mathrm{G}}(X_i)$, and the Poisson WSG distribution function $P^{\mathrm{WSG}}(X_i)$. There is a visible divergence between the Gaussian and the other two distribution functions. We measured the element-wise KL divergence $K(X_i)$, as defined in equation (3.16), for $P^{\mathrm{G}}(X_i)$ compared to the MD data and the Poisson WSG distribution function as shown in figure 2*b*. The results suggest that even though the Gaussian and the Poisson WSG probability distribution functions have the same second moment, their deviations in the fourth- and higher-order moments influence strongly the form of the distribution function. In §5, we show how these deviations effect the LBM equilibrium distribution function.

4. Simulations set-up

All measured data, from probability distribution functions of the displacements $P^{\mathrm{MD}}(X_i)$ to the equilibrium distribution function $f_i^{\mathrm{eq,MD}}$ depicted in figures 3–5, are obtained from MD simulations. To perform the MD simulations we used the open-source MD framework LAMMPS [18,19] developed by Sandia National Laboratories. The LAMMPS package uses the Velocity-Verlet integration scheme. The MD simulations consist of particles interacting with the standard 6-12 LJ intermolecular potential given by

$$V_{\mathrm{LJ}} = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right], \tag{4.1}$$

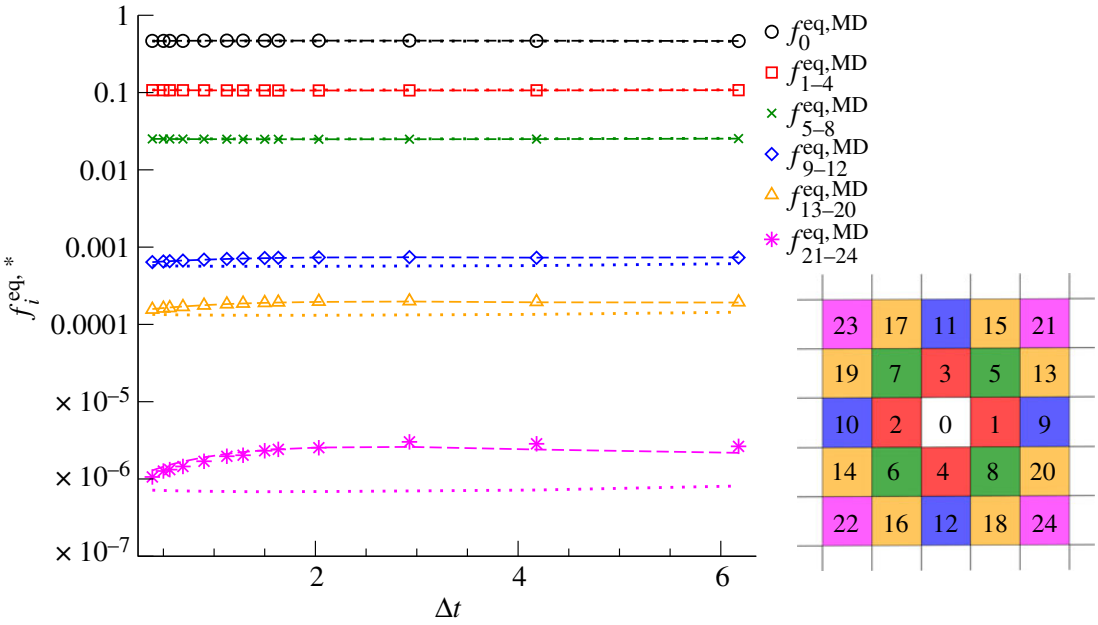


Figure 3. (*a*) Estimated equilibrium distribution functions $f_i^{\mathrm{eq},*}$ obtained either from MD simulation data ($f_i^{\mathrm{eq,MD}}$) depicted with symbols, theoretical solution using a single Gaussian distribution function ($f_i^{\mathrm{eq,G}}$) depicted with dotted lines or theoretical solution using Poisson WSG ($f_i^{\mathrm{eq,WSG}}$) depicted with dashed lines. (*b*) Our numbering for the velocities in a D2Q25 lattice. The equilibrium distribution function $f_i^{\mathrm{eq},*}$ values are colour coded and each colour represents one of the six sets of equilibrium distribution function contributions. Here, the asterisk (*) corresponds to the variety of methods used to obtain an equilibrium distribution function: measured from MD simulation, single Gaussian analytical solution and Poisson WSG analytical solution. Note that by using a simple-minded direct comparison on a log-scale (rather than a Kulbeck–Leibler measure) practically irrelevant errors for very small occupation numbers stand out here. (Online version in colour.)

with σ being the distance at which the inter-particle potential goes to zero, r is the distance between two particles, and ε is the potential well depth. The particle mass and the LJ particle diameter are set to $m = 1$ and $\sigma = 1$, respectively. The LJ timescale is given by the time needed for a particle with kinetic energy of half the potential energy well ε to traverse one diameter σ of an LJ particle. This can be also expressed as

$$\tau_{\mathrm{LJ}} = \sqrt{\frac{m\sigma^2}{\varepsilon}}. \tag{4.2}$$

The thermal time scale corresponds to 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

$$\tau_{\mathrm{th}} = \sqrt{\frac{m\sigma^2}{k_B T}}. \tag{4.3}$$

We executed MD simulations with temperature of 20 in the LJ units defined above. This corresponds to a thermal time scale smaller than the LJ time scale τ_{LJ} by factor of $1/\sqrt{20} \approx 0.22$.

The number of particles in each simulation has been fixed to $N = 99\,856$ which fills a two-dimensional square with length $L = 1000\sigma$. The area fraction ϕ of the domain is calculated from the area of the circular LJ particles multiplied by the number of particles divided by the area of the domain, where the diameter of the circular LJ particle is given by σ. The MD simulations considered in this publication have an area fraction of $\phi = 0.078387$. We initialized the simulations using homogeneously distributed particles with kinetic energy corresponding to temperature equal to 20 in LJ units. This corresponds to a dilute gas with high temperature. The temperature is way above the critical temperature for liquid-gas coexistence of $T_c = 1.3120(7)$, and the density is way below the critical density $\rho_c = 0.316(1)$ [20]. We focus our attention to MD simulations of a fairly dilute gas in equilibrium, since the assumption that the collision times is Poisson distributed is correct only for dilute systems.

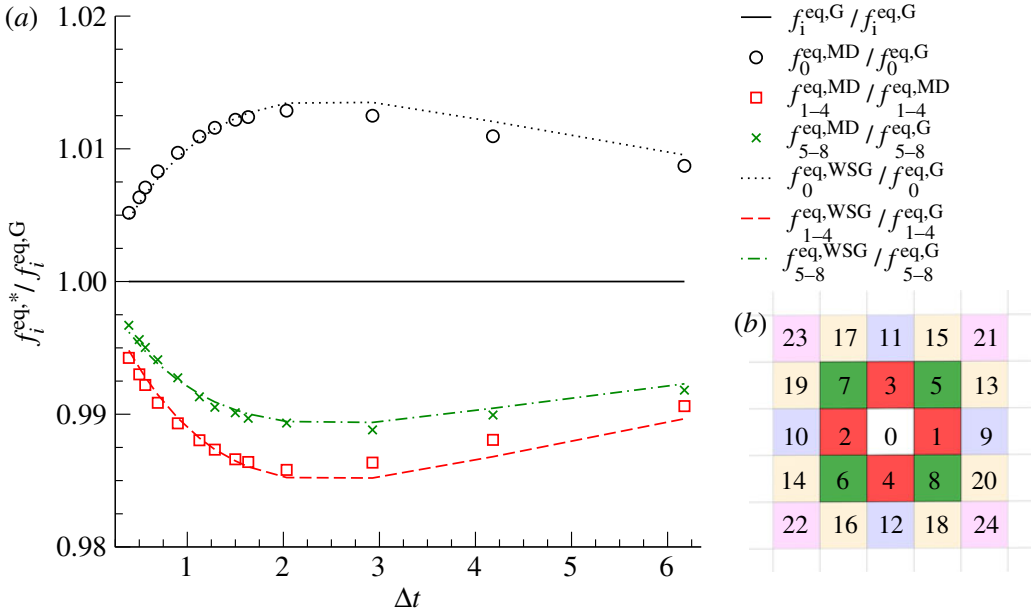


Figure 4. (*a*) First layer equilibrium distribution functions $f_{0-8}^{eq,*}$ scaled to the Gaussian equilibrium distribution function. The equilibrium distribution functions are obtained either from MD simulation data ($f_{0-8}^{eq,MD}$), theoretical solution using a single Gaussian distribution function ($f_{0-8}^{eq,G}$) or theoretical solution using Poisson WSG ($f_{0-8}^{eq,WSG}$). (*b*) Schematic representation of the D2Q25 lattice. The equilibrium distribution function $f_i^{eq,*}$ values are colour coded and each colour represents one of the six sets of equilibrium distribution function contributions. Here, the asterisk (*) corresponds to the variety of methods used to obtain an equilibrium distribution function: measured from MD simulation, single Gaussian analytical solution and Poisson WSG analytical solution. (Online version in colour.)

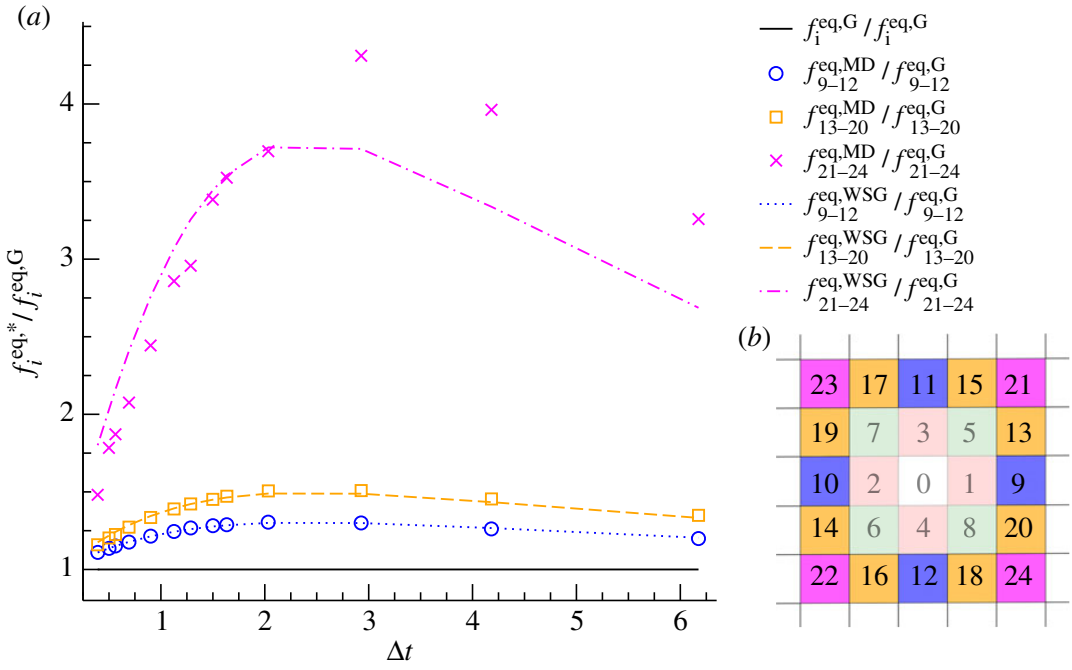


Figure 5. (*a*) Second layer equilibrium distribution functions $f_{9-24}^{eq,*}$ scaled to the Gaussian equilibrium distribution function. The equilibrium distribution functions are obtained either from MD simulation data ($f_{9-24}^{eq,MD}$), theoretical solution using a single Gaussian distribution function ($f_{9-24}^{eq,G}$) or theoretical solution using Poisson WSG ($f_{9-24}^{eq,WSG}$). (*b*) Schematic representation of the D2Q25 lattice. The equilibrium distribution function $f_i^{eq,*}$ values are colour coded and each colour represents one of the six sets of equilibrium distribution function contributions. Here, the asterisk (*) corresponds to the variety of methods used to obtain an equilibrium distribution function: measured from MD simulation, single Gaussian analytical solution and Poisson WSG analytical solution. (Online version in colour.)

Table 1. Initialization parameters of the molecular dynamics simulations performed using LAMMPS framework. For all MD simulations, the MD step size is fixed to $0.0001\tau_{\mathrm{LJ}}$ and the number of coarse-grained iterations is 2000.

Δt	Δx	lx	MD output frequency $(1/\tau_{\mathrm{LJ}})$	total MD time (τ_{LJ})
0.3911	4	250	3911	782.2
0.5000	5	200	5000	1000.0
0.5626	5.5	180	5626	1125.2
0.6927	6.6(6)	150	6927	1385.4
0.9009	8.3(3)	120	9009	1801.8
1.1261	10	100	11 261	2252.2
1.4994	12.5	80	14 994	2998.8
1.6342	13.3(3)	75	16 342	3268.4
2.0338	15.625	64	20 338	4067.6
2.9280	20	50	29 280	5856.0
4.1821	25	40	41 821	8364.2
6.1751	31.25	32	61 751	12 350.2

Since the MD simulations correspond to a dilute high temperature gas, the particle velocities will also be larger than for a typical MD simulation. Thus, we set the MD step size to $0.0001\ \tau_{\mathrm{LJ}}$ which is considerably smaller to ensure high accuracy of the MD data. We define a dimensionless coarse-grained time step Δt being the product of the MD step size and the MD output frequency shown in table 1. The time step Δt is chosen such that the MD simulations are restricted to the ratio of the mean-squared displacement and the squared lattice size being set to

$$a^2 = \frac{\langle(\delta x)^2\rangle}{(\Delta x)^2} \approx 0.1611, \tag{4.4}$$

this corresponds to the parameter a^2 given in equation (3.5), which has also been used in earlier publications [1,21]. By fixing the value, we ensure that most of the LJ particles in equilibrium will travel up to one lattice space which corresponds to a D2Q9 LBM. To verify that the Poisson WSG equilibrium distribution function $f_i^{\mathrm{eq,WSG}}$ approximated the MD data better than the single Gaussian equilibrium distribution function $f_i^{\mathrm{eq,G}}$ across the length scale, from ballistic to diffusive regime, we vary the coarse-grained time step $\Delta t \in [0.3911, 6.1751]$ and the lattice size $\Delta x \in [4, 31.25]$ of the executed simulations. An overview of the MD simulation set-up is given in table 1. The number of lattice points lx varies from 250 to 32 depending on the coarse-grained time step Δt. For each coarse-grained time step Δt we performed 2 000 iterations which corresponds to total MD time of 782.2 τ_{LJ} to 12 350.2 τ_{LJ} for the smallest and largest coarse-grained time step Δt, respectively. In order to bring the MD simulations to equilibrium state before we start collecting data, the initial 3 000 000 iterations of each simulation were discarded. The discarded iterations are not included in table 1 for clarity.

The MD simulation set-up characterizes a hot dilute gas in equilibrium with average velocity u_α fixed to zero

$$Nu_\alpha = \sum_{j=1}^{N} v_{j,\alpha} = 0, \tag{4.5}$$

where N is the number of LJ particles.

We performed standard MD simulations without thermostat. In the LAMMPS framework, this is called NVE integration. The microcanonical ensamble NVE is characterized by constant number of particles (N), constant volume (V) and constant energy (E).

5. Results

In order to obtain a measured equilibrium distribution function, we post-process the collected MD data using the MDLG analysis tool. The MD domain is overlapped with a lattice and we trace the migration of the particles over time from one lattice to another. By doing this, we obtain the MDLG occupation numbers $n_i(x,t)$ as defined in equation (2.1) which after sufficient averaging deliver the MDLB equilibrium distribution function $f_i^{\mathrm{eq,MD}}$ as defined in equation (2.6).

The analytical models of the equilibrium distribution function defined in §3 depend only on the choice of the one-particle displacement distribution function. Since we define two different one-particle distributions, we expect to see also changes in the respective equilibrium distribution function derived from them, even though their second-order moments are equivalent. However, a non-trivial question remains how the migration of particles from one node to another changes within a lattice.

To gain a better understanding, we calculate the equilibrium distribution function for an extended D2Q25 lattice which corresponds to two neighbouring cells in X- and Y-directions for a two-dimensional domain. A schematic of the D2Q25 lattice is given in figure 3*b*. In equilibrium state with zero initial velocity, one distinguishes six sets of equilibrium distribution function contributions: $f_0^{\mathrm{eq},*}$, $f_{1-4}^{\mathrm{eq},*}$, $f_{5-8}^{\mathrm{eq},*}$, $f_{9-12}^{\mathrm{eq},*}$, $f_{13-20}^{\mathrm{eq},*}$ and $f_{21-24}^{\mathrm{eq},*}$, where each set has a unique displacement length from the central lattice. When measuring the equilibrium distribution function $f_i^{\mathrm{eq,MD}}$ from the MD simulations, we average over the number of lattices for each set to obtain a symmetric probability distribution function. It is worth mentioning that the deviations of the $f_i^{\mathrm{eq,MD}}$ values within each set are relatively small.

The MDLG analysis was introduced for a D2Q49 lattice including a third layer of neighbouring cells, however, the number of considered neighbouring layers depends solely on the problem at hand. For a simulation in equilibrium with zero velocity, and a parameter a^2 as defined in equation (4.4) being set to approximately 0.1611, we obtain an equilibrium distribution function which is symmetric and has significant contributions up to D2Q25 lattice nodes.

The estimated equilibrium distribution function $f_i^{\mathrm{eq},*}$ for a variety of coarse-grained time steps $\Delta t \in [0.3911, 6.1751]$ is shown in figure 3*a*. The equilibrium distribution function $f_i^{\mathrm{eq},*}$, as mentioned above, is obtained from three different methods: $f_i^{\mathrm{eq,MD}}$ is measured from an MD simulation, $f_i^{\mathrm{eq,G}}$ is theoretically estimated using a single Gaussian distribution function and $f_i^{\mathrm{eq,WSG}}$ is theoretically estimated from a Poisson WSG distribution function. The theoretical equilibrium distribution function models are described in detail in §3a and 3b, respectively.

In figure 3*a*, one can see that the largest equilibrium distribution function contributions are coming from the first layer neighbours $f_{0-8}^{\mathrm{eq},*}$. These nodes are approximated very well by both theoretical models, please refer to figure 4*a* for a detailed comparison of the measured and the theoretical $f_{0-8}^{\mathrm{eq},*}$. The next equilibrium distribution function groups $f_{9-12}^{\mathrm{eq},*}$ and $f_{13-20}^{\mathrm{eq},*}$ are significantly smaller than $f_{0-8}^{\mathrm{eq},*}$ with one to two order of magnitude. For $f_{9-12}^{\mathrm{eq},*}$ and $f_{13-20}^{\mathrm{eq},*}$, we see that the deviations of the measured and the theoretical single Gaussian model become larger. The Poisson WSG $f_{9-20}^{\mathrm{eq},*}$ shows a very good agreement with the measured equilibrium distribution function. The diagonal nodes in the second layer $f_{21-25}^{\mathrm{eq},*}$ are even smaller and their value could be considered negligible. However, the measured equilibrium distribution function $f_i^{\mathrm{eq,MD}}$ shows a good agreement with the theoretical Poisson WSG $f_i^{\mathrm{eq,WSG}}$ even for very small contributions such as $f_{21-25}^{\mathrm{eq},*}$. This suggests that these contributions even though really small are not just noise but theoretically justified.

Figures 4*a* and 5*a* depict the equilibrium distribution functions scaled to the single Gaussian equilibrium function. They show how the measured from MD simulation and the novel Poisson WSG equilibrium distribution functions deviate from the single Gaussian. The first layer equilibrium distribution function values are shown in figure 4*a*. These nodes have the largest contribution to the total equilibrium distribution function.

Figure 4*a* shows more particles staying at node zero and a depression for the first neighbouring layer (nodes 1 to 8). This very same feature repeats itself in figure 2*b*. The $\mathrm{P}^{\mathrm{WSG}}_{\lambda 2}\log(\mathrm{P}^{\mathrm{WSG}}_{\lambda 2}/\mathrm{P}^{\mathrm{G}})$ values depicted in blue show that the number of small displacements is enhanced $X_i/\Delta x \in [0, 0.3]$ while the number of $X_i/\Delta x \in [0.3, 1.0]$ displacements is suppressed.

The second layer equilibrium distribution function values are depicted in figure 5*a*. As one can see in figure 2*b*, there is an enhanced probability of large displacements $X_i/\Delta x \in [0.9, 1.6]$ which corresponds to the larger values of $f^{\mathrm{eq,WSG}}_{9-24}$ in figure 5*a*. The deviations (up to approx. 4.5%) from the theoretical single Gaussian equilibrium distribution function are also larger compared to the first layer nodes $f^{\mathrm{eq,WSG}}_{0-8}$. Since the $f^{\mathrm{eq,WSG}}_{9-24}$ true values are smaller by multiple orders of magnitude than the first layer neighbours $f^{\mathrm{eq,WSG}}_{0-8}$ these deviations are almost irrelevant for the total equilibrium distribution function, even though they are larger. Nevertheless, figure 5*a* shows clearly that the Poisson WSG equilibrium distribution function captures the MD data more precisely.

6. Outlook

In this article, we have derived a better approximation for the MDLG equilibrium distribution function. It deviates from the previous best approximation by Parsa *et al.* [1] in a broad transition region between the ballistic and diffusive regime of random particle displacements.

Despite the fact that these deviations are small, we expect them to be of great importance in the analysis of non-equilibrium systems, particularly systems not too far from equilibrium, as is typical in hydrodynamic systems. What we have outlined here is the equilibrium behaviour of the MDLG mapping of an MD simulation onto an LG. The key interest, however, lies in the non-equilibrium predictions of this mapping. In future research, we will investigate MDLG predictions for LG and lattice Boltzmann collision operators. In such systems, we expect to find only small deviations from local equilibrium, and to quantify these small deviations it is essential to have a very good understanding of the equilibrium behaviour of the MDLG mapping.

Data accessibility. This manuscript has no further supporting data.

Authors' contributions. A.J.W. supervised the research, contributed to it and revised the manuscript. A.P. contributed to the research, embedded the proposed model, set up the test cases, performed the data analysis and wrote the manuscript. All authors read and approved the manuscript.

Competing interests. The authors declare that they have no competing interests.

Funding. A.P. is partially supported by the Center for Nonlinear Studies (CNLS) and the Laboratory Directed Research and Development (LDRD) program at Los Alamos National Laboratory (LANL), and the German Federal Ministry of Education and Research (BMBF) in the scope of the project Aerotherm (reference numbers: 01IS16016A-B).

References

1. Parsa MR, Wagner AJ. 2017 Lattice gas with molecular dynamics collision operator. *Phys. Rev. E* **96**, 013314. (doi:10.1103/PhysRevE.96.013314)
2. Parsa MR, Wagner AJ. 2020 Large fluctuations in nonideal coarse-grained systems. *Phys. Rev. Lett.* **124**, 234501. (doi:10.1103/PhysRevLett.124.234501)
3. Blommel T, Wagner AJ. 2018 Integer lattice gas with Monte Carlo collision operator recovers the lattice Boltzmann method with Poisson-distributed fluctuations. *Phys. Rev. E* **97**, 023310. (doi:10.1103/PhysRevE.97.023310)
4. Frisch U, Hasslacher B, Pomeau Y. 1986 Lattice-gas automata for the Navier-Stokes equation. *Phys. Rev. Lett.* **56**, 1505–1508. (doi:10.1103/PhysRevLett.56.1505)

5. He X, Luo L-S. 1997 Theory of the lattice Boltzmann method: from the Boltzmann equation to the lattice Boltzmann equation. *Phys. Rev. E* **56**, 6811. (doi:10.1103/PhysRevE.56.6811)
6. Qian YH, D'Humières D, Lallemand P. 1992 Lattice BGK models for Navier-Stokes Equation. *Europhys. Lett. (EPL)* **17**, 479–484. (doi:10.1209/0295-5075/17/6/001)
7. Adhikari R, Stratford K, Cates ME, Wagner AJ. 2005 Fluctuating lattice Boltzmann. *Europhys. Lett. (EPL)* **71**, 473–479. (doi:10.1209/epl/i2004-10542-5)
8. Dünweg B, Schiller UD, Ladd AJC. 2007 Statistical mechanics of the fluctuating lattice Boltzmann equation. *Phys. Rev. E* **76**, 036704. (doi:10.1103/PhysRevE.76.036704)
9. Ladd AJC. 1993 Short-time motion of colloidal particles: numerical simulation via a fluctuating lattice-Boltzmann equation. *Phys. Rev. Lett.* **70**, 1339–1342. (doi:10.1103/PhysRevLett.70.1339)
10. Wagner AJ, Strand K. 2016 Fluctuating lattice Boltzmann method for the diffusion equation. *Phys. Rev. E* **94**, 033302. (doi:10.1103/PhysRevE.94.033302)
11. He X, Chen S, Doolen GD. 1998 A novel thermal model for the lattice Boltzmann method in incompressible limit. *J. Comput. Phys.* **146**, 282–300. (doi:10.1006/jcph.1998.6057)
12. McNamara GR, Garcia AL, Alder BJ. 1995 Stabilization of thermal lattice Boltzmann models. *J. Stat. Phys.* **81**, 395–408. (doi:10.1007/BF02179986)
13. Briant A, Wagner A, Yeomans J. 2004 Lattice Boltzmann simulations of contact line motion. I. Liquid-gas systems. *Phys. Rev. E* **69**, 031602. (doi:10.1103/PhysRevE.69.031602)
14. Osborn W, Orlandini E, Swift MR, Yeomans J, Banavar JR. 1995 Lattice Boltzmann study of hydrodynamic spinodal decomposition. *Phys. Rev. Lett.* **75**, 4031. (doi:10.1103/PhysRevLett.75.4031)
15. Shan X, Doolen G. 1995 Multicomponent lattice-Boltzmann model with interparticle interaction. *J. Stat. Phys.* **81**, 379–393. (doi:10.1007/BF02179985)
16. Pachalieva A, Wagner AJ. 2020 Non-Gaussian distribution of displacements for Lennard-Jones particles in equilibrium. *Phys. Rev. E* **102**, 053310. (doi:10.1103/PhysRevE.102.053310)
17. Kullback S, Leibler RA. 1951 On information and sufficiency. *Ann. Math. Stat.* **22**, 79–86. (doi:10.1214/aoms/1177729694)
18. LAMMPS Official Website: http://lammps.sandia.gov.
19. Plimpton S. 1995 Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **117**, 1–19. (doi:10.1006/jcph.1995.1039)
20. Potoff JJ, Panagiotopoulos AZ. 1998 Critical point and phase behavior of the pure fluid and a Lennard-Jones mixture. *J. Chem. Phys.* **109**, 10914–10920. (doi:10.1063/1.477787)
21. Parsa MR, Pachalieva A, Wagner AJ. 2019 Validity of the molecular-dynamics-lattice-gas global equilibrium distribution function. *Int. J. Mod. Phys. C* **30**, 1941007. (doi:10.1142/S0129183119410079)

APPENDIX E

Publication 4: Connecting lattice Boltzmann methods to physical reality by coarse-graining molecular dynamics simulations

The success of lattice Boltzmann methods has been attributed to their mesoscopic nature as a method derivable from a physically consistent microscopic model. Original lattice Boltzmann methods were Boltzmann averages of an underlying lattice gas. In the transition to modern lattice Boltzmann method, this link was broken, and the frequently used over-relaxation to achieve high Reynolds numbers has been seen as lacking physical motivation. While this approach has undeniable utility, it appeared to break the link to any underlying physical reality putting into question the special place of lattice Boltzmann methods among fluid simulation methods. In this letter, we show that over-relaxation arises naturally from physical lattice gases derived as a coarse-graining of molecular dynamics simulations thereby re-affirming the firm foundation of lattice Boltzmann methods in physical reality.

Preprint of an article published in

arXiv:2109.05009, 2021 [86]

Submitted to Physical Review Letters

Connecting lattice Boltzmann methods to physical reality by coarse-graining Molecular Dynamics simulations

Aleksandra Pachalieva[1, 2, 3, *] and Alexander J. Wagner[4, †]

[1] *Computational Earth Science Group, Earth and Environmental Sciences Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA*

[2] *Center for Nonlinear Studies, Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA*

[3] *Department of Mechanical Engineering, Technical University of Munich, 85748 Garching, Germany*

[4] *Department of Physics, North Dakota State University, Fargo, ND 58108, USA*

(Dated: December 2, 2021)

The success of lattice Boltzmann methods has been attributed to their mesoscopic nature as a method derivable from a physically consistent microscopic model. Original lattice Boltzmann methods were Boltzmann averages of an underlying lattice gas. In the transition to modern lattice Boltzmann method, this link was broken, and the frequently used over-relaxation to achieve high Reynolds numbers has been seen as lacking physical motivation. While this approach has undeniable utility, it appeared to break the link to any underlying physical reality putting into question the special place of lattice Boltzmann methods among fluid simulation methods. In this letter, we show that over-relaxation arises naturally from physical lattice gases derived as a coarse-graining of molecular dynamics simulations thereby re-affirming the firm foundation of lattice Boltzmann methods in physical reality.

Lattice Boltzmann methods originated from the revolutionary lattice gas method developed by Frisch, Hasslacher and Pomeau [1]. In lattice gases, local collisions redistribute particles according to collision rules that conserve particle number and local momentum. The effectiveness of collisions can be tuned to some degree [2, 3], and the viscosity is reduced as a result. However, such collisions bring the simulations at most to a state of local equilibrium.

To derive the macroscopic behavior of these systems, one first takes a non-equilibrium ensemble average of the lattice gas method, which results in a lattice Boltzmann equation [2]. This averaged lattice Boltzmann equation can be simulated directly, resulting in a noise-free simulation method [4]. This increases computational efficiency, since additional averaging over lattice gas results is no longer required, which counteracts the higher computational cost arising from transitioning from a Boolean lattice gas to a lattice Boltzmann method requiring real numbers. Instead of directly averaging the lattice gas collision terms, as was done by McNamara *et al.* [4], one can relax the distributions towards local equilibrium distribution function [5], resulting in the BGK approach.

The original lattice gas models were Boolean lattice gases, *i.e.* only zero or one particle were allowed per occupation number, leading to Fermi-Dirac, rather than Boltzmann equilibrium distribution [2]. This implied that the resulting hydrodynamic equations had mildly Galilean invariance violating terms [2]. Lattice Boltzmann methods, that abandoned their direct connection to the underlying lattice gas by imposing a Maxwell-Boltzmann equilibrium distribution with the BGK collision term, removed those velocity dependent terms in the Navier-Stokes equation [6]. They are given by

$$f_i(\mathbf{r} + \mathbf{v}_i\Delta t, t + \Delta t) = f_i(\mathbf{r}, t) + \Omega_i. \tag{1}$$

The BGK collision operator is

$$\Omega_i = \sum_j \Lambda_{ij}[f_j^{\mathrm{eq}} - f_j(\mathbf{r}, t)], \tag{2}$$

where the f_i are continuous densities associated with a lattice velocity $\mathbf{v}_i$ that represent an expectation value for the number of particles moving from lattice site $\mathbf{r} - \mathbf{v}_i\Delta t$ to lattice site $\mathbf{r}$ at time t. The BGK collision term redistributes those densities and relaxes them towards an imposed local equilibrium distribution f_i^{eq}. In the simplest case, the relaxation matrix Λ_{ij} has a single relaxation time $\Lambda_{ij} = (1/\tau)\delta_{ij}$, where $\tau = 1$ implies that local equilibrium is reached in one time step. For these methods the viscosity is

$$\nu = (\tau - 0.5)/3, \tag{3}$$

where the offset of 0.5 is a result of recombining terms from the Taylor expansion of the occupation probabilities with the terms obtained from the continuous Boltzmann equation. A general Λ_{ij} leads to multiple relaxation times, which is unimportant for this letter, since only one relaxation time turns out to be relevant for simple shear.

BGK lattice Boltzmann methods can no longer be justified as ensemble averages of the Boolean lattice gas models, and it became necessary to consider an alternative way of deriving the lattice Boltzmann method [7]. This was achieved by deriving lattice Boltzmann directly as a discretization of the continuous Boltzmann equation. Decades later, it was realized that it is still possible to

derive the BGK lattice Boltzmann methods from lattice gases with integer occupation numbers [8].

However, any of these derivations require the relaxation time in Eq. (3) to be $\tau > 1$, *i.e.* the averaged collisions bring the distribution functions at most to local equilibrium. He, Chen and Doolen [9] originally postulated that deriving lattice Boltzmann directly from the continuous Boltzmann equation could recover over-relaxation. Later Bösch and Karlin [10] showed that it was only an uncontrolled approximation in their derivation that lead to this result, whereas an exact analysis showed that the regime of over-relaxation is disconnected from the kinetic theory domain. Despite this apparent disconnect between over-relaxation and physical theory, over-relaxing the densities, *i.e.* using $0.5 \leq \tau < 1$, is extremely useful in obtaining lower viscosities, and is frequently used in practical applications.

Clearly, lattice Boltzmann methods with over-relaxation can no longer be related to lattice gas methods by a statistical average, since the local collisions can only achieve equilibrium, but never over-relax. Deriving lattice Boltzmann methods directly from discretizations of the Boltzmann equation equally fails to justify the usage of over-relaxation as was shown by Bösch and Karlin [10].

One could argue that is not important if lattice Boltzmann methods can be connected to some underlying physical model, as long as the method performs well. We believe this to be a shortsighted view. Seeing the lattice Boltzmann method as just another way of discretizing the Navier-Stokes equations misses the key ingredient allowing lattice Boltzmann to outperform classical Computational Fluid Dynamics (CFD) approaches in a number of areas. A stunning example is that lattice Boltzmann methods have all but displaced classical CFD from the modeling of automotive hydrodynamics and are making significant inroads in the aerospace industry [11]. The success of the lattice Boltzmann method must be firmly attributed to its grounding in some physical reality, and the inability of linking the frequently used over-relaxation to a physical underpinning is a worrying shortcoming.

In this letter, we show how this shortcoming can be overcome by a novel way of deriving lattice Boltzmann methods using a direct mapping approach from an average over Molecular Dynamics (MD) simulations onto lattice Boltzmann method. This approach has its roots in the Molecular Dynamics Lattice Gas (MDLG) method, pioneered by Parsa *et al.* [12].

Briefly the MDLG method consists of overlaying a square lattice with lattice spacing Δx onto an MD simulation. We define lattice displacements vectors c_i connecting different lattice sites, using the index i to enumerate the possible displacements. After fixing a time step Δt, we identify the number of particles that move from cell $\mathbf{r} - \mathbf{c}_i$ at time $t - \Delta t$ to lattice cell $\mathbf{r}$ at time t with $n_i(\mathbf{r}, t)$ lattice gas occupation number. This procedure maps the MD simulation onto a lattice gas [12] as shown in Fig. 1a. The idea of the Molecular Dynamics Lattice Boltzmann (MDLB) is then to average over an ensemble of MD simulations of the same macroscopic state to obtain the lattice Boltzmann densities

$$f_i(\mathbf{r}, t) = \langle n_i(\mathbf{r}, t) \rangle. \tag{4}$$

Once we have $f_i(x, t)$, we can determine the lattice Boltzmann collision operator

$$\Omega_i = f_i(\mathbf{r} + \mathbf{c}_i, t + \Delta t) - f_i(\mathbf{r}, t). \tag{5}$$

The focus of this letter are the properties of the MDLB collision operator and its ability to exhibit over-relaxation.

While the above described procedure is general and can, in principle, be applied to any flow, the numerical cost of averaging over a large number of MD simulations can be considerable. Instead, we investigate the simplest non-equilibrium situation: a simple shear flow where the averaged velocities are given by

$$u_x = \dot{\gamma} y; \quad u_y = 0, \tag{6}$$

with $\dot{\gamma}$ being the shear rate, $\mathbf{r} = (x, y)^T$ is the position vector, and the density remains constant. Since, this flow is invariant under translation in the x-direction, and shifts in the y-direction can be related by a simple Galilean transformation to the $y = 0$ position, all points are in this sense equivalent. Thus, we can average over all lattice points at all times, allowing for ample statistical averaging.

The MD simulations are executed using LAMMPS framework [13, 14] developed by Sandia National Laboratories. The system consists of particles interacting with the standard 6-12 Lennard-Jones (LJ) intermolecular potential. The particle mass m and diameter σ are set to one. Each simulation contains $N = 99\,856$ particles in a two-dimensional square with length L = 1000 LJ units referring to an area fraction of $\phi = 0.078387$. The area fraction ϕ for circular LJ particles with van der Waals radius $r = \sigma/2$ is defined as the product of the particle surface area and the number of particles, divided by the square length of the simulation box. We initialised the simulations using homogeneously distributed particles with kinetic energy equal to 20 in LJ units, which corresponds to a dilute gas. We use the LAMMPS `nvt/sllod` thermostat to generate the desired non-equilibrium dynamics. The lattice Boltzmann discretizations in time and space (Δt and Δx) are fixed so that $\langle (\delta x)^2 \rangle^{\mathrm{eq}} / (\Delta x)^2 \approx \frac{1}{6}$, where $\langle (\delta x)^2 \rangle^{\mathrm{eq}}$ is the equilibrium mean-squared displacement. This ratio ensures that the particle displacements are essentially limited to a neighborhood touching the central cell as shown in Fig. 1a. This is referred to as an D2Q9 model since it resides in two dimensions and requires nine lattice velocities. We perform a wide range of simulations – from

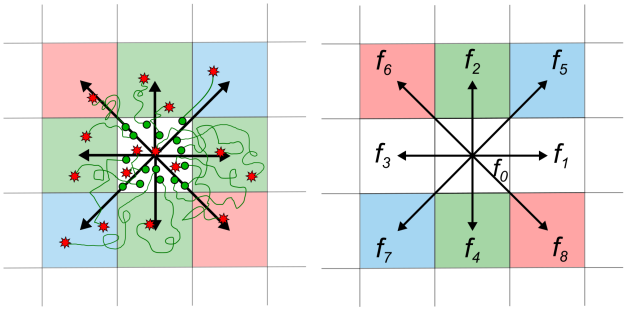


FIG. 1. (Color online) (a) Sketch of the MDLG dynamics for D2Q9 lattice. A lattice is imposed onto the MD simulation domain and the movement of the particles is tracked from the central node using their MD trajectories. The circles (green) represent the particle position at time $t - \Delta t$ and the stars (red) are their respective positions at time t. The arrows (black) depict the lattice velocities. (b) Schematic representation of the D2Q9 lattice showing the numbering convention for f_i. The symmetries in the lattice are color-coded.

simulations, where mean free time (*i.e.* the time between collisions) is much larger than Δt (ballistic regime) to simulations, where Δt is much larger than the mean free time (diffusive regime). The data is collected after the simple shear has reached a steady state. For further information, please, refer to the supplemental material and the LAMMPS documentation [14].

The symmetry of the simple shear flow puts significant constraints on the collision term Ω_i defined in Eq. (5). The degrees of freedom for the collision operator at the point $y = 0$, where the mean velocity is zero, are sketched in Fig. 1b. The point symmetry about the center of the lattice implies $f_2 = f_4$, $f_5 = f_7$, and $f_6 = f_8$. Translational symmetry in the x-direction implies that f_0, f_1, and f_3 are unchanged by the collision. Therefore, symmetry leaves only three independent values for an D2Q9 velocity set in the collision term Ω_i, which is reduced to two because mass conservation adds the additional constraint $\sum_i \Omega_i = 0$. Therefore, the D2Q9 collision operator is determined by two terms that we choose as

$$\Omega^\alpha = \Omega_6 - \Omega_5 + \Omega_8 - \Omega_7, \tag{7}$$

$$\Omega^\beta = \Omega_2 + \Omega_4, \tag{8}$$

with $\Omega^\alpha \gg \Omega^\beta$ for a simple shear. In this letter, we focus on the dominant collision contribution Ω^α. Now, we can define the moment before the collision as a function of the probability distribution function f_i

$$M^\alpha = f_6 - f_5 + f_8 - f_7, \tag{9}$$

and the moment after the collision

$$M^{\alpha,*} = M^\alpha + \Omega^\alpha. \tag{10}$$

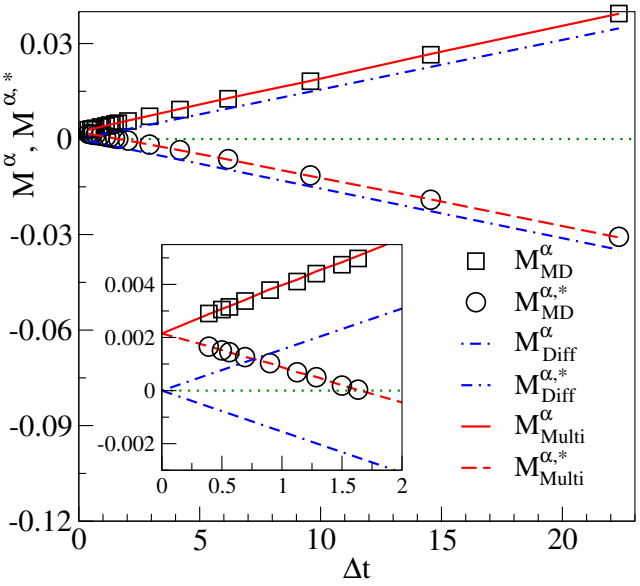


FIG. 2. (Color online) The pre- and post-collision moments (M^α and $M^{\alpha,*}$) are shown as a function of Δt. The sign change between M^α and $M^{\alpha,*}$ indicates the transition from under- to over-relaxation. Three sets of data are shown: the measured from MD depicted as symbols (black); analytical solution using the multivariate Gaussian with diffusive moments depicted as dash-dotted lines (blue); analytical solution using the multivariate Gaussian with measured moments depicted as full and dashed lines (red). The zoomed plot shows the sign change of $M^{\alpha,*}_{\rm MD}$ and $M^{\alpha,*}_{\rm Multi}$ at $\Delta t \approx 1.6$.

In equilibrium, we have $M^{\alpha,\rm eq} = 0$ due to symmetry. The signature of over-relaxation is, therefore, a sign change between M^α and $M^{\alpha,*}$. The measured values of these quantities are shown as symbols in Fig. 2 as a function of Δt. For small Δt both the $M^\alpha_{\rm MD}$ and $M^{\alpha,*}_{\rm MD}$ are positive, but $M^{\alpha,*}_{\rm MD}$ changes sign for $\Delta t \gtrapprox 1.6$. Hence, the MDLB procedure predicts that for larger coarse-graining the relaxation towards equilibrium is replaced by an over-relaxation.

In terms of the relaxation time τ, we have

$$M^{\alpha,*} = M^\alpha + \frac{1}{\tau}(M^{\alpha,eq} - M^\alpha), \tag{11}$$

with $M^{\alpha,eq} = 0$, the relaxation time can be expressed as

$$\tau = \frac{M^\alpha}{M^\alpha - M^{\alpha *}}. \tag{12}$$

In Fig. 3, we show $(\tau - 0.5)$ as a function of Δt.

The remainder of this letter focuses on the origin of the observed over-relaxation. In the MDLG coarse-graining, f_i can be expressed in terms of the one-particle displacement function $P(\mathbf{r}, \delta\mathbf{r}, t)$ which gives the probability of finding a particle at position r at time t that was at position r $-$ δr at time $t - \Delta t$ [12]

$$f_i(\xi, t) = \int\int P(\mathbf{r}, \delta\mathbf{r}) \Delta_{\xi - \mathbf{c}_i}(\mathbf{r} - \delta\mathbf{r}) \Delta_\xi(\mathbf{r}) d\mathbf{r} d\delta\mathbf{r}, \tag{13}$$

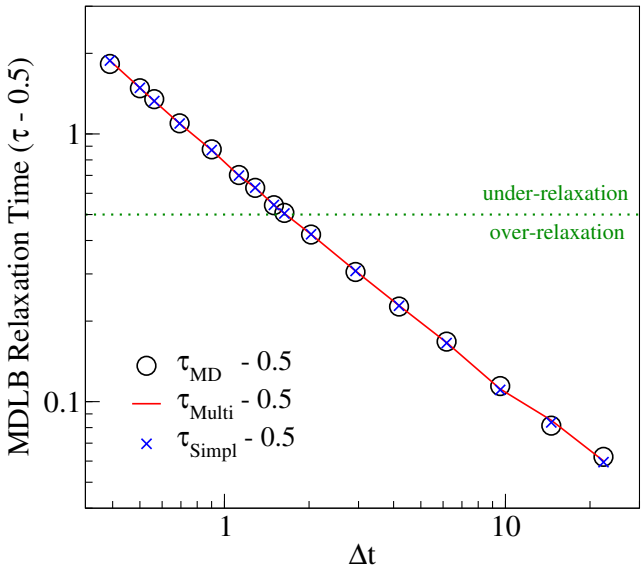


FIG. 3. (Color online) The relaxation time $(\tau - 0.5)$ as a function of Δt (logarithmic scale). Three sets of data are shown: measured from MD simulation depicted as circles (black); multivariate Gaussian with the MD depicted as a line (red); simplified multivariate Gaussian with $\langle(\delta x)^2\rangle = \langle(\delta y)^2\rangle = \langle(\delta x)^2\rangle^{\rm eq}$ depicted as crosses (blue). For $(\tau - 0.5)$ above the dotted line (green), the collision operator under-relaxes, while for $(\tau - 0.5)$ below that line, the collision over-relaxes.

where $\Delta_\xi(\mathbf{r})$ is one, if $\mathbf{r}$ resides in the lattice site ξ and zero otherwise, and $\mathbf{c}_i = \mathbf{v}_i\Delta t$ is the lattice displacement. This reduces the problem of finding f_i to the problem of finding the one-particle displacement function, which can be very challenging for arbitrary flows. In the diffusive limit, *i.e.* when the mean free path is small and particles effectively undergo Brownian motion, an analytical solution exist [15–17]. In this case, the one-particle displacement distribution function is given by a multivariate Gaussian probability distribution

$$P(x,y,\delta x,\delta y) = \frac{\sqrt{-\frac{\sigma_x}{\sigma_{xy}^2}+\frac{4}{\sigma_y}}}{2\pi\sqrt{\sigma_x}} \times \exp\left(-\frac{(\delta x - y\dot\gamma\Delta t)^2}{\sigma_x} - \frac{(\delta x - y\dot\gamma\Delta t)\delta y}{\sigma_{xy}} - \frac{(\delta y)^2}{\sigma_y}\right) \tag{14}$$

with the moments

$$\begin{aligned}\sigma_x &= \langle(\delta x)^2\rangle^{\rm eq}(1+\frac{\dot\gamma^2\Delta t^2}{3}),\\ \sigma_{xy} &= \frac{\langle(\delta x)^2\rangle^{\rm eq}\Delta t\dot\gamma}{2},\\ \sigma_y &= \langle(\delta x)^2\rangle^{\rm eq},\end{aligned} \tag{15}$$

where $\langle(\delta x)^2\rangle^{\rm eq}$ is the measured equilibrium mean-squared displacement as defined in [18, 19]. Note that a Galilean transformation is applied to the x-displacements that are at $y \neq 0$. Using Eqs. (13)-(15) we calculate f_i and obtain $M^{\alpha}_{\rm Diff}$, and $M^{\alpha,*}_{\rm Diff}$, which are shown as dash-dotted lines (blue) in Fig. 2. The trend is very similar to the MD measurements but the results obtained using the diffusive moments are offset by a constant. The analytical result is entirely symmetric around the origin, leading to a relaxation time of 0.5 for all time steps.

If we instead use a multivariate Gaussian with moments measured in the MD simulation

$$\sigma_x = \langle(\delta x)^2\rangle, \quad \sigma_{xy} = \langle\delta x\delta y\rangle, \quad \sigma_y = \langle(\delta y)^2\rangle, \tag{16}$$

we obtain the predictions for $M^{\alpha}_{\rm MD}$, shown as solid line (red), and $M^{\alpha,*}_{\rm MD}$, shown as dashed line (red), in Fig. 2. They are in excellent agreement with the measured values. In Fig. 3, we show that the resulting relaxation time $(\tau_{\rm Multi} - 0.5)$ is likewise in excellent agreement with the measurement $(\tau_{\rm MD} - 0.5)$.

To understand the physical origin of the transition from under- to over-relaxation let us make a few observations: for the modest shear considered here with $(\dot\gamma\Delta t)^2 \ll 3$ in Eq. (15), we have $\langle(\delta x)^2\rangle \approx \langle(\delta y)^2\rangle$ and both are approximately given by the equilibrium mean-squared displacement $\langle(\delta x)^2\rangle^{\rm eq}$. The key change occurs in the off-diagonal moment $\langle\delta x\delta y\rangle$. In Fig. 4, we show $\langle\delta x\delta y\rangle/\langle(\delta x)^2\rangle^{\rm eq}$ as a function of Δt. We depict the off-diagonal moment measured from the MD simulation with symbols (black) and the one calculated using the diffusive moments in Eq. (15) with a line (red). The off-diagonal moment $\langle\delta x\delta y\rangle_{\rm MD}$ changes sign at $\Delta t \approx 4$ and otherwise behaves similar to the diffusive theory, albeit with an offset. The qualitative behavior in the diffusive case is straight forward: as particles diffuse into the positive y-direction they get carried away with the flow, and obtain an additional positive x-displacement leading to a positive correlation between x- and y-displacements. This means that any memory is quickly lost in frequent collisions. In the ballistic case, however, collisions are rare, and particles carry a memory of their history over larger distances. In particular, particles that move into the positive y-direction will typically have last collided at a position with negative y. In these regions, the average velocity is negative, so these particles will carry the average negative x-velocity prevalent in the region of their last collision to the regions of larger y. This leads to an anti-correlation between the x- and y-displacement. As we are looking at larger Δt, collisions become more frequent, and eventually the diffusive behavior becomes dominant.

The predication of the relaxation time $(\tau_{\rm Simpl} - 0.5)$ in Fig. 3, is calculated using a simple model with $\sigma_x = \sigma_y = \langle(\delta x)^2\rangle^{\rm eq}$ and the measured off-diagonal moment $\langle\delta x\delta y\rangle$. We see that this simple model is also in excellent agreement with the measured values, showing that the off-diagonal moment is indeed responsible for the transition from under- to over-relaxation. Note, however, that it is not simply the sign change that determines this

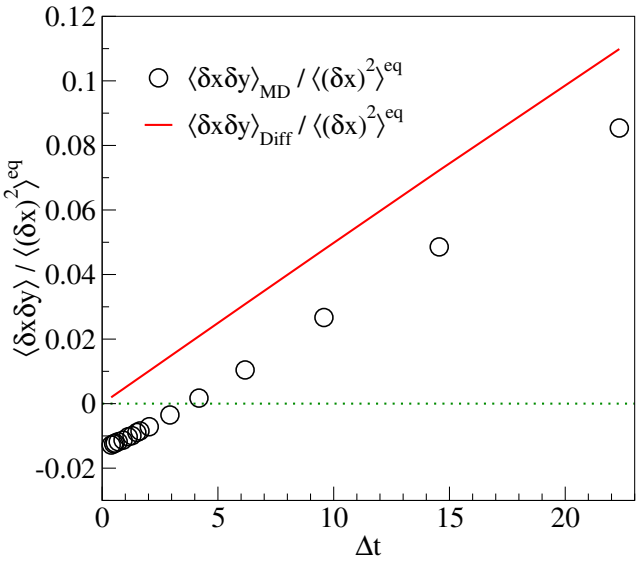


FIG. 4. (Color online) The off-diagonal moment $\langle \delta x \delta y \rangle$ bears the main change from under- to over-relaxation, shown here as the moment is normalized by the equilibrium mean-squared displacement $\langle (\Delta x)^2 \rangle^{\text{eq}}$. The circles (black) depict $\langle \delta x \delta y \rangle_{\text{MD}}$ measured from the MD simulation. The line (red) depicts $\langle \delta x \delta y \rangle_{\text{Diff}}$ obtained using the multivariate Gaussian with diffusive moments. The sign change of $\langle \delta x \delta y \rangle$ is a key feature of the transition from under- to over-relaxation, however, it is not the only factor since the over-relaxation occurs at $\Delta t \approx 1.6$ and the sign change happens at $\Delta t \approx 4$.

transition as the sign change occurs at $\Delta t \approx 4$ whereas the transition from under- to over-relaxation occurs at $\Delta t \approx 1.6$.

In conclusion, in this letter we have shown that a lattice Boltzmann collision operator can be directly derived from one-particle displacement probability distribution, which can be obtained from an MD simulation. This approach shows that such lattice Boltzmann collision operators naturally transition from under- to over-relaxation. Thus, the over-relaxation in lattice Boltzmann methods can be derived from first principles and is a consequence of the coarse-grained representation of a lattice Boltzmann method.

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

[1] U. Frisch, B. Hasslacher, and Y. Pomeau, Phys. Rev. Lett. **56**, 1505 (1986).
[2] U. Frisch, D. d'Humieres, B. Hasslacher, P. Lallemand, Y. Pomeau, and J.-P. Rivet, Complex Systems **1**, 649 (1987).
[3] M. Hénon, Complex Systems **1**, 763 (1987).
[4] G. R. McNamara and G. Zanetti, Phys. Rev. Lett. **61**, 2332 (1988).
[5] F. J. Higuera and J. Jiménez, EPL (Europhysics Letters) **9**, 663 (1989).
[6] Y.-H. Qian, D. d'Humières, and P. Lallemand, EPL (Europhysics Letters) **17**, 479 (1992).
[7] X. He and L.-S. Luo, Phys. Rev. E **56**, 6811 (1997).
[8] T. Blommel and A. J. Wagner, Phys. Rev. E **97**, 023310 (2018).
[9] X. He, S. Chen, and G. D. Doolen, Journal of Computational Physics **146**, 282 (1998).
[10] F. Bösch and I. V. Karlin, Phys. Rev. Lett. **111**, 090601 (2013).
[11] H. Chen, personal communication (2018).
[12] M. R. Parsa and A. J. Wagner, Phys. Rev. E **96**, 013314 (2017).
[13] S. Plimpton, Journal of Computational Physics **117**, 1 (1995).
[14] LAMMPS Official Website: http://lammps.sandia.gov.
[15] D. E. Elrick, Australian Journal of Physics **15**, 283 (1962).
[16] T. G. M. Van de Ven, Journal of Colloid and Interface Science **62**, 352 (1977).
[17] R. T. Foister and T. G. M. Van De Ven, Journal of Fluid Mechanics **96**, 105 (1980).
[18] A. Pachalieva and A. J. Wagner, Phys. Rev. E **102**, 053310 (2020).
[19] A. Pachalieva and A. J. Wagner, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **379**, 20200404 (2021).

Supplemental material to "Connecting lattice Boltzmann methods to physical reality by coarse-graining Molecular Dynamics simulations"

Aleksandra Pachalieva[1, 2, 3, *] and Alexander Wagner[4, †]

[1] *Computational Earth Science Group (EES-16), Earth and Environmental Sciences Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA*

[2] *Center for Nonlinear Studies, Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA*

[3] *Department of Mechanical Engineering, Technical University of Munich, 85748 Garching, Germany*

[4] *Department of Physics, North Dakota State University, Fargo, ND 58108, USA*

(Dated: December 2, 2021)

In order to ensure reproducibility of the results, we provide additional information about the complete workflow to obtain the results shown in the accompanying letter. First, we give details about the derivation of the MDLB collision operator, which can be expressed as a function of the one-particle displacements probability distribution function. Then, we include the molecular dynamics (MD) simulation setup and the LAMMPS scripts, which we execute to obtain the MD particle positions and trajectories. Next, we describe the Molecular Dynamics Lattice Gas (MDLG) and Lattice Boltzmann (MDLB) toolbox, which consists of an C/C++ framework, three Mathematica scripts and two Excel worksheets.

DERIVATION OF THE MDLB COLLISION OPERATOR

In this section, we take a closer look at the derivation of the MDLB collision operator. To do this, first we define the probability distribution function as

$$\begin{aligned} f_i(\vec{x},t) &= \langle n_i(\vec{x},t)\rangle^{\text{neq}} \\ &= \int dx_1 \int d\delta_1 \cdots \int dx_N \int d\delta_N P^N(x_1,\delta_1,\ldots,x_N,\delta_N,t) \sum_j \Delta_{\vec{x}}(x_j)\Delta_{\vec{x}-\vec{c}_i}(x_j-\delta_j), \end{aligned} \tag{1}$$

with $\langle n_i(\vec{x},t)\rangle^{\text{neq}}$ being the non-equilibrium ensemble average and $P^N(x_1,\delta_1,\ldots,x_N,\delta_N,t)$ being the N-particle displacement distribution function. $P^N(x_1,\delta_1,\ldots,x_N,\delta_N,t)$ is symmetric in all particles, thus, we can simplify Eq. (1) to

$$\begin{aligned} f_i(\vec{x},t) &= \langle n_i(\vec{x},t)\rangle^{\text{neq}} \\ &= N \int dx_1 \int d\delta_1 P^1(x_1,\delta_1,t_1)\Delta_{\vec{x}}(x_1)\Delta_{\vec{x}-\vec{c}_1}(x_1,\delta_1), \end{aligned} \tag{2}$$

where N is the total number of particles and $P^{(1)}$ is the one-particle displacement distribution function. Thus, the MDLB collision operator Ω_i is given by

$$\begin{aligned} \Omega_i &= \langle \Xi_i \rangle \\ &= \int dx_1 \int d\delta_1 \cdots \int dx_N \int d\delta_N P^N(x_1,\delta_1,\ldots,x_N,\delta_N,t+\Delta t)\Delta_{\vec{x}+\vec{c}_i}(x_j)\Delta_{\vec{x}}(x_j-\delta_j) \\ &\quad - \int dx_1 \int d\delta_1 \cdots \int dx_N \int d\delta_N P^N(x_1,\delta_1,\ldots,x_N,\delta_N,t)\Delta_{\vec{x}}(x_j)\Delta_{\vec{x}-\vec{c}_i}(x_j-\delta_j) \\ &= N \int dx_1 \int d\delta_1 P^1(x_1,\delta_1,t+\Delta t)\Delta_{\vec{x}+\vec{c}_i}(x_1)\Delta_{\vec{x}}(x_1-\delta_1) \\ &\quad - N \int dx_1 \int d\delta_1 P^1(x_1,\delta_1,t)\Delta_{\vec{x}}(x_1)\Delta_{\vec{x}-\vec{c}_i}(x_1-\delta_1) \end{aligned} \tag{3}$$

For stationary problems the one-particle distribution function does not depend on time. In such cases, we obtain

$$\begin{aligned}\Omega_i &= \langle \Xi_i \rangle \\ &= N \int dx_1 \int d\delta_1 P^1(x_1, \delta_1) \Delta_{\vec{x}+\vec{c}_i}(x_1) \Delta_{\vec{x}}(x_1 - \delta_1) \\ &\quad - N \int dy_1 \int d\gamma_1 P^1(y_1, \gamma_1) \Delta_{\vec{x}}(y_1) \Delta_{\vec{x}-\vec{c}_i}(y_1 - \gamma_1)\end{aligned} \tag{4}$$

For a stationary processes with non-equilibrium dynamics, the MDLB formulation allows us to express $f_i(\vec{x}, t)$ and Ω_i as quantities which depend solely on the one-particle displacements Probability Distribution Function (PDF). Thus, knowing the analytical solution for $P^1(\vec{x}, \delta, t)$ allows us to obtain the Ω_i also analytically.

MOLECULAR DYNAMICS SIMULATIONS

We executed a simple shear flow simulation with velocity profile given by

$$u_x = \dot{\gamma} y; \quad u_y = 0, \tag{5}$$

with $\dot{\gamma}$ being the shear rate, the position vector is given by $\vec{x} = (x, y)^T$ and the density remains constant. The shear rate of the system $\dot{\gamma}$ is set to 0.01. A sketch of the simulation is given in Fig. 1 below.

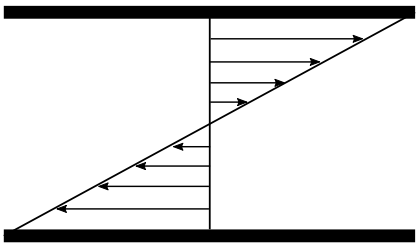


FIG. 1. Sketch of the MD simple shear flow simulation.

The molecular dynamics simulations were performed using the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) package developed by Sandia National Laboratories [1, 2]. The system of interest consists of particles interacting with the standard 6-12 Lennard-Jones (LJ) intermolecular potential given by

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

The particle mass m and the LJ particle diameter σ are set to one. Each simulation included $N = 99\,856$ number of particles in a two-dimensional (2D) square with length $L = 1000$ LJ units referring to an area fraction of $\phi = 0.078387$. The area fraction ϕ for circular LJ particles with van der Waals radius $r = \sigma/2$ 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. We initialised the simulations using homogeneously distributed particles with kinetic energy corresponding to 20 in LJ units. To ensure that the initial system is in equilibrium, we execute 200 000 MD iterations as shown below in Steps 0 and 1. After that, we apply a force field to the particle velocities to initialize a simple shear flow and we use the LAMMPS `nvt/sllod` thermostat to generate the desired non-equilibrium molecular dynamics as shown in Steps 2 and 3. The time and space discretization ratio is fixed to $\langle (\delta x)^2 \rangle / (\Delta x)^2 \approx \frac{1}{6}$, where $\langle (\delta x)^2 \rangle$ is mean-squared displacement and Δx is the lattice spacing. We chose a very small MD step size, equal to 0.0001, to ensure high accuracy of the MD simulation data. We perform a wide range of simulations – from simulations where mean free time (*i.e.* the time between collisions) is much larger than the time-step Δt (ballistic regime) to simulations where the time step Δt is much larger than the mean free time (diffusive regime). The simple shear flow simulations are executed with a variety of time steps $\Delta t \in [0.3911, 22.3372]$ as given here in Table I. For each time step Δt we saved 2 000 coarse-grained iterations. The simple shear flow does not change after reaching a steady state; however, it still describes non-equilibrium dynamics. However, we run the simulations for large number of iterations to collect enough data such that sufficient averaging is ensured. For further information, please, refer to the LAMMPS documentation [2].

TABLE I. Simulation parameters

Δt	Δx	lx
0.3911	4.0	250
0.5000	5.0	200
0.5626	5.5	180
0.6927	6.6(6)	150
0.9009	8.3(3)	120
1.1261	10.0	100
1.4994	12.5	80
1.6342	13.3(3)	75
2.0338	15.625	64
2.9280	20.0	50
4.1821	25.0	40
6.1751	31.25	32
9.5793	40.0	25
14.5580	50.0	20
22.3372	62.5	16

LAMMPS SCRIPTS

In order to obtain the final MD data set, we perform the following steps for each group of simulation parame-

ters as shown below in Table I.

Step 0: Initialization

We have generated an input data field with homogeneous distribution of particles. The length of the two-dimensional domain is $L = 1000$ LJ units and we used $N = 99\,856$ number of particles.

Step 1: Equilibrium Simulation

To initialize our MD simulation, we use the homogeneously distributed particles as described in Step 0. The kinetic temperature is set to a temperature of 20 in Lennard-Jones (LJ) units and the LAMMPS script is given below:

```
# 2d Lennard-Jones
# Equilibrium Simulation
dimension       2
units           lj
atom_style      atomic
timestep        0.0001
variable        out_folder universe /output/
boundary        pp pp pp

# Initial data structure
read_data       initStruct_1000.input
mass            1 1.0
velocity        all create 20 87287 loop geom
pair_style      lj/cut 2.5
pair_coeff      1 1 1.0 1.0 2.5
neigh_modify    delay 0 every 1 check no
fix             1 all nve

compute         rtemp all temp/deform
thermo          1000
thermo_style    custom step dt pe ke c_rtemp

run             100000

velocity        all create 20 87287 loop geom

run             100000

write_restart   init.restart
```

We run the equilibrium simulation for 200 000 MD time steps and output an LAMMPS restart file.

Step 2: Simple Shear Flow

The restart file obtained from Step 1 is transformed into a text file using the following command:

```
lmp -restart2data init.restart init.data
```

We read the resulting ASCII file `init.data` line by line and add a velocity profile in x-direction as defined earlier in Eq. (5).

Step 3: Non-equilibrium Simulation

In the MD context, the simple shear flow is implemented using Lees-Edwards boundary conditions. To obtain this in LAMMPS, we use the `fix deform` command, accompanying with the `nvt/sllod` thermostat. The complete LAMMPS script is given below:

```
# 2d Lennard-Jones
# Non-equilibrium Simulation
dimension       2
units           lj
atom_style      atomic
timestep        0.0001
variable        out_folder universe /output
# See Table 1
variable        out_iter equal "11261"
variable        out_dump equal
    ${out_iter}*2000
boundary        pp pp pp

# Initial data structure
read_data       initEqVelStruct_1000.input
mass            1 1.0
pair_style      lj/cut 2.5
pair_coeff      1 1 1.0 1.0 2.5
neigh_modify    delay 0 every 1 check no

# Lees-Edwards Boundary Conditions
fix             1 all deform 1 xy erate 0.01
    remap v
fix             2 all nvt/sllod temp 20.0
    20.0 20.0

compute         rtemp all temp/deform
thermo          20000
thermo_style    custom step dt pe ke c_rtemp

dump            1 all custom ${out_iter}
    ${out_folder}/02_Neq_nvt_${out_iter}.relax
    id type xs ys zs vx vy vz
dump_modify     1 sort id
dump_modify     1 format line "%d %d %.8f %.8f
    %f %.8f %.8f %f"

run             ${out_dump}
```

The number of iterations `out_iter` varies from $\Delta t \in [0.3911, 22.3372]$ as given in Table I.

MDLG ANALYSIS TOOL

The MDLG analysis tool can be found in [3]. The framework is implemented in C/C++. It reads and post-process the molecular dynamics data obtained from

the previously described LAMMPS scripts. By coarse-graining the MD particle positions and trajectories, we obtain key features of the lattice Boltzmann method. A complementary Graphical User Interface (GUI) developed by A. J. Wagner [4] is used for visualization purposes. For more information on how to compile and run the code, please, contact the authors.

MATHEMATICA SCRIPTS

The Mathematica scripts include the calculation of the probability distribution function f_i from the one-particle multivariate Gaussian distribution function as defined in Eq. (14) in the letter. The calculation is carried on for a variety of time steps and discretization sizes as shown in Table I. The three Mathematica scripts include the following:

1. Calculation of the f_i using the diffusive moments as defined in Eq. (15) in the letter;

2. Calculation of the f_i using the measured MD moments as defined in Eq. (16) in the letter;

3. Calculation of the f_i using $\langle(\delta x)^2\rangle = \langle(\delta y)^2\rangle$ set to the equilibrium values of the mean-squared displacement $\langle(\delta x)^2\rangle^{\mathrm{eq}}$, and the off-diagonal moment $\langle\delta x\delta y\rangle$ set to the measured one from the MD simulation.

We used Mathematica v12.1. More information can be found within the Mathematica scripts.

EXCEL WORKSHEETS

In the excel worksheets, we calculate the collision operator Ω_i, the pre- and post-collision moments M^α and $M^{\alpha,*}$ and the relaxation time τ as defined in the letter Eqs. (7)-(10) and Eq. (12). These scripts use the probability distribution functions f_i obtained from the MDLG framework, which are based on the MD data, and the Multivariate Gaussian probability distribution function with diffusive and measured moments as given in Eqs. (14)-(16) in the letter.

In the second worksheet of the excel file, the collision operator Ω_i, the pre- and post-collision moments M^α and $M^{\alpha,*}$ and the relaxation time τ are calculated using a more simple definition of the moments. We fixed the values of $\langle(\delta x)^2\rangle$ and $\langle(\delta y)^2\rangle$ to the equilibrium values of the mean-squared displacement $\langle(\delta x)^2\rangle^{\mathrm{eq}}$ and used the off-diagonal moment $\langle\delta x\delta y\rangle$ to the measured moment from the MD simulation.

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

[1] S. Plimpton, Journal of Computational Physics **117**, 1 (1995).
[2] "LAMMPS Official Website: http://lammps.sandia.gov,".
[3] A. Pachalieva and A. J. Wagner, "Molecular Dynamics Lattice Gas GUI," `https://gitlab.lrz.de/ga35pak/MDLG_GUI` (2021).
[4] A. J. Wagner, "Graphical User Interface," `https://www.ndsu.edu/pubweb/~carswagn/GUI/index.html` (2016).

Bibliography

[1] Sauro Succi. *The lattice Boltzmann equation: For complex states of flowing matter*. Oxford University Press, 2018.

[2] Hudong Chen. personal communication, 2018.

[3] Christoph A. Niedermeier and Thomas Indinger. Altair Engineering GmbH. personal communication, 2016-2019.

[4] Shuling Hou, J. Sterling, Shiyi Chen, and Gary D. Doolen. A lattice Boltzmann subgrid model for high Reynolds number flows. *Journal of Computational Physics*, 118(329), 1995.

[5] Christopher M. Teixeira. Incorporating turbulence models into the lattice-Boltzmann method. *International Journal of Modern Physics C*, 9(08):1159–1175, 1998.

[6] Orestis Malaspinas and Pierre Sagaut. Consistent subgrid scale modelling for lattice boltzmann methods. *Journal of Fluid Mechanics*, 700:514–542, 2012.

[7] W. R. Osborn, E. Orlandini, Michael R. Swift, J. M. Yeomans, and Jayanth R. Banavar. Lattice Boltzmann study of hydrodynamic spinodal decomposition. *Physical Review Letters*, 75(22):4031, 1995.

[8] Xiaowen Shan and Gary D. Doolen. Multicomponent lattice-Boltzmann model with interparticle interaction. *Journal of Statistical Physics*, 81(1-2):379–393, 1995.

[9] Xiaoyi He and Li-Shi Luo. Theory of the lattice Boltzmann method: From the Boltzmann equation to the lattice Boltzmann equation. *Physical Review E*, 56(6):6811, 1997.

[10] A. J. Briant, Alexander J. Wagner, and J. M. Yeomans. Lattice Boltzmann simulations of contact line motion. I. Liquid-gas systems. *Physical Review E*, 69(3):031602, 2004.

[11] Xiaoyi He, Shiyi Chen, and Gary D. Doolen. A novel thermal model for the lattice Boltzmann method in incompressible limit. *Journal of Computational Physics*, 146(1):282–300, 1998.

[12] Guy R. McNamara, Alejandro L. Garcia, and Berni J. Alder. Stabilization of thermal lattice Boltzmann models. *Journal of Statistical Physics*, 81(1-2):395–408, 1995.

[13] Sauro Succi and Roberto Benzi. Lattice Boltzmann equation for quantum mechanics. *Physica D: Nonlinear Phenomena*, 69(3-4):327–332, 1993.

[14] Silvia Palpacelli and Sauro Succi. The quantum lattice Boltzmann equation: Recent developments. *Communications in Computational Physics*, 4(5):980–1007, 2008.

[15] Paul J. Dellar, Denis Lapitski, Silvia Palpacelli, and Sauro Succi. Isotropy of three-dimensional quantum lattice boltzmann schemes. *Physical Review E*, 83(4):046706, 2011.

[16] Miller Mendoza, Bruce M. Boghosian, Hans Jürgen Herrmann, and Sauro Succi. Fast lattice Boltzmann solver for relativistic hydrodynamics. *Physical review letters*, 105(1):014502, 2010.

[17] R. Peyret and T. D. Taylor. Computational methods for fluid flow. *New York*, 1985.

[18] G. K. Batchelor. *An Introduction to Fluid Dynamics*. Cambridge University Press, 2000.

[19] Joel H. Ferziger, Milovan Perić, and Robert L. Street. *Computational methods for fluid dynamics*, volume 3. Springer, 2002.

[20] Henk Kaarle Versteeg and Weeratunge Malalasekera. *An introduction to computational fluid dynamics: The finite volume method*. Pearson education, 2007.

[21] Suhas Patankar. *Numerical Heat Transfer and Fluid Flow*. Taylor & Francis, 2018.

[22] Timm Krüger, Halim Kusumaatmaja, Alexandr Kuzmin, Orest Shardt, Goncalo Silva, and Erlend Magnus Viggen. *The Lattice Boltzmann Method: Principles and Practice*. Graduate Texts in Physics. Springer International Publishing, Cham, 2017.

[23] Jean Pierre Boon and Sidney Yip. *Molecular hydrodynamics*. Courier Corporation, 1991.

[24] Daan Frenkel and Berend Smit. *Understanding molecular simulation: From algorithms to applications*, volume 1. Elsevier, 2001.

[25] Mark E. Tuckerman and Glenn J. Martyna. Understanding modern molecular dynamics: Techniques and applications. *The Journal of Physical Chemistry B*, 104(2):159–178, 2000.

[26] Carlo Cercignani. *The Boltzmann Equation and Its Applications*, volume 67. Springer Science & Business Media, 2012.

[27] Jean-Pierre Rivet and Jean Pierre Boon. *Lattice gas hydrodynamics*. 2001.

[28] L. D. Landau and E. M. Lifshitz. Fluid mechanics. translated from the russian by jb sykes and wh reid. *Course of Theoretical Physics*, 6, 1987.

[29] Dieter A. Wolf-Gladrow. *Lattice-gas cellular automata and lattice Boltzmann models: An introduction*. Springer, 2004.

[30] Gary D. Doolen. *Lattice gas methods for partial differential equations*. CRC Press, 2019.

[31] Stephen Wolfram. Cellular automaton fluids 1: Basic theory. *Journal of statistical physics*, 45(3):471–526, 1986.

[32] Kerson Huang. *Statistical Mechanics, Second Edition*. Wiley, 1987.

[33] Uriel Frisch, Dominique d'Humières, Brosl Hasslacher, Pierre Lallemand, Yves Pomeau, and Jean-Pierre Rivet. Lattice gas hydrodynamics in two and three dimensions. *Complex Systems*, 1:649–707, 1987.

[34] Fabian Bösch and Ilya V. Karlin. Exact lattice Boltzmann equation. *Physical Review Letters*, 111(9):090601, 2013.

[35] M. Reza Parsa and Alexander J. Wagner. Lattice gas with molecular dynamics collision operator. *Physical Review E*, 96(1):013314, July 2017.

[36] Ilya V. Karlin. Tutorial on lattice Boltzmann Method. presented at the DSFD 2018, 2018.

[37] P. L. Bhatnagar, E. P. Gross, and M. Krook. A model for collision processes in gases. I. Small amplitude processes in charged and neutral one-component systems. *Physical Review*, 94(3):511–525, May 1954.

[38] Dominique d'Humières. Multiple-relaxation-time lattice Boltzmann models in three dimensions. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 360(1792):437–451, 2002.

[39] Rui Du, Baochang Shi, and Xingwang Chen. Multi-relaxation-time lattice Boltzmann model for incompressible flow. *Physics Letters A*, 359(6):564–572, 2006.

[40] A. Kuzmin, A. A. Mohamad, and S. Succi. Multi-relaxation time lattice Boltzmann model for multiphase flows. *International Journal of Modern Physics C*, 19(06):875–902, 2008.

[41] Goetz Kaehler and Alexander J. Wagner. Derivation of hydrodynamics for multi-relaxation time lattice Boltzmann using the moment approach. *Communications in Computational Physics*, 13(3):614–628, 2013.

[42] Irina Ginzburg, Frederik Verhaeghe, and Dominique d'Humières. Two-relaxation-time lattice Boltzmann scheme: About parametrization, velocity, pressure and mixed boundary conditions. *Communications in computational physics*, 3(2):427–478, 2008.

[43] Irina Ginzburg, Frederik Verhaeghe, and Dominique d'Humières. Study of simple hydrodynamic solutions with the two-relaxation-times lattice Boltzmann scheme. *Communications in computational physics*, 3(3):519–581, 2008.

[44] Jonas Latt. *Hydrodynamic limit of lattice Boltzmann equations*. PhD thesis, University of Geneva, 2007.

[45] Jonas Latt and Bastien Chopard. Lattice Boltzmann method with regularized pre-collision distribution functions. *Mathematics and Computers in Simulation*, 72(2-6):165–168, 2006.

[46] Raoyang Zhang, Xiaowen Shan, and Hudong Chen. Efficient kinetic method for fluid simulation beyond the Navier-Stokes equation. *Physical Review E*, 74(4):046703, 2006.

[47] Bruce M. Boghosian, Jeffrey Yepez, Peter V. Coveney, and Alexander J. Wager. Entropic lattice Boltzmann methods. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 457(2007):717–766, 2001.

[48] Santosh Ansumali and Ilya V. Karlin. Single relaxation time model for entropic lattice Boltzmann methods. *Physical Review E*, 65(5):056312, 2002.

[49] Shyam S. Chikatamarla, Santosh Ansumali, and Ilya V. Karlin. Entropic lattice Boltzmann models for hydrodynamics in three dimensions. *Physical review letters*, 97(1):010201, 2006.

[50] Nicolò Frapolli, Shyam S. Chikatamarla, and Ilya V. Karlin. Entropic lattice Boltzmann model for gas dynamics: Theory, boundary conditions, and implementation. *Physical Review E*, 93(6):063302, 2016.

[51] Martin Geier, Andreas Greiner, and Jan G. Korvink. Cascaded digital lattice Boltzmann automata for high reynolds number flow. *Physical Review E*, 73(6):066705, 2006.

[52] Yang Ning, Kannan N Premnath, and Dhiraj V Patil. Numerical study of the properties of the central moment lattice Boltzmann method. *International Journal for Numerical Methods in Fluids*, 82(2):59–90, 2016.

[53] Martin Geier, Martin Schönherr, Andrea Pasquali, and Manfred Krafczyk. The cumulant lattice Boltzmann equation in three dimensions: Theory and validation. *Computers & Mathematics with Applications*, 70(4):507–547, 2015.

[54] Martin Geier, Andrea Pasquali, and Martin Schönherr. Parametrization of the cumulant lattice Boltzmann method for fourth order accurate diffusion part I: Derivation and validation. *Journal of Computational Physics*, 348:862–888, 2017.

[55] Martin Geier, Andrea Pasquali, and Martin Schönherr. Parametrization of the cumulant lattice Boltzmann method for fourth order accurate diffusion part II: Application to flow around a sphere at drag crisis. *Journal of Computational Physics*, 348:889–898, 2017.

[56] Anthony J. C. Ladd. Short-time motion of colloidal particles: Numerical simulation via a fluctuating lattice-Boltzmann equation. *Physical Review Letters*, 70(9):1339, 1993.

[57] Alexander J. Wagner. *Theory and applications of the lattice Boltzmann method*. PhD thesis, Theoretical Physics, University of Oxford, 1997.

[58] R. Adhikari, Kyle Stratford, M. E. Cates, and Alexander J. Wagner. Fluctuating lattice Boltzmann. *EPL (Europhysics Letters)*, 71(3):473, 2005.

[59] Burkhard Dünweg, Ulf D. Schiller, and Anthony J. C. Ladd. Statistical mechanics of the fluctuating lattice Boltzmann equation. *Physical Review E*, 76(3):036704, 2007.

[60] Renwei Mei, Dazhi Yu, Wei Shyy, and Li-Shi Luo. Force evaluation in the lattice Boltzmann method involving curved geometry. *Physical Review E*, 65(4):041203, 2002.

[61] Zhaoli Guo, Chuguang Zheng, and Baochang Shi. Discrete lattice effects on the forcing term in the lattice Boltzmann method. *Physical review E*, 65(4):046308, 2002.

[62] AA Mohamad and A Kuzmin. A critical evaluation of force term in lattice Boltzmann method, natural convection problem. *International Journal of Heat and Mass Transfer*, 53(5-6):990–996, 2010.

[63] Shiyi Chen and Gary D. Doolen. Lattice Boltzmann method for fluid flows. *Annual review of fluid mechanics*, 30(1):329–364, 1998.

[64] Cyrus K. Aidun and Jonathan R. Clausen. Lattice-Boltzmann method for complex flows. *Annual review of fluid mechanics*, 42:439–472, 2010.

[65] Xuewen Yin and Junfeng Zhang. An improved bounce-back scheme for complex boundary conditions in lattice Boltzmann method. *Journal of Computational Physics*, 231(11):4295–4303, 2012.

[66] Shyam S. Chikatamarla and Ilya V. Karlin. Entropic lattice Boltzmann method for turbulent flow simulations: Boundary conditions. *Physica A: Statistical Mechanics and its Applications*, 392(9):1925–1930, 2013.

[67] Michael P. Allen and Dominic J. Tildesley. *Computer simulation of liquids*. Oxford university press, 2017.

[68] David F Rogers. *Laminar flow analysis*. Cambridge University Press, 1992.

[69] Sandip Mazumder. *Numerical methods for partial differential equations: finite difference and finite volume methods*. Academic Press, 2015.

[70] Grégoire Allaire. *Numerical analysis and optimization: an introduction to mathematical modelling and numerical simulation*. OUP Oxford, 2007.

[71] Uriel Frisch, Brosl Hasslacher, and Yves Pomeau. Lattice-gas automata for the Navier-Stokes equation. *Physical Review Letters*, 56(14):1505, 1986.

[72] J. Hardy, Yves Pomeau, and O. De Pazzis. Time evolution of a two-dimensional model system. I. Invariant states and time correlation functions. *Journal of Mathematical Physics*, 14(12):1746–1759, 1973.

[73] Daniel H. Rothman and Stiphane Zaleski. *Lattice-gas cellular automata*. 2004.

[74] Michael Hénon. Viscosity of a lattice gas. *Complex Systems*, 1:763–789, 1987.

[75] Bruce M. Boghosian, Jeffrey Yepez, Francis J. Alexander, and Norman H. Margolus. Integer lattice gases. *Physical Review E*, 55:4137–4147, Apr 1997.

[76] Thomas Blommel and Alexander J. Wagner. Integer lattice gas with Monte Carlo collision operator recovers the lattice Boltzmann method with Poisson-distributed fluctuations. *Physical Review E*, 97(2):023310, February 2018.

[77] Guy R. McNamara and Gianluigi Zanetti. Use of the Boltzmann equation to simulate lattice-gas automata. *Physical Review Letters*, 61:2332–2335, Nov 1988.

[78] F. J. Higuera, S. Succi, and R. Benzi. Lattice gas dynamics with enhanced collisions. *EPL (Europhysics Letters)*, 9(4):345, 1989.

[79] F. J. Higuera and Javier Jiménez. Boltzmann approach to lattice gas simulations. *EPL (Europhysics Letters)*, 9(7):663, 1989.

[80] Yue-Hong Qian, Dominique d'Humières, and Pierre Lallemand. Lattice BGK models for Navier-Stokes equation. *EPL (Europhysics Letters)*, 17(6):479, 1992.

[81] Jun Li. Chapman-Enskog expansion in the lattice Boltzmann method. *arXiv preprint arXiv:1512.02599*, 2015.

[82] Alexander J. Wagner. A practical introduction to the lattice Boltzmann method. *Adt. Notes for Statistical Mechanics*, 463:663, 2008.

[83] Aleksandra Pachalieva and Alexander J. Wagner. Non-gaussian distribution of displacements for Lennard-Jones particles in equilibrium. *Physical Review E*, 102:053310, November 2020.

[84] Aleksandra Pachalieva and Alexander J. Wagner. Molecular dynamics lattice gas equilibrium distribution function for Lennard-Jones particles. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379(2208):20200404, 2021.

[85] M. Reza Parsa, Aleksandra Pachalieva, and Alexander J. Wagner. Validity of the molecular-dynamics-lattice-gas global equilibrium distribution function. *International Journal of Modern Physics C*, 30(10):1941007, October 2019.

[86] Aleksandra Pachalieva and Alexander J. Wagner. Connecting lattice Boltzmann methods to physical reality by coarse-graining molecular dynamics simulations. *arXiv preprint arXiv:2109.05009*, 2021.

[87] Donald L. Ermak and J. Andrew McCammon. Brownian dynamics with hydrodynamic interactions. *The Journal of Chemical Physics*, 69(4):1352–1360, 1978.

[88] P. J. Hoogerbrugge and J. M. V. A. Koelman. Simulating microscopic hydrodynamic phenomena with dissipative particle dynamics. *EPL (Europhysics Letters)*, 19(3):155, 1992.

[89] T. Ihle and D. M. Kroll. Stochastic rotation dynamics: A Galilean-invariant mesoscopic model for fluid flow. *Physical Review E*, 63(2):020201, 2001.

[90] M. Reza Parsa and Alexander J. Wagner. Large fluctuations in nonideal coarse-grained systems. *Physical Review Letters*, 124:234501, Jun 2020.