



ACADEMIC
PRESS

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Journal of Computational Physics 187 (2003) 274–297

JOURNAL OF
COMPUTATIONAL
PHYSICS

www.elsevier.com/locate/jcp

Statistical error in particle simulations of hydrodynamic phenomena

Nicolas G. Hadjiconstantinou^{a,*}, Alejandro L. Garcia^{b,1}, Martin Z. Bazant^c,
Gang He^a

^a Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

^b Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, Livermore, CA 94551, USA

^c Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Received 17 July 2002; received in revised form 12 February 2003; accepted 17 February 2003

Abstract

We present predictions for the statistical error due to finite sampling in the presence of thermal fluctuations in molecular simulation algorithms. Specifically, we establish how these errors depend on Mach number, Knudsen number, number of particles, etc. Expressions for the common hydrodynamic variables of interest such as flow velocity, temperature, density, pressure, shear stress, and heat flux are derived using equilibrium statistical mechanics. Both volume-averaged and surface-averaged quantities are considered. Comparisons between theory and computations using direct simulation Monte Carlo for dilute gases, and molecular dynamics for dense fluids, show that the use of equilibrium theory provides accurate results.

© 2003 Elsevier Science B.V. All rights reserved.

Keywords: Statistical error; Sampling; Fluctuations; Monte Carlo; Molecular dynamics; Hydrodynamics

1. Introduction

Recently much attention has been focused on the simulation of hydrodynamic problems at small scales using molecular simulation methods such as Molecular Dynamics (MD) [1,2] or the direct simulation Monte Carlo (DSMC) [3,4]. Molecular Dynamics is generally used to simulate liquids while DSMC is a very efficient algorithm for simulating dilute gases. In molecular simulation methods the connection to macroscopic observable fields, such as velocity and temperature, is achieved through averaging appropriate microscopic properties. The simulation results are therefore inherently statistical and statistical errors due to finite sampling need to be fully quantified.

* Corresponding author.

E-mail address: ngh@mit.edu (N.G. Hadjiconstantinou).

¹ Permanent address: Department of Physics, San Jose State University, San Jose, CA 95192-0106, USA.

Though confidence intervals may be estimated by measuring the variance of these sampled quantities, this additional computation can be burdensome and thus is often omitted. Furthermore, it would be useful to estimate confidence intervals a priori so that one could predict the computational effort required to achieve a desired level of accuracy. For example, it is well known that obtaining accurate hydrodynamic fields (e.g., velocity profiles) is computationally expensive in low Mach number flows so it is useful to have an estimate of the computational effort required to reach the desired level of accuracy.

In this paper, we present expressions for the magnitude of statistical errors due to thermal fluctuations in molecular simulations for the typical observables of interest, such as velocity, density, temperature, and pressure. We also derive expressions for the shear stress and heat flux in the dilute gas limit. Both volume averaging and flux averaging is considered. Although we make use of expressions from equilibrium statistical mechanics, the non-equilibrium modifications to these results are very small, even under extreme conditions [5]. This is verified by the good agreement between our theoretical expressions and the corresponding measurements in our simulations.

In addition to direct measurements of hydrodynamic fields, this analysis will benefit algorithmic applications by providing a framework where statistical fluctuations can be correctly accounted for. One example of such application is the measurement of temperature for temperature-dependent collision rates [15]; another example is the measurement of velocity and temperature for the purpose of imposing boundary conditions [18]. Additional examples include hybrid methods [11,14] where the coupling between continuum and molecular fields requires both averaging of finite numbers of particles in a sequence of molecular realizations as well as the generation of molecular realizations, based on continuum fields, with relatively small numbers of particles (e.g., buffer cells).

In Section 2 the theoretical expressions for the statistical error due to thermodynamic fluctuations are derived. These expressions are verified by molecular simulations, as described in Section 3. The effect of correlations between samples in dilute gases is briefly discussed in Section 4 and concluding remarks appear in Section 5.

2. Statistical error due to thermal fluctuations

2.1. Volume-averaged quantities

We first consider the fluid velocity. In a particle simulation, the flow field is obtained by measuring the instantaneous center of mass velocity, \mathbf{u} , for particles in a statistical cell volume. The statistical mean value of the local fluid velocity, $\langle \mathbf{u} \rangle_s$, is estimated over M independent samples. For steady flows, these may be sequential samples taken in time; for transient flows these may be samples from an ensemble of realizations. The average fluid velocity, $\langle \mathbf{u} \rangle$, is defined such that $\langle \mathbf{u} \rangle_s \rightarrow \langle \mathbf{u} \rangle$ as $M \rightarrow \infty$; for notational convenience we also write $\langle \mathbf{u} \rangle = \mathbf{u}_0$. Define $\delta u_x \equiv u_x - u_{x0}$ to be the instantaneous fluctuation in the x -component of the fluid velocity; note that all three components are equivalent. From equilibrium statistical mechanics [6],

$$\langle (\delta u_x)^2 \rangle = \frac{kT_0}{mN_0} = \frac{a^2}{\gamma Ac^2 N_0}, \tag{1}$$

where N_0 is the average number of particles in the statistical cell, T_0 is the average temperature, m is the particle mass, k is Boltzmann’s constant, a is the sound speed, and $\gamma = c_p/c_v$ is the ratio of the specific heats. The acoustic number $Ac = a/a^i$ is the ratio of the fluid’s sound speed to the sound speed of a “reference” ideal gas at the same temperature

$$a^i = \sqrt{\gamma kT_0/m}. \tag{2}$$

Note that this reference ideal gas has a ratio of specific heats (γ^i) equal to the *original fluid* specific heat ratio, that is $\gamma^i = \gamma$ as shown in Eq. (2).

We may define a “signal-to-noise” ratio as the average fluid velocity over its standard deviation; from the above

$$\frac{|u_{x0}|}{\sqrt{\langle(\delta u_x)^2\rangle}} = Ac Ma \sqrt{\gamma N_0}, \quad (3)$$

where $Ma = |u_{x0}|/a$ is the local Mach number based on the velocity component of interest. This result shows that for fixed Mach number, in a dilute gas simulation ($Ac = 1$), the statistical error due to thermal fluctuations cannot be ameliorated by reducing the temperature. However, when the Mach number is small enough for compressibility effects to be negligible, favorable relative statistical errors may be obtained by performing simulations at an increased Mach number (to a level where compressibility effects are still negligible).

The one-standard-deviation error bar for the sample estimate $\langle u_x \rangle_s$ is $\sigma_u = \sqrt{\langle(\delta u_x)^2\rangle}/\sqrt{M}$ and the fractional error in the estimate of the fluid velocity is

$$E_u = \frac{\sigma_u}{|u_{x0}|} = \frac{1}{\sqrt{MN_0}} \frac{1}{Ac Ma \sqrt{\gamma}}, \quad (4)$$

yielding

$$M = \frac{1}{\gamma Ac^2 N_0 Ma^2 E_u^2}. \quad (5)$$

For example, with $N_0 = 100$ particles in a statistical cell, if a one percent fractional error is desired in a $Ma = 1$ flow, about $M = 100$ independent statistical samples are required (assuming $Ac \approx 1$). However, for a $Ma = 10^{-2}$ flow, about 10^6 independent samples are needed, which quantifies the empirical observation that the resolution of the flow velocity is computationally expensive for low Mach number flows.

Next we turn our attention to the density. From equilibrium statistical mechanics, the fluctuation in the number of particles in a cell is

$$\langle(\delta N)^2\rangle = -N^2 \frac{kT_0}{V^2} \left(\frac{\partial V}{\partial P} \right)_T = \kappa_T N_0^2 \frac{kT_0}{V}, \quad (6)$$

where V is the volume of the statistics cell and $\kappa_T \equiv -V^{-1}(\partial V/\partial P)_T$ is the isothermal compressibility. Note that for a dilute gas $\kappa_T = 1/P$ so $\langle(\delta N)^2\rangle = N$ and, in fact, N is Poisson random variable. The fractional error in the estimate of the density is

$$E_\rho = \frac{\sigma_\rho}{\rho_0} = \frac{\sigma_N}{N_0} = \frac{\sqrt{\langle(\delta N)^2\rangle}}{N_0 \sqrt{M}} = \frac{\sqrt{\kappa_T kT_0}}{\sqrt{MV}} = \frac{\sqrt{\kappa_T/\kappa_T^i}}{\sqrt{MN_0}}, \quad (7)$$

where $\kappa_T^i = V/N_0 kT_0$ is the isothermal compressibility of the reference dilute gas ($\gamma^i = \gamma$) at the same density and temperature. Since $a \propto 1/\sqrt{\kappa_T}$,

$$E_\rho = \frac{1}{\sqrt{MN_0}} \frac{1}{Ac}. \quad (8)$$

Note that for fixed M and N_0 , the error decreases as the compressibility decreases (i.e., as the sound speed increases) since the density fluctuations are smaller.

Let us now consider the measurement of temperature. First, we should remark that the measurement of instantaneous temperature is subtle, even in a dilute gas. But given that temperature is measured correctly, equilibrium statistical mechanics gives the variance in the temperature fluctuations to be

$$\langle(\delta T)^2\rangle = \frac{kT_0^2}{c_V N_0} \tag{9}$$

where c_V is the heat capacity per particle at constant volume. The fractional error in the estimate of the temperature is

$$E_T = \frac{\sigma_T}{T_0} = \frac{\sqrt{\langle(\delta T)^2\rangle}}{T_0\sqrt{M}} = \frac{1}{\sqrt{MN_0}} \sqrt{\frac{k}{c_V}}. \tag{10}$$

Because the fluctuations are smaller, the error in the temperature is smaller when the heat capacity is large. Note that the temperature associated with various degrees of freedom (translational, vibrational, rotational) may be separately defined and measured. For example, if we consider only the measurement of the translational temperature, then the appropriate heat capacity is that of an ideal gas with three degrees of freedom, i.e., $c_V = (3/2)k$, corresponding to the three translational components.

Finally, the variance in the pressure fluctuations is

$$\langle(\delta P)^2\rangle = -kT_0 \left(\frac{\partial P}{\partial V}\right)_S = \frac{\gamma kT_0}{V\kappa_T} \tag{11}$$

so the fractional error in the estimate of the pressure is

$$E_P = \frac{\sigma_P}{P_0} = \frac{\sqrt{\langle(\delta P)^2\rangle}}{P_0\sqrt{M}} = \frac{P_0^i}{P_0} \frac{Ac\sqrt{\gamma}}{\sqrt{MN_0}}, \tag{12}$$

where $P_0^i = N_0kT_0/V$ is the pressure of an ideal gas under the same conditions. Note that the error in the pressure is proportional to the acoustic number while the error in the density, Eq. (8), goes as Ac^{-1} .

2.2. Shear stress and heat flux for dilute gases

The thermodynamic results in the previous section are general; in this section we consider transport quantities and restrict our analysis to dilute gases. In a dilute gas, the shear stress and heat flux are defined as

$$\tau_{xy0} = \langle\tau_{xy}\rangle = \langle\rho c_x c_y\rangle \tag{13}$$

and

$$q_{x0} = \langle q_x \rangle = \left\langle \frac{1}{2} \rho c_x c^2 \right\rangle, \tag{14}$$

respectively, where \underline{c} is the molecular velocity. Under the assumption of a dilute gas, the fluctuation in the (equilibrium) shear stress and heat flux in a volume V containing N_0 particles can be calculated using the Maxwell–Boltzmann distribution. Note that in equilibrium the expected values of the shear stress and the heat flux are zero.

Using the definitions of shear stress and heat flux in terms of moments of the velocity distribution, direct calculation of the variance of the x – y component of the stress tensor based on a single-particle distribution function gives

$$\langle \tau_{xy}^2 \rangle = \langle (\rho c_x c_y)^2 \rangle, \quad (15)$$

$$= \rho_0^2 \langle (c_x c_y)^2 \rangle, \quad (16)$$

$$= \frac{1}{4} \rho_0^2 \left(\frac{2kT_0}{m} \right)^2 = P_0^2. \quad (17)$$

In obtaining the second equation we used the fact that $\langle c_x \rangle = \langle c_y \rangle = \langle (\rho - \rho_0) c_x c_y \rangle = 0$. For the x component of the heat flux vector, we find

$$\langle q_x^2 \rangle = \left\langle \left(\frac{1}{2} \rho c_x c^2 \right)^2 \right\rangle, \quad (18)$$

$$= \rho_0^2 \left\langle \left(\frac{1}{2} c_x c^2 \right)^2 \right\rangle, \quad (19)$$

$$= \frac{35}{32} \rho_0^2 \left(\frac{2kT_0}{m} \right)^3 = \frac{35}{8} c_m^2 P_0^2, \quad (20)$$

where $c_m = \sqrt{2kT_0/m}$ is the most probable particle speed and we have used the fact that $\langle c_x \rangle = \langle c_y \rangle = \langle (\rho - \rho_0) c^2 \rangle = 0$. Note that in equilibrium for a cell containing N_0 particles, the variance of the mean is given by $\langle (\delta\tau_{xy})^2 \rangle = \langle \tau_{xy}^2 \rangle / N_0$ and $\langle (\delta q_x)^2 \rangle = \langle q_x^2 \rangle / N_0$ for the shear stress and heat flux, respectively.

In order to derive expressions for the relative fluctuations we need expressions for the magnitude of the fluxes. We are only able to provide closed form expressions for the latter in the continuum regime where

$$\tau_{xy} = \mu \left(\frac{\partial u_x}{\partial y} + \frac{\partial u_y}{\partial x} \right) \quad (21)$$

and

$$q_x = -\kappa \frac{\partial T}{\partial x}, \quad (22)$$

where μ is the coefficient of viscosity and κ is the thermal conductivity. Above Knudsen numbers of $Kn \approx 0.1$, it is known that these continuum expressions fail and more accurate results can be obtained using more elaborate formulations from kinetic theory. Here, the Knudsen number is defined as $Kn = \lambda/\ell$, where λ is a viscosity-based mean free path

$$\lambda = \frac{8}{5\sqrt{\pi}} \frac{c_m \mu}{P_0} \quad (23)$$

and ℓ is a characteristic length scale. Note that this expression for the mean free path simplifies to the hard sphere result when the viscosity is taken to be that of hard spheres.

Using (21) and (22) we find that in *continuum flows*, the relative fluctuations in the shear stress and heat flux are given by

$$E_\tau = \frac{\sqrt{\langle (\delta\tau_{xy})^2 \rangle}}{|\tau_{xy0}| \sqrt{M}} = \frac{16}{5\sqrt{2\pi\gamma}} \frac{1}{Kn Ma_*} \frac{1}{\sqrt{N_0 M}} \left| \frac{\partial u_x^*}{\partial y^*} + \frac{\partial u_y^*}{\partial x^*} \right|^{-1} \quad (24)$$

and

$$E_q = \frac{\sqrt{\langle(\delta q_x)^2\rangle}}{|q_{x0}|\sqrt{M}} = \frac{8\sqrt{35}}{5\sqrt{2\pi}} \frac{Pr(\gamma-1)}{\gamma} \frac{T}{\Delta T} \frac{1}{Kn} \frac{1}{\sqrt{N_0M}} \left| \frac{\partial T^*}{\partial x^*} \right|^{-1}, \tag{25}$$

respectively. Here, stars denote non-dimensional quantities: $u^* = u_0/\tilde{u}$, $T^* = T/\Delta T$, and $x^* = x/\ell$, where \tilde{u} and ΔT are a characteristic velocity and temperature variation, while Pr is the Prandtl number. The Mach number Ma_* is defined with respect to the characteristic velocity \tilde{u} rather than the local velocity used to define Ma .

If viscous heat generation is responsible for the temperature differences characterized by ΔT , then it is possible to express Eq. (25) in the following form:

$$E_q = \frac{\sqrt{\langle(\delta q_x)^2\rangle}}{|q_{x0}|\sqrt{M}} = \frac{8\sqrt{35}}{5\gamma\sqrt{2\pi}} \frac{Br}{KnMa_*^2} \frac{1}{\sqrt{N_0M}} \left| \frac{\partial T^*}{\partial x^*} \right|^{-1}. \tag{26}$$

The Brinkman number,

$$Br = \frac{\mu\tilde{u}^2}{\kappa\Delta T}, \tag{27}$$

is the relevant non-dimensional group that compares temperature differences due to viscous heat generation to the characteristic temperature differences in the flow. (It follows that if viscous heat generation is responsible for the temperature changes, $Br \approx 1$.)

It is very instructive to extend the above analysis to Eq. (10). If we define the relative error in temperature with respect to the temperature changes rather than the absolute temperature, we obtain

$$E_{\Delta T} = \frac{\sigma_T}{\Delta T} = \frac{\sqrt{\langle(\delta T)^2\rangle}}{\Delta T\sqrt{M}} = \frac{T_0}{\Delta T\sqrt{MN_0}} \sqrt{\frac{k}{c_V}}, \tag{28}$$

$$= \frac{1}{Pr(\gamma-1)} \frac{Br}{Ma_*^2} \frac{1}{\sqrt{MN_0}} \sqrt{\frac{k}{c_V}}, \tag{29}$$

where again, if viscous heat generation is the only source of heat, $Br \approx 1$. The above development shows that resolving the temperature *differences* or heat flux due to viscous heat generation in the continuum regime is very computationally inefficient for low speed flow since for a given expected error $E_{\Delta T}$ we find that the number of samples scales as $M \propto Ma_*^{-4}$.

Comparison of Eqs. (4) and (24) and Eqs. (26) and (29) reveals that, in the continuum regime,

$$E_\tau \sim \frac{E_u}{Kn} \tag{30}$$

and

$$E_q \sim \frac{E_{\Delta T}}{Kn} \tag{31}$$

since the non-dimensional gradients will be of order one. As the above equations were derived for the continuum regime ($Kn < 0.1$), it follows that the relative error in these moments is significantly higher. This will also be shown to be the case in the next section when the shear stress and heat flux are evaluated as fluxal (surface) quantities. This has important consequences in hybrid methods [11,14], as coupling in terms of state (Dirichlet) conditions is subject to less variation than coupling in terms of flux conditions.

2.3. Fluxal quantities

In this section, we predict the relative errors in the fluxes of mass and heat as well as the components of the stress tensor, when calculated as fluxes across a reference surface. Our analysis is based on the assumption of an infinite, ideal gas in equilibrium. In this case, it is well known that the number of particles, N_i , in any infinitesimal cell i with dimensions $\Delta x_i, \Delta y_i, \Delta z_i$ is an independent Poisson random variable with mean and variance

$$\langle N_i \rangle = \langle (\delta N_i)^2 \rangle = n \Delta x_i \Delta y_i \Delta z_i, \tag{32}$$

where $n = \rho/m$ is the particle number density.

In Appendix A, it is shown that, more generally, all number fluctuations are independent and Poisson distributed. In particular, the number of particles, N_{ij}^+ , leaving a cell i at position (x_i, y_i, z_i) , crossing a surface, and arriving in another cell j at position (x_j, y_j, z_j) after a time Δt is also an independent Poisson random variable. Its mean and variance are given by

$$\langle N_{ij}^+ \rangle = \langle (\delta N_{ij}^+)^2 \rangle = P_{ij} \langle N_i \rangle, \tag{33}$$

where

$$P_{ij} = P(x_{ij}, y_{ij}, z_{ij}) \Delta x_j \Delta y_j \Delta z_j \tag{34}$$

is the probability that a particle makes the transition, which depends only upon the relative displacements, $x_{ij} = x_j - x_i$, $y_{ij} = y_j - y_i$, and $z_{ij} = z_j - z_i$. These properties hold exactly for an infinite ideal gas in equilibrium (see Fig. 1), but they are also excellent approximations for most finite, dilute gases, even far from equilibrium, e.g., as shown by the DSMC result in Fig. 2.

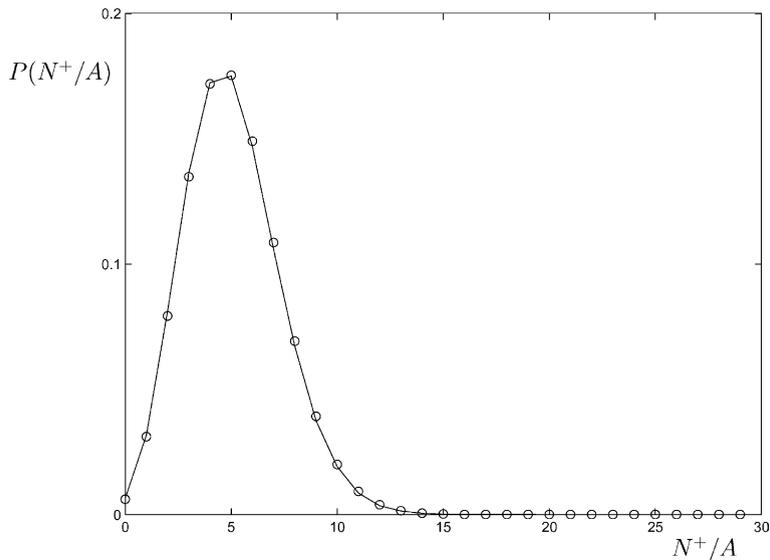


Fig. 1. Probability distribution of number of particles per unit area crossing a surface in the gas in one direction in time $\Delta t \rightarrow 0$. The circles denote equilibrium DSMC results and the solid line shows a Poisson distribution with mean based on the equilibrium result $\langle N^+ / A \rangle = n c_m / (2\sqrt{\pi}) \Delta t$.

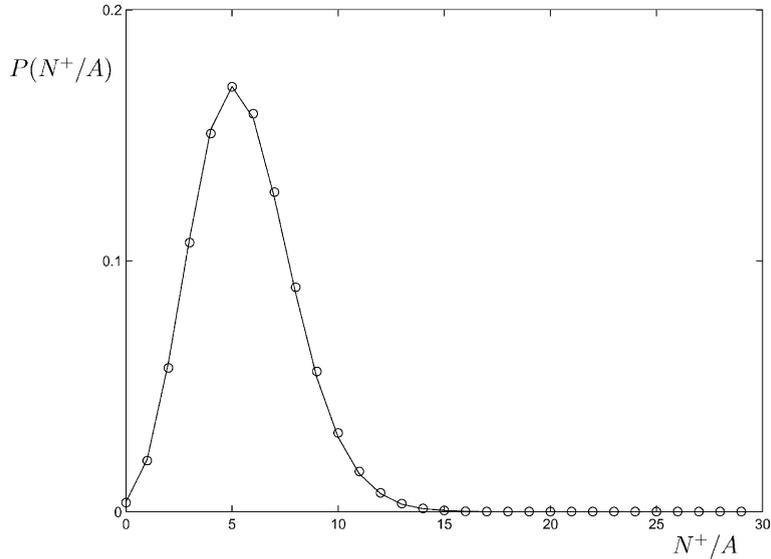


Fig. 2. Probability distribution of number of particles crossing a surface in the gas in one direction in time $\Delta t \rightarrow 0$, in the presence of mean flow normal to the surface of interest. The circles denote DSMC results for shear flow and the solid line shows a Poisson distribution with mean based on the particle flux as calculated by Eq. (40) by taking account of the local mean velocity due to the shear flow and an equilibrium velocity distribution.

Without loss of generality we place a large planar surface of area A at $x = 0$ (see Fig. 3). The flux in the positive direction, J_h^+ , of any velocity-dependent quantity, $h(c_x, c_y, c_z)$, can be expressed as a sum of independent, random contributions from different cells i and j on opposite sides of the surface

$$J_h^+ = \frac{1}{A\Delta t} \sum_{x_i < 0} \sum_{x_j > 0} \sum_{y_i} \sum_{y_j} \sum_{z_i} \sum_{z_j} \bar{h}(x_{ij}, y_{ij}, z_{ij}) N_{ij}^+, \tag{35}$$

where $\bar{h}(x_{ij}, y_{ij}, z_{ij}) = h(c_x, c_y, c_z)$ is the same quantity expressed in terms of relative displacements during the time interval, Δt . We begin by calculating the mean flux. Taking expectations in Eq. (35), we obtain

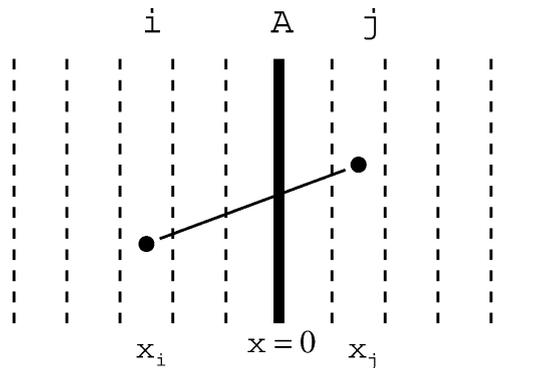


Fig. 3. Schematic particle motion crossing surface A .

$$\langle J_h^+ \rangle = \frac{1}{A\Delta t} \sum_{x_i < 0} \sum_{x_j > 0} \sum_{y_i} \sum_{y_j} \sum_{z_i} \sum_{z_j} \bar{h}(x_{ij}, y_{ij}, z_{ij}) \langle N_i \rangle P_{ij}, \tag{36}$$

$$\sim \frac{1}{\Delta t} \sum_{x_i < 0} \sum_{x_j > 0} \sum_{y_j} \sum_{z_j} \bar{h}(x_{ij}, y_j, z_j) n \Delta x_i P(x_{ij}, y_j, z_j) \Delta x_j \Delta y_j \Delta z_j, \tag{37}$$

where without loss of generality we have taken $y_i = z_i = 0$ and also separately performed the sums parallel to the surface,

$$\sum_{y_i} \sum_{z_i} \Delta y_i \Delta z_i = A,$$

by invoking translational invariance and neglecting any edge effects, since $\langle c_y \rangle \Delta t, \langle c_z \rangle \Delta t = O(\sqrt{A})$ in the limit $\Delta t \rightarrow 0$ (taken below).

Passing to the continuum limit in Eq. (37) and letting $A \rightarrow \infty$, we arrive at an integral expression for the mean flux through an infinite flat surface

$$\begin{aligned} \langle J_h^+ \rangle &= \frac{n}{\Delta t} \int_{-\infty}^0 dx' \int_0^{\infty} dx \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} dz \bar{h}(x - x', y, z) P(x - x', y, z) \\ &= \frac{n}{\Delta t} \int_0^{\infty} dx \int_x^{\infty} ds \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} dz \bar{h}(s, y, z) P(s, y, z). \end{aligned} \tag{38}$$

Since the integrand does not depend on x , switching the order of integration

$$\begin{aligned} \langle J_h^+ \rangle &= \frac{n}{\Delta t} \int_0^{\infty} ds \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} dz \bar{h}(s, y, z) P(s, y, z) \int_0^s dx \\ &= n \int_0^{\infty} ds \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} dz \bar{h}(s, y, z) P(s, y, z) \frac{s}{\Delta t} \end{aligned} \tag{39}$$

produces a simple formula for the mean flux as a conditional average over the velocity distribution (with $c_x > 0$),

$$\langle J_h^+ \rangle = n \langle c_x h(c_x, c_y, c_z) \rangle^+ \tag{40}$$

in the limit $\Delta t \rightarrow 0$. For $h(c_x, c_y, c_z) = m, mc_x, mc_y, mc_z, (1/2)mc^2$ in the absence of a mean flow, we obtain:

$$\langle J_m^+ \rangle = n \langle c_x m \rangle^+ = \frac{\rho c_m}{2\sqrt{\pi}}, \tag{41}$$

$$\langle J_{xx}^+ \rangle = n \langle c_x mc_x \rangle^+ = \frac{\rho k T_0}{2m} = \frac{P_0}{2}, \tag{42}$$

$$\langle J_{xy}^+ \rangle = n \langle c_x mc_y \rangle^+ = 0, \quad \langle J_{xz}^+ \rangle = n \langle c_x mc_z \rangle^+ = 0, \tag{43}$$

$$\langle J_e^+ \rangle = n \left\langle c_x \frac{1}{2} mc^2 \right\rangle^+ = \frac{\rho c_m^3}{2\sqrt{\pi}}, \tag{44}$$

which are the well-known results for the expected one-sided fluxes of mass, momentum, and energy, respectively.

Using the same formalism as above, we now derive a general formula for the variance of the flux, $\langle(\delta J_h^+)^2\rangle$. For simplicity, we assume that the mean flow is zero normal to the reference surface, but our results below are easily extended to the case of a non-zero mean flow through the surface since, as shown in Fig. 2, Eq. (33) still holds in that case. Taking the variance of Eq. (35) and using the independence of $\{N_{ij}^+\}$, we obtain

$$\begin{aligned} \langle(\delta J_h^+)^2\rangle &= \frac{1}{A^2\Delta t^2} \sum_{x_i < 0} \sum_{x_j > 0} \sum_{y_i} \sum_{y_j} \sum_{z_i} \sum_{z_j} \bar{h}(x_{ij}, y_{ij}, z_{ij})^2 \langle(\delta N_{ij}^+)^2\rangle \\ &= \frac{1}{A^2\Delta t^2} \sum_{x_i < 0} \sum_{x_j > 0} \sum_{y_i} \sum_{y_j} \sum_{z_i} \sum_{z_j} \bar{h}(x_{ij}, y_{ij}, z_{ij})^2 \langle N_i \rangle P_{ij}. \end{aligned} \tag{45}$$

Following the same steps above leading from Eq. (36) to Eq. (40) we arrive at the simple formula

$$\langle(\delta J_h^+)^2\rangle = \frac{n}{A\Delta t} \langle c_x h(c_x, c_y, c_z)^2 \rangle^+ \tag{46}$$

in the continuum limit as $\Delta t \rightarrow 0$. The variance of the total flux is

$$\langle(\delta J_h)^2\rangle = \langle(\delta J_h^+ + \delta J_h^-)^2\rangle = \langle(\delta J_h^+)^2\rangle + \langle(\delta J_h^-)^2\rangle = 2\langle(\delta J_h^+)^2\rangle \tag{47}$$

since the one-sided fluxes through the surface, J_h^+ and J_h^- , are independent and identically distributed (with opposite sign).

Using this general result, we can evaluate the standard deviations of the fluxes above:

$$\sqrt{\langle(\delta J_m)^2\rangle} = \sqrt{\frac{\rho c_m}{\sqrt{\pi}}} \sqrt{\frac{m}{A\Delta t}}, \tag{48}$$

$$\sqrt{\langle(\delta J_{xy})^2\rangle} = \sqrt{\frac{\rho c_m^3}{2\sqrt{\pi}}} \sqrt{\frac{m}{A\Delta t}}, \tag{49}$$

$$\sqrt{\langle(\delta J_e)^2\rangle} = \sqrt{\frac{3\rho c_m^5}{2\sqrt{\pi}}} \sqrt{\frac{m}{A\Delta t}}. \tag{50}$$

These formulae may be simplified by noting that

$$\langle J_m^+ \rangle = \frac{mN^+}{A\Delta t} = \frac{\rho c_m}{2\sqrt{\pi}}, \tag{51}$$

where N^+ is the mean total number of particles crossing the reference surface in one direction in time Δt , which yields:

$$\sqrt{\langle(\delta J_m)^2\rangle} = \frac{\rho c_m}{\sqrt{2\pi}} \frac{1}{\sqrt{N^+}}, \tag{52}$$

$$\sqrt{\langle(\delta J_{xy})^2\rangle} = \frac{\rho c_m^2}{2\sqrt{\pi}} \frac{1}{\sqrt{N^+}}, \tag{53}$$

$$\sqrt{\langle(\delta J_e)^2\rangle} = \sqrt{\frac{3}{\pi}} \frac{\rho c_m^3}{2} \frac{1}{\sqrt{N^+}}. \tag{54}$$

We may relate these to the results from the previous section by identifying $\langle(\delta \tau_{xy}^f)^2\rangle = \langle(\delta J_{xy})^2\rangle$ and $\langle(\delta q_x^f)^2\rangle = \langle(\delta J_e)^2\rangle$, and additionally include the effect of M (independent) samples in time. The superscript f

denotes fluxal measurement. By noting that transport fluxes are defined with respect to the rest frame of the fluid, it can be easily verified that the above relations hold in the case where a mean flow in directions parallel to the measuring surface exists, under the assumption of a local equilibrium distribution.

We can derive expressions for the relative expected error in the continuum regime in which models exist for the shear stress and heat flux. In this regime we find

$$E_{\tau}^f = \frac{\sqrt{\langle (\delta \tau_{xy}^f)^2 \rangle}}{|\tau_{xy0}^f| \sqrt{M}} = \frac{16}{5\pi\sqrt{2}\gamma} \frac{1}{KnMa_*} \frac{1}{\sqrt{MN^+}} \left| \frac{\partial u^*}{\partial y^*} + \frac{\partial v^*}{\partial x^*} \right|^{-1} \quad (55)$$

and

$$E_q^f = \frac{\sqrt{\langle (\delta q_x^f)^2 \rangle}}{|q_{x0}^f| \sqrt{M}} = \frac{16\sqrt{3}}{5\pi} \frac{Pr(\gamma-1)}{\gamma} \frac{T}{\Delta T} \frac{1}{Kn} \frac{1}{\sqrt{MN^+}} \left| \frac{\partial T}{\partial x^*} \right|^{-1}, \quad (56)$$

$$= \frac{16\sqrt{3}}{5\gamma\pi} \frac{Br}{KnMa_*^2} \frac{1}{\sqrt{MN^+}} \left| \frac{\partial T}{\partial x^*} \right|^{-1}. \quad (57)$$

Comparing (55) with the corresponding expressions for volume-averaged stress tensor, (24), one finds that, aside from the numerical coefficients, the expressions differ only in the number of particles used, either N^+ or N_0 ; one finds a similar result for the heat flux.

2.4. Connection to fluctuating hydrodynamics

Fluctuating hydrodynamics as developed by Landau approximates the stress tensor and heat flux as white noises, with variances fixed by the equilibrium fluctuations. In this section, we identify the connection between Landau's theory and the variances of fluxes obtained in Section 2.2.

Landau introduced fluctuations into the hydrodynamic equations by adding white noise terms to the stress tensor and heat flux [16] (in the spirit of Langevin's theory of Brownian motion). The amplitudes of these noises are fixed by evaluating the resulting variances of velocity and temperature and matching those with the results from equilibrium statistical mechanics [17]. For example, in Landau's formulation the total heat flux in the x -direction is $q_x^L = -\kappa \partial T / \partial x + g_x$ where the first term on the r.h.s. is the deterministic part of the flux and the second is the white noise term. The latter has mean zero and time correlation given by the following expression:

$$\langle g_x(t) g_x(t') \rangle = \frac{2k\kappa T^2}{V} \delta(t - t'). \quad (58)$$

Note that at a steady state the deterministic part is constant so $\langle \delta q_x^L(t) \delta q_x^L(t') \rangle = \langle g_x(t) g_x(t') \rangle$. On the other hand, recall from Eq. (20),

$$\langle (\delta q_x)^2 \rangle = \frac{35}{32} \frac{\rho^2 c_m^6}{N_0}. \quad (59)$$

The question naturally arises: how does one reconcile (58) and (59)? Note that the fluctuating hydrodynamics expression contains the thermal conductivity, which depends on the particle interaction (e.g., for hard-spheres κ depends on the particle diameter) while the kinetic theory expression is independent of this interaction.

The key lies in identifying the δ -function with a decay time, t_d , that is vanishingly small at hydrodynamic scales. Specifically, we may write

$$\delta(t - t') \rightarrow \begin{cases} t_d^{-1}, & |t - t'| < t_d, \\ 0 & \text{otherwise} \end{cases} \tag{60}$$

so

$$\langle (g_x(t))^2 \rangle = \frac{2k\kappa T^2}{t_d V} = \frac{\kappa V}{2kN_0 c_m^2 t_d} \frac{\rho^2 c_m^6}{N_0}. \tag{61}$$

Comparing with the above gives

$$t_d = \frac{16}{35} \frac{\kappa V}{kN_0 c_m^2}. \tag{62}$$

For a hard sphere gas, $\kappa = (5/2)c_V\mu$ so using (23) we may write this as

$$t_d = \frac{15\sqrt{\pi}}{28} \frac{\lambda}{c_m}. \tag{63}$$

For other particle interactions the coefficients will be slightly different but in general $t_d \approx \lambda/c_m$, thus it is approximately equal to the molecular collision time. In conclusion, the two formulations are compatible once the white noise approximation in fluctuating hydrodynamics is justified by the fact that the hydrodynamic time scale is much longer than the kinetic (i.e., collisional) time scale. Landau’s construction provides a useful hydrodynamic approximation for g_x but (59) is the actual variance of the heat flux.

3. Simulations

3.1. Dilute gases

We performed DSMC simulations to verify the validity of the expressions derived above. Standard DSMC techniques [3,4] were used to simulate flow of gaseous argon (molecular mass $m = 6.63 \times 10^{-26}$ kg, hard sphere diameter $\sigma = 3.66 \times 10^{-10}$ m) in a two-dimensional channel (length L and height H). The simulation was periodic in the x -direction (along the channel axis). The two walls at $y = -H/2$ and $y = H/2$ were fully accommodating and flat. The simulation was also periodic in the third (homogeneous) direction.

The average gas density was $\rho_0 = 1.78$ kg/m³ and in all calculations over 40 particles per cell were used. The cell size was $\Delta x = \lambda_0/3$ where λ_0 is the reference mean free path. The time step was $\Delta t = \lambda_0/(7.5c_m)$. For a discussion of the errors resulting from finite cell sizes and time steps, see [8–10]. The fractional error in the simulations is obtained from the standard deviation of cell values in the x - and z -directions. To ensure that the samples were independent, samples were taken only once every 250 time steps. To ensure that the system was in its steady state the simulation was run for 10^6 time steps before sampling was started.

A constant acceleration was applied to the particles to produce Poiseuille flow in the x -direction with maximum velocity at the centerline $u_0^{\max} \approx 2$ m/s. Figs. 4–6 show good agreement between the theoretical expressions from Section 2 and simulation measurements for the fractional error in velocity, density, and temperature, respectively. The fractional error in the velocity measurement is minimum at the centerline since the Poiseuille velocity profile is parabolic and maximum at the centerline (see Fig. 4). The density and temperature were nearly constant across the system so the fractional errors in these quantities are also nearly constant.

The expressions for shear stress and heat flux were verified using Couette (walls at equal temperature with different velocities) and “Temperature Couette” (walls at zero velocity with different temperatures) calculations, respectively. In these calculations, very small cell sizes ($\Delta x = \lambda/6$) and time steps

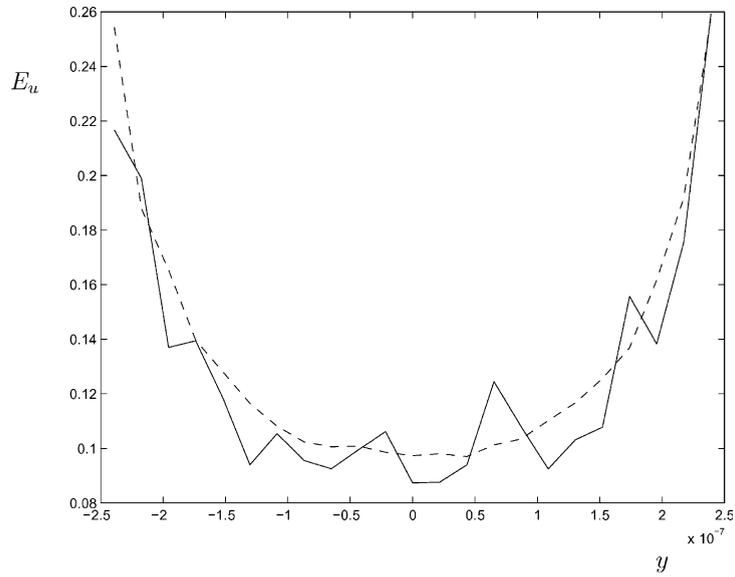


Fig. 4. Fractional error in velocity for Poiseuille flow in a channel as a function of the transverse channel coordinate, y (measured in meters). The dashed line denotes Eq. (4) and the solid line denotes DSMC simulation results.

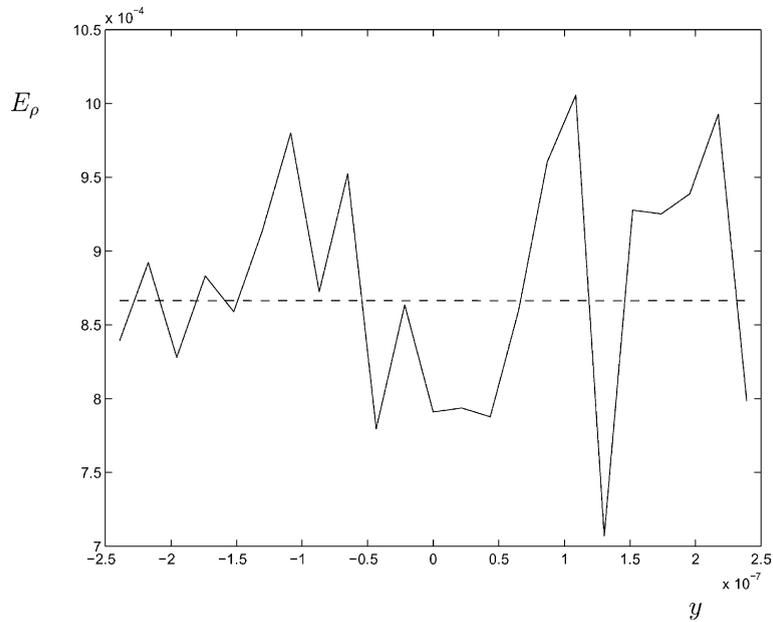


Fig. 5. Fractional error in density for Poiseuille flow in a channel as a function of the transverse channel coordinate, y (measured in meters). The dashed line denotes Eq. (8) and the solid line indicates DSMC simulation results.

($\Delta t = \lambda/(30c_m)$) were used. The system was equilibrated for 10^6 time steps and samples were taken every 50 time steps. The momentum and energy fluxes de-correlate faster than the conserved hydrodynamic variables, such as density, so independent samples are obtained after fewer time steps (see Section 4). Good

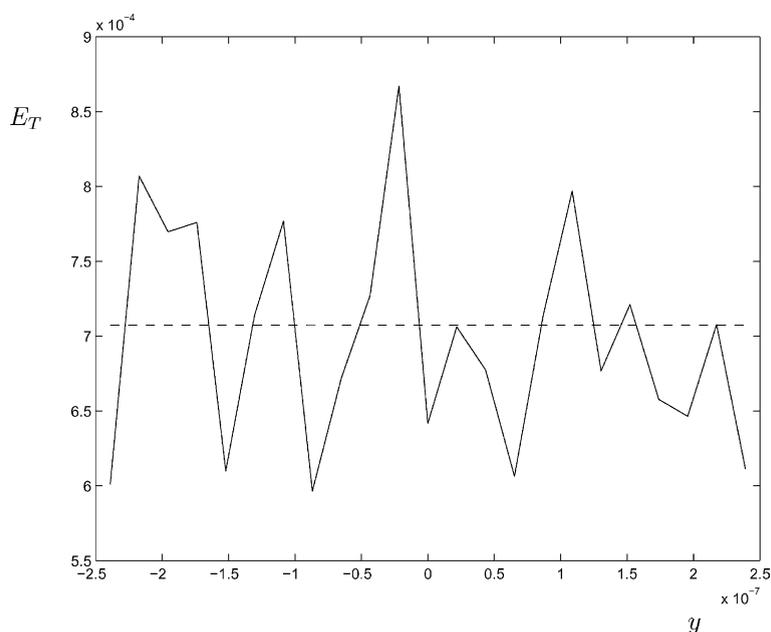


Fig. 6. Fractional error in temperature for Poiseuille flow in a channel as a function of the transverse channel coordinate, y (measured in meters). The dashed line denotes Eq. (10) and the solid line indicates DSMC simulation results.

agreement is found between the theoretical results and simulation measurements for volume-averaged and fluxal quantities, as shown in Figs. 7–10.

A final note: In DSMC simulations one considers each particle as “representing” a number of molecules, \mathcal{N} , in the physical system. Typically $\mathcal{N} \gg 1$ although this value is only restricted by the available computer resources (given sufficient memory, \mathcal{N} can even be less than one). Despite its stochastic nature, DSMC has been shown [5] to reproduce the correct hydrodynamic fluctuations, both under equilibrium and non-equilibrium conditions, that is, the Monte Carlo element of the algorithm does not introduce any additional fluctuations. However, the molecule-to-particle rescaling does affect the magnitude of the fluctuations with the variance increased by a factor of \mathcal{N} relative to that in the physical system. In all the expressions given above for estimating statistical errors, N_0 and N^+ are the number of particles used by the simulation. Therefore, the relative error can be reduced by using larger numbers of particles in a simulation (i.e., lowering the molecule-to-particle ratio, \mathcal{N}).

3.2. Dense fluids

We performed molecular dynamics simulations to test the validity of Eqs. (4), (8) and (10) for dense fluids. A similar geometry to the dilute gas simulations described above was used but at a significantly higher density. In particular, we simulated liquid argon ($\sigma_{\text{LJ}} = 3.4 \times 10^{-10}$ m, $\varepsilon_{\text{LJ}} = 119.8k_b$) at $T = 240$ K and $\rho = 860$ kg/m³ in a two-dimensional channel with the x - and z -directions periodic. The channel height was $H = 69.7\sigma_{\text{LJ}}$. The wall molecules were connected to fcc lattice sites through springs and interacted with the fluid through a Lennard–Jones potential with the same parameters. The spring constant $k_s = 460\varepsilon\sigma^{-2}$ was chosen in such a way that the root mean square displacement of wall atoms around their equilibrium position at the simulated temperature was well below the Lindemann criterion for the melting point of a solid. The length and depth of the system was $28\sigma_{\text{LJ}}$ and $29.1\sigma_{\text{LJ}}$ in the x - and z -directions, respectively.

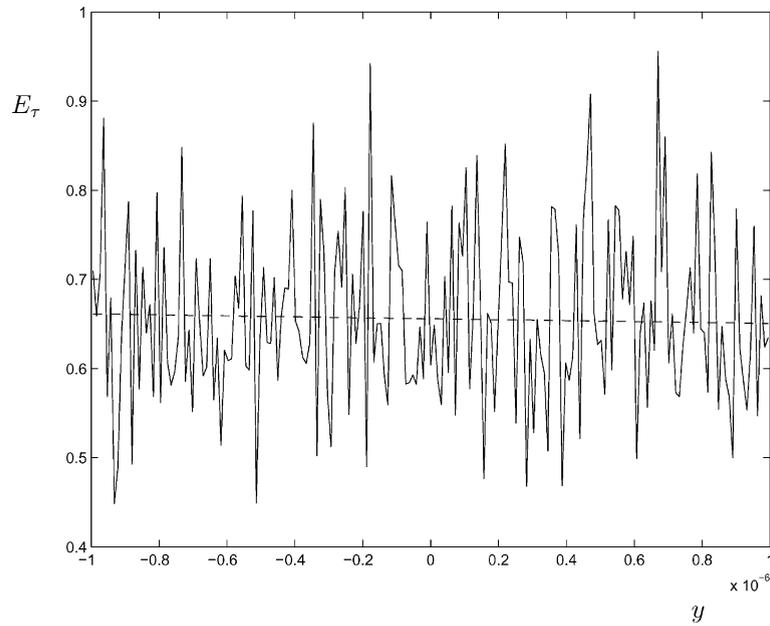


Fig. 7. Fractional error in the shear stress τ_{xy} for Couette flow in a channel as a function of the transverse channel coordinate, y (measured in meters). The dashed line denotes Eq. (24) and the solid line indicates DSMC simulation results.

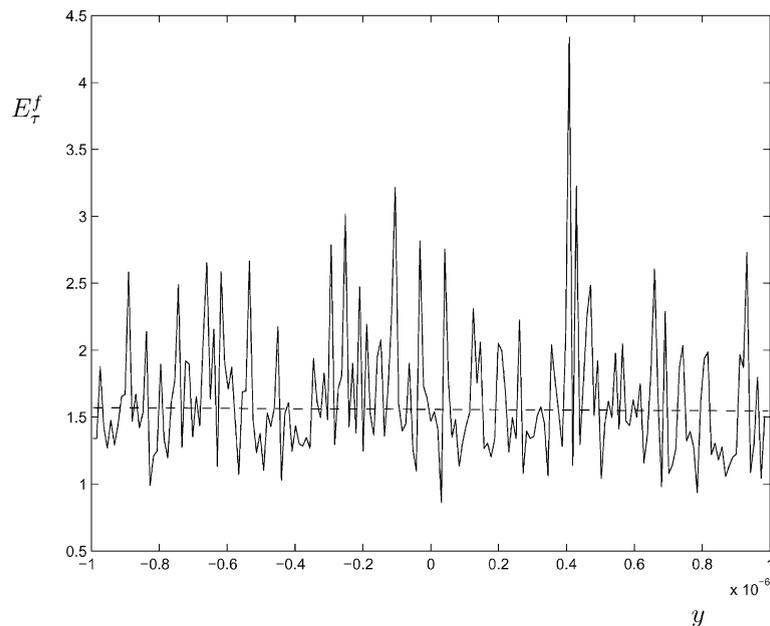


Fig. 8. Fractional error in the shear stress for Couette flow in a channel as a function of the transverse channel coordinate, y (measured in meters). The dashed line denotes Eq. (55) and the solid line indicates DSMC simulation results.

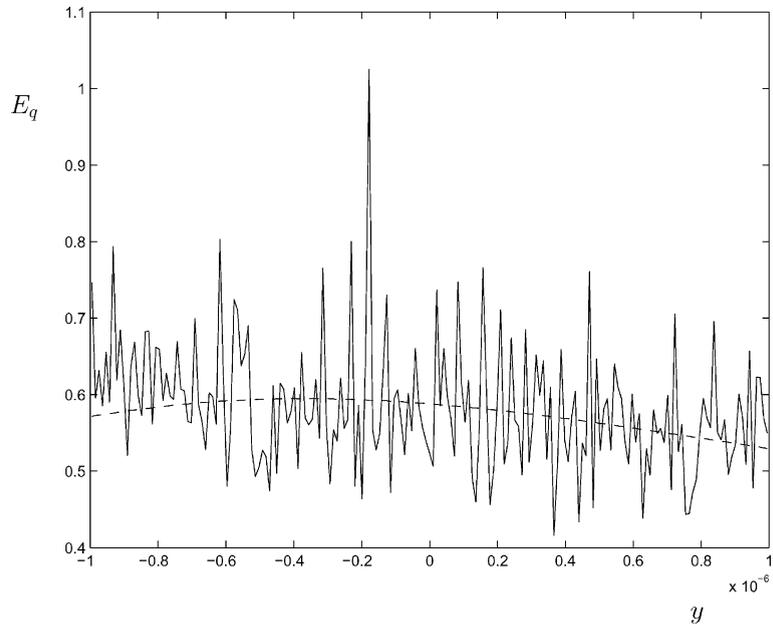


Fig. 9. Fractional error in the heat flux for “Temperature Couette” in a channel as a function of the transverse channel coordinate, y (measured in meters). The dashed line denotes Eq. (25) and the solid line indicates DSMC simulation results.

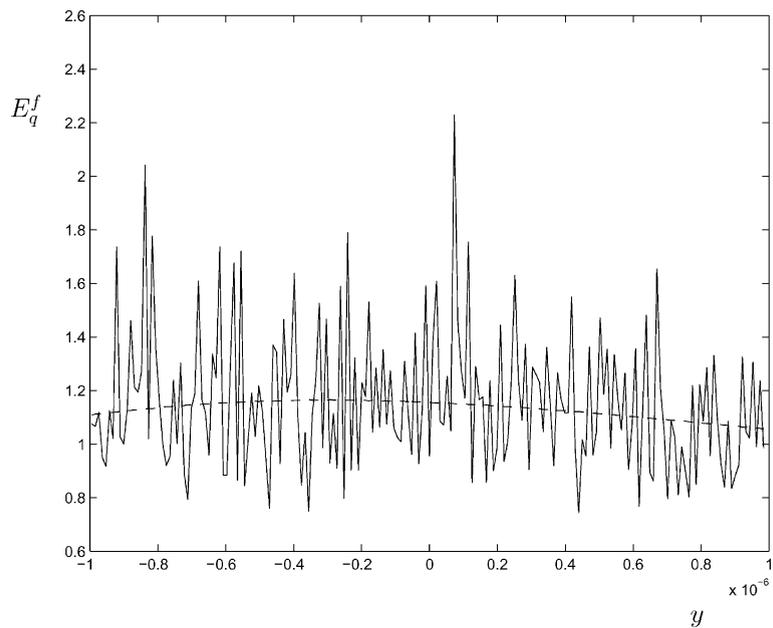


Fig. 10. Fractional error in the heat flux for “Temperature Couette” in a channel as a function of the transverse channel coordinate, y (measured in meters). The dashed line denotes Eq. (57) and the solid line indicates DSMC simulation results.

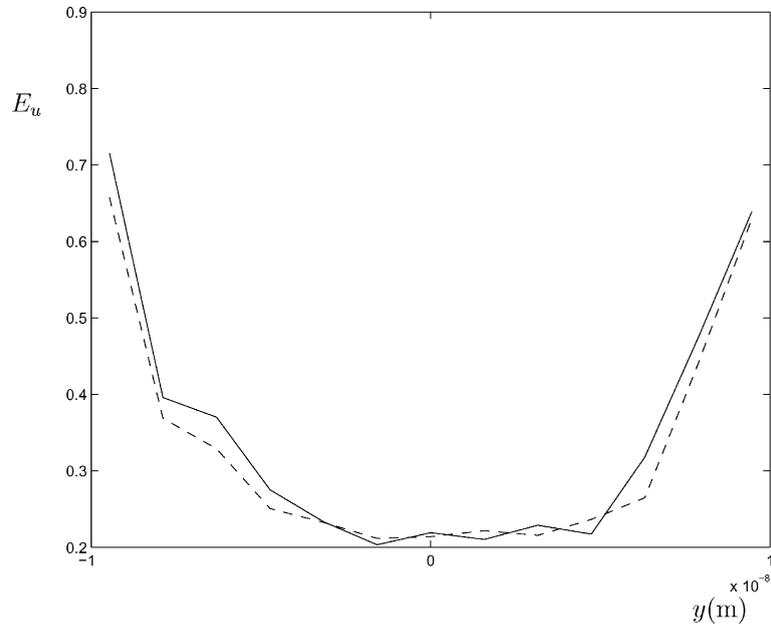


Fig. 11. Fractional error in velocity for dense-fluid Poiseuille flow in a channel as a function of the transverse channel coordinate, y . The dashed line denotes Eq. (4) and the solid line denotes MD simulation results.

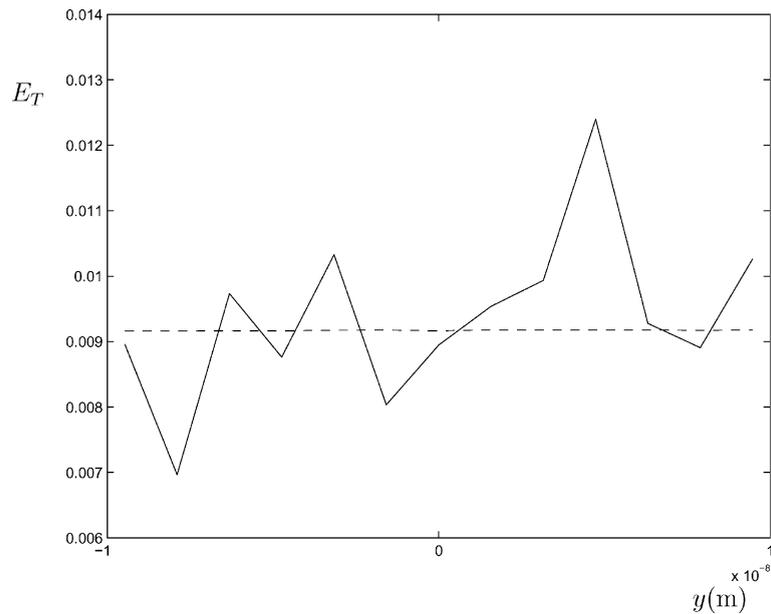


Fig. 12. Fractional error in temperature for dense-fluid Poiseuille flow in a channel as a function of the transverse channel coordinate, y . The dashed line denotes Eq. (10) and the solid line indicates MD simulation results.

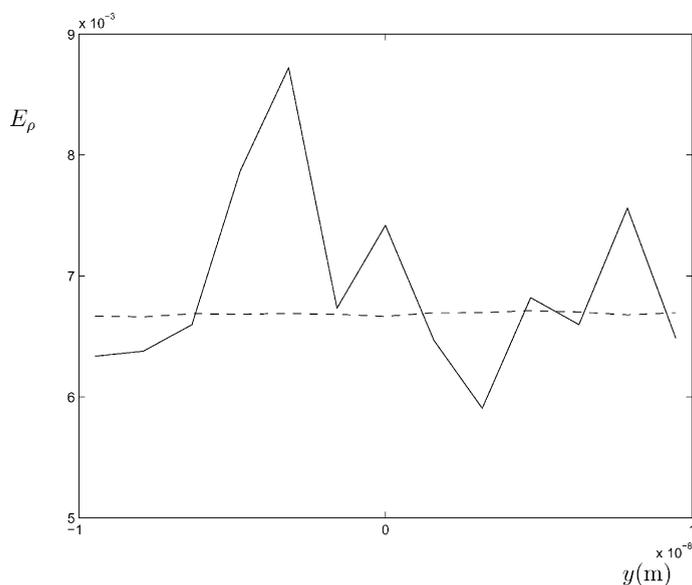


Fig. 13. Fractional error in density for dense-fluid Poiseuille flow in a channel as a function of the transverse channel coordinate, y . The dashed line denotes Eq. (8) and the solid line indicates MD simulation results.

A constant force $f = 8 \times 10^{-5} \varepsilon / \sigma_{LJ}$ per particle was used to generate a velocity field with a maximum velocity of approximately 13 m/s.

In order to calculate the fluctuation of density, temperature, and velocity, we divided the simulation cell into 13 layers in the y -direction with a height $\Delta y = 4.6396\sigma$. We further divided each layer into 49 cells, 7 in each of the x - and z -directions. The density, temperature, and velocity in each cell were calculated every 2000 timesteps ($0.005t_{LJ}$), where $t_{LJ} = \sqrt{m_{LJ}\sigma_{LJ}^2/\varepsilon_{LJ}}$. We have checked that this time interval is longer than the system's correlation time such that samples taken between such intervals are independent. For each cell, 200 samples are used to calculate the average density, temperature, and velocity. The fluctuation was calculated for each layer using the 49 equivalent cells in the x - z plane.

Due to the sensitivity of the compressibility κ_T on the interaction cutoff r_c , a rather conservative value of $r_c = 4.0\sigma_{LJ}$ was used. We also introduced a correction for the still-finite cutoff which used the compressibility predictions of the Modified Benedict–Webb–Rubin equation of state [13]. The agreement between the theoretical predictions and the simulations is good (see Figs. 11–13).

4. Independent samples and correlations

The results in Figs. 7–10 suggest that volume-averaged measurements provide a superior performance due to a smaller relative error. This conclusion, however, is not necessarily correct because our results are based on arbitrary choices of “measurement spacing”, in the sense that the only consideration was to eliminate correlations in the data since the theoretical formulation in Section 2 is based on the assumption of uncorrelated samples. The two methods of sampling are in fact linked by a very interesting interplay between the roles of time and space: fluxal sampling is a measurement at a *fixed* position in space for a period of time, whereas volume sampling is performed over some region of space at a *fixed* time. The theoretical performance of each method can be increased by extending the respective window of observation. However, by increasing the time of observation, a fluxal measurement becomes correlated with

neighboring measurements if the same particle crosses more than one measuring station within the period of observation. Similarly, by increasing the region of measurement, subsequent volume measurements will suffer from time correlations if previously interrogated particles do not have sufficient time to leave the measurement volume. This relation between spatial and temporal samplings and the role of the particle characteristic velocity is also manifested in the theoretical predictions. Enforcing equality of the variances of the respective volume and fluxal measurements (Eqs. (24) and (55), and Eqs. (25) and (57)) yields $\Delta x = \beta c_m \Delta t$, where $\beta \approx 1$. The generalization of this work to include time and spatial correlations is the subject of future work.

The effect of time correlation in volume measurements can be approximated using the theory of the “persistent random walk” [19], first introduced by Fürth [20] and Taylor [7]. A persistent random walk is one in which each step displacement, $u(t_\eta) = u(\eta\Delta t)$, is identically distributed and has a positive correlation coefficient with the previous step

$$\alpha = \frac{\langle \delta u(t_\eta) \delta u(t_{\eta-1}) \rangle}{\langle (\delta u)^2 \rangle} \quad (64)$$

($0 < \alpha < 1$). Models such as these have been proposed to model diffusion in a turbulent fluid [7,21]. This assumption implies that step correlations decay exponentially in time

$$\frac{\langle \delta u(t_\eta) \delta u(0) \rangle}{\langle (\delta u)^2 \rangle} = \alpha^\eta = e^{-t_\eta/t_c}, \quad (65)$$

where $t_c = -\Delta t / \log \alpha$ is the correlation time beyond which the steps are essentially independent. The position, $U(t_\eta)$, of the random walker after η steps is the sum of these correlated random displacements, $U(t_\eta) = \sum_{j=1}^{\eta} u(t_j)$. Following Taylor, it is straightforward to show that for times long compared to the correlation time ($t_\eta \gg t_c$), the usual diffusive scaling holds

$$\langle (\delta U(t_\eta))^2 \rangle \sim \langle (\delta u)^2 \rangle \eta \left(\frac{1 + \alpha}{1 - \alpha} \right), \quad (66)$$

where the bare diffusion coefficient is modified by the term in parentheses due to correlations.

There is a natural connection with the measurement of statistical averages. If we view each step in the persistent random walk as a correlated sample of some quantity in the gas, then the position of the walker (divided by the number of samples) corresponds to the sample average. Thus the variance of a set of sequentially correlated random variables $\{u(t_\eta)\}$, $\text{Var}^c(u)$, may be written as

$$\text{Var}^c(u) = \text{Var}(u) \frac{1 + \alpha}{1 - \alpha}, \quad (67)$$

where $\text{Var}(u)$ is the variance of the uncorrelated samples ($\alpha = 0$). The theory above implies that the sample variance is amplified by the presence of correlations, because effectively fewer independent samples have been taken compared to the uncorrelated case, $\alpha = 0$.

Note that a sequence of correlated random variables, $\{u(t_\eta)\}$, satisfying Eqs. (64) and (65) can be explicitly constructed from a sequence of independent, identically distributed variables, $\{\tilde{u}(t_\eta)\}$, by letting $u(t_\eta)$ equal the previous value $u(t_{\eta-1})$ with probability α or a new value $\tilde{u}(t_\eta)$ with probability $1 - \alpha$. This allows us to interpret α as the probability that a sample is the same as the previous one, which is precisely the source of correlations when volume-sampling dilute gases.

There are two distinct ways that a new sample of a dilute gas can actually provide new information: (i) either some new particles have entered the sampling cell, or (ii) some particles previously inside the cell have changed their properties due to collisions. Regarding (i), the probability that a particle will remain in a cell of size Δx after time step Δt is given by

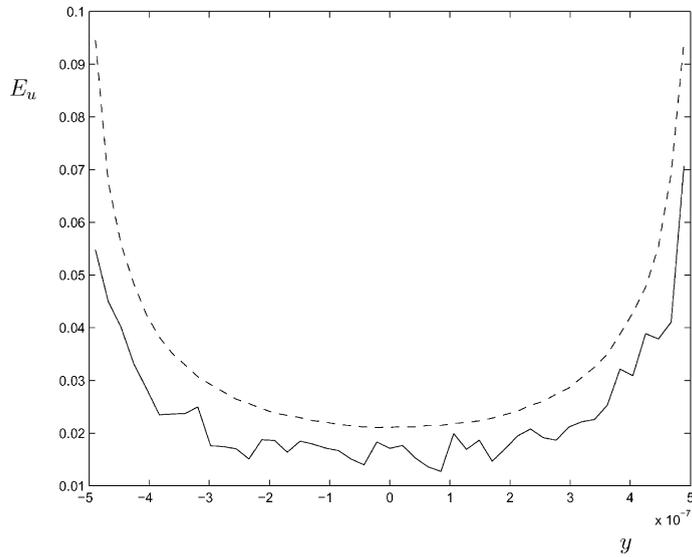


Fig. 14. Fractional error in velocity for Poiseuille flow in a channel as a function of the transverse channel coordinate, y (measured in meters). The dashed line denotes the theoretical prediction including correlations and the solid line denotes DSMC simulation results.

$$\alpha_l = \left(\frac{1}{\Delta x} \int_{-\Delta x/2}^{\Delta x/2} \int_{-\Delta x/2}^{\Delta x/2} f\left(\frac{x - \tilde{x}}{\Delta t}\right) dx d\tilde{x} \right)^d, \tag{68}$$

where d is the dimensionality of the cell and $f(c_i)$ is the probability distribution function for the velocity of particle i , and thus $f((x - \tilde{x})/\Delta t)$ represents the probability that a particle originating from location \tilde{x} is

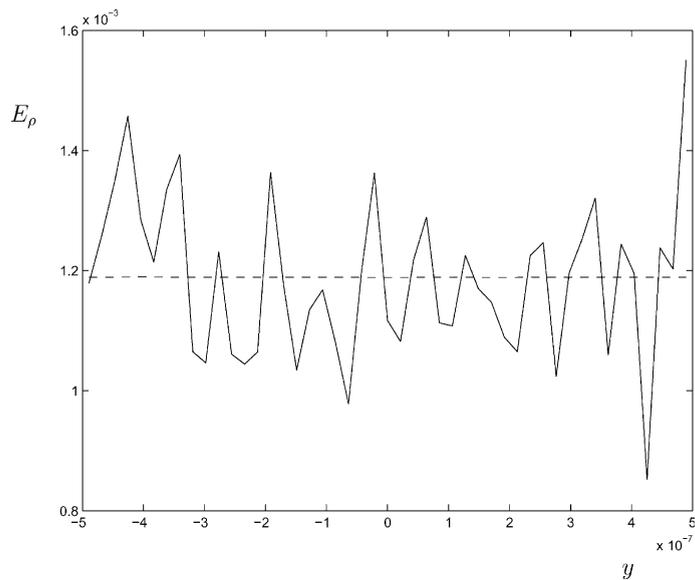


Fig. 15. Fractional error in density for Poiseuille flow in a channel as a function of the transverse channel coordinate, y (measured in meters). The dashed line denotes the theoretical prediction including correlations and the solid line indicates DSMC simulation results.

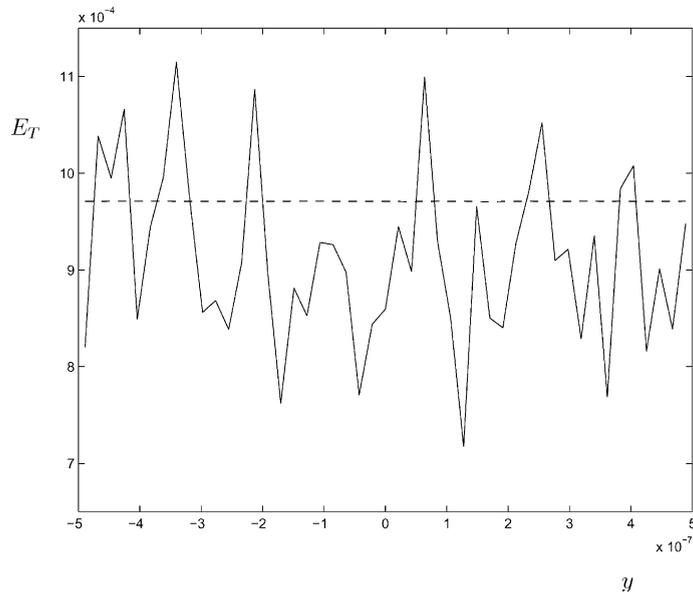


Fig. 16. Fractional error in temperature for Poiseuille flow in a channel as a function of the transverse channel coordinate, y (measured in meters). The dashed line denotes the theoretical prediction including correlations and the solid line indicates DSMC simulation results.

found at x after a time interval Δt . Note that the above expression holds when the components of the particle velocity in different directions are uncorrelated and that the particle spatial distribution inside the cell is initially uniform. Regarding (ii), if the single-particle auto-correlation function decays exponentially, the “renewal” probability $1 - \alpha_c$ for collisional de-correlation can, at least in principle, be inferred from simulations using Eq. (65). The net correlation coefficient, giving the probability of an “identical” sample, is $\alpha = \alpha_c \alpha_l$.

DSMC simulations show that the single-particle velocity correlation coefficient is $\alpha_c \approx 0.94$ for $\Delta t = \lambda / (15c_m)$ by fitting Eq. (65) to data; the same value for α_c was found in both equilibrium and non-equilibrium simulations. In the comparison of Figs. 14–16, however, we use $\alpha_c = 1$ since mass, momentum, and energy are conserved *in a cell* during collisions. We find that this analysis produces very good results in the case of density and temperature (see Figs. 15 and 16) and acceptable results for the case of mean velocity (see Fig. 14). Use of $\alpha_c = 0.94$ would tend to make the agreement better but no theoretical justification exists for it within our simple framework. In our comparison, the value of α_l was directly calculated by assuming an equilibrium distribution in Eq. (68). For our two-dimensional calculations with $\Delta x / (c_m \Delta t) = 2.5$, we find $\alpha_l \approx 0.6$. The effect of a mean flow is very small if the Mach number is small. This was verified through both direct evaluation of Eq. (68) using a local equilibrium distribution function and DSMC simulations.

5. Conclusions

We have presented expressions for the statistical error in estimating the velocity, density, temperature, and pressure in molecular simulations. These expressions were validated for flow of a dilute gas and dense liquid in a two-dimensional channel using the direct simulation Monte Carlo and Molecular Dynamics, respectively. Despite the non-equilibrium nature of the validation experiments, good agreement is found

between theory and simulation, verifying that modifications to non-equilibrium results are very small. The agreement with equilibrium theory is particularly remarkable in the dense fluid case where significant non-equilibrium due to a shear of the order of $5.5 \times 10^8 \text{ s}^{-1}$ exists. We thus expect our results to hold for general non-equilibrium applications of interest.

Predictions were also presented for the statistical error in estimating the shear stress and heat flux in dilute gases through cell averaging and surface averaging. Comparison with direct Monte Carlo simulations shows that the equilibrium assumption is justified. Within the same sets of assumptions, we were able to show that the distribution of particles leaving a cell is Poisson-distributed.

It was found that the fluctuation in state variables is significantly smaller compared to flux variables in the continuum regime $Kn \rightarrow 0$. This is important for the development of hybrid methods. Although a direct comparison was only presented between volume-averaged quantities, we find that the fluxal measurements for the shear stress and heat flux perform similarly to the volume-averaged counterparts (regarding scaling with the Knudsen number).

Acknowledgements

The authors thank M. Malek-Mansour and B. Alder for helpful discussions. The authors also thank X. Garaizar for making this work possible through the computer resources made available to them. This work was also supported, in part, by a grant from the University of Singapore, through the Singapore–MIT alliance. This work was also supported, in part, by the U.S. Department of Energy through Lawrence Livermore National Laboratory, University of California (contract No. ~W-7405-Eng-48).

Appendix A. Number fluctuations in dilute gases

Consider an infinite ideal gas in equilibrium with mean number density, n . By definition, each infinitesimal volume element, dV , contains a particle with probability $n dV$ and each such event is independent (i.e., particles are distributed according to a spatial Poisson process). From these assumptions, it is straightforward to show that the number of particles N in an arbitrary volume V is a Poisson random variable [12]

$$\text{Prob}(N = m) = \frac{e^{-\langle N \rangle} \langle N \rangle^m}{m!}$$

with mean and variance given by

$$\langle N \rangle = \langle (\delta N)^2 \rangle = nV.$$

Note that this result does not strictly apply to a finite ideal gas because the probabilities of finding particles in different infinitesimal volumes are no longer independent, due to the global constraint of a fixed total number of particles. Nevertheless, it is an excellent approximation for most dilute gases, even finite non-ideal gases far from equilibrium (e.g., as demonstrated by simulations in the main text).

Now consider a property A which each particle in a volume V may possess independently with probability, P_A . In this section, we show that the distribution of the number of such particles, N_A , is also a Poisson random variable with mean and variance given by

$$\langle N_A \rangle = \langle (\delta N_A)^2 \rangle = P_A \langle N \rangle = P_A nV.$$

For example, in the main text we require the number of particles, N_{ij} , which travel from one region, i , to another region, j , in a time interval, Δt . The proof given here, however, is much more general and applies to

arbitrary number fluctuations of an infinite ideal gas in equilibrium, such as the number of particles in a certain region, moving in a certain direction, of a certain “color”, with speeds above a certain threshold, etc.

We begin by expressing N_A as a random sum of random variables

$$N_A = \sum_{i=1}^N \chi_i, \quad (\text{A.1})$$

where

$$\chi_i = \begin{cases} 1 & \text{if } A \text{ occurs,} \\ 0 & \text{otherwise} \end{cases}$$

is an indicator function for particle i to possess property A , which is a Bernoulli random variable with mean, P_A . It is convenient to introduce probability generating functions

$$f_\chi(z) = \sum_{m=0}^1 \text{Prob}(\chi = m)z^m = (1 - P_A) + P_A z$$

and

$$f_N(z) = \sum_{m=0}^{\infty} \text{Prob}(N = m)z^m = \sum_{m=0}^{\infty} \frac{e^{-\langle N \rangle} \langle N \rangle^m z^m}{m!} = e^{\langle N \rangle(z-1)}, \quad (\text{A.2})$$

because the generating function for a random sum of random variables, as in Eq. (A.1), is simply given by a composition of the generating functions for the summand and the number of terms [12]

$$f_{N_A}(z) = \sum_{m=0}^{\infty} \text{Prob}(N_A = m)z^m = f_N(f_\chi(z)).$$

Combining these expressions we have

$$f_{N_A}(z) = e^{\langle N \rangle(1 - P_A + P_A z - 1)} = e^{P_A \langle N \rangle(z-1)}.$$

Comparing with Eq. (A.2) completes the proof that N_A is a Poisson random variable with mean, $P_A \langle N \rangle$.

References

- [1] M.P. Allen, D.J. Tildesley, *Computer Simulation of Liquids*, Clarendon Press, Oxford, 1987.
- [2] D. Frenkel, B. Smit, *Understanding Molecular Simulation, From Algorithms to Applications*, Academic Press, San Diego, 2002.
- [3] G.A. Bird, *Molecular Gas Dynamics and the Direct Simulation of Gas Flows*, Clarendon Press, Oxford, 1994.
- [4] F.J. Alexander, A.L. Garcia, The direct simulation Monte Carlo, *Comp. Phys.* 11 (1997) 588–593.
- [5] M. Malek-Mansour, A.L. Garcia, G.C. Lie, E. Clementi, Fluctuating hydrodynamics in a dilute gas, *Phys. Rev. Lett.* 58 (1987) 874–877;
A.L. Garcia, M. Malek-Mansour, G. Lie, M. Mareschal, E. Clementi, Hydrodynamic fluctuations in a dilute gas under shear, *Phys. Rev. A* 36 (1987) 4348–4355.
- [6] L.D. Landau, E.M. Lifshitz, *Statistical Physics, Part 1*, Pergamon Press, Oxford, 1980 (Chapter XII).
- [7] G.I. Taylor, Diffusion by continuous movements, *Proc. London Math. Soc.* 20 (1920) 196–211.
- [8] F.J. Alexander, A.L. Garcia, B.J. Alder, Cell size dependence of transport coefficients in stochastic particle algorithms, *Phys. Fluids* 10 (1998) 1540;
Erratum: *Phys. Fluids* 12 (2000) 731.
- [9] N.G. Hadjiconstantinou, Analysis of discretization in the direct simulation Monte Carlo, *Phys. Fluids* 12 (2000) 2634.

- [10] A.L. Garcia, W. Wagner, Time step truncation error in direct simulation Monte Carlo, *Phys. Fluids* 12 (2000) 2621.
- [11] N.G. Hadjiconstantinou, Hybrid atomistic-continuum formulations and the moving contact-line problem, *J. Comput. Phys.* 154 (1999) 245–265.
- [12] W. Feller, *An Introduction to Probability Theory and its Applications*, vol. 1, third ed., Wiley, New York, 1968 (Chapter 12).
- [13] J.K. Johnson, J.A. Zollweg, K.E. Gubbins, The Lennard–Jones equation of state revisited, *Mol. Phys.* 78 (1993) 591–618.
- [14] A.L. Garcia, J.B. Bell, Wm.Y. Crutchfield, B.J. Alder, Adaptive mesh and algorithm refinement using direct simulation Monte Carlo, *J. Comp. Phys.* 154 (1999) 134.
- [15] F. Alexander, A. Garcia, B. Alder, The consistent Boltzmann algorithm for the van der Waals equation of state, *Physica A* 240 (1997) 196.
- [16] L.D. Landau, E.M. Lifshitz, *Statistical Physics, Part 2*, Pergamon Press, Oxford, 1980 (Section 88).
- [17] L.D. Landau, E.M. Lifshitz, *Statistical Physics, Part 1*, Pergamon Press, Oxford, 1980 (Section 112).
- [18] W.W. Liu, Y.C. Fang, Implicit boundary conditions for direct simulation Monte Carlo method in MEMS flow predictions, *CMES* 1 (2000) 119.
- [19] G.H. Weiss, Some applications of persistent random walks and the telegrapher’s equation, in preparation.
- [20] R. Fürth, *Schwankungerscheinungen in der Physik*, Sammlung Verlag, Braunschweig, 1920.
- [21] S. Ghosal, J.B. Keller, A hyperbolic equation for turbulent diffusion, *Nonlinearity* 13 (2000) 1855–1866.