Gaseous diffusion as a correlated random walk

Santos Bravo Yuste ◉,* Rubén Gómez González ◉,† and Vicente Garzó ◉‡

Departamento de Física and Instituto de Computación Científica Avanzada (ICCAEx),
Universidad de Extremadura, E-06006 Badajoz, Spain

The mean-square displacement per collision of a molecule immersed in a gas at equilibrium is given by its mean-square displacement between two consecutive collisions (mean-square free path) corrected by a prefactor in the form of a series. The nth term of the series is proportional to the mean value of the scalar product $\mathbf{r}_1 \cdot \mathbf{r}_n$, where $\mathbf{r}_i$ is the displacement of the molecule between the $(i-1)$-th and ith collisions. Simple arguments are used to obtain approximate expressions for each term. The key finding is that the ratio of consecutive terms in the series closely approximates the so-called mean persistence ratio. Exact expressions for the terms in the series are considered and their ratios for several consecutive terms are calculated for the case of hard spheres, showing an excellent agreement with the mean persistence ratio. These theoretical results are confirmed by solving the Boltzmann equation by means of the direct simulation Monte Carlo method. By summing the series, the mean-square displacement and the diffusion coefficient can be determined using only two quantities: the mean-square free path and the mean persistence ratio. A simple and an improved expression for the diffusion coefficient D are considered and compared with the so-called first and second Sonine approximations to D as well as with computer simulations of the Boltzmann equation. It is found that the improved diffusion coefficient shows very good agreement with simulation results over all intruder and molecule mass ranges. When the intruder mass is smaller than that of the gas molecules, the improved formula even outperforms the first Sonine approximation.

I. INTRODUCTION

The Chapman-Enskog solution of the Boltzmann equation is the most successful and well-known method for determining the transport coefficients of molecular gases [1]. This well-studied procedure assumes the existence of a normal solution where all the space and time dependence of the velocity distribution function f occurs only through a functional dependence on the hydrodynamic fields. This functional dependence can be made more explicit when the spatial gradients are small. In this situation, f is written as a series. expansion in powers of the spatial gradients of the hydrodynamic fields (density, mean flow velocity, and temperature): $f = f^{(0)} + f^{(1)} + f^{(2)} + \cdots$. Here $f^{(0)}$ is the local version of the Maxwell-Boltzmann distribution function (zeroth order in gradients), $f^{(1)}$ is of first-order in gradients (Navier-Stokes order), $f^{(2)}$ is of second order in gradients (Burnett order), and so on.

Unfortunately, within the Chapman-Enskog procedure, one has to be ready to pay the price of dealing with a mathematically complex and not very intuitive method. On the other hand, since the beginning of the kinetic theory of gases with Maxwell, Stefan, Boltzmann, and Meyer [2], there has been

a primary approach to address the study of transport properties of gases, namely, the mean-free-path (MFP) theory. This theory is much simpler and more intuitive than the Chapman-Enskog theory. This explains its use in many textbooks as an introductory method for studying transport properties in gases [3,4] and why is sometimes referred to as the *elementary* kinetic theory of gases.

In the elementary kinetic theory, the molecules are treated as masses of negligible volume (point particles) in constant random motion. They experience negligible intermolecular forces except during collisions, which justifies the concept of a free path between collisions. However, this approach suffers from some shortcomings: (i) its results are not always accurate, (ii) its arguments are in many occasions qualitative and highly debatable, and (iii) there is no general systematic way to improve its approximations [1,5]. Notwithstanding, numerous researchers have shown that many of these limitations can be lessened and that a free path approach can yield good predictions on the diffusion of gases while retaining its ability to offer insights into the underlying physical mechanisms [5–9].

In this paper we consider the problem of diffusion of an intruder in a dilute gas at equilibrium (intruder and particles of the gas are in general mechanically different, i.e., they can differ in mass and size). In our approach, we treat the diffusion process of a molecule (or intruder) in the gas as a random-walk problem. In other words, we view this diffusion problem as the result of a random process in which the molecule is, or behaves like, a random walker moving with steps $\mathbf{r}_i$ of random size and direction. These steps (also referred to as flights)

---

*Contact author: santos@unex.es; https://fisteor.cms.unex.es/investigadores/santos-bravo-yuste/

†Contact author: ruben@unex.es

‡Contact author: vicenteg@unex.es; https://fisteor.cms.unex.es/investigadores/vicente-garzo-puertos/

simply represent displacements of the molecule (or walker) between consecutive collisions. Alternatively, employing the classic terminology, they are recognized as the free paths. In this paper we calculate the diffusion coefficient by evaluating the mean-square displacement of the molecule (random walker). Authors such as Yang [6], Furry [5], Pitkanen [7], and Monchick [8] explored this avenue many years ago. In particular, if one assumes that the steps $\mathbf{r}_i$ are *uncorrelated* [i.e., $\langle \mathbf{r}_i \cdot \mathbf{r}_{i+k} \rangle = 0$, where $\langle \cdots \rangle$ denotes an average], then one finds a very simple expression for the diffusion coefficient given by [3,4]

$$D \equiv D_{\text{eKT}} = \frac{\lambda \bar{v}}{d}, \tag{1}$$

where λ and $\bar{v}$ are the mean free path and the average velocity between two successive collisions, respectively, of the molecule, and d is the dimensionality of the system. In Eq. (1), the subscript eKT means that the diffusion coefficient has been evaluated within the elementary kinetic theory. Unfortunately, the expression (1) is quite inaccurate. This is primarily due to the assumption that the steps are uncorrelated, which is a significant oversimplification. On the contrary, when molecules collide they exhibit a strong tendency to continue in a direction similar to the one they had before the collision (i.e., $\langle \mathbf{r}_i \cdot \mathbf{r}_{i+k} \rangle > 0$). This is the "persistence" of collisions, a fact employed in the past to improve the expression (1) [10,11].

Thus, for the reasons mentioned before, we will consider in this paper the diffusion of an intruder in a gas at equilibrium as a random walk with *correlated* steps (flights). We will show how the inclusion of these sort of correlations in an approximate and simple way yields an expression for the diffusion coefficient D that is fully consistent with the classical formula derived by Jeans [11] taking into account the persistence. We will subsequently extend our discussion and propose an improved expression of D. This new form of the diffusion coefficient turns out to be equivalent to an expression mentioned but not developed in the work of Jeans [11]. It should be noted that the formulas for the coefficient D reported in this paper have been derived in a different, independent, and, in our opinion, more compelling approach than the ones displayed in the work of Jeans [11]. This is one of the new added values of the present contribution. The key point in our methodology is the evaluation of the mean value of the projection of $\mathbf{r}_{i+k}$ over $\mathbf{r}_i$, i.e, $\langle \mathbf{r}_i \cdot \mathbf{r}_{i+k} \rangle$, which constitute the terms appearing in the series that leads to the diffusion coefficient. The evaluation of $\langle \mathbf{r}_i \cdot \mathbf{r}_{i+k} \rangle$ will allow us to estimate the decorrelation length; it represents the distance that the random walker has to travel to lose the memory of the size and direction of its initial step. In this work, we focus on a hard-sphere gas as the specific system to apply our methodology. However, any other type of interaction could have been considered as long as the conditions for a free path description are met.

It is also interesting to note that, as shown by Monchik and Mason [8,12], there exists a very close relationship between the results derived from the Chapman-Enskog and MFP theories. It turns out that an integral equation for the velocity distribution function *coincides* with the integral equation verifying the Chapman-Enskog approximation $f^{(1)}$ when considering the limit of infinitely many collisions in the MFP

theory. However, Monchik and Mason [12] pointed out that the velocity distributions corresponding to the two theories differ *slightly* because of the use of slightly different auxiliary conditions.

The plan of the paper is as follows. In Sec. II, we present the diffusion of an intruder in a dilute gas at equilibrium as a random-walk problem. We calculate the mean-square displacement and the diffusion coefficient as an infinite series (the collisional series) with terms of the form $\langle \mathbf{r}_1 \cdot \mathbf{r}_k \rangle$. In Sec. III, we make some simple approximations to estimate the value of each term $\langle \mathbf{r}_1 \cdot \mathbf{r}_k \rangle$ in this series. We then sum the series and obtain the formula proposed by Jeans [11] for the diffusion coefficient D. The expression of D is improved in Sec. IV by incorporating the exact value of the mean-square free path $\langle r_1^2 \rangle$. The general formulas for evaluating the averages $\langle \mathbf{r}_1 \cdot \mathbf{r}_k \rangle$ in terms of the collision frequency and the so-called transition rate are developed in Sec. V. We use these expressions to compute the first few terms of the collisional series for hard spheres. Notably, we find that the quotient of the successive terms of this series is very well approximated by the mean persistence ratio. To gauge the accuracy of our expressions for D, we compare in Sec. VI this random-walk expression with the theoretical forms of D obtained from the Chapman-Enskog method in the so-called first- and second-Sonine approximations and with the results obtained by numerically solving the Boltzmann equation by means of the direct simulation Monte Carlo (DSMC) method [13]. We end the paper in Sec. VII with some conclusions and remarks.

II. THE MOLECULE AS A RANDOM WALKER: MEAN-SQUARE DISPLACEMENT

We consider a gas in equilibrium with no external fields. The velocity distribution of the molecules is then a Maxwell-Boltzmann distribution. We single out a molecule that can be either identical to or distinct from the other molecules. We refer to this molecule as the random walker or intruder (a term commonly used in kinetic theory of granular gases [14]). We are interested in understanding how far this molecule will be from its initial position $\mathbf{R}(t)$ after a given time t. With no external field, the average of this quantity is equal to zero: $\langle \mathbf{R}(t) \rangle = 0$. However, the second moment $\langle R^2(t) \rangle$, or mean-square displacement (MSD), is the relevant quantity because provides information about the extent to which these intruders or random walkers spread (diffuse) within the gas. In particular, the diffusion coefficient D of the molecule in a three-dimensional gas can be obtained from the long-time (i.e., very large number of collisions) value of the quotient $\langle R^2(t) \rangle / 6t$, namely

$$D = \lim_{t \to \infty} \frac{\langle R^2(t) \rangle}{6t}. \tag{2}$$

The diffusion of a molecule in a gas at equilibrium is normal so that the limit of Eq. (2) exists.

A. Evaluation of $\langle R_N^2 \rangle$

To evaluate D or $\langle R^2(t) \rangle$, it is convenient to first consider the displacement of the random walker (the intruder) after N

steps (collisions):

$$\mathbf{R}_N = \sum_{i=1}^{N} \mathbf{r}_i, \qquad (3)$$

where $\mathbf{r}_i$ is ith displacement (flight, free path) of the molecule. From Eq. (3), one finds

$$R_N^2 = \sum_{i=1}^{N} \mathbf{r}_i \cdot \mathbf{r}_i + \sum_{i \neq j}^{N} \mathbf{r}_i \cdot \mathbf{r}_j \qquad (4)$$

$$= \sum_{i=1}^{N} \mathbf{r}_i \cdot \mathbf{r}_i + 2 \sum_{i=1}^{N-1} \mathbf{r}_i \cdot \mathbf{r}_{i+1} + 2 \sum_{i=1}^{N-2} \mathbf{r}_i \cdot \mathbf{r}_{i+2} + \cdots. \qquad (5)$$

At equilibrium, the probability density function (pdf) of the step sizes does not depend on time, so that

$$\langle R_N^2 \rangle = N \langle \mathbf{r}_1 \cdot \mathbf{r}_1 \rangle + 2(N-1) \langle \mathbf{r}_1 \cdot \mathbf{r}_2 \rangle + 2(N-2) \langle \mathbf{r}_1 \cdot \mathbf{r}_3 \rangle$$

$$+ \cdots. \qquad (6)$$

It is convenient to write $\langle R_N^2 \rangle$ as

$$\langle R_N^2 \rangle = N \ell_e^2, \qquad (7)$$

where ℓ_e can be identified as an *effective* free path. The quantity $\ell_e^2 = \langle R_N^2 \rangle / N$, is the MSD per collision and can be approximated by

$$\ell_e^2 \approx \langle r_1^2 \rangle + 2 \langle \mathbf{r}_1 \cdot \mathbf{r}_2 \rangle + 2 \langle \mathbf{r}_1 \cdot \mathbf{r}_3 \rangle + \cdots \qquad (8)$$

for sufficiently large N. Here we have assumed that the steps decorrelate fast enough so that the error in the approximation

$$(N - k - 1) \langle \mathbf{r}_1 \cdot \mathbf{r}_k \rangle \approx N \langle \mathbf{r}_1 \cdot \mathbf{r}_k \rangle \qquad (9)$$

can be ignored even for large k. For example, we will see in Sec. III that, for elastic hard spheres where the mass of the intruder m_1 is equal to the mass m_2 of the molecules of the gas, the quantity $\langle \mathbf{r}_1 \cdot \mathbf{r}_{1+k} \rangle$ decays exponentially fast, approximately as 0.406^k. The validity of this approximation will be discussed in Sec. VI.

It is convenient to rewrite Eq. (8) as

$$\frac{\ell_e^2}{\langle r_1^2 \rangle} \approx 1 + \frac{2 \langle \mathbf{r}_1 \cdot \mathbf{r}_2 \rangle}{\langle r_1^2 \rangle} + \frac{2 \langle \mathbf{r}_1 \cdot \mathbf{r}_3 \rangle}{\langle r_1^2 \rangle} + \cdots \equiv \sum_{n=1}^{\infty} c_n. \qquad (10)$$

We will call the series in Eq. (8) the *collisional series* while the series $\sum_{n=1}^{\infty} c_n$ appearing in Eq. (10) will be referred to as the *reduced* collisional series. In summary, Eqs. (7) and (10) tell us that the mean-square displacement of the intruder per collision, $\langle R_N^2 \rangle / N$, is just the mean-square free path, $\langle r_1^2 \rangle$, corrected by the reduced collisional series.

The expression (7) for the MSD is exact for a random walk with *isotropic* and *uncorrelated* steps of a fixed size ℓ_e. This effective random walk [where ℓ_e is determined by Eq. (8)] is then *equivalent* to the actual random walk in the sense that both lead to the same MSD.

B. Evaluation of $\langle R^2(t) \rangle$ from $\langle R_N^2 \rangle$

The pdf of the position $\mathbf{R}$ of the intruder at time t can be written as

$$P(\mathbf{R}, t) = \sum_{N=0}^{\infty} P_N(\mathbf{R}) \chi_N(t), \qquad (11)$$

where $P_N(\mathbf{R})$ is the pdf of the position of the walker after N steps and $\chi_N(t)$ is the probability of taking exactly N steps up to time t. In terms of $P(\mathbf{R}, t)$, the MSD is

$$\langle R^2(t) \rangle = \int d\mathbf{R} \, P(\mathbf{R}, t) R^2. \qquad (12)$$

Substitution of Eq. (11) into Eq. (12) leads to

$$\langle R^2(t) \rangle = \sum_{N=0}^{\infty} \chi_N(t) \int P_N(\mathbf{r}) R^2 \, d\mathbf{R} = \sum_{N=0}^{\infty} \chi_N(t) \langle R_N^2 \rangle. \qquad (13)$$

On obtaining Eq. (13) summation and integration have been interchanged. Taking into account Eq. (7), we find

$$\langle R^2(t) \rangle = \ell_e^2 \sum_{N=0}^{\infty} \chi_N(t) N = \langle N(t) \rangle \, \ell_e^2, \qquad (14)$$

$\langle N(t) \rangle$ being the average number of steps given (collisions suffered) by the random walker up to time t. From Eqs. (2) and (14), one achieves the result

$$D = \lim_{t \to \infty} \frac{\langle N(t) \rangle}{6t} \ell_e^2. \qquad (15)$$

When the gas is at equilibrium, the collision frequency does not depend of the position and time. Therefore, $\langle N(t) \rangle / t$ represents the average collision frequency $\langle \mu \rangle$ (recall that it is assumed that t is large in this expression, implying a significant number of collisions). Because the average time between collisions $\langle \tau \rangle$ (collision interval) is the inverse of the collision frequency [3,6,15], then Eq. (15) can be written as

$$D = \frac{\langle \mu \rangle \ell_e^2}{6} = \frac{\ell_e^2}{6 \langle \tau \rangle}. \qquad (16)$$

It is convenient to rewrite Eq. (16) in terms of the mean free path λ and the elementary kinetic theory diffusion coefficient D_{eKT}:

$$D = D_{\text{eKT}} \frac{\ell_e^2}{2 \lambda^2}. \qquad (17)$$

On writing Eq. (17), use has been made of Eq. (1) and the fact that the average speed $\bar{v}$ of the molecules is $\bar{v} = \lambda / \langle \tau \rangle = \lambda \langle \mu \rangle$ [3,6,15]. Using Eq. (8), Eq. (17) can be finally written as

$$D = D_{\text{eKT}} \frac{\langle r_1^2 \rangle}{2 \lambda^2} \left(1 + \frac{2 \langle \mathbf{r}_1 \cdot \mathbf{r}_2 \rangle}{\langle r_1^2 \rangle} + \frac{2 \langle \mathbf{r}_1 \cdot \mathbf{r}_3 \rangle}{\langle r_1^2 \rangle} + \cdots \right). \qquad (18)$$

Assuming that λ and $\langle \mu \rangle$ are known, the only quantity to evaluate for estimating D is the effective mean free path ℓ_e^2. This is tantamount to evaluate the averages $\langle \mathbf{r}_1 \cdot \mathbf{r}_k \rangle$ [see Eq. (8)]. Since this task is not simple (as discussed later in Sec. V), we can make some simple approximations. First, we take $\langle r_1^2 \rangle \approx 2 \lambda^2$, which would be true if the gas molecules (except the intruder) were motionless [3,4,15]. Second, we neglect correlations between successive steps (i.e., $\langle \mathbf{r}_1 \cdot \mathbf{r}_k \rangle = 0$ for $k > 0$). Combining both approximations, we find that $\ell_e^2 = 2 \lambda^2$ and so, the diffusion coefficient is given by the elementary kinetic theory [7], Eq. (1).

III. CORRELATIONS, PERSISTENCE, AND MSD: JEANS FORMULA

In the previous Sec. II, we have discussed how the evaluation of the intruder's MSD (or, equivalently, its diffusion coefficient) can be reduced to the evaluation of the effective free path ℓ_e [see Eq. (17)], or the evaluation of $\langle \mathbf{r}_1 \cdot \mathbf{r}_k \rangle$ [see Eq. (8)]. The precise computation of these quantities will be addressed in Sec. V. In the present section we show how to approximate them in terms of the mean free path λ and the so-called persistence of the collisions. We employ relatively straightforward and simple arguments.

The evaluation of $\langle \mathbf{r}_1 \cdot \mathbf{r}_k \rangle$ for $k = 1$ (that is, $\langle r_1^2 \rangle$) is the easiest one. For elastic hard spheres, an elementary computation yields $\langle r_1^2 \rangle \approx 2\lambda^2$. This comes from the simple (approximate) expression $P(\ell) = \exp(-\ell/\lambda)/\lambda$ for the distribution of the free path length ($\ell \equiv r_1$) [4,15]. This expression is derived under the simplifying assumption that the intruder moves within a gas where its particles are "frozen." The rigorous result is more complex and we will discuss this in Sec. V B.

For estimating the average $\langle \mathbf{r}_1 \cdot \mathbf{r}_k \rangle = \langle r_1 r_k \cos \theta_{1,k} \rangle$ with $k \geqslant 2$ we assume that the size r_1 of the step 1 and the value of the projection of the step $\mathbf{r}_k$ over $\mathbf{r}_1$ are uncorrelated. Namely, we make the approximation

$$\langle \mathbf{r}_1 \cdot \mathbf{r}_k \rangle \approx \langle r_1 \rangle \langle r_k \cos \theta_{1,k} \rangle = \lambda \langle r_k \cos \theta_{1,k} \rangle, \quad (19)$$

where $\theta_{i,j}$ denotes the angle between the displacements $\mathbf{r}_i$ and $\mathbf{r}_j$. It is important to note that correlations indeed exist due to the fact that the size of displacements (free paths) before and after the collision ($\mathbf{r}_i$ and $\mathbf{r}_{i+1}$, respectively) as well as the angle formed by these displacements depend on the corresponding velocities $\mathbf{v}_i$ and $\mathbf{v}_{i+1}$, which are correlated.

Let us now see how to estimate the term $\langle r_k \cos \theta_{1,k} \rangle$ by using the spherical cosine law that relates the cosines of the angles $\theta_{1,k}$, $\theta_{1,2}$, and $\theta_{2,k}$:

$$\cos \theta_{1,k} = \cos \theta_{1,2} \cos \theta_{2,k} + \sin \theta_{1,2} \sin \theta_{2,k} \cos \varphi_{1,2,k}, \quad (20)$$

where $\varphi_{1,2,k}$ is the angle between the plane generated by $\{\mathbf{r}_1, \mathbf{r}_2\}$ and the plane generated by $\{\mathbf{r}_2, \mathbf{r}_k\}$. Due to the rotational symmetry of collisions along the direction of the precollisional displacements (or velocities), all values of $\varphi_{1,2,k}$ are equally probable. This implies that $\langle r_k \sin \theta_{1,2} \sin \theta_{2,k} \cos \varphi_{1,2,k} \rangle = 0$, and then

$$\langle r_k \cos \theta_{1k} \rangle = \langle r_k \cos \theta_{1,2} \cos \theta_{2,k} \rangle. \quad (21)$$

The relation (21) can be rewritten as

$$\langle r_k \cos \theta_{1k} \rangle = \left\langle r_1 \frac{r_2}{r_1} \cos \theta_{1,2} \frac{r_k}{r_2} \cos \theta_{2,k} \right\rangle \quad (22)$$

or, neglecting correlations, as

$$\langle r_k \cos \theta_{1k} \rangle \approx \lambda \left\langle \frac{r_2}{r_1} \cos \theta_{1,2} \right\rangle \left\langle \frac{r_k}{r_2} \cos \theta_{2,k} \right\rangle. \quad (23)$$

We repeat this procedure for estimating the value of the last factor $\langle (r_k/r_2) \cos \theta_{2,k} \rangle$. We use again the spherical cosine law that relates $\theta_{2,k}$ with $\theta_{2,3}$ and $\theta_{3,k}$. As before,

$\langle (r_k/r_2) \sin \theta_{2,3} \sin \theta_{k-1,k} \cos \varphi_{2,3,k} \rangle = 0$ and so

$$\left\langle \frac{r_k}{r_2} \cos \theta_{2,k} \right\rangle = \left\langle \frac{r_3}{r_2} \cos \theta_{2,3} \frac{r_k}{r_3} \cos \theta_{3,k} \right\rangle$$
$$\approx \left\langle \frac{r_3}{r_2} \cos \theta_{2,3} \right\rangle \left\langle \frac{r_k}{r_3} \cos \theta_{3,k} \right\rangle. \quad (24)$$

Repeating this procedure we get

$$\langle r_k \cos \theta_{1k} \rangle \approx \lambda \left\langle \frac{r_2}{r_1} \cos \theta_{1,2} \right\rangle \left\langle \frac{r_3}{r_2} \cos \theta_{2,3} \right\rangle \cdots$$
$$= \lambda \prod_{i=1}^{k-1} \left\langle \frac{r_{i+1}}{r_i} \cos \theta_{i,i+1} \right\rangle. \quad (25)$$

On the other hand, at equilibrium, the averages of the form $\langle (r_{i+1}/r_i) \cos \theta_{i,i+1} \rangle$ do not depend on i (i.e., on time), and then Eq. (25) yields

$$\langle r_k \cos \theta_{1k} \rangle \approx \lambda \left\langle \frac{r_2}{r_1} \cos \theta_{1,2} \right\rangle^{k-1}. \quad (26)$$

Following this, we approximate

$$\left\langle \frac{r_2}{r_1} \cos \theta_{1,2} \right\rangle \approx \left\langle \frac{v_2}{v_1} \cos \theta_{1,2} \right\rangle. \quad (27)$$

Note that this relation would be exact if all the flight times (times between collisions) τ_i were the same for all i as $r_i = v_i \tau_i$. Recall that $r_2 \cos \theta_{1,2}$ is the component of the postcollisional displacement $\mathbf{r}_2$ along the direction of the precollisional displacement $\mathbf{r}_1$. One could then see $r_2 \cos \theta_{1,2}/r_1$ as the fraction of the precollisional displacement that is not lost, that persists, after the collision. Similarly, $v_2 \cos \theta_{1,2}/v_1$ is the fraction of $\mathbf{v}_1$ that persists after a collision. The average of this quantity is the so-called mean persistence ratio ω:

$$\left\langle \frac{v_2}{v_1} \cos \theta_{1,2} \right\rangle \equiv \omega. \quad (28)$$

The mean persistence ω is a well-studied quantity in the kinetic theory of gases (see, for example, Sec. 5.5 of Ref. [1]). This explains our interest in constructing the factorizations of Eqs. (22), (24), and (25).

Collecting Eqs. (19), (26), and (27), we finally find:

$$\langle \mathbf{r}_1 \cdot \mathbf{r}_k \rangle \approx \lambda^2 \, \omega^{k-1}. \quad (29)$$

An alternative deduction of this formula is presented in Appendix A. Equation (29) together with $\langle r_1^2 \rangle \approx 2\lambda^2$ and Eq. (8) leads to

$$\ell_e^2 \approx 2\lambda^2 (1 + \omega + \omega^2 + \cdots)$$
$$= \frac{2\lambda^2}{1 - \omega}, \quad (30)$$

which implies [see Eq. (14)]

$$\langle R^2(t) \rangle = \langle N(t) \rangle \frac{2\lambda^2}{1 - \omega} \quad (31)$$

or [see Eq. (17)]

$$D = \frac{D_{\text{eKT}}}{1 - \omega} = \frac{\lambda \bar{v}}{3(1 - \omega)} \equiv D_J. \quad (32)$$

This simple and nice expression D_J is no more than the classic expression derived by Jeans (see the second formula of section 169 of Ref. [11]), which takes into account the persistence of the molecules to improve the elementary kinetic expression D_{eKT}.

A full assessment of these formulas for elastic hard spheres will be given in Sec. VI when we compare them against DSMC simulations. Nonetheless, we can provide here some numbers that show the large improvement that the consideration of the persistence of the molecules leads to. For elastic hard spheres, one has $\omega \simeq 0.406$ for $m_1/m_2 = 1$, $\omega \simeq 0.245$ for $m_1/m_2 = 0.5$, and $\omega \simeq 0.785$ for $m_1/m_2 = 5$. Here m_1 and m_2 are the masses of the intruder and the gas, respectively. From Eq. (32), one finds that D/D_{eKT} is 1.68 for $m_1/m_2 = 1$, 1.32 for $m_1/m_2 = 0.5$, and 4.66 for $m_1/m_2 = 5$. The corresponding DSMC simulation values for D/D_{eKT} are 1.40, 1.82, and 5.37, respectively. The inclusion of correlations between displacements through the mean persistence ratio clearly improves the D_{eKT} values. More details will be given in Sec. VI.

A long time ago it was realized (see Ref. [11], for example) that one way to improve the accuracy of the MFP theory predictions was to take into account that, after a collision, the exit direction of the colliding molecule is not completely isotropic. Instead, there is a certain probability that the exit direction is similar to the direction of incidence. This is known as the persistence of velocities after collision [1,11]. MFP-type formulas incorporating this aspect offer better and more accurate predictions. However, the arguments used to incorporate this fact into the equations were not entirely convincing. Perhaps for this reason, the consideration of arguments based on the persistence of collisions in the MFP formulation has not gained much success or relevance, remaining limited essentially to the contributions of Jeans [10,11]. This is despite the clear and intuitive understanding that persistence must play a role in the transport properties, since these properties necessarily depend on the manner in which collisions occur. In the present section, we have shown how to include the persistence in the calculation of the diffusion coefficient using a different approach. We will further refine this random-walk approach in the subsequent sections.

IV. AN IMPROVED FORMULA FOR THE DIFFUSION COEFFICIENT

One of the limitations of free path theory is its inability to control the reliability of its assumptions and approximations. This issue has been frequently highlighted, particularly in the book of Chapman and Cowling (see Sec. 6.5 of Ref. [1]). However, it is important to recognize that the standard approach advocated by the Chapman-Enskog method [1] is not entirely immune to this limitation, as demonstrated by the widespread use of DSMC and molecular dynamics simulations to ensure the results of the kinetic theory.

In this section, we examine the assumptions underlying Eqs. (30) and (32). Equation (30) was obtained from Eq. (29) by assuming two key approximations: first, $\langle r_1^2 \rangle \approx 2\lambda^2$ and, second, $\langle \mathbf{r}_1 \cdot \mathbf{r}_k \rangle \approx \lambda^2 \omega^{k-1}$ for $k \geqslant 2$ [see Eq. (29)] We can (partially) eliminate the first approximation by using the exact value of the mean-square free path $\langle r_1^2 \rangle$. In this case, from

Eqs. (8) and (29), one has

$$
\begin{aligned}
\ell_e^2 &\approx \langle r_1^2 \rangle + 2\lambda^2(\omega + \omega^2 + \cdots) \\
&= \langle r_1^2 \rangle \left[1 + \frac{2\lambda^2}{\langle r_1^2 \rangle}(\omega + \omega^2 + \cdots) \right] \\
&\approx \langle r_1^2 \rangle (1 + \omega + \omega^2 + \cdots) \\
&= \frac{\langle r_1^2 \rangle}{1 - \omega},
\end{aligned}
\tag{33}
$$

where in the third line of Eq. (33) we have replaced $2\lambda^2$ by $\langle r_1^2 \rangle$. Substitution of Eq. (33) into Eq. (14) yields

$$
\langle R^2(t) \rangle = \langle N(t) \rangle \frac{\langle r_1^2 \rangle}{1 - \omega},
\tag{34}
$$

or, from Eq. (16), one gets

$$
D = \frac{\langle r_1^2 \rangle}{6\langle \tau \rangle} \frac{1}{1 - \omega}.
\tag{35}
$$

It is convenient to rewrite this equation in terms of D_{eKT} as

$$
D = D_{eKT} \frac{\kappa}{1 - \omega} \equiv D_{RW},
\tag{36}
$$

where

$$
\kappa = \frac{\langle r_1^2 \rangle}{2\lambda^2}.
\tag{37}
$$

On writing Eq. (36), use has been made of Eq. (1) and the relation $\bar{v} = \lambda/\langle \tau \rangle$.

The nice thing about Eqs. (34) and (36) is that they are *equivalent* to the formula

$$
D = \frac{\langle v_1^2 \tau_1 \rangle}{d} \frac{1}{1 - \omega},
\tag{38}
$$

which is obtained from Eq. (32) if one replaces in this equation $\lambda \bar{v}$ by $\langle r_1 v_1 \rangle = \langle v_1^2 \tau_1 \rangle$. This replacement, mentioned by Jeans in Ref. [11] (see the last equation of page 203 of this reference), was not explored further, leaving its implications uncharted. Equations (36) and (38) are equivalent provided that $\langle r_1^2 \rangle / \langle \tau \rangle = 2\langle v_1^2 \tau_1 \rangle$. This is proved in Sec. V B [see Eq. (49)].

It is important to note that the validity of Eq. (33) [and, equivalently, Eqs. (34) and (36)], rests in the validity of the approximation

$$
c_{1+k} = \frac{2\langle \mathbf{r}_1 \cdot \mathbf{r}_{1+k} \rangle}{\langle r_1^2 \rangle} \approx \omega^k, \quad k \geqslant 1.
\tag{39}
$$

As mentioned in Sec. II, this result implies that Eq. (9) holds unless $m_1/m_2 \to \infty$ because (see at the end of Sec. B) $\omega \to 1$ in this limiting case.

The relationship $\langle \mathbf{r}_1 \cdot \mathbf{r}_{1+k} \rangle \sim \omega^k$ indicates that the correlations between successive free paths decay exponentially. These short-range correlations imply that the diffusion of the molecule is normal [16]: At large times, $\langle R^2 \rangle$ increases linearly with time and the distribution of the position $R(t)$ is Gaussian. The fast exponential decay of the correlations means that after a few collisions almost no memory of the initial displacement $\mathbf{r}_1$ remains, and the preferred direction of movement $\mathbf{r}_1$ becomes barely noticeable. Let us assume our

measurement resolution for $\langle \mathbf{r}_1 \cdot \mathbf{r}_{1+k} \rangle$ is ϵ. This implies that after $n = \ln(\epsilon)/\ln \omega$ collisions we are not longer able to detect correlations between successive free paths $\mathbf{r}_{1+k}$ and the initial displacement $\mathbf{r}_1$. Within this resolution, $\langle R^2(n) \rangle = n\ell_e^2$ holds, and therefore $\langle R^2(N) \rangle/N = \langle R^2(n) \rangle/n = \ell_e^2$. In other words, if we observe the intruder movement with a spatial resolution of the order $\langle R^2(n) \rangle^{1/2}$ (or, equivalently, with a temporal resolution of around $n\langle \tau \rangle$), then what we see is a self-similar random walk. Obviously, this self-similarity does not hold if we scrutinize the particle's movement with a finer spatial or temporal resolution. In that case, we perceive persistence (i.e., the correlations between successive steps), something that goes unnoticed when the motion is viewed at larger scales.

On shorter spatial or temporal scales, the "diffusion" of the intruder is not normal, i.e., $\langle R^2(n) \rangle$ is not proportional to the number n of steps (collisions). This conclusion can be easily obtained from Eqs. (6) and (39):

$$\langle R^2(n) \rangle \approx \langle r_1^2 \rangle \sum_{k=0}^{n-1} (n-k)\omega^k = \frac{\langle r_1^2 \rangle}{1-\omega} \left[n - \frac{\omega(1-\omega^n)}{1-\omega} \right]. \tag{40}$$

Equation (40) clearly shows that $\langle R^2(n) \rangle$ is not a linear function of n. However, after a relatively small number n of collisions, the exponential corrective term becomes negligible with respect to n, and one eventually finds that the MSD of the molecule grows linearly with n. The number n of collisions for which the correction term becomes negligible is smaller when ω is smaller, i.e., the smaller the mass of the intruder (this is because ω is an increasing function of the mass of the intruder). In this normal diffusive regime, the slope $d\langle R^2(n) \rangle/dn$ of the MSD does not depend on n. In our case, the slope of the MSD for all n [in units of the asymptotic long-term slope $\langle r_1^2 \rangle/(1-\omega)$] is approximately given by this nonconstant function:

$$1 + \frac{\omega^{1+n} \ln \omega}{1-\omega} \equiv s(n, \omega). \tag{41}$$

The separation of $s(n, \omega)$ from 1 is a measure of how far the intruder's displacement is from the long-time asymptotic normal diffusion regime. The function $s(n, \omega)$ is plotted in Fig. 1 versus n for several values of ω. We see that $s(n, \omega)$ approaches 1 very quickly, meaning that the normal diffusive regime [where $s(n, \omega) = 1$] is reached after a small number of collisions. For example, for $m_1/m_2 = 0.5$, 1, and 5, it takes only about 4, 7, and 28 collisions, respectively, to reach a regime where the slope of $\langle R^2(n) \rangle$ differs by less than one part in a thousand from its long-term asymptotic value $\langle r_1^2 \rangle/(1-\omega)$. In other words, if one observes the motion of the intruder on time and/or space scales larger than those corresponding to 4, 7, and 28 collisions, then the motion of the intruder with $m_1/m_2 = 0.5$, 1, and 5, respectively, will look like normal diffusion. Of course, these numbers depend on the criterion used to decide when normal diffusive behavior is finally achieved. For example, if we use an alternative criterion and decide that the molecule reaches normal diffusion when the value of $d\ln\langle R^2(n) \rangle/d\ln n$ is no more than, say, 5% away from 1, then the corresponding numbers of collisions would be 7, 14, and 77, respectively.
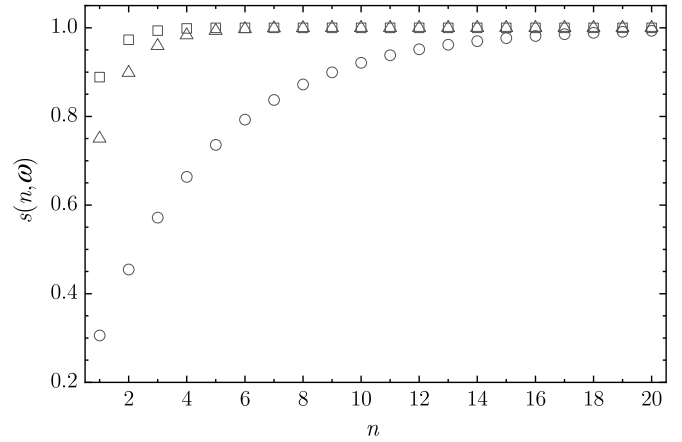

FIG. 1. Plot of $s(n, \omega)$ (the slope of the MSD in units of the asymptotic long-time slope) vs the number of collisions n for $m_1/m_2 = 0.5$ ($\omega \simeq 0.245$; square), $m_1/m_2 = 1$ ($\omega \simeq 0.406$; triangles), and $m_1/m_2 = 5$ ($\omega \simeq 0.785$; circles).

Finally, since Eq. (39) plays a key role in our approximations, verifying its validity is a must. To do this, we have to evaluate the average $\langle \mathbf{r}_1 \cdot \mathbf{r}_{1+k} \rangle$. This will be the subject of Sec. V.

V. EVALUATION OF $\langle \mathbf{r}_1 \cdot \mathbf{r}_k \rangle$ AND DIFFUSION COEFFICIENT

A. Probability of a given path of N steps

Let $P_N(\mathbf{v}_1, \tau_1; \mathbf{v}_2, \tau_2; \ldots; \mathbf{v}_N, \tau_N)d\mathbf{v}_1 d\mathbf{v}_2 \cdots d\mathbf{v}_N \ d\tau_N$ be the probability of finding molecule 1 (the intruder) with any velocity which is deflected by collision into $\mathbf{v}_1$ and $\mathbf{v}_1 + d\mathbf{v}_1$ (let's denote it as $\mathbf{v}_1, d\mathbf{v}_1$) and travels freely for a time τ_1, is then deflected again into $\mathbf{v}_2, d\mathbf{v}_2$ and continues to travel freely for another time τ_2, and so on, until it is deflected into $\mathbf{v}_N, d\mathbf{v}_N$ before finally colliding in the time interval $\tau_N, d\tau_N$. Note that, because $\mathbf{r}_i = \mathbf{v}_i\tau_i$, this probability can also be seen as the probability that a molecule follows the path $\mathbf{r}_1 + \mathbf{r}_2 + \cdots + \mathbf{r}_N$ with velocities $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N\}$. The pdf $P_N(\mathbf{v}_1, \tau_1; \mathbf{v}_2, \tau_2; \ldots; \mathbf{v}_N, \tau_N) \equiv P_N(1, 2, \ldots, N)$ is given by [6]

$$P_N(1, 2, \ldots; N) = f(1)\mu(1) \prod_{i=1}^{N-1} W(i, i+1)$$

$$\times \prod_{j=1}^{N} e^{-\mu(j)\tau_j} \mu(N), \tag{42}$$

where $\mu(i) \equiv \mu(\mathbf{v}_i)$ is the collision frequency for particles of velocity $\mathbf{v}_i$ and

$$f(1) \equiv f(\mathbf{v}_1) = \left(\frac{m_1}{2\pi k_B T} \right)^{3/2} e^{-m_1 v_1^2/2k_B T} \tag{43}$$

is the equilibrium pdf of the velocity of the intruder [17]. In addition, k_B is the Boltzmann's constant and the transition rate $W(i, i+1) \equiv W(\mathbf{v}_i, \mathbf{v}_{i+1})$ is the probability that a molecule with precollisional velocity $\mathbf{v}_1$ is deflected into $\mathbf{v}_2, d\mathbf{v}_2$

(postcollisional velocity). With the help of the equation

$$\int f(\mathbf{v}_1) W(\mathbf{v}_1, \mathbf{v}_2) d\mathbf{v}_1 = f(\mathbf{v}_2) \mu(\mathbf{v}_2) \tag{44}$$

(which embodies the principle of detailed balancing [6]), it can be proved that P_N is not normalized to the unity, but to the average collision rate:

$$\int P_N(1, 2, \ldots; N) \, d\mathbf{v}_1 \cdots d\mathbf{v}_N d\tau_N = \langle \mu \rangle. \tag{45}$$

We will use the notation $\tilde{P}_N = P_N / \langle \mu \rangle$ for the pdf normalized to the unity.

B. The mean-square and the pdf of the free path for hard spheres

From the definition of $\tilde{P}_1$, the mean-square free path is defined as

$$\langle r_1^2 \rangle = \int_0^\infty d\tau_1 \int d\mathbf{v}_1 \, \tilde{P}_1(\mathbf{v}_1, \tau_1)(v_1 \tau_1)^2. \tag{46}$$

When $N = 1$, Eq. (42) leads to

$$P_1(\mathbf{v}_1, \tau_1) = f(v_1) \mu^2(v_1) e^{-\mu(v_1)\tau_1}. \tag{47}$$

Using the expression (47) into Eq. (46) one easily finds

$$\langle r_1^2 \rangle = \frac{2}{\langle \mu \rangle} \int d\mathbf{v}_1 f(\mathbf{v}_1) \frac{v_1^2}{\mu(v_1)} = \frac{8\pi}{\langle \mu \rangle} \int_0^\infty dv_1 f(v_1) \frac{v_1^4}{\mu(v_1)}. \tag{48}$$

Since the term $v_1^2/\mu(v_1)$ can be rewritten as $v_1^2\tau_1$ in Eq. (48), then $\langle r_1^2 \rangle$ is the average over the velocity distribution $f(v_1)$ of $2v_1^2\tau_1/\langle \mu \rangle$:

$$\langle r_1^2 \rangle = 2 \frac{\langle v_1^2 \tau_1 \rangle}{\langle \mu \rangle} = 2\langle v_1^2 \tau_1 \rangle \langle \tau_1 \rangle. \tag{49}$$

This justifies our claim in Sec. IV that the random-walk formula (36) and the Jeans formula (38) are equivalent.

Introducing the expressions of f_1 and μ for hard spheres given by Eqs. (43) and (B1) into Eq. (48), one finds that the mean-square free path $\langle r_1^2 \rangle$ (in units of $2\lambda^2$) is

$$\kappa \equiv \frac{\langle r_1^2 \rangle}{2\lambda^2} = 2\sqrt{\pi} \left(\frac{m_1}{m_2}\right)^2 \left(1 + \frac{m_1}{m_2}\right)^{1/2} I_{1,1}(m_1/m_2), \tag{50}$$

where

$$I_{1,1}(m) = \int_0^\infty e^{-my^2} \frac{y^4}{E(y)} dy. \tag{51}$$

No explicit expression for $I_{1,1}(m)$ is known, but it can be easily evaluated numerically. The function κ is plotted in Fig. 2 as a function of the mass ratio m_1/m_2. We see that the larger m_1/m_2, the larger is the difference of κ from 1. The largest difference is for $m_1/m_2 \to \infty$. In this case $\kappa = 3\pi/8 \simeq 1.1781$ [18].

Note that

$$P_1(\mathbf{v}_1, \tau_1) d\mathbf{v}_1 d\tau_1 = (1/v_1) P_1(\mathbf{v}_1, r_1/v_1) d\mathbf{v}_1 dr_1. \tag{52}$$

Using the standard notation ℓ for the length r_1 of the free path, and taking into account Eqs. (47) and (52), we find that the pdf
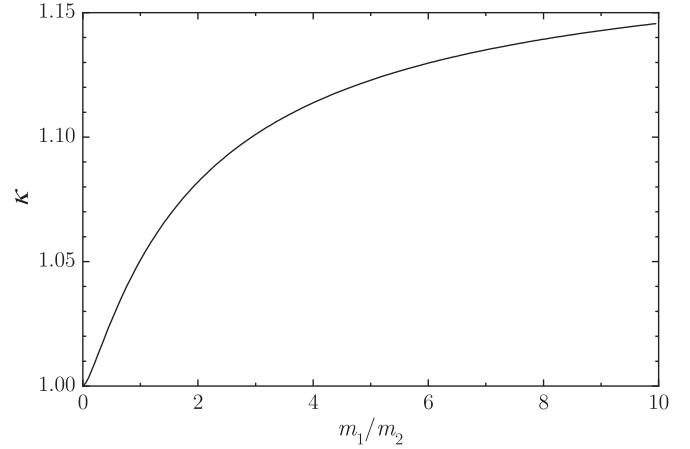

FIG. 2. Plot of $\kappa = \langle r_1^2 \rangle / 2\lambda^2$ vs the mass ratio m_1/m_2.

of a free path of length ℓ (whatever its corresponding velocity) is given by

$$P(\ell) = \int \frac{\tilde{P}_1(\mathbf{v}, \ell/v)}{v} d\mathbf{v} = \frac{1}{\langle \mu \rangle} \int f(\mathbf{v}) \frac{\mu^2(v)}{v} e^{-\mu(v)\ell/v} d\mathbf{v}. \tag{53}$$

From this expression, we can evaluate $\langle \ell^2 \rangle \equiv \langle r_1^2 \rangle = \int \ell^2 P(\ell) d\ell$. Changing the order of integration, the integral over ℓ is immediate and we reobtain Eq. (48). A different way of obtaining Eq. (53) was used in Ref. [19]. This way involves the so-called on-collision pdf velocity distribution $f_{\text{coll}} = f(v)\mu(v)/\langle \mu \rangle$ (see Refs. [15,19,20]). An interesting short discussion of $P(\ell)$ can be found in Sec. II C of Ref. [19] (see also Ref. [15]).

It must be remarked that the quantity κ would be 1 for all m_1/m_2 if the simple approximation $P(\ell) \approx \exp(-\ell/\lambda)/\lambda$ for the free path length distribution Eq. (53) would hold. This was the approximation made to obtain D_{eKT} (see the end of Sec. II) as well as the Jeans formula (32) for the diffusion coefficient D_J (see the beginning of Sec. III). Of course, this is only an approximation. The exact value of $P(\ell)$ for hard spheres can be obtained from Eq. (53) as

$$\lambda P(\ell) = \frac{2(m_1/m_2)^2}{\pi(1 + m_1/m_2)} I_\ell(\ell/\lambda, m_1/m_2). \tag{54}$$

Here the function $I_\ell(x, m)$ is given by

$$I_\ell(x, m) = \int_0^\infty dy \, y \, E^2(y) \exp\left[-my^2 - \frac{E(y)}{y\sqrt{\pi(1+m)}} x\right], \tag{55}$$

where the quantity $E(y)$ is defined by Eq. (B3).

The (dimensionless) function $\lambda P(\ell)$ is plotted in Fig. 3 as a function of ℓ/λ for several values of the mass ratio m_1/m_2. It is quite apparent that Fig. 3 confirms that the expression $\exp(-\ell/\lambda)/\lambda$ is only an approximation of the true length distribution $P(\ell)$, as expected. Figure 3 also highlights that the approximation $P(\ell) \approx \exp(-\ell/\lambda)/\lambda$ becomes more accurate as the mass ratio m_1/m_2 decreases. From a physical point of view, this is an expected trend since $P(\ell) \approx \exp(-\ell/\lambda)/\lambda$ characterizes the free path distribution of an intruder moving in a gas composed of very massive molecules, which are, from
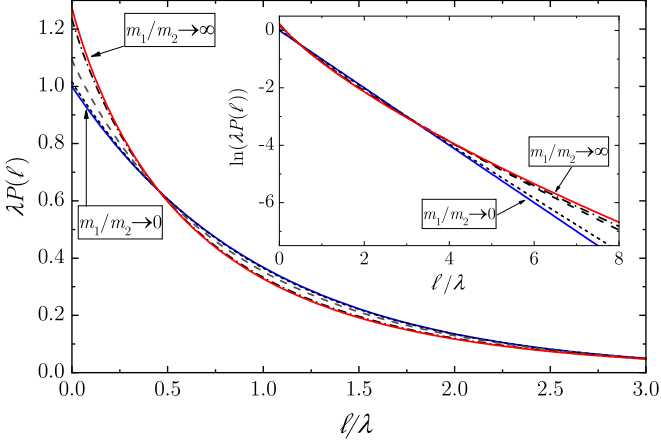
FIG. 3. Plot of $\lambda P(\ell)$ for hard spheres for $m_1/m_2 = 0.2$ (dotted line), $m_1/m_2 = 1$ (dashed line), and $m_1/m_2 = 10$ (dashed-dotted line). The approximation valid for a very light intruder, $\exp(-\ell/\lambda)/\lambda$, and for a very massive intruder, Eq. (56), are the solid lines. Same in the inset but as a semi-log plot.

the intruder's perspective, almost at rest. Mathematically, this result is obtained from the asymptotic expansion of $P(\ell)$ for $m \equiv m_1/m_2 \to 0$ (see Appendix C). On the other hand, the asymptotic expansion of $P(\ell)$ for $m \to \infty$ (very heavy intruder) leads to (see Appendix C)

$$\lambda P(\ell) \sim \frac{4}{\pi^{3/2}} G_{0,3}^{3,0}\left(0, \frac{1}{2}, 1 \,\middle|\, \frac{\ell^2}{\pi\lambda^2}\right), \tag{56}$$

where $G_{0,3}^{3,0}$ is a Meijer G function [21]. We have also plotted this expression in Fig. 3. It is remarkable that the line corresponding to the expression (56) is almost indistinguishable from the one obtained from Eq. (54) for $m = 10$.

C. The average $\langle \mathbf{r}_1 \cdot \mathbf{r}_k \rangle$ and the diffusion coefficient for hard spheres

We aim to evaluate the average $\langle r_1 r_k \cos \theta_{1,k} \rangle$. Following the same type of steps as those carried in Sec. III [using the spherical law of cosines, Eq. (20)], one finds the result

$$\langle \mathbf{r}_1 \cdot \mathbf{r}_k \rangle = \langle r_1 r_k \cos \theta_{1,2} \cos \theta_{2,3} \ldots \cos \theta_{k-1,k} \rangle. \tag{57}$$

More explicitly,

$$\langle \mathbf{r}_1 \cdot \mathbf{r}_k \rangle = \int \tilde{P}_k(1, \ldots, k) \, v_1 v_k \, \tau_1 \tau_k \cos \theta_{1,2} \ldots \cos \theta_{k-1,k}$$
$$\times \, d\mathbf{v}_1 d\tau_1 \ldots d\mathbf{v}_k d\tau_k. \tag{58}$$

Using Eq. (42) and integrating over all τ_i, Eq. (58) becomes

$$\langle \mathbf{r}_1 \cdot \mathbf{r}_k \rangle = \frac{1}{\langle \mu \rangle} \int f(\mathbf{v}_1) \frac{W(\mathbf{v}_1, \mathbf{v}_2) \cdots W(\mathbf{v}_{k-1}, \mathbf{v}_k)}{\mu(v_1) \cdots \mu(v_k)}$$
$$\times \, v_1 v_k \, \cos \theta_{1,2} \cdots \cos \theta_{k-1,k} \, d\mathbf{v}_1 \cdots d\mathbf{v}_k. \tag{59}$$

Using the collision rate (B1) and the transition rate (B4) of hard spheres in Eq. (59) and working in spherical coordinates, one finds that $\langle \mathbf{r}_1 \cdot \mathbf{r}_k \rangle$ can be written as

$$\langle \mathbf{r}_1 \cdot \mathbf{r}_k \rangle = \frac{2^{2-k}}{\pi^{3/2}} \left(\frac{m_1}{m_2}\right)^2 \left(1 + \frac{m_1}{m_2}\right)^{2k-5/2} \frac{1}{\left(n_2 \sigma_{12}^2\right)^2}$$
$$\times I_{1,k}(m_1/m_2), \tag{60}$$

for $k \geqslant 2$, where

$$I_{1,k}(m) = \int_0^\infty \cdots \int_0^\pi e^{-my_1^2} \frac{\widetilde{W}(\mathbf{y}_1, \mathbf{y}_2) \cdots \widetilde{W}(\mathbf{y}_{k-1}, \mathbf{y}_k)}{E(y_1) \cdots E(y_k)}$$
$$\times y_1^3 y_2^2 \cdots y_{k-1}^2 y_k^3 \cos \theta_{12} \cdots \cos \theta_{k-1,k}$$
$$\times \sin \theta_{12} \cdots \sin \theta_{k-1,k} dy_1 \cdots dy_k d\theta_{12} \cdots d\theta_{k-1,k}. \tag{61}$$

In units of λ^2, one gets

$$\frac{\langle \mathbf{r}_1 \cdot \mathbf{r}_k \rangle}{\lambda^2} = 2^{2-k} \sqrt{\pi} \left(\frac{m_1}{m_2}\right)^2 \left(1 + \frac{m_1}{m_2}\right)^{2k-3/2}$$
$$\times I_{1,k}(m_1/m_2), \tag{62}$$

for $k \geqslant 2$. Therefore, according to Eq. (10), we find that the collisional series can be written as

$$\ell_e^2 = \sum_{k=1}^\infty \frac{2^{3-k}}{\pi^{3/2}} \left(\frac{m_1}{m_2}\right)^2 \left(1 + \frac{m_1}{m_2}\right)^{2k-5/2} \frac{I_{1,k}(m_1/m_2)}{\left(n_2 \sigma_{12}^2\right)^2}. \tag{63}$$

Alternatively, taking into account Eqs. (50) and (62), the kth term of the reduced collisional series $c_k = 2\langle \mathbf{r}_1 \cdot \mathbf{r}_k \rangle / \langle r_1^2 \rangle$ is given by

$$c_k = \frac{\langle \mathbf{r}_1 \cdot \mathbf{r}_k \rangle}{\kappa \lambda^2}$$
$$= 2^{1-k} \left(1 + \frac{m_1}{m_2}\right)^{2(k-1)} \frac{I_{1,k}(m_1/m_2)}{I_{1,1}(m_1/m_2)}. \tag{64}$$

Note that the expression (64) also applies for $k = 1$. Hence, the reduced collisional series (10) can be written as

$$\frac{\ell_e^2}{\langle r_1^2 \rangle} = \sum_{k=1}^\infty 2^{1-k} \left(1 + \frac{m_1}{m_2}\right)^{2k-2} \frac{I_{1,k}(m_1/m_2)}{I_{1,1}(m_1/m_2)}. \tag{65}$$

From Eq. (18), the diffusion coefficient D is

$$D = D_{\text{eKT}} \frac{\langle r_1^2 \rangle}{2\lambda^2} \sum_{k=1}^\infty 2^{2-k} \left(1 + \frac{m_1}{m_2}\right)^{2k-2} \frac{I_{1,k}(m_1/m_2)}{I_{1,1}(m_1/m_2)}. \tag{66}$$

Of course, the problem with the series (66) is that the kth term requires the evaluation of $I_{1,k}$, namely, a $2k - 1$-dimensional integral which is unapproachable for large k. However, the numerical evaluation of the first few terms is accessible. We show this in Sec. VI.

Before finishing this section it is worthwhile to compare the integral forms of $\langle \mathbf{r}_1 \cdot \mathbf{r}_k \rangle$ and the mean persistence ratio ω. We start noticing that the integrand of Eq. (59) has an interesting form. Let us denote this integrand by the symbol $A_{1,k}$, that is,

$$\langle \mathbf{r}_1 \cdot \mathbf{r}_k \rangle = \int A_{1,k} d\mathbf{v}_1 \cdots d\mathbf{v}_k. \tag{67}$$

We see that the integrand of $\langle \mathbf{r}_1 \cdot \mathbf{r}_k \rangle$ and the integrand of $\langle \mathbf{r}_1 \cdot \mathbf{r}_{k-1} \rangle$ just differ by the factor $p_{k-1,k}$, that is $A_{1,k} = A_{1,k-1} p_{k-1,k}$, where

$$p_{k-1,k} = \frac{W(k-1, k)}{\mu(k)} \frac{v_k}{v_{k-1}} \cos \theta_{k-1,k}. \tag{68}$$

The form of $A_{1,k} = A_{1,k-1} p_{k-1,k}$ suggests that $\langle \mathbf{r}_1 \cdot \mathbf{r}_{k-1} \rangle$, $\langle \mathbf{r}_1 \cdot \mathbf{r}_k \rangle$, etc., might follow a sequence with a ratio between

TABLE I. Numerical and DSMC simulation values of $\langle \mathbf{r}_1 \cdot \mathbf{r}_k \rangle$, with $k = 1, 2, 3, 4$, in units of λ^2 for several values of m_1/m_2. For each value of the mass ratio m_1/m_2, the first row are the numerical values and the second row are the DSMC values.

m_1/m_2	$\langle r_1^2 \rangle$	$\langle \mathbf{r}_1 \cdot \mathbf{r}_2 \rangle$	$\langle \mathbf{r}_1 \cdot \mathbf{r}_3 \rangle$	$\langle \mathbf{r}_1 \cdot \mathbf{r}_4 \rangle$
0.2	2.0179	0.1198	0.0145	0.0018
	2.0174	0.1198	0.0145	0.0018
0.5	2.0529	0.2621	0.0686	0.0183
	2.0524	0.2619	0.0686	0.0183
1	2.1012	0.4331	0.1818	0.0772
	2.1009	0.4329	0.1818	0.0772
2	2.1640	0.6386	0.3799	0.2269
	2.1636	0.6383	0.3798	0.2268
5	2.2458	0.8838	0.6963	0.5489
	2.2454	0.8833	0.6960	0.5487
10	2.2915	1.0103	0.8910	0.7858
	2.2911	1.0100	0.8907	0.7856

consecutive terms related to the integration of $p_{k-1,k}$. Note the closeness of $p_{k-1,k}$ to the integrand of the integral

$$\omega = \int f(v_{k-1}) \frac{W(v_{k-1}, v_k)}{\langle \mu \rangle} \frac{v_k}{v_{k-1}} \cos \theta_{k-1,k} d\mathbf{v}_{k-1} d\mathbf{v}_k \quad (69)$$

that provides the mean persistence ratio. This observation suggests the proportionality relationship between the ratio of two consecutive terms of the collisional series and the mean persistence ratio that we discussed in Secs. III and IV.

VI. COMPARISON WITH DSMC SIMULATIONS AND BOLTZMANN RESULTS FOR HARD SPHERES

In this section we estimate the terms c_k of the reduced collisional series by numerically integrating $I_{1,k}$; this result will be assessed via a comparison with DSMC simulations. Additionally, the theoretical predictions for the diffusion coefficient given by Eqs. (32) and (36) and by the first- and second-Sonine approximations (one and two terms in the Sonine polynomial expansion of the velocity distribution function) to the Chapman-Enskog solution [1] to the Boltzmann equation will be also compared with simulation data obtained from the DSMC method [13]. It is important to remark that both DSMC simulations and Sonine approximations are derived through the Boltzmann kinetic equation, which plays a pivotal role in modeling the kinetic behavior of gaseous particles. This approach provides a solid theoretical foundation for our comparisons, enabling us to assess the validity and accuracy of our predictions in relation to well-established and widely accepted methods for describing transport phenomena in gaseous systems.

In Table I, we present numerical estimates for the first four terms of the collisional series (8) for various representative values of the mass ratio m_1/m_2. These estimates require the numerical integration of $I_{1,k}$ [see Eqs. (50), (51), (60), and (61)]. As k increases, the numerical integration becomes progressively challenging, leading to less precise results, particularly for small values of m_1/m_2. We also include in Table I values obtained through DSMC simulations at very low densities. Some technical details on the implementation of the

TABLE II. The first three ratios c_2/c_1, c_3/c_2, and c_4/c_3 of the coefficients of the reduced collisional series (10) for several values of m_1/m_2 evaluated by means of Eq. (64). DSMC simulations provide the same results for all cases with the exception of c_4/c_3 for $m_1/m_2 = 0.2$ and $m_1/m_2 = 1$; in these cases the simulation results are 0.127 and 0.424, respectively. The last column is the value of the mean persistence ratio ω for hard spheres [see Eq. (B11)].

m_1/m_2	c_2/c_1	c_3/c_2	c_4/c_3	ω
0.2	0.119	0.121	*0.126*	0.107
0.5	0.255	0.262	0.267	0.245
1	0.412	0.420	*0.425*	0.406
2	0.590	0.595	0.597	0.587
5	0.787	0.788	0.788	0.785
10	0.882	0.882	0.882	0.881

DSMC method to the evaluation of the averages $\langle \mathbf{r}_1 \cdot \mathbf{r}_k \rangle$ are provided in the Appendix D. Specifically, we consider approximately 10^5 molecules, conducting up to 5×10^5 runs, each comprising an average of 500 proposed candidates to collide in every Monte Carlo step, with the final number of collisions depending on the mass ratio m_1/m_2. Table I highlights the excellent agreement between DSMC simulations and the numerical results. From these results, the values of the first ratios between successive terms of the reduced collisional series can be derived. These values are given in Table II. Additionally, we provide the corresponding values of the mean persistence ratio ω. The proximity of the coefficients c_k/c_{k-1} to ω is quite remarkable. Note that this closeness improves for heavier intruders. In Fig. 4 we compare the first three ratios of the reduced collisional series, c_2/c_1, c_3/c_2 and c_4/c_3, with ω for a large set of mass ratios. The agreement is stunning, providing strong support for the approximation $c_{1+k}/c_k \approx \omega$ or, equivalently, for $c_{1+k} \approx \omega^k$ (recall that $c_1 = 1$). This result is the cornerstone of the approximations (31) and (34) for the MSD, and of the approximations (32) and (36) for the
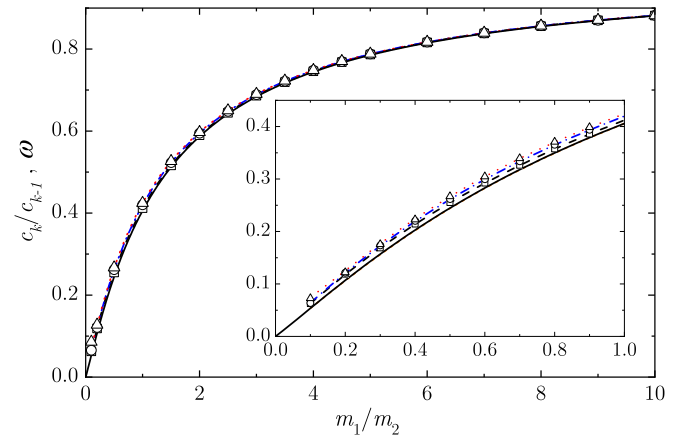

FIG. 4. Plot of c_2/c_1 (dashed line, squares), c_3/c_2 (dashed-dotted line, circles), and c_4/c_3 (dotted line, triangles) vs the mass ratio m_1/m_2. The lines are obtained by numerically evaluating the expression (64) for each mass ratio while the symbols are DSMC simulation results. The solid line is the mean persistence ratio ω. Inset: The same when the intruder is lighter than the molecules of the gas.

diffusion coefficient. These findings highlight the reliability of the random-walk approach.

It has long been recognized that a connection exists between the random-walk formulation of the diffusion in gases and the Jeans formulation, which involves the persistence of velocity. It is therefore surprising that the possibility (as demonstrated in this paper) of deriving Jeans' diffusion coefficients from a random-walk description of molecular diffusion has gone unnoticed. We suspect that this oversight may be attributed to the apparent disparity between the two formulations (Jeans and random-walk approaches) when relying on Yang's (very rough and poor) estimate of $\langle \mathbf{r}_i \cdot \mathbf{r}_{i+k} \rangle$ for hard spheres with $k = 1, 2$ [6]. For $m_1 = m_2$, the values of c_2/c_1 and c_3/c_2 that one obtains from Yang's results (see Appendix C of Ref. [6]) would be 0.72 and 0.47, respectively. These values are very different from each other (and from the true values 0.412 and 0.420 we give in Table II) and from the mean persistence ratio 0.406 corresponding to this case. These imprecise estimates by Yang are understandable given the complexity of the integrals $I_{1,k}$, and that Yang lacked the computational means to numerically evaluate them. Had Yang been able to compute the first few terms $\langle \mathbf{r}_1 \cdot \mathbf{r}_{1+k} \rangle$ accurately, it is likely that he or others (Monchick, for example: See Refs. [8] and [12], specifically compare Sec. IV of Ref. [12] with the last paragraph of Sec. IV of Ref. [8]) would have recognized the close resemblance of the ratio between these terms to the mean persistence ratio. Such a realization would have unveiled the connection between Yang's random-walk formulas and those of Jeans.

From the relation $c_2 \approx c_{1+k}/c_k \approx \omega$, we have obtained $c_{1+k} \approx \omega^k$. Employing this result in Eqs. (10) and (18), we achieve the expressions (33) and (36) in Sec. IV. Note, however, that the approximation $c_2 \approx c_{1+k}/c_k$ also implies $c_{1+k} \approx c_2^k$. The use of this result in Eqs. (10) and (18) yields the expressions

$$\langle R^2(t) \rangle = \langle N(t) \rangle \frac{\langle r_1^2 \rangle}{1 - c_2}, \tag{70}$$

and

$$D = D_{\text{eKT}} \frac{\kappa}{1 - c_2} = \frac{\lambda \bar{v}}{d} \frac{\kappa}{1 - c_2} \equiv D_{\text{RW}}'. \tag{71}$$

Certainly, choosing for ω as the estimate for c_{1+k}/c_k offers the advantage of having the closed explicit expression (B11), while the evaluation of c_2 requires numerical computation. On the other hand, as we will see next, D_{RW}' turns out to be somewhat better than D_{RW}.

In Table III we compare the diffusion coefficients provided by the random-walk approach, Eqs. (32), (36), and (71), with the diffusion coefficients provided by two standard approximations of the kinetic theory of gases, namely the first and second Sonine approximations. In terms of D_{eKT}, the diffusion coefficient $D_{\text{KT}}^{[1]}$ for hard spheres obtained from the first Sonine approximation is given by (see Sec. 14.2 of Ref. [1])

$$\frac{D_{\text{KT}}^{[1]}}{D_{\text{eKT}}} = \frac{9\pi}{32} \left(1 + \frac{m_1}{m_2} \right). \tag{72}$$

TABLE III. The diffusion coefficients D_J, D_{RW}, D_{RW}', $D_{\text{KT}}^{[1]}$, and $D_{\text{KT}}^{[2]}$, as well as DSMC simulations values in units of $D_{KT}^{[0]}$ for several values of m_1/m_2.

m_1/m_2	D_J	D_{RW}	D_{RW}'	$D_{\text{KT}}^{[1]}$	$D_{\text{KT}}^{[2]}$	DSMC
0.2	1.12	1.13	1.14	1.06	1.12	1.15
0.5	1.32	1.36	1.38	1.33	1.37	1.40
1	1.68	1.77	1.79	1.77	1.80	1.82
2	2.42	2.62	2.64	2.65	2.67	2.67
5	4.66	5.23	5.27	5.30	5.31	5.34
10	8.40	9.63	9.70	9.72	9.72	9.84

The second-Sonine approximation $D_{\text{KT}}^{[2]}$ is (see Sec. 14.3 of Ref. [1])

$$\frac{D_{\text{KT}}^{[2]}}{D_{\text{eKT}}} = \frac{9\pi}{32} \left(1 + \frac{m_1}{m_2} \right) \frac{13 + 16m_1/m_2 + 30(m_1/m_2)^2}{12 + 16m_1/m_2 + 30(m_1/m_2)^2}. \tag{73}$$

We plot in Fig. 5 the diffusion coefficients D_J, D_{RW}, $D_{\text{KT}}^{[1]}$, and $D_{\text{KT}}^{[2]}$ as a function of the mass ratio m_2/m_1 when the intruder is heavier than the molecules of the gas ($m_1 > m_2$). The complementary case of an intruder lighter than the molecules of the gas ($m_1 < m_2$) is considered in Fig. 6. In Figs. 5 and 6 we show the results of DSMC simulations. We see that the simplest formula with persistence, D_J [Eq. (32)] is a significant improvement over the diffusion coefficient of the elementary kinetic theory D_{eKT} and qualitatively captures the effect of the mass ratio of the random-walk coefficient. However, at a quantitative level, this approach is not fully satisfactory. On the contrary, the improved random-walk formula [given by the coefficient D_{RW}, see Eq. (36)] provides excellent results. In fact, when the intruder is lighter than the molecule of the gas, it is even better than the first Sonine approximation. In the opposite case, when the intruder is heavier than the molecule of the gas, the results are quite good, slightly worse
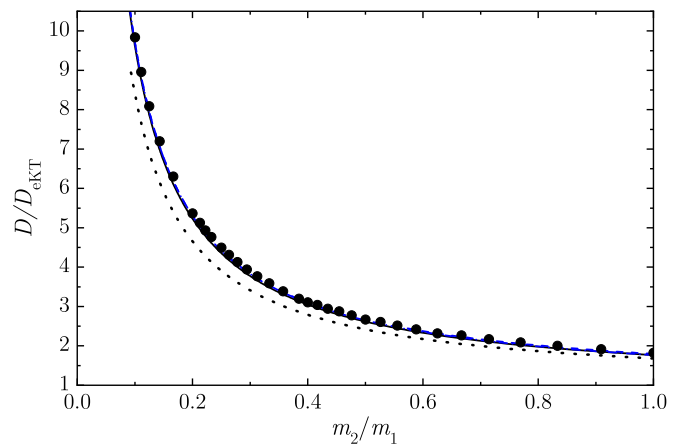

FIG. 5. Plot of the reduced diffusion coefficient D/D_{eKT} as a function of the mass ratio m_2/m_1 when the intruder is heavier than the molecules of the gas. The lines correspond to $D = D_J$ (dotted line), $D = D_{\text{RW}}$ (solid line), $D = D_{\text{KT}}^{[1]}$ (dashed-dotted line), and $D = D_{\text{KT}}^{[2]}$ (dashed line). The circles are DSMC simulation values.
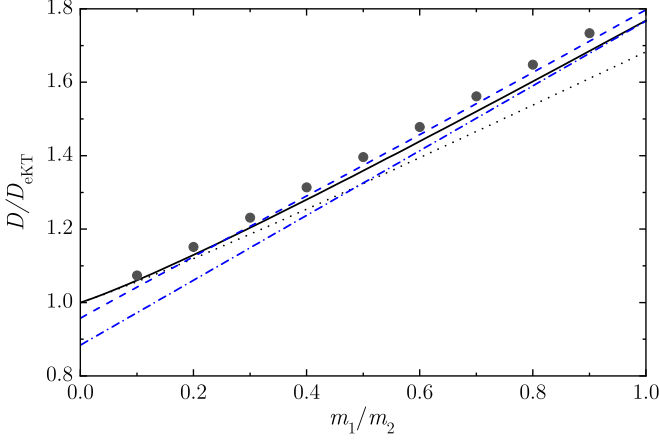
FIG. 6. Plot of the reduced diffusion coefficient D/D_{eKT} as a function of the mass ratio m_1/m_2 when the intruder is lighter than the molecules of the gas. The lines correspond to $D = D_J$ (dotted line), $D = D_{RW}$ (solid line), $D = D_{KT}^{[1]}$ (dashed-dotted line), and $D = D_{KT}^{[2]}$ (dashed line). The circles are DSMC simulation values.

than, but comparable, to the results obtained with the Sonine approximations.

VII. CONCLUSIONS

In this paper we have used the fact that mean-square displacement of a molecule in a gas at time t can be written as the mean-square displacement of a molecule with isotropic collisions, $\langle R^2(t) \rangle = \langle N(t) \rangle \langle r_1^2 \rangle$, corrected by the prefactor (the reduced collisional series) $\sum_{k=1}^{\infty} c_k = 1 + 2 \sum_{k=2}^{\infty} \langle \mathbf{r}_1 \cdot \mathbf{r}_k \rangle / \langle r_1^2 \rangle$. The key result of our paper is the realization that c_{1+k} can be well approximated by ω^k, leading to the result

$$\sum_{k=1}^{\infty} c_k = \sum_{k=0}^{\infty} \omega^k = \frac{1}{1 - \omega}, \qquad (74)$$

and, hence,

$$\langle R^2(t) \rangle = \langle N(t) \rangle \frac{\langle r_1^2 \rangle}{1 - \omega}. \qquad (75)$$

This implies that, to estimate the MSD of a molecule in a gas at equilibrium, only two quantities need to be known: the mean-square free path $\langle r_1^2 \rangle$, and the mean persistence ratio ω. Expressed as $\langle R^2(t) \rangle / (2dt)$, the diffusion coefficient is the value associated with purely isotropic collisions (i.e., $\lambda \bar{v}/d$), corrected by the factor $(1 - \omega)^{-1}$. We have justified this result in two ways: first, through simple and intuitive arguments (that involve some uncontrolled approximations) and, second, by rigorously evaluating the first few terms c_k of the reduced collisional series and realizing that, for hard spheres, these terms form an almost perfect geometric sequence with a common ratio very close to ω.

The formulas we have derived for the diffusion coefficient, Eqs. (32) and (36), are found to be identical to those proposed by Jeans long ago through a different pathway. While Jeans' arguments were insightful and demonstrated a mastery of free path arguments, his deduction of the corrective terms was somewhat hand-waving and lacked a clear identification of their nature—specifically, that these corrective terms are

$2\langle \mathbf{r}_k \cdot \mathbf{r}_{1+k} \rangle / \langle r_1^2 \rangle$. In contrast, our approach not only identifies the nature of the corrective terms but also allows us to quantify the validity of the approximations, at least for the first few terms. We have verified this both numerically and through DSMC simulations for the particular case of hard spheres.

For a dilute gas of hard spheres, we have compared the theoretical predictions of the diffusion coefficient obtained from the random-walk and kinetic theory (first and second Sonine approximations to the Chapman-Enskog solution to the Boltzmann equation) approaches with those obtained by numerically solving the Boltzmann equation by means of the DSMC method [13]. We have found that the results provided by the improved random-walk formula were quite good, surpassing even the first Sonine approximation for intruders lighter than the gas molecules. Furthermore, the improved random-walk results closely approximate the second Sonine coefficient for a wide range of molecule mass values.

Finally, we note that the generalization of the present random-walk approach to mixtures with arbitrary concentration and/or other type of interactions and dimensions can be easily performed. As a different project, we also plan to extend the present results to the case of granular gases modeled as a gas of smooth hard spheres with inelastic collisions [14]. Work along these lines is currently in progress.

APPENDIX A: AN ALTERNATIVE DEDUCTION OF EQ. (29)

Here we present an alternative approach to derive Eq. (29) in Sec. III. We begin by assuming that the sizes of the steps 1 and k, as well as their angles, are uncorrelated, so that

$$\langle \mathbf{r}_1 \cdot \mathbf{r}_k \rangle \approx \langle r_1 \rangle \langle r_k \rangle \langle \cos \theta_{1,k} \rangle = \lambda^2 \langle \cos \theta_{1,k} \rangle. \qquad (A1)$$

Next, using the spherical cosine law Eq. (20) and considering that $\varphi_{1,2,k}$ is equiprobable, we find $\langle \cos \theta_{1k} \rangle \approx \langle \cos \theta_{1,2} \cos \theta_{2,k} \rangle$. Now we assume that the dispersion angles of $\theta_{1,2}$ and $\theta_{2,k}$ for two successive collisions are uncorrelated (or weakly correlated), so that $\langle \cos \theta_{1k} \rangle \approx \langle \cos \theta_{1,2} \rangle \langle \cos \theta_{2,k} \rangle$. We estimate $\langle \cos \theta_{2,k} \rangle$ as we did with $\langle \cos \theta_{1,k} \rangle$, and so on. The final result is

$$\langle \cos_{1k} \rangle = \prod_{i=1}^{k-1} \langle \cos \theta_{i,i+1} \rangle. \qquad (A2)$$

At equilibrium, the value of i is not relevant and this equation becomes $\langle \cos \theta_{1k} \rangle = \langle \cos \theta_{1,2} \rangle^{k-1}$. Now we use Eq. (27) and take the approximation $\omega \approx \langle v_2/v_1 \rangle \langle \cos \theta \rangle$. We know that $\langle v_2 \rangle = \langle v_1 \rangle$. This prompts us to make the approximation $\langle v_2/v_1 \rangle \approx 1$. Thus, $\langle \cos \theta_{1,2} \rangle \approx \omega$ and Eq. (A1) becomes Eq. (29).

APPENDIX B: SOME BASIC QUANTITIES FOR HARD SPHERES

In this Appendix, for reference, we provide results on some basic quantities (mean collision time, mean and variance of the free path, free path distribution, and mean persistence ratio) that appear in the random-walk approach for the particular case of a gas of hard spheres at equilibrium. Recall that in this case the distribution of velocities $f(\mathbf{v})$ is the Maxwell-Boltzmann distribution (43).

To evaluate averages with respect to the trajectory of a molecule, we need to know the pdf P_N. From Eq. (42) we see that this requires knowing both the collision rate μ and the transition rate W. For hard spheres, the collision rate is given by [1,6]

$$\mu(\mathbf{v}) = n_2 \sigma_{12}^2 \left(\frac{2\pi k_B T}{m_2} \right)^{1/2} E(y), \tag{B1}$$

where

$$\mathbf{y} = \left(\frac{m_2}{2k_B T} \right)^{1/2} \mathbf{v}, \tag{B2}$$

and [22]

$$E(y) = e^{-y^2} + \left(2y + \frac{1}{y} \right) \frac{\sqrt{\pi}}{2} \mathrm{Erf}(y). \tag{B3}$$

Here T is the temperature of the gas, n_2 is the number density of molecules of the gas, and $\sigma_{12} = (\sigma_1 + \sigma_2)/2$, where σ_1 and σ_2 are the diameters of the intruder and the molecules of the gas, respectively, and Erf is the error function.

Formula (B1) and those we will give hereafter are valid for dilute gases, i.e., for very small n_2. One can extend the validity of these formulas to larger values of n_2 by simply realizing that the local density of molecules with which the intruder collides is $g(\sigma_{12})n_2$, with $g(r)$ being the radial distribution function of the gas. This implies substituting n_2 for $g(\sigma_{12})n_2$ in all expressions where n_2 appears (which we will not do in this article). Precise values of the radial distribution function at contact, $g(\sigma_{12})$, for mixtures of elastic hard spheres are well known (see Ref. [23], for example).

The transition rate $W(\mathbf{v}_1, \mathbf{v}_2)$ for hard spheres is [6,19]

$$W(\mathbf{v}_1, \mathbf{v}_2) = \frac{n_2 \sigma_{12}^2}{\sqrt{\pi}} \left(\frac{m_2}{2k_B T} \right) \left(\frac{m_1 + m_2}{2m_2} \right)^2 \widetilde{W}(\mathbf{y}_1, \mathbf{y}_2) \tag{B4}$$

with

$$\widetilde{W}(\mathbf{y}_1, \mathbf{y}_2) = \frac{\exp \left\{ - \left[\frac{m_1+m_2}{2m_2} |\mathbf{y}_2 - \mathbf{y}_1| + \mathbf{y}_1 \cdot \frac{\mathbf{y}_2 - \mathbf{y}_1}{|\mathbf{y}_2 - \mathbf{y}_1|} \right]^2 \right\}}{|\mathbf{y}_2 - \mathbf{y}_1|}. \tag{B5}$$

From Eq. (B1) one finds the well-known values of the average collision rate,

$$\langle \mu \rangle = \int d\mathbf{v} \, f(\mathbf{v}) \mu(\mathbf{v})$$
$$= 2\sqrt{\pi} \left(\frac{m_1 + m_2}{m_1} \right)^{1/2} \left(\frac{2k_B T}{m_2} \right)^{1/2} n_2 \sigma_{12}^2, \tag{B6}$$

and the mean free path,

$$\lambda \equiv \langle r_1 \rangle = \int_0^\infty d\tau_1 \int d\mathbf{v}_1 \, \tilde{P}_1(\mathbf{v}_1)(v_1 \tau_1)$$
$$= \frac{1}{\pi (1 + m_1/m_2)^{1/2} n_2 \sigma_{12}^2}. \tag{B7}$$

Here the relation $r_1 = v_1 \tau_1$ has been used. From these results and taking into account Eq. (1) and that $\bar{v} = \lambda \langle \mu \rangle = (8k_B T/\pi m_1)^{1/2}$, one gets

$$D_{\mathrm{eKT}} = \frac{1}{3\pi (1 + m_1/m_2)^{1/2} n_2 \sigma_{12}^2} \sqrt{\frac{8k_B T}{\pi m_1}} \tag{B8}$$

for hard spheres.

With the expression (B4) at hand, one can evaluate the mean persistence ratio as

$$\omega = \int \tilde{P}_2(1, 2) \frac{v_2}{v_1} \cos \theta_{12} d\mathbf{v}_1 d\mathbf{v}_2 d\tau_1 d\tau_2, \tag{B9}$$

or, equivalently, as

$$\omega = \left\langle \frac{v_2}{v_1} \cos \theta_{12} \right\rangle$$
$$= \frac{1}{\langle \mu \rangle} \int f(\mathbf{v}_1) W(\mathbf{v}_1, \mathbf{v}_2) \frac{v_2}{v_1} \cos \theta_{12} d\mathbf{v}_1 d\mathbf{v}_2. \tag{B10}$$

The integral (B10) can be evaluated explicitly (see Eq. (5.51,2) of Ref. [1]) and the result is

$$\omega = \frac{m_1/m_2 + \widetilde{\omega}(m_1/m_2)}{1 + m_1/m_2}, \tag{B11}$$

where

$$\widetilde{\omega}(m) = \frac{m}{2} \left[\frac{m}{\sqrt{1+m}} \ln \left(\frac{1 + \sqrt{1+m}}{\sqrt{m}} \right) - 1 \right]. \tag{B12}$$

Figure 4 shows the mean persistence ratio ω (apart from the ratios c_k/c_{k-1}) as a function of the mass ratio m_1/m_2. We see that ω is smaller than unity unless $m_1/m_2 \to \infty$. This implies that the approximation Eq. (39) is excellent, except when m_1/m_2 becomes very large. The fact that $\omega \to 1$ when $m_1 \gg m_2$ makes sense since when the intruder is heavier than the particles of the gas, it retains more of its trajectory and velocity after collisions. Conversely, if $m_1/m_2 \to 0$, then we see that $\omega \to 0$. This is because the lighter the intruder, the less capacity it has to keep its trajectory and velocity after collisions.

APPENDIX C: ASYMPTOTIC EXPRESSIONS OF THE PDF $P(\ell)$ OF THE FREE PATH

In this Appendix we provide the asymptotic expressions for the pdf $P(\ell)$ given in Eq. (54) when $m \to 0$ and/or $m \to \infty$ (for ℓ finite) and when $\ell \to \infty$ (for m finite). The expression of $P(\ell)$ in the case of a light intruder moving in a sea of very massive molecules can be obtained from the asymptotic evaluation of the integral $I_\ell(x, m)$ of Eq. (55) for $m \to 0$. In this limiting case, only large values of y are relevant in the integrand of $I_\ell(x, m)$. Thus, when $y \to \infty$, $E(y) \sim \sqrt{\pi} y$ and

$I_\ell(x, m \to 0)$ behaves as

$$I_\ell(x, m \to 0) \sim \pi e^{-x} \int_0^\infty y^3 e^{-my^2} dy$$

$$\sim \frac{\pi}{2m^2} e^{-x}. \tag{C1}$$

Introducing this result into Eq. (54), one gets

$$\lambda P(\ell) \sim \exp(-\ell/\lambda) \tag{C2}$$

for $m \to 0$.

An asymptotic expression in the opposite limit of a very massive intruder moving in a sea of very light particles ($m \to \infty$) can also be obtained when one realizes that for $m \to \infty$ only small values of y are relevant in the integrand of $I_\ell(x, m)$. Taking into account that $E(y) \sim 2$ for $y \to 0$, one finds

$$I_\ell(x, m) \sim 4 \int_0^\infty y \exp\left[-my^2 - \frac{2x}{y\sqrt{\pi(1+m)}}\right] dy$$

$$\sim \frac{2x^2}{(1+m)\pi^{3/2}} G_{0,3}^{3,0}\left[-1, -\frac{1}{2}, 0 \,\middle|\, \frac{mx^2}{(1+m)\pi}\right], \tag{C3}$$

where $G_{0,3}^{3,0}$ is a Meijer G function. Substituting Eq. (C3) into Eq. (54), and taking the limit $m \to \infty$ in the resulting expression, one finds

$$\lambda P(\ell) \sim \frac{4}{\pi^{3/2}} G_{0,3}^{3,0}\left(0, \frac{1}{2}, 1 \,\middle|\, \frac{\ell^2}{\pi\lambda^2}\right). \tag{C4}$$

From the behavior of the Meijer G function $G_{0,3}^{3,0}$ for small and large arguments [21], one gets

$$\lambda P(\ell) \sim \frac{4}{\pi} \tag{C5}$$

for small ℓ/λ, and

$$\lambda P(\ell) \sim \frac{8}{\sqrt{3}\pi^{2/3}} \left(\frac{\ell}{\lambda}\right)^{1/3} \exp\left[-\frac{3}{\pi^{1/3}} \left(\frac{\ell}{\lambda}\right)^{2/3}\right] \tag{C6}$$

for large ℓ/λ. Finally, for $\ell/\lambda \to \infty$ but finite m, only large values of y are relevant in the integrand of $I_\ell(x, m)$. In this case,

$$I_\ell(x, m) \sim \pi e^{-x/\sqrt{1+m}} \int_0^\infty y^3 e^{-my^2} dy$$

$$\sim \frac{\pi}{2m^2} e^{-x/\sqrt{1+m}} \tag{C7}$$

and, hence, $\lambda P(\ell)$ behaves as

$$\lambda P(\ell) \sim \frac{1}{1+m} \exp\left(-\frac{\ell/\lambda}{\sqrt{1+m}}\right). \tag{C8}$$

For $m = 1$ this expression reduces to that obtained in Ref. [19]. When $m = 0$, the asymptotic form of $P(\ell)$ for $m \to 0$ is recovered.

APPENDIX D: DIRECT SIMULATION MONTE CARLO METHOD

The DSMC method is employed in this paper to assess the various approximations made for calculating both the

coefficients of the reduced collisional series c_k and the diffusion coefficient D_{RW}.

In the context of a rarefied gas regime dominated by collisions, the DSMC method emerges as an alternative yet complementary approach for solving the Boltzmann equation. Originally introduced by Bird [24], classical DSMC simulations were specifically designed to address rarefied gas flows that are computationally inaccessible via molecular dynamics simulations [13,25]. In this work, we follow similar steps as those proposed by Pöschel and Schwager [26] to solve the Boltzmann equation for a mixture composed by hard spheres.

The simulation is initiated by drawing the particle velocities from the corresponding Maxwell distribution f_i [see Eq. (43)] at an initial temperature T_0 for each species i. The discretized distribution function $f_i^{(\mathcal{N}_i)}$ of species i is derived from the velocities $\{v_k\}$ of $\mathcal{N}_i$ "virtual particles" as given by:

$$f_i^{(\mathcal{N}_i)}(\mathbf{v}; t) \to \frac{n_i}{N_i} \sum_{k=1}^{\mathcal{N}_i} \delta[\mathbf{v} - \mathbf{v}_k(t)], \tag{D1}$$

where δ is the Dirac delta function. For low-density regimes, since the collisions are assumed to be instantaneous, the free flight of particles and the collision stage can be temporally decoupled. The DSMC method maintains this assumption, leading to a division into two distinct steps: the convective and collision stages. However, since our simulations assume equilibrium for the molecular gas, only the collision stage is detailed here. The "general" procedure can be summarized as follows:

(1) To simulate collisions between particles of species i and j, a required number of $N_{ij}^{\Delta t}$ candidate pairs for collision in a time Δt is selected. This number is determined by [26]:

$$\mathcal{N}_{ij}^{\Delta t} = \pi \mathcal{N}_i n_j \sigma_{ij}^2 g_{ij}^{\max} \Delta t, \tag{D2}$$

where $\mathcal{N}_i$ is the total number of simulated particles of species i, and $g_{ij}^{\max}$ is an upper bound for the average relative velocity between two particles. A suitable estimate is $g_{ij}^{\max} = C v_{ij}^{\text{th}}$, where $v_{ij}^{\text{th}} = \sqrt{2T_0/\overline{m}}$ is the mean thermal velocity, $\overline{m} = (m_i + m_j)/2$, and C is a constant, e.g., $C = 5$ [13].

(2) A colliding direction $\widehat{\boldsymbol{\sigma}}_{k\ell}$ for a pair of colliding particles labeled as k and ℓ is randomly selected with equal probability.

(3) The collision is accepted if

$$|\widehat{\boldsymbol{\sigma}}_{k\ell} \cdot \mathbf{g}_{k\ell}| \equiv |\widehat{\boldsymbol{\sigma}}_{k\ell} \cdot (\mathbf{v}_k - \mathbf{v}_\ell)| > \mathcal{R}(0, 1) g_{ij}^{\max}, \tag{D3}$$

where $\mathcal{R}(0, 1)$ is a uniformly distributed random number in the interval [0,1].

(4) If the collision is accepted, then the velocities of particles are updated according to the scattering rules [14]:

$$\mathbf{v}_k \to \mathbf{v}_k - 2\mu_{ji}(\mathbf{g}_{k\ell} \cdot \widehat{\boldsymbol{\sigma}}_{k\ell})\widehat{\boldsymbol{\sigma}}_{k\ell},$$

$$\mathbf{v}_\ell \to \mathbf{v}_\ell + 2\mu_{ij}(\mathbf{g}_{k\ell} \cdot \widehat{\boldsymbol{\sigma}}_{k\ell})\widehat{\boldsymbol{\sigma}}_{k\ell}, \tag{D4}$$

being $\mu_{ij} = m_i/(m_i + m_j)$.

(5) Repeat the procedure for all the permutations of $i = 1$, 2 and $j = 1, 2$.

Since our interest lies in studying a mixture where one species is present in tracer concentration, certain modifications are made to the algorithm. Assuming $i \equiv 1$ (intruders)

and $j \equiv 2$ (molecular gas), collisions 1-1 are neglected, and when a collision 1-2 or 2-1 occurs, only the velocity of the intruder is modified according to the scattering rules given in step 4. Although the molecular gas is inherently in equilibrium, described by the Maxwell distribution, it is crucial to consider collisions between molecules (labeled as 2-2) in the algorithm. This consideration helps prevent correlations stemming from a distribution biased by the initial velocities assigned to the finite number of molecules $\mathcal{N}_2$.

We measure the diffusion coefficient D of an intruder immersed in a dilute gas at equilibrium by employing the Einstein relation:

$$D = \frac{1}{6\Delta t}[\langle R^2(t+\Delta t)\rangle - \langle R^2(t)\rangle]. \quad (D5)$$

Here $\langle \cdots \rangle$ denotes the average over the $\mathcal{N}_1$ intruders.

1. Coefficients of the reduced collisional series c_k

To measure the projection over a free path of the subsequent nth free path, we must track the trajectory of the intruders between collisions to find $\mathbf{r}_n$. What one would be tempted to do would be to perform averaging of $\mathbf{r}_1 \cdot \mathbf{r}_k$ over the subset of realizations where the algorithm detects the very first collision of the intruder with a molecule of the gas. However, this first accepted collision is likely to be one in which the relative velocity between the colliding particles is quite large, biasing the distribution function of the particles involved in the averaged quantities [see Eq. (D3)]. Therefore, to avoid that, the coefficients c_k are computed by considering the position of the intruder after every collision n by

$$c_k = 2\frac{\langle \mathbf{r}_n \cdot \mathbf{r}_{n+k}\rangle}{\langle r_n^2\rangle}. \quad (D6)$$

This ensures that f_i is the correct one.

[1] S. Chapman and T. G. Cowling, *The Mathematical Theory of Nonuniform Gases* (Cambridge University Press, Cambridge, UK, 1970).
[2] C. Cercignani, *Ludwig Boltzmann. The Man Who Trusted Atoms* (Oxford University Press, Oxford, 1998).
[3] F. Reif, *Fundamentals of Statistical and Thermal Physics* (McGraw–Hill, Tokyo, 1965).
[4] D. A. McQuarrie, *Statistical Mechanics* (Harper & Row, New York, 1975).
[5] W. H. Furry, On the elementary explanation of diffusion phenomena in gases, Am. J. Phys. **16**, 63 (1948).
[6] L. M. Yang, Kinetic theory of diffusion in gases and liquids I. Diffusion and the Brownian motion, Proc. R. Soc. Lond. A **198**, 94 (1949).
[7] W. H. Furry and P. H. Pitkanen, Gaseous diffusion as a random process, J. Chem. Phys. **19**, 729 (1951).
[8] L. Monchick, Equivalence of the Chapman-Enskog and the mean-free-path theory of gases, Phys. Fluids **5**, 1393 (1962).
[9] N. D. Kosov, Elementary kinetic theory of diffusion in gases, J. Eng. Phys. **42**, 181 (1982).
[10] J. Jeans, LXX. The persistence of molecular velocities in the kinetic theory of gases, Lond. Edinb. Philos. Mag. J. Sci. **8**, 700 (1904).
[11] J. Jeans, *An Introduction to the Kinetic Theory of Gases* (Cambridge University Press, Cambridge, UK, 2009).
[12] L. Monchick and E. A. Mason, Free-flight theory of gas mixtures, Phys. Fluids **10**, 1377 (1967).
[13] G. A. Bird, *Molecular Gas Dynamics and the Direct Simulation Monte Carlo of Gas Flows* (Clarendon Press, Oxford, 1994).
[14] V. Garzó, *Granular Gaseous Flows* (Springer Nature, Cham, 2019).
[15] S. T. Paik, Is the mean free path the mean of a distribution? Am. J. Phys. **82**, 602 (2014).
[16] J.-P. Bouchaud and A. Georges, Anomalous diffusion in disordered media: Statistical mechanisms, models and physical applications, Phys. Rep. **195**, 127 (1990).
[17] Note that the meaning of $f(\mathbf{v})$ here differs from that in Ref. [6]: $f(\mathbf{v})$ is what is called $f_1(\mathbf{v})/n_1$ in Ref. [6], n_1 being the number density of molecule 1.
[18] To evaluate κ for $m_1/m_2 \to \infty$ one first notes that, due to the term $\exp(-my^2)$ in the integrand of $I_{1,1}(m)$, the only relevant values of y for the evaluation of this integral are those around zero. But $\lim_{y\to 0} E(y) = 2$ so that the dominant term of $I_{1,1}$ is given by half the integral of $y^4 \exp(-my^2)$. This integral is equal to $3\sqrt{\pi}/8m^{5/2}$, and the value of $\kappa = 3\pi/8$, for $m_1/m_2 \to \infty$, follows.
[19] P. Visco, F. van Wijland, and E. Trizac, Collisional statistics of the hard-sphere gas, Phys. Rev. E **77**, 041117 (2008).
[20] L. Lue, Collision statistics, thermodynamics, and transport coefficients of hard hyperspheres in three, four, and five dimensions, J. Chem. Phys. **122**, 044513 (2005).
[21] Meijer G-Function (The Wolfram Functions Site) [http://functions.wolfram.com/07.34.02.0001.01].
[22] Be aware that this function E is not the function with this name in Ref. [1] (see Eq. 5.4.9 there). The E function in Ref. [1] is y times the function E defined here.
[23] A. Santos, S. B. Yuste, and M. López de Haro, Structural and thermodynamic properties of hard-sphere fluids, J. Chem. Phys. **153**, 120901 (2020).
[24] G. A. Bird, Approach to translational equilibrium in a rigid sphere gas, Phys. Fluids **6**, 1518 (1963).
[25] G. A. Bird, Direct simulation and the Boltzmann equation, Phys. Fluids **13**, 2676 (1970).
[26] T. Pöschel and T. Schwager, *Computational Granular Dynamics: Models and Algorithms* (Springer Science & Business Media, Berlin, 2005).