Variance reduction for Monte Carlo solutions of the Boltzmann equation

Lowell L. Baker and Nicolas G. Hadjiconstantinou
Mechanical Engineering Department, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 28 December 2004; accepted 23 February 2005; published online 28 April 2005)

We show that by considering only the deviation from equilibrium, significant computational savings can be obtained in Monte Carlo evaluations of the Boltzmann collision integral for flow problems in the small Mach number (Ma) limit. The benefits of this variance reduction approach include a significantly reduced statistical uncertainty when the deviation from equilibrium is small, and a flow-velocity signal-to-noise ratio that remains approximately constant with Ma in the $\mathrm{Ma} \ll 1$ limit. This results in stochastic Boltzmann solution methods whose computational cost for a given signal-to-noise ratio is essentially independent of Ma for $\mathrm{Ma} \ll 1$; our numerical implementation demonstrates this for Mach numbers as low as 10^{-5}. These features are in sharp contrast to current particle-based simulation techniques in which statistical sampling leads to computational cost that is proportional to Ma^{-2}, making calculations at small Ma very expensive. © *2005 American Institute of Physics*. [DOI: 10.1063/1.1899210]

Interest in numerical solution of the Boltzmann equation[1,2] has recently been revived in connection with modeling gaseous flows is small-scale devices (microelectromechanical systems) where the Navier–Stokes description is no longer valid.[3,4] This new regime of interest is typically characterized by problems exhibiting small deviations from equilibrium; a typical example, and one which we will use throughout this Letter as an archetypal problem, is low speed—i.e., low Mach number—flow. Although these flows are in general more amenable to Boltzmann equation analysis due to the possibility of linearized approaches, they present significant challenges to the prevalent Boltzmann simulation tool, known as the direct simulation Monte Carlo (DSMC). DSMC is a stochastic simulation method[5,6] for solving[7] the *nonlinear* Boltzmann equation. Unfortunately, DSMC is not well suited to the simulation of low Mach number flows: the slow convergence of the statistical sampling of macroscopic observables (in this case the flow velocity) results in a rapid increase in the number of samples required as the magnitude of these quantities decreases.[8] On the other hand, DSMC is very attractive due to the simplicity of its intuitive particle-tracking formulation, which not only appeals to users, but also endows this method with a significant efficiency advantage (see below).

The objective of the work presented here is to develop an approach which addresses the limitations of Monte Carlo approaches in the case of low-speed flows commonly found in small scale devices. Our approach has focused on efficient methods for evaluating the collision integral since, in current approaches, it is by far the most time-consuming part of the calculation. In particular, we focused on preserving the ingredients we feel make DSMC such a powerful and successful approach, namely, simplicity and the efficiency stemming from the evaluation of the collision integral by importance sampling, while improving upon the performance of DSMC and similar approaches. An example of such a "similar" approach is the Δ-ϵ method of Tan and Varghese,[9] which addresses high-speed flows; in this method, the collision integral is evaluated by a Monte Carlo sampling of representative collisions in a fashion which closely resembles the collision process in DSMC.

In this Letter we report on significant computational improvements obtained by exploiting the fact that for low-speed flows the deviation from equilibrium is small. More specifically, we can construct methods that are significantly more efficient by focusing the computational effort on calculating the value of the collision integral due to the *deviation* from equilibrium. This approach falls in the broad category of variance reduction[10] techniques.

Solving for the deviation from equilibrium has been considered by Cheremisin[11] in a different context, namely, as a method for removing the stiffness in explicit time integration of *deterministic* discrete velocity approximations of the Boltzmann equation in the limit of small mean-free path. Cheremisin's deterministic approach, coupled to his interest in high-speed flows where little, if any, computational gain is obtained by this decomposition is, perhaps, the reason that the potential of considering the deviation from equilibrium within an importance sampling framework has not been realized before.

Here we consider a hard-sphere gas, although our approach can be easily extended to other interaction models (e.g., variable hard-sphere model[5]). Let $f(\mathbf{r},\mathbf{c},t)$ be the velocity distribution function normalized by a reference number density n_o and most probable molecular speed, $c_m = \sqrt{2kT_o/m}$, where k is Boltzmann's constant, m is the molecular mass, and T_o is a reference temperature. Here $\mathbf{r} = (x,y,z)$ is the position vector in physical space, $\mathbf{c} = (c_x,c_y,c_z)$ is the molecular velocity vector, and t is time. These quantities are nondimensionalized by the mean-free path $\lambda = 1/(\sqrt{2}\pi n_o d^2)$, most probable molecular speed, and the molecular collision time $\tau = \sqrt{\pi}\lambda/(2c_m)$, respectively, where d is the hard-sphere molecular diameter.[5] In the ab-

sence of external body forces we write the Boltzmann equation in the following form:

$$\frac{\partial f}{\partial t} + \frac{\sqrt{\pi}}{2}\mathbf{c}\cdot\frac{\partial f}{\partial \mathbf{r}} = \left[\frac{df}{dt}\right]_{\text{coll}}(\mathbf{r},\mathbf{c},t), \tag{1}$$

$$\left[\frac{df}{dt}\right]_{\text{coll}}(\mathbf{r},\mathbf{c},t) = \frac{\sqrt{\pi}}{4}\int\int\int(\delta_1' + \delta' - \delta_1 - \delta)ff_1 \times g\sigma d^2\Omega d^3\mathbf{v}_1 d^3\mathbf{v}, \tag{2}$$

where $\sigma=(4\sqrt{2}\pi)^{-1}$ is the nondimensional differential collision cross section for hard spheres, $f_1=f(\mathbf{r},\mathbf{c}_1,t)$, $\delta=\delta^3(\mathbf{v}-\mathbf{c})$, $\delta_1=\delta^3(\mathbf{v}_1-\mathbf{c})$, $\delta_1'=\delta^3(\mathbf{v}_1'-\mathbf{c})$, $\delta'=\delta^3(\mathbf{v}'-\mathbf{c})$, $\mathbf{v},\mathbf{v}_1$ are the precollision velocities, $g=|\mathbf{v}-\mathbf{v}_1|$ is the magnitude of the relative velocity vector, and $\mathbf{v}',\mathbf{v}_1'$ are the postcollision velocities, related to the precollision velocities through the scattering angle Ω. Integration in velocity space extends from $-\infty$ to ∞ unless otherwise stated; similarly, the solid angle integration is over the surface of the unit sphere. The formulation given above can be obtained by considering the weak form of the collision integral[2] and choosing the delta function $\delta^3(\mathbf{v}-\mathbf{c})$ as a test function.[9]

Equation (2) and its importance sampling interpretation motivate a number of Monte Carlo solution schemes[9] including particle schemes (such as DSMC) and the method presented here. Let us write Eq. (2) for the collision integral as

$$\left[\frac{df}{dt}\right]_{\text{coll}}(\mathbf{r},\mathbf{c},t) = \mathcal{N}^2\frac{\sqrt{\pi}}{4}\int\int\int(\delta_1' + \delta' - \delta_1 - \delta)\frac{ff_1}{\mathcal{N}^2} \times g\sigma d^2\Omega d^3\mathbf{v}_1 d^3\mathbf{v}, \tag{3}$$

where $\mathcal{N}=\int f d^3\mathbf{v}$. Noting that $f/\mathcal{N}$ is a *normalized* probability distribution function, expression (3) lends itself to Monte Carlo evaluation using importance sampling,[10] which is typically significantly more efficient than the more straightforward Monte Carlo evaluations. Using importance sampling, the collision integral can be approximated by

$$\left[\frac{df}{dt}\right]_{\text{coll}}(\mathbf{r},\mathbf{c},t) = \frac{4\pi\mathcal{N}^2}{M}\frac{\sqrt{\pi}}{4}\sum_{i=1}^{M}(\delta_{1,i}' + \delta_i' - \delta_{1,i} - \delta_i)g_i\sigma_i, \tag{4}$$

where the precollision velocities $\mathbf{v}_i$ and $\mathbf{v}_{1,i}$ are chosen independently with probability $f/\mathcal{N}$ and $f_1/\mathcal{N}$, respectively, within a *finite* phase space volume ν; here, M is the number of Monte Carlo samples. Considering only a finite volume of velocity space is justified[2,12] in low-speed flows, provided this volume is sufficiently large. The scattering angle Ω_i [which does not appear explicitly in (4), but affects the values of the collision cross section and the postcollision velocities] is chosen with uniform probability on the unit sphere. The analogy with the collision process in DSMC is apparent.

We now discuss the variance reduction approach proposed here. We begin by considering an *arbitrary* Maxwell–Boltzmann distribution f^{MB} and defining $f^d\equiv f-f^{\text{MB}}$. Upon substitution into Eq. (4), we obtain

$$\left[\frac{df}{dt}\right]_{\text{coll}}(\mathbf{r},\mathbf{c},t) = \frac{\sqrt{\pi}}{4}\int\int\int(\delta_1' + \delta' - \delta_1 - \delta)(2f^{\text{MB}}f_1^d + f^d f_1^d)g\sigma d^2\Omega d^3\mathbf{v}_1 d^3\mathbf{v}, \tag{5}$$

since the integral involving $f^{\text{MB}}f_1^{\text{MB}}$ is identically zero (f^{MB} is an equilibrium distribution) and interchanging $\mathbf{v}$ and $\mathbf{v}_1$ has no effect on the physical situation (i.e., the integrals involving $f^{\text{MB}}f_1^d$ and $f^d f_1^{\text{MB}}$ are equal). Separating terms gives

$$\left[\frac{df}{dt}\right]_{\text{coll}}(\mathbf{r},\mathbf{c},t) = \frac{\sqrt{\pi}}{4}\int\int\int(\delta_1' + \delta' - \delta_1 - \delta)(2f^{\text{MB}} + f^d)f_1^d g\sigma d^2\Omega d^3\mathbf{v}_1 d^3\mathbf{v}. \tag{6}$$

Using the same approach as in the preceding section, we can approximate integral (6) by the following single sum:

$$\left[\frac{df}{dt}\right]_{\text{coll}}(\mathbf{r},\mathbf{c},t) = \frac{\pi^{3/2}}{M}\hat{\mathcal{N}}\sum_{i=1}^{M}(\delta_{i,1}' + \delta_i' - \delta_{i,1} - \delta_i) \times \text{sgn}(2f_i^{\text{MB}} + f_i^d)\text{sgn}(f_{1,i}^d)g_i\sigma_i, \tag{7}$$

which employs importance sampling by selecting the velocity $\mathbf{v}$ with probability $|2f^{\text{MB}}(\mathbf{v}) + f^d(\mathbf{v})|/\int|2f^{\text{MB}}(\mathbf{v}) + f^d(\mathbf{v})|d^3\mathbf{v}$ and $\mathbf{v}_1$ with probability $|f^d(\mathbf{v}_1)|/\int|f^d(\mathbf{v})|d^3\mathbf{v}$. Here $\hat{\mathcal{N}} \equiv \int|2f^{\text{MB}}(\mathbf{v}) + f^d(\mathbf{v})|d^3\mathbf{v}\times\int|f^d(\mathbf{v})|d^3\mathbf{v}$ and $\text{sgn}(x\gtrless 0)=\pm 1$.

As will be seen below, the variance reduction in cases where $f^d \ll f^{\text{MB}}$ is considerable to the extent that at $\text{Ma}\approx O(0.1)$ this approach provides considerable computational savings compared to direct methods which use (4). This is to be expected, since this approach allows one to avoid considering a large number of physically occurring collisions (the vast majority for $f^d \ll f^{\text{MB}}$) with *no net effect*. Also note that the above relations hold for *any* Maxwell–Boltzmann distribution. In other words, f^{MB} may be chosen to vary as a function of space and time so as to maximize computational efficiency by minimizing $f-f^{\text{MB}}$. Applying this approach to a flow where $f^d \not\ll f^{\text{MB}}$ (e.g., shock wave) or choosing the "wrong" Maxwell–Boltzmann distribution should *not* affect the accuracy of the solution, only degrade its efficiency.

One particularly desirable feature of this approach is that the degree of variance reduction is larger for distributions that are closer to the equilibrium distribution (i.e., when f^d is smaller in magnitude), leading to a method that can practically capture very small deviations from equilibrium. This is in contrast to current methods (such as DSMC) where as $f^d \to 0$, f^{MB} dominates the integrand landscape and thus leads to a statistical noise which is independent of f^d and a signal-to-noise ratio that decreases linearly[8] with decreasing Mach number ($\text{Ma}=\sqrt{2/\gamma}U$ where γ is the ratio of specific heats and U is the nondimensional local velocity). On the other hand, in the method presented here, the integrand landscape and consequently the statistical error[10] scale with f^d; consequently, as $f^d \to 0$, the statistical error decreases linearly with the signal leading to a constant signal-to-noise ratio.

We now proceed with some numerical examples and a brief description of our numerical implementation; a comprehensive discussion of numerical implementation details will be given in a future communication. In the interest of simplicity, Eq. (7) is evaluated by choosing precollision veloci-

ties to lie on computational nodes; this constraint can be relaxed by using a suitable interpolation (accounting for discontinuities in velocity space). The δ functions in Eq. (7) are approximated by

$$\delta(\mathbf{v}) \approx \phi(\mathbf{v}) \equiv \begin{cases} 1/h^3 & \text{if } \|\mathbf{v}\|_\infty \leq h/2, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

We see that if we take h equal to the nodal spacing, this will result in each "collision event" contributing to the value of the collision integral at four nodes, namely the nodes that are closest to each of the two precollision and postcollision velocities. In other words, the collision integral at each node is taken to be the average of the collision integral for velocities near that node; this can be seen more clearly by considering the weak form of the collision integral with ϕ as a test function. This approach can also be interpreted as a method akin to kernel density estimation, a technique for nonparametric probability distribution estimation.[13,14] Higher accuracy can be obtained by making h smaller; for all results discussed in this Letter, we use the approximation (8) with h equal to the nodal spacing. Although the fine spacing used here results in a sufficient degree of conservation, it is important to note that the above approximation will, in general, lead to a nonconservative scheme. This is a common issue associated with Boltzmann solution methods[15] and a number of approaches to remedy this have been proposed; some of the more notable ones include a recent approach[11] in which nodal contributions are distributed according to weights which ensure conservation of mass, momentum, and energy. Schemes which address lack of conservation by applying corrections to the newly updated distribution function have also been proposed.[15,16] This topic is not a central point of this work, so we will use the simpler method proposed above.

Equation (1) is solved by splitting the time evolution into a "homogeneous" collision and a collisionless advection step.[15] The advection step is treated by a first-order finite-volume discretization[17] (subject to a Courant stability condition); within the collision step the distribution function is updated by a first-order Euler step. Fluxal boundary conditions follow from specifying the distribution function for velocities satisfying $\mathbf{n}\cdot\mathbf{c} > 0$, where $\mathbf{n}$ is the wall normal pointing into the gas (i.e., corresponding to particles leaving the wall), subject to the constraint of zero net mass flux to the wall.[12] A diffuse Maxwell reflection was assumed.

A number of higher order schemes for integrating the advection equation are available, including a fourth-order accurate finite difference scheme[9] and a number of implicit schemes;[17] an extensive discussion of this topic can be found in the latter reference. Second-order-accurate (in time) splitting schemes are also possible by appropriate symmetrization.[18] As will be seen below, the first-order schemes used here provide adequate accuracy for the validation purposes of this work. An iterative method for steady-state problems,[15] based on the same spatial discretization, and suitable for the present collision-integral evaluation method has also been developed and will be presented in a future communication. This method was used to solve the pressure driven flow problem discussed below.

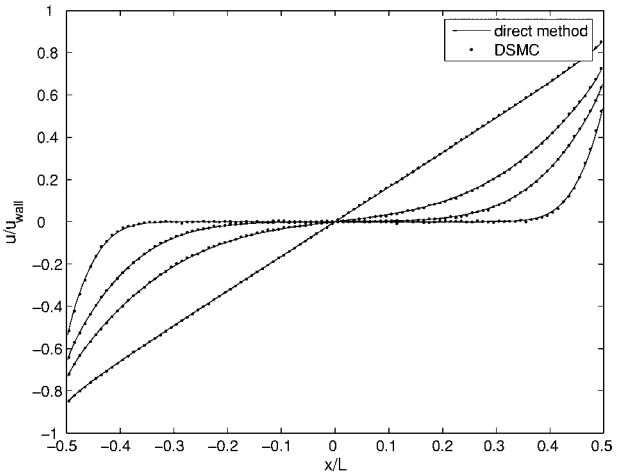

FIG. 1. Normalized velocity profile at $t=1,3,6$ and steady-state for unsteady Couette flow with wall velocity of $u_{\text{wall}} = \pm 0.1$. Comparison between the method proposed here and DSMC. The gas is initially at equilibrium.

The approach for evaluating the collision integral proposed here can be applied to higher spatial dimensions directly since it in no way depends on the dimensionality of the advection operator. Depending on the method used, however, discretization of the advection operator may require more care since discontinuities in the distribution function propagate into the gas in the case of convex boundaries.[19]

Figure 1 shows a comparison between the method proposed here and DSMC for a time-dependent Couette flow. In this test case, the gas is at equilibrium with zero velocity at time $t=0$ when the walls at $x=\pm 5$ are impulsively accelerated to a velocity of ± 0.1. Velocity profiles are shown for $t=1,3,6$ and at steady state. This figure shows an excellent level of agreement with the velocity profiles calculated using DSMC. Our results also show that higher moments of the distribution function are also accurately captured; these results will be presented in a future communication.

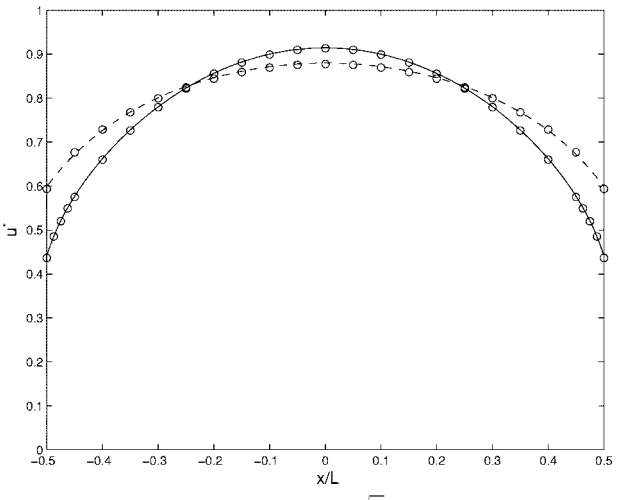

FIG. 2. Pressure-driven flow at $\text{Kn} = 0.8/\sqrt{\pi}$ (solid line) and $\text{Kn} = 4/\sqrt{\pi}$ (dotted line). Comparison between our results and the numerical solutions (Ref. 12) of the linearized Boltzmann equation (circles). u^* is the nondimensional flow velocity normalized by the nondimensional logarithmic pressure gradient $(-1/P)(dP/dx)L$.

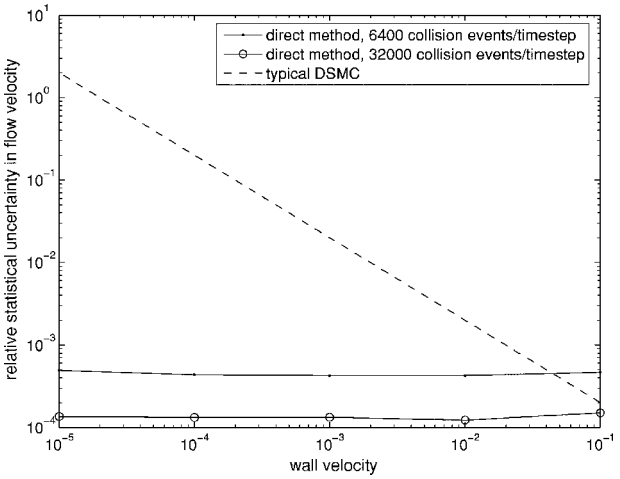

FIG. 3. Relative statistical uncertainty in flow velocity (averaged over the flow domain) as a function of (nondimensional) wall velocity in Couette flow. Note that the number of samples required to make the statistical uncertainty between the two methods the same scales with the *square* of the ratio of statistical uncertainties.

Figure 2 shows our results for pressure driven flow in a two-dimensional channel of transverse dimension L. This problem was formulated in one physical dimension by using Cercignani's[20] linearization procedure (small pressure gradient); although immaterial for this problem, we retained our full nonlinear formulation for the collision integral. Our results are compared to numerical solutions of the linearized Boltzmann equation[12] for two transition-regime Knudsen numbers ($\mathrm{Kn}=\lambda/L$). The agreement is very good. Note that the pressure gradient in our calculations was such that the nondimensional flow speed was of the order of 10^{-5}. Essentially noise-free calculations such as shown in Fig. 2 at flow speeds of this order would have been impossible to perform using DSMC with our current computational resources.

As explained above, the most desirable feature of the method proposed here is the removal of the statistical uncertainty associated with the equilibrium part of the distribution function. This endows the method with the following desirable properties: first, a significant reduction of relative statistical error compared to other direct simulation methods at Mach numbers as high as $O(0.1)$; second, a statistical uncertainty which scales with the deviation from equilibrium leading to a relative statistical error (defined as the one standard deviation in uncertainty divided by the characteristic velocity[8]) which remains essentially constant with Mach number. This is illustrated in Fig. 3. Note that the level of stochastic error depends on the number of samples used in evaluating sum (7); two sampling levels are shown in this figure. In contrast, in DSMC the relative statistical error in velocity due to statistical noise scales as Ma^{-1} for small Mach numbers;[8] this implies that the number of samples required to obtain a given level of relative statistical error scales as Ma^{-2}. Typical DSMC scaling performance is shown for comparison on the same figure. Note that the crossover point is implementation dependent and difficult to determine, especially if complicating factors such as sample correlations are taken into account.

Our approach eliminates the most important disadvantage associated with the use of current Monte Carlo approaches for microscale flows, namely, the poor noise to signal scaling with decreasing Mach number. More specifically, the approach proposed here exhibits an approximately constant relative uncertainty as the Mach number decreases, implying that calculations at arbitrarily small Mach numbers are possible at no extra cost. Computational savings compared to other direct solution methods can be expected when the deviation from equilibrium is smaller (in the sense that the function variance[10] is smaller) than the full distribution function, a condition which is satisfied by a wide variety of flows. As a direct outcome of the present research, we envision a modified DSMC-like procedure which simulates only the deviation from equilibrium.

This work was supported by Sandia National Laboratory. The authors are grateful to Dr. M. A. Gallis for hosting L.L.B. during his visit at Sandia National Laboratory and for many useful comments and discussions. The authors would also like to thank Dr. S. Kempka for project support.

[1] W. G. Vincenti and C. H. Kruger, *Introduction to Physical Gas Dynamics* (Krieger, Florida, 1965).
[2] C. Cercignani, *The Boltzmann Equation and its Applications* (Springer, New York, 1988).
[3] C. M. Ho and Y. C. Tai, "Micro-electro-mechanical systems (MEMS) and fluid flows," Annu. Rev. Fluid Mech. **30**, 579 (1998).
[4] G. E. Karniadakis and A. Beskok, *Microflows: Fundamentals and Simulation* (Springer, Berlin, 2001).
[5] G. A. Bird, *Molecular Gas Dynamics and the Direct Simulation of Gas Flows* (Clarendon, Oxford, 1994).
[6] F. J. Alexander and A. L. Garcia, "The direct simulation Monte Carlo method," Comput. Phys. **11**, 588 (1997).
[7] W. Wagner, "A convergence proof for bird's direct simulation Monte Carlo method for the Boltzmann equation," J. Stat. Phys. **66**, 1011 (1992).
[8] N. G. Hadjiconstantinou, A. L. Garcia, M. Z. Bazant, and G. He, "Statistical error in particle simulations of hydrodynamic phenomena," J. Comput. Phys. **187**, 274 (2003).
[9] Z. Tan and P. L. Varghese, "The Δ-ϵ method for the Boltzmann equation," J. Comput. Phys. **110**, 327 (1994).
[10] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, 2nd ed. (Cambridge University Press, Cambridge, 1992).
[11] F. G. Cheremisin, "Solving the Boltzmann equation in the case of passing to the hydro-dynamic flow regime," Dokl. Phys. **45**, 401 (2000).
[12] T. Ohwada, Y. Sone, and K. Aoki, "Numerical analysis of the Poiseuille and thermal transpiration flows between two parallel plates on the basis of the Boltzmann equation for hard-sphere molecules," Phys. Fluids A **1**, 2042 (1989).
[13] D. W. Scott, *Multivariate Density Estimation: Theory Practice and Visualization* (Wiley, New York, 1992).
[14] *Encyclopedia of Statistical Sciences*, edited by S. Kotz, N. L. Johnson, and C. B. Read (Wiley, New York, 1982), Vol. 2.
[15] S. M. Yen, "Numerical solution of the nonlinear Boltzmann equation for nonequilibrium gas flow problems," Annu. Rev. Fluid Mech. **16**, 67 (1984).
[16] A. E. Beylich, "Solving the kinetic equation for all Knudsen numbers," Phys. Fluids **12**, 444 (2000).
[17] V. V. Aristov, *Direct Methods for Solving the Boltzman Equation and Study of Nonequilibrium Flows* (Kluwer, Dordrecht, 2001).
[18] F. Filbet and G. Russo, "High order numerical methods for the space non-homogeneous Boltzmann equation," J. Comput. Phys. **186**, 457 (2003).
[19] S. Takata, Y. Sone, and K. Aoki, "Numerical analysis of a uniform flow of a rarefied gas past a sphere on the basis of the Boltzmann equation for hard-sphere molecules," Phys. Fluids A **5**, 716 (1993).
[20] C. Cercignani and A. Daneri, "Flow of a rarefied gas between two parallel plates," J. Appl. Phys. **34**, 3509 (1963).