

# NUMERICAL SIMULATION OF MICROFLOWS USING HERMITE SPECTRAL METHODS\*

ZHICHENG HU<sup>†</sup>, ZHENNING CAI<sup>‡</sup>, AND YANLI WANG<sup>§</sup>

**Abstract.** We propose a Hermite spectral method for the spatially inhomogeneous Boltzmann equation. For the inverse-power-law model, we generalize a class of approximate quadratic collision operators defined in the normalized and dimensionless setting to operators for arbitrary distribution functions. An efficient algorithm with a fast transform is introduced to discretize the new collision operators. The method is tested for one- and two-dimensional benchmark microflow problems.

**Key words.** Boltzmann equation, Hermite spectral method, microflow

**AMS subject classification.** 76P05

**DOI.** 10.1137/18M120066X

**1. Introduction.** Rarefied gas dynamics studies the gas flows when the mean free path of the gas molecules is comparable to the characteristic length of the problem we are concerned about. Typical cases include the gas dynamics in astronautics (large mean free path) and the micro-electro-mechanical systems (small characteristic length). In these cases, continuum fluid models such as Euler equations and Navier–Stokes equations are no longer accurate; on the other hand, molecular dynamics is still too expensive to solve these problems due to the huge number of gas molecules. Therefore, people usually adopt the method in statistical physics to obtain the mesoscopic kinetic models for rarefied gas dynamics. One of the most important models is the Boltzmann equation derived from molecular chaos assumption, which reads

$$(1.1) \quad \frac{\partial f}{\partial t} + \nabla_{\mathbf{x}} \cdot (\mathbf{v}f) = Q[f, f], \quad t \in \mathbb{R}^+, \quad \mathbf{x} \in \Omega \subset \mathbb{R}^3, \quad \mathbf{v} \in \mathbb{R}^3.$$

Here the unknown function  $f(t, \mathbf{x}, \mathbf{v})$  is the distribution function, which describes the number density of molecules in the joint position-velocity  $(\mathbf{x}, \mathbf{v})$  space at time  $t$ . The right-hand side of (1.1) models the collision between gas molecules. It usually takes the quadratic form

$$(1.2) \quad \begin{aligned} & Q[f, f](t, \mathbf{x}, \mathbf{v}) \\ &= \int_{\mathbb{R}^3} \int_{\mathbf{n} \perp \mathbf{g}} \int_0^\pi [f(t, \mathbf{x}, \mathbf{v}'_1) f(t, \mathbf{x}, \mathbf{v}') - f(t, \mathbf{x}, \mathbf{v}_1) f(t, \mathbf{x}, \mathbf{v})] B(|\mathbf{g}|, \chi) d\chi d\mathbf{n} d\mathbf{v}_1. \end{aligned}$$

\*Submitted to the journal's Computational Methods in Science and Engineering section July 16, 2018; accepted for publication (in revised form) September 27, 2019; published electronically January 9, 2020.

<https://doi.org/10.1137/18M120066X>

**Funding:** The work of the first author was partially supported by the National Natural Science Foundation of China (11601229), and the Natural Science Foundation of Jiangsu Province of China (BK20160784). The work of the second author was supported by National University of Singapore Startup Fund grant R-146-000-241-133. The work of the third author was supported by the National Natural Science Foundation of China (11501042), and the Postdoctoral Science Foundation of China (2018M631233). The computational resources were supported by the high performance computing platform of Peking University, China.

<sup>†</sup>Department of Mathematics, College of Science, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China (huzhicheng@nuaa.edu.cn).

<sup>‡</sup>Department of Mathematics, National University of Singapore, Level 4, Block S17, 10 Lower Kent Ridge Road, Singapore 119076 (matcz@nus.edu.sg).

<sup>§</sup>Department of Engineering, Peking University, Beijing 100871, China (wang-yanli@pku.edu.cn).

Here  $\mathbf{n}$  is a unit vector perpendicular to the relative velocity  $\mathbf{g} = \mathbf{v} - \mathbf{v}_1$ , and  $\mathbf{v}', \mathbf{v}'_1$  are postcollisional velocities

$$(1.3) \quad \begin{aligned} \mathbf{v}' &= \cos^2(\chi/2)\mathbf{v} + \sin^2(\chi/2)\mathbf{v}_1 - |\mathbf{g}| \cos(\chi/2) \sin(\chi/2)\mathbf{n}, \\ \mathbf{v}'_1 &= \cos^2(\chi/2)\mathbf{v}_1 + \sin^2(\chi/2)\mathbf{v} + |\mathbf{g}| \cos(\chi/2) \sin(\chi/2)\mathbf{n}. \end{aligned}$$

The collision kernel  $B(\cdot, \cdot)$  is a nonnegative function describing the interaction between molecules.

It is generally accepted that the Boltzmann equation provides solutions with enough accuracy in rarefied gas dynamics. The DSMC (direct simulation of Monte Carlo) method [3], a stochastic numerical solver for the Boltzmann equation, has been widely used in the simulation. The DSMC method is usually efficient in solving highly rarefied flows and steady-state problems, whereas in this work, we are interested in deterministic solvers, which are expected to be better at flows in the hydrodynamic regime and dynamic problems. Meanwhile, we also anticipate smoother numerical results and a higher order of convergence using deterministic solvers.

Obviously, the most complicated part of the Boltzmann equation is the collision term (1.2), which is also supposed to be the most expensive part in the numerical method. One classical method to discretize (1.2) is the discrete velocity method [17], which turns out to be inefficient due to its low convergence order [27]. A much more efficient method is the Fourier spectral method [28, 26, 15], and some two-dimensional and three-dimensional simulations have been carried out based on this method [14, 23]. To seek higher numerical efficiency, the Hermite spectral method has been introduced in [16, 31] to solve the spatially homogeneous Boltzmann equation. The idea of the Hermite spectral method can be traced back to Grad's classical paper [18]. Grad's method is based on the fact that the collision operator vanishes when the distribution function takes the form of the Maxwellian  $f(\mathbf{v}) = \rho \mathcal{M}_{\mathbf{u}, \theta}(\mathbf{v})$ , where  $\rho, \theta > 0$ ,  $\mathbf{u} \in \mathbb{R}^3$ , and

$$(1.4) \quad \mathcal{M}_{\mathbf{u}, \theta}(\mathbf{v}) := \frac{1}{m(2\pi\theta)^{3/2}} \exp\left(-\frac{|\mathbf{v} - \mathbf{u}|^2}{2\theta}\right),$$

where  $m$  is the mass of a single gas molecule. Therefore for smaller Knudsen number, the distribution function is expected to be closer to the Maxwellian. By such a property, it is natural to consider the expansion of the distribution function using orthogonal polynomials with the weight function  $\mathcal{M}_{\mathbf{u}, \theta}$ . These polynomials are just the Hermite polynomials. Using this expansion, the explicit formulas for all the equations in Galerkin's method are derived for the first time in [31], and for inverse-power-law models, numerical tests have been carried out. Meanwhile, a modeling technique is introduced therein to simplify the collision term so that the computational cost can be reduced.

However, the work in [31] is not ready for the simulation of the spatially inhomogeneous Boltzmann equation. The major reason is that the simplified collision model has a simple form only when the distribution function is represented as a "Grad's series," which means that the parameters  $\mathbf{u}$  and  $\theta$  in the weight function  $\mathcal{M}_{\mathbf{u}, \theta}$  are, respectively, the local mean velocity and the local temperature in energy units. When considering spatially inhomogeneous problems, these parameters vary spatially, resulting in different basis functions at different spatial locations. Consequently, the discretization of the spatial derivative becomes nontrivial.

There are two possible ways to resolve this issue. One is to introduce projections to deal with operations between distribution functions represented by different basis

functions, which essentially introduces nonlinearity into the discrete convection term. This approach is used in [10] for the linearized collision operator, from which it can be seen that the implementation is rather involved. The other way is to use uniform basis functions for all spatial grid points, and find an appropriate representation for the simplified collision term proposed in [31]. This paper will follow the second idea and it will turn out that the implementation is relatively easier.

Numerical simulations will be done for one-dimensional benchmark problems in microflows including Couette flows and Fourier flows. For such problems, numerical results in [7, 13] have shown that the Bhatnagar–Gross–Krook (BGK)-type models cannot provide reliable predictions when the Knudsen number is large. Results in [10] show that even for linearized collision models, an obvious deviation can be observed when compared with DSMC results. In this work, we are going to show better agreement with DSMC results using our method. Additionally, to test the efficiency of our method, some preliminary two-dimensional tests for lid-driven cavity flows are also carried out.

The rest of this paper is organized as follows: section 2 is a review of the background of our method, and the description of our algorithm is mostly given in section 3, with the discussion on boundary conditions left to section 4. Numerical results are exhibited in section 5, and the paper ends with a conclusion in section 6.

**2. Preliminaries.** In this section, we are going to provide some preliminary knowledge for our further discussion, including the discretization of the distribution function introduced in a previous work [6], and the collision kernels we are going to consider later. A brief review of these topics will be given in the following two subsections.

**2.1. Discretization of the distribution function.** In most cases, when solving the Boltzmann equation, we are not interested in the distribution function itself. What we are really concerned about is usually the macroscopic physical quantities such as the density  $\rho$ , momentum  $\mathbf{m}$ , and energy  $E$ . These quantities are in fact the moments of the distribution function, and are related to the distribution function  $f(t, \mathbf{x}, \mathbf{v})$  by

$$\begin{aligned} \rho(t, \mathbf{x}) &= m \int_{\mathbb{R}^3} f(t, \mathbf{x}, \mathbf{v}) \, d\mathbf{v}, \\ \mathbf{m}(t, \mathbf{x}) &= m \int_{\mathbb{R}^3} \mathbf{v} f(t, \mathbf{x}, \mathbf{v}) \, d\mathbf{v}, \\ E(t, \mathbf{x}) &= m \int_{\mathbb{R}^3} \frac{|\mathbf{v}|^2}{2} f(t, \mathbf{x}, \mathbf{v}) \, d\mathbf{v}, \end{aligned} \quad (2.1)$$

where  $m$  is the mass of a single gas molecule. Due to the importance of the moments, Grad [18] proposed an expansion of the distribution function in the velocity space which has easy access to these moments, and the approximation of the distribution function is a truncation of the series. Here we adopt the equivalent notation used in [6]:

$$f(t, \mathbf{x}, \mathbf{v}) = \sum_{\alpha \in \mathbb{N}^3} \tilde{f}_{\alpha}^{[\bar{\mathbf{u}}, \bar{\theta}]}(t, \mathbf{x}) \mathcal{H}_{\alpha}^{[\bar{\mathbf{u}}, \bar{\theta}]}(\mathbf{v}), \quad (2.2)$$

where  $\bar{\mathbf{u}} \in \mathbb{R}^3$  and  $\bar{\theta} \in \mathbb{R}_+$  have, respectively, the same dimensions as  $\mathbf{v}$  and  $|\mathbf{v}|^2$ , and they can be chosen such that the convergence of the series is fast.<sup>1</sup> For any multi-

<sup>1</sup>In [18, 6], the parameters  $\bar{\mathbf{u}}$  and  $\bar{\theta}$  are chosen to be the local velocity and scaled temperature of the gas. Here we are discussing a more general form of the expansion.

index  $\alpha = (\alpha_1, \alpha_2, \alpha_3) \in \mathbb{N}^3$ , its norm is defined as  $|\alpha| = \alpha_1 + \alpha_2 + \alpha_3$ , and the basis function  $\mathcal{H}_\alpha^{[\bar{\mathbf{u}}, \bar{\theta}]}(\mathbf{v})$  is defined by

$$(2.3) \quad \mathcal{H}_\alpha^{[\bar{\mathbf{u}}, \bar{\theta}]}(\mathbf{v}) = \bar{\theta}^{-\frac{|\alpha|}{2}} H_\alpha \left( \frac{\mathbf{v} - \bar{\mathbf{u}}}{\sqrt{\bar{\theta}}} \right) \mathcal{M}_{\bar{\mathbf{u}}, \bar{\theta}}(\mathbf{v})$$

with  $H_\alpha$  being the Hermite polynomial

$$(2.4) \quad H_\alpha(\mathbf{v}) = (-1)^{|\alpha|} \exp \left( \frac{|\mathbf{v}|^2}{2} \right) \frac{\partial^{|\alpha|}}{\partial v_1^{\alpha_1} \partial v_2^{\alpha_2} \partial v_3^{\alpha_3}} \left[ \exp \left( -\frac{|\mathbf{v}|^2}{2} \right) \right].$$

An advantage of this expansion is that the coefficients are also “moments” of the distribution function. For example, when  $\alpha = 0$ , the coefficient  $\tilde{f}_\alpha^{[\bar{\mathbf{u}}, \bar{\theta}]}$  is just the density of the distribution function  $\rho$ . Other moments can also be easily represented by these coefficients. More details will be revealed later in this section.

Our discretization of the distribution function is simply a truncation of the series (2.2):

$$(2.5) \quad f_M(t, \mathbf{x}, \mathbf{v}) := \sum_{|\alpha| \leq M} \tilde{f}_\alpha^{[\bar{\mathbf{u}}, \bar{\theta}]}(t, \mathbf{x}) \mathcal{H}_\alpha^{[\bar{\mathbf{u}}, \bar{\theta}]}(\mathbf{v}) \in \mathcal{F}_M(\bar{\mathbf{u}}, \bar{\theta}), \quad M \in \mathbb{N},$$

where the finite-dimensional function space  $\mathcal{F}_M(\bar{\mathbf{u}}, \bar{\theta})$  is

$$(2.6) \quad \mathcal{F}_M(\bar{\mathbf{u}}, \bar{\theta}) = \text{span}\{\mathcal{H}_\alpha^{[\bar{\mathbf{u}}, \bar{\theta}]}(\mathbf{v}) \mid |\alpha| \leq M\}.$$

Apparently, for every  $t$  and  $\mathbf{x}$ ,  $f_M$  is an approximation of  $f$  in the function space  $\mathcal{F}_M(\bar{\mathbf{u}}, \bar{\theta})$ . For simplicity, from now on, we will consider  $\bar{\mathbf{u}}$  and  $\bar{\theta}$  as constants. The coefficients  $\tilde{f}_\alpha^{[\bar{\mathbf{u}}, \bar{\theta}]}$  will be shortened to  $\tilde{f}_\alpha$ , and the basis function  $\mathcal{H}_\alpha^{[\bar{\mathbf{u}}, \bar{\theta}]}$  will be shortened to  $\mathcal{H}_\alpha$ . For example, (2.5) is simplified to

$$(2.7) \quad f_M(t, \mathbf{x}, \mathbf{v}) := \sum_{|\alpha| \leq M} \tilde{f}_\alpha(t, \mathbf{x}) \mathcal{H}_\alpha(\mathbf{v}) \in \mathcal{F}_M(\bar{\mathbf{u}}, \bar{\theta}),$$

which looks more concise. However, when we use parameters other than  $\bar{\mathbf{u}}$  and  $\bar{\theta}$  in (2.2), the parameters will still be explicitly written out.

One advantage of the approximation (2.7) is that the truncation preserves low-order moments. Precisely speaking, using the canonical unit vectors  $e_1$ ,  $e_2$ , and  $e_3$  to denote the multi-indices  $(1, 0, 0)$ ,  $(0, 1, 0)$ , and  $(0, 0, 1)$ , we have

$$(2.8) \quad \rho = \tilde{f}_0, \quad \mathbf{m} = \rho \bar{\mathbf{u}} + (\tilde{f}_{e_1}, \tilde{f}_{e_2}, \tilde{f}_{e_3})^T, \quad E = \mathbf{m} \cdot \bar{\mathbf{u}} - \frac{1}{2} \rho |\bar{\mathbf{u}}|^2 + \frac{3}{2} \rho \bar{\theta} + \sum_{d=1}^3 \tilde{f}_{2e_d},$$

which can be derived from the orthogonality of Hermite polynomials

$$(2.9) \quad \int_{\mathbb{R}^3} H_\alpha(\mathbf{v}) H_\beta(\mathbf{v}) \exp \left( -\frac{|\mathbf{v}|^2}{2} \right) d\mathbf{v} = (2\pi)^{3/2} \alpha! \delta_{\alpha_1 \beta_1} \delta_{\alpha_2 \beta_2} \delta_{\alpha_3 \beta_3},$$

where  $\alpha! = \alpha_1! \alpha_2! \alpha_3!$ . With these moments, we can also obtain the mean velocity  $\mathbf{u}$  and the temperature  $T$  by

$$(2.10) \quad \mathbf{u} = \frac{\mathbf{m}}{\rho}, \quad T = \frac{m}{k_B} \frac{2E - \rho |\mathbf{u}|^2}{3\rho},$$

where  $k_B$  is the Boltzmann constant. Following the convention, we define

$$(2.11) \quad \theta = \frac{k_B}{m} T = \frac{2E - \rho|\mathbf{u}|^2}{3\rho}.$$

It can be seen that all the quantities from (2.8) to (2.11) are not changed by the truncation (2.7) if  $M \geq 2$ . More generally, by the orthogonality (2.9), we can obtain the coefficients  $\tilde{f}_\alpha$  from the distribution function by

$$(2.12) \quad \tilde{f}_\alpha(t, \mathbf{x}) = \frac{m}{\alpha!} \bar{\theta}^{\frac{|\alpha|}{2}} \int_{\mathbb{R}^3} H_\alpha \left( \frac{\mathbf{v} - \bar{\mathbf{u}}}{\sqrt{\bar{\theta}}} \right) f(t, \mathbf{x}, \mathbf{v}) d\mathbf{v}, \quad |\alpha| \leq M.$$

With (2.12), other interesting moments such as the stress tensor  $\sigma_{ij}$  and the heat flux  $q_i$ , which are defined by

$$\begin{aligned} \sigma_{ij} &= m \int_{\mathbb{R}^3} \left( (v_i - u_i)(v_j - u_j) - \frac{1}{3} \delta_{ij} |\mathbf{v} - \mathbf{u}|^2 \right) f d\mathbf{v}, \quad i, j = 1, 2, 3, \\ q_i &= \frac{1}{2} m \int_{\mathbb{R}^3} |\mathbf{v} - \mathbf{u}|^2 (v_i - u_i) f d\mathbf{v}, \quad i = 1, 2, 3, \end{aligned}$$

can also be easily related to the first few coefficients as follows:

$$\begin{aligned} \sigma_{ij} &= (1 + \delta_{ij}) \tilde{f}_{e_i + e_j} + \delta_{ij} \rho (\bar{\theta} - \theta) - \rho (\bar{u}_i - u_i)(\bar{u}_j - u_j), \\ q_i &= 2 \tilde{f}_{3e_i} + (\bar{u}_i - u_i) \tilde{f}_{2e_i} + |\bar{\mathbf{u}} - \mathbf{u}|^2 \tilde{f}_{e_i} \\ &\quad + \sum_{k=1}^3 \left[ \tilde{f}_{2e_k + e_i} + (\bar{u}_k - u_k) \tilde{f}_{e_k + e_i} + (\bar{u}_i - u_i) \tilde{f}_{2e_k} \right]. \end{aligned}$$

The above expressions involve only the parameters  $\bar{\mathbf{u}}$ ,  $\bar{\theta}$ , and the coefficients with index norm less than or equal to 3.

In our numerical method,  $f_M(t, \mathbf{x}, \mathbf{v})$  defined in (2.7) will be used as the semi-discrete distribution function. The choice of  $\bar{\mathbf{u}}$  and  $\bar{\theta}$  is based on the a priori understanding of the problem. In general,  $\bar{\mathbf{u}}$  should be selected so that  $\mathbf{u}(t, \mathbf{x})$  is not too far away from  $\bar{\mathbf{u}}$  for all  $t$  and  $\mathbf{x}$ , and  $\bar{\theta}$  should be selected large enough so that the tail of the distribution function can be well captured. Practically, this can be estimated by a rough simulation beforehand with a small number of grids. Once the parameters are fixed, the major difficulty in the discretization of the equation lies in the collision operator, which will be discussed in detail in the following sections.

**2.2. Collision kernels.** In order to apply (2.7) to the numerical scheme, we first need to define the collision kernel  $B(\cdot, \cdot)$  (see (1.2)). In this paper, we are interested in the inverse-power-law (IPL) model, for which

$$(2.13) \quad B(|\mathbf{g}|, \chi) = \left( \frac{2\kappa}{m} \right)^{\frac{2}{\eta-1}} |\mathbf{g}|^{\frac{\eta-5}{\eta-1}} W_0 \left| \frac{dW_0}{d\chi} \right|, \quad \eta > 3.$$

Here  $\eta$  is an index indicating the decay rate of the repulsive force between gas molecules when their distance increases, and  $\kappa$  is a constant indicating the intensity of the potential. The angle  $\chi$  and dimensionless impact parameter  $W_0$  are related by

$$(2.14) \quad \chi = \pi - 2 \int_0^{W_1} \left[ 1 - W^2 - \frac{2}{\eta-1} \left( \frac{W}{W_0} \right)^{\eta-1} \right]^{-1/2} dW$$

with  $W_1$  being a positive real number satisfying

$$(2.15) \quad 1 - W_1^2 - \frac{2}{\eta - 1} \left( \frac{W_1}{W_0} \right)^{\eta-1} = 0.$$

Details about this model can be found in [3]. The IPL model works well for a wide range of gases around room temperature. We refer readers to [12] for more details.

Another related model is the variable-hard-sphere (VHS) model, which is proposed by Bird in [2] as an approximation of the IPL model. The collision kernel of the VHS model is

$$(2.16) \quad B(|\mathbf{g}|, \chi) = \frac{1}{4} d_{\text{ref}}^2 g_{\text{ref}}^{2\nu} |\mathbf{g}|^{1-2\nu} \sin \chi, \quad 0 \leq \nu < 1,$$

where  $d_{\text{ref}}$  is the reference molecular diameter and  $g_{\text{ref}}$  is the reference speed. When approximating the IPL model with index  $\eta$ , the parameter  $\nu$  is chosen as  $2/(\eta - 1)$ . For the “hard potential” case where  $\eta > 5$ , the region for  $\nu$  is  $0 \leq \nu < \frac{1}{2}$ .

**3. Hermite spectral method for the Boltzmann equation.** Now we are ready to find the evolution equations for the coefficients  $\tilde{f}_\alpha$  in (2.7), which contains the discretization of the convection term and the collision term of the Boltzmann equation (1.1). Based on the idea of Galerkin’s method, we need to expand both the convection term and the collision term with the same basis functions  $\mathcal{H}_\alpha$  as in (2.7). The major difficulty lies in the collision term, which will be discussed first below.

**3.1. Series expansions of the IPL collision terms.** In what follows, we will study the series expansion of the quadratic collision term  $Q[f, f]$  defined in (1.2). Precisely speaking, for any given  $\bar{\mathbf{u}}$  and  $\bar{\theta}$ , the binary collision term  $Q[f, f]$  is to be expanded as

$$(3.1) \quad Q[f, f](t, \mathbf{x}, \mathbf{v}) = \sum_{\alpha \in \mathbb{N}^3} \tilde{Q}_\alpha(t, \mathbf{x}) \mathcal{H}_\alpha(\mathbf{v}).$$

By the orthogonality of Hermite polynomials (2.9), the coefficients can be evaluated by

$$(3.2) \quad \tilde{Q}_\alpha(t, \mathbf{x}) = \frac{m}{\alpha!} \bar{\theta}^{\frac{|\alpha|}{2}} \int_{\mathbb{R}^3} H_\alpha \left( \frac{\mathbf{v} - \bar{\mathbf{u}}}{\sqrt{\bar{\theta}}} \right) Q[f, f](t, \mathbf{x}, \mathbf{v}) d\mathbf{v}.$$

Since we are focusing on the collision operator, in sections 3.1 and 3.2 below, the variables  $t$  and  $\mathbf{x}$  will be temporarily omitted.

In [31], the authors have proposed an algorithm to find the values of these coefficients for the dimensionless Boltzmann collision operator in a special case  $\bar{\mathbf{u}} = 0$  and  $\bar{\theta} = 1$ . Thus, in order to make use of the result in [31], we will first apply the nondimensionalization by defining  $\hat{\mathbf{v}}$  and  $h(\hat{\mathbf{v}})$  as

$$(3.3) \quad \mathbf{v} = \bar{\mathbf{u}} + \sqrt{\bar{\theta}} \hat{\mathbf{v}}, \quad h(\hat{\mathbf{v}}) = \frac{m \bar{\theta}^{3/2}}{\rho} f(\bar{\mathbf{u}} + \sqrt{\bar{\theta}} \hat{\mathbf{v}}),$$

where  $\rho$  is the density defined in (2.1). From (2.7), one can derive the series expansion of  $h(\hat{\mathbf{v}})$  as

$$(3.4) \quad h(\hat{\mathbf{v}}) = \sum_{\alpha \in \mathbb{N}^3} \tilde{h}_\alpha \hat{\mathcal{H}}_\alpha(\hat{\mathbf{v}}),$$

where  $\hat{\mathcal{H}}_\alpha(\hat{\mathbf{v}}) = m\mathcal{H}_\alpha^{[0,1]}(\hat{\mathbf{v}})$  is the dimensionless basis function (see (2.3) for the definition of  $\mathcal{H}_\alpha^{[0,1]}$ ), and the coefficients

$$(3.5) \quad \tilde{h}_\alpha = \rho^{-1} \bar{\theta}^{-\frac{|\alpha|}{2}} \tilde{f}_\alpha, \quad \alpha \in \mathbb{N}^3,$$

are also dimensionless. Using the above definitions, the collision term  $Q[f, f]$  (1.2) changes to

$$(3.6) \quad \begin{aligned} Q[f, f](\bar{\mathbf{u}} + \sqrt{\bar{\theta}}\hat{\mathbf{v}}) \\ = \frac{\rho^2}{m^2 \bar{\theta}^{\frac{3}{2}}} \int_{\mathbb{R}^3} \int_{\mathbf{n} \perp \hat{\mathbf{g}}} \int_0^\pi B(\sqrt{\bar{\theta}}|\hat{\mathbf{g}}|, \chi) [h(\hat{\mathbf{v}}'_1)h(\hat{\mathbf{v}}') - h(\hat{\mathbf{v}}_1)h(\hat{\mathbf{v}})] \, d\chi \, d\mathbf{n} \, d\hat{\mathbf{v}}_1, \end{aligned}$$

where  $\hat{\mathbf{g}} = \hat{\mathbf{v}} - \hat{\mathbf{v}}_1$ . Specifically, for the IPL model, it is convenient to define the dimensionless collision kernel  $\hat{B}$  by

$$(3.7) \quad \hat{B}(|\hat{\mathbf{g}}|, \chi) = |\hat{\mathbf{g}}|^{\frac{\eta-5}{\eta-1}} W_0 \left| \frac{dW_0}{d\chi} \right|,$$

and then the IPL collision term turns out to be

$$(3.8) \quad Q[f, f](\bar{\mathbf{u}} + \sqrt{\bar{\theta}}\hat{\mathbf{v}}) = \frac{\rho^2}{m^2 \bar{\theta}^{\frac{3}{2}}} \left( \frac{2\kappa}{m} \right)^{\frac{2}{\eta-1}} \bar{\theta}^{\frac{\eta-5}{2(\eta-1)}} \hat{Q}[h, h](\hat{\mathbf{v}}),$$

where  $\hat{Q}[h, h]$  is the dimensionless collision operator

$$(3.9) \quad \hat{Q}[h, h](\hat{\mathbf{v}}) = \int_{\mathbb{R}^3} \int_{\mathbf{n} \perp \hat{\mathbf{g}}} \int_0^\pi \hat{B}(|\hat{\mathbf{g}}|, \chi) [h(\hat{\mathbf{v}}'_1)h(\hat{\mathbf{v}}') - h(\hat{\mathbf{v}}_1)h(\hat{\mathbf{v}})] \, d\chi \, d\mathbf{n} \, d\hat{\mathbf{v}}_1.$$

Inserting (3.8) into (3.2), we obtain

$$(3.10) \quad \tilde{Q}_\alpha = \frac{\rho^2}{m} \left( \frac{2\kappa}{m} \right)^{\frac{2}{\eta-1}} \bar{\theta}^{\frac{\eta-5}{2(\eta-1)} + \frac{|\alpha|}{2}} \frac{1}{\alpha!} \int_{\mathbb{R}^3} H_\alpha(\hat{\mathbf{v}}) \hat{Q}[h, h](\hat{\mathbf{v}}) \, d\hat{\mathbf{v}}, \quad \alpha \in \mathbb{N}^3.$$

For the IPL model, the integral in (3.10) has been deeply studied in [31]. The general result is

$$(3.11) \quad \frac{1}{\alpha!} \int_{\mathbb{R}^3} H_\alpha(\hat{\mathbf{v}}) \hat{Q}[h, h](\hat{\mathbf{v}}) \, d\hat{\mathbf{v}} = \sum_{\beta \in \mathbb{N}^3} \sum_{\gamma \in \mathbb{N}^3} A_\alpha^{\beta, \gamma} \tilde{h}_\beta \tilde{h}_\gamma, \quad \alpha \in \mathbb{N}^3.$$

The coefficients  $A_\alpha^{\beta, \gamma}$  are constants for a given collision model, and an algorithm to compute these coefficients is given [31] for all IPL models. The algorithm uses an explicit expression of  $A_\alpha^{\beta, \gamma}$  which involves only a one-dimensional integral, and this integral is evaluated by adaptive numerical integration. Such an algorithm can provide very accurate values for these coefficients, and has been verified to be reliable in the computation of homogeneous Boltzmann equation. Due to the lengthy expressions involved in the algorithm, we are not going to repeat the details in this paper. We would just like to mention that in order to find the values of all  $A_\alpha^{\beta, \gamma}$  with  $|\alpha|, |\beta|, |\gamma| \leq M$ , the computational cost is proportional to  $M^{12}$ . Although the time complexity is high, all these coefficients can be precomputed and stored. Readers are referred to [31] for the details of the algorithm.

Substituting (3.5) and (3.11) into (3.10), we finally get

$$(3.12) \quad \tilde{Q}_\alpha = \frac{1}{m} \left( \frac{2\kappa}{m} \right)^{\frac{2}{\eta-1}} \bar{\theta}^{\frac{\eta-5}{2(\eta-1)}} \sum_{\beta \in \mathbb{N}^3} \sum_{\gamma \in \mathbb{N}^3} \bar{\theta}^{\frac{1}{2}(|\alpha|-|\beta|-|\gamma|)} A_{\alpha}^{\beta, \gamma} \tilde{f}_\beta \tilde{f}_\gamma, \quad \alpha \in \mathbb{N}^3.$$

As mentioned at the beginning of section 3.1, the parameters  $\bar{\mathbf{u}}$  and  $\bar{\theta}$ , which affect the coefficients  $\tilde{f}_\beta$  and  $\tilde{f}_\gamma$  implicitly, can be arbitrarily chosen. A special choice is  $\bar{\mathbf{u}} = \mathbf{u}$  and  $\bar{\theta} = \theta$  (see (2.1) (2.10), and (2.11) for the definitions), which leads to

$$(3.13) \quad \tilde{Q}_\alpha^{[\mathbf{u}, \theta]} = \frac{c\theta}{\mu} \sum_{\beta \in \mathbb{N}^3} \sum_{\gamma \in \mathbb{N}^3} \theta^{\frac{1}{2}(|\alpha|-|\beta|-|\gamma|)} A_{\alpha}^{\beta, \gamma} \tilde{f}_\beta^{[\mathbf{u}, \theta]} \tilde{f}_\gamma^{[\mathbf{u}, \theta]}, \quad \alpha \in \mathbb{N}^3,$$

where  $\mu$  is the viscosity coefficient (see [3, eq. (3.62)])

$$(3.14) \quad \mu = \frac{5m(\theta/\pi)^{1/2}(2m\theta/\kappa)^{2/(\eta-1)}}{8A_2(\eta)\Gamma(4-2(\eta-1))}, \quad A_2(\eta) = \int_0^{+\infty} W_0 \sin^2 \chi \, dW_0,$$

and  $c$  is a constant given by<sup>2</sup>

$$(3.15) \quad c = -\left(A_{\zeta}^{0, \zeta} + A_{\zeta}^{\zeta, 0}\right)^{-1}, \quad \zeta = (1, 1, 0).$$

Such a special case will be used in the next section when we reduce the computational cost by simplifying the collision term.

**3.2. Approximation to the Boltzmann collision term.** The previous section establishes the basic theory for discretization of the collision term. By Galerkin's method, the equations for  $\tilde{f}_\alpha$  should hold the form

$$(3.16) \quad \frac{\partial \tilde{f}_\alpha}{\partial t} + \cdots = \tilde{Q}_\alpha, \quad |\alpha| \leq M,$$

where  $\cdots$  denotes the corresponding convection term to be discussed in section 3.3, and  $\tilde{Q}_\alpha$  is the coefficient of the collision term defined by

$$(3.17) \quad \tilde{Q}_\alpha(t, \mathbf{x}) = \frac{m}{\alpha!} \bar{\theta}^{\frac{|\alpha|}{2}} \int_{\mathbb{R}^3} H_\alpha \left( \frac{\mathbf{v} - \bar{\mathbf{u}}}{\sqrt{\bar{\theta}}} \right) Q[f_M, f_M](t, \mathbf{x}, \mathbf{v}) \, d\mathbf{v}$$

with  $f_M$  being the discrete distribution function defined in (2.7). Note that this definition is slightly different from (3.2) in which the parameters of  $Q$  are  $f$  instead of  $f_M$ . In other words, (3.17) is the discrete version of (3.2). Correspondingly, to compute (3.17), we just need to truncate the infinite sums in (3.12). Thereby, we know that the total computational cost for evaluating all  $\tilde{Q}_\alpha$  with  $|\alpha| \leq M$  is  $O(M^9)$ , which is unacceptable for a large  $M$ , especially for spatially inhomogeneous problems. The aim of this section is to build new collision models and derive the corresponding  $\tilde{Q}_\alpha$  following the method proposed in [31].

<sup>2</sup>When  $\bar{\mathbf{u}} = \mathbf{u}$  and  $\bar{\theta} = \theta$ , the expansion (2.2) is identical to the one proposed by Grad in [18], where the expansion of the collision term is also considered. For example, the equation (A3.56) is  $J_{ij}^{(2)} = -\frac{6}{m} B_1^{(2)} \rho a_{ij}^{(2)} + \cdots$ . When  $i = 1$  and  $j = 2$ , it can be translated to our language:  $\tilde{Q}_\zeta^{[\mathbf{u}, \theta]} = -\frac{6}{m} B_1^{(2)} \rho \tilde{f}_\zeta^{[\mathbf{u}, \theta]} + \cdots$ , by using  $\tilde{Q}_\zeta^{[\mathbf{u}, \theta]} = \rho \theta J_{12}^{(2)}$  and  $\tilde{f}_\zeta^{[\mathbf{u}, \theta]} = \rho \theta a_{12}^{(2)}$ . Comparing this equation with (3.12), we find  $\frac{1}{m} \left( \frac{2\kappa}{m} \right)^{\frac{2}{\eta-1}} \bar{\theta}^{\frac{\eta-5}{2(\eta-1)}} (A_{\zeta}^{0, \zeta} + A_{\zeta}^{\zeta, 0}) = -\frac{6}{m} B_1^{(2)}$ . By  $\mu = \frac{m\theta}{6B_1^{(2)}}$  ([18, equation (5.30)]), we obtain the coefficients in front of the sums in (3.13).



A strategy to reduce the computational cost has been proposed in [31] based on the dimensionless and normalized settings. To demonstrate the result, we consider again the dimensionless distribution function  $h(\hat{\mathbf{v}})$  defined in the previous section. When  $h(\hat{\mathbf{v}})$  satisfies

$$(3.18) \quad \int_{\mathbb{R}^3} \hat{\mathbf{v}} h(\hat{\mathbf{v}}) d\hat{\mathbf{v}} = 0, \quad \frac{1}{3} \int_{\mathbb{R}^3} |\hat{\mathbf{v}}|^2 h(\hat{\mathbf{v}}) d\hat{\mathbf{v}} = 1,$$

the dimensionless collision operator  $\hat{Q}[h, h]$  is approximated by

$$(3.19) \quad \hat{Q}^*[h, h](\hat{\mathbf{v}}) = \sum_{|\alpha| \leq M_0} \sum_{|\beta| \leq M_0} \sum_{|\gamma| \leq M_0} A_{\alpha}^{\beta, \gamma} \tilde{h}_{\beta} \tilde{h}_{\gamma} \hat{\mathcal{H}}_{\alpha}(\hat{\mathbf{v}}) - \sum_{|\alpha| > M_0} \nu_{M_0} \tilde{h}_{\alpha} \hat{\mathcal{H}}_{\alpha}(\hat{\mathbf{v}}),$$

where  $M_0$  is an arbitrarily chosen positive integer, and  $\nu_{M_0}$  gives the decay rate of the higher-order coefficients. The idea is to apply the quadratic collision operator only to the first few coefficients, and for the remaining coefficients, we adopt the idea of the BGK-type operator and simply let it decay to zero exponentially at a constant rate. Unfortunately, the conditions (3.18) do not hold in general. By (3.3), it can be found that only in the special case  $\bar{\mathbf{u}} = \mathbf{u}$ ,  $\bar{\theta} = \theta$ , are the equalities (3.18) true, and thus we can recover the dimensions from (3.19) by mimicking (3.8):

$$(3.20) \quad \begin{aligned} Q^*[f, f](\mathbf{v}) &= \frac{\rho^2}{m^2 \bar{\theta}^{\frac{3}{2}}} \left( \frac{2\kappa}{m} \right)^{\frac{2}{\eta-1}} \bar{\theta}^{\frac{\eta-5}{2(\eta-1)}} \hat{Q}^*[h, h](\hat{\mathbf{v}}) \\ &= \frac{c\theta}{\mu} \sum_{|\alpha| \leq M_0} \sum_{|\beta| \leq M_0} \sum_{|\gamma| \leq M_0} \theta^{\frac{1}{2}(|\alpha| - |\beta| - |\gamma|)} A_{\alpha}^{\beta, \gamma} \tilde{f}_{\beta}^{[\mathbf{u}, \theta]} \tilde{f}_{\gamma}^{[\mathbf{u}, \theta]} \mathcal{H}_{\alpha}^{[\mathbf{u}, \theta]}(\mathbf{v}) \\ &\quad - \frac{c\theta}{\mu} \sum_{|\alpha| > M_0} \nu_{M_0} \rho \tilde{f}_{\alpha}^{[\mathbf{u}, \theta]} \mathcal{H}_{\alpha}^{[\mathbf{u}, \theta]}(\mathbf{v}), \end{aligned}$$

where we have used (3.3) and (3.5). When  $\bar{\mathbf{u}} = \mathbf{u}$  or  $\bar{\theta} = \theta$  does not hold, we cannot use the same way to construct the approximate collision operators, i.e., we cannot just remove the superscripts  $[\mathbf{u}, \theta]$  in (3.20), since the resulting operator has no guarantee that it will vanish for Maxwellians, which is a fundamental property of the Boltzmann equation.

To overcome such a difficulty, we will apply (3.20) in our numerical method, where the coefficients  $\tilde{f}_{\alpha}^{[\mathbf{u}, \theta]}$  will be computed based on the knowledge of all the coefficients  $\tilde{f}_{\alpha}$ . The algorithm to obtain  $\tilde{f}_{\alpha}^{[\mathbf{u}, \theta]}$  is inspired by the method proposed in [5], and can be stated by the following theorem.

**THEOREM 3.1.** *Suppose the function  $\phi(\mathbf{v})$  satisfies*

$$(3.21) \quad \int_{\mathbb{R}^3} (1 + |\mathbf{v}|^M) |\phi(\mathbf{v})| d\mathbf{v} < +\infty$$

*for some positive integer  $M$ . Given  $\mathbf{w}, \mathbf{w}^* \in \mathbb{R}^3$ , and  $\eta, \eta^* > 0$ , for any  $\alpha \in \mathbb{N}^3$  satisfying  $|\alpha| \leq M$ , define*

$$(3.22) \quad \tilde{\phi}_{\alpha} = \frac{1}{\alpha!} \eta^{\frac{|\alpha|}{2}} \int_{\mathbb{R}^3} H_{\alpha} \left( \frac{\mathbf{v} - \mathbf{w}}{\sqrt{\eta}} \right) \phi(\mathbf{v}) d\mathbf{v}, \quad \tilde{\phi}_{\alpha}^* = \frac{1}{\alpha!} (\eta^*)^{\frac{|\alpha|}{2}} \int_{\mathbb{R}^3} H_{\alpha} \left( \frac{\mathbf{v} - \mathbf{w}^*}{\sqrt{\eta^*}} \right) \phi(\mathbf{v}) d\mathbf{v}.$$

*Then*

$$(3.23) \quad \tilde{\phi}_{\alpha}^* = \sum_{k=0}^{|\alpha|} \tilde{\phi}_{\alpha}^{(k)},$$

where  $\tilde{\phi}_\alpha^{(k)}$  is recursively defined by

$$(3.24) \quad \tilde{\phi}_\alpha^{(k)} = \begin{cases} \tilde{\phi}_\alpha & \text{if } k = 0, \\ \frac{1}{k} \sum_{j=1}^3 \left( (w_d^* - w_d) \tilde{\phi}_{\alpha - e_j}^{(k-1)} + \frac{1}{2} (\eta^* - \eta) \tilde{\phi}_{\alpha - 2e_j}^{(k-1)} \right) & \text{if } 1 \leq k \leq |\alpha|. \end{cases}$$

Here the terms with negative values in the subscript indices are regarded as zero.

*Proof.* For  $\tau \in [0, 1]$ , define the functions

$$(3.25) \quad \mathbf{w}(\tau) = (1 - \tau)\mathbf{w} + \tau\mathbf{w}^*, \quad \eta(\tau) = (1 - \tau)\eta + \tau\eta^*,$$

and

$$(3.26) \quad \tilde{\phi}_\alpha(\tau) = \frac{1}{\alpha!} [\eta(\tau)]^{\frac{|\alpha|}{2}} \int_{\mathbb{R}^3} H_\alpha \left( \frac{\mathbf{v} - \mathbf{w}(\tau)}{\sqrt{\eta(\tau)}} \right) \phi(\mathbf{v}) \, d\mathbf{v}, \quad |\alpha| \leq M.$$

The condition (3.21) ensures that  $\tilde{\phi}_\alpha(\tau)$  exists for any  $\tau \in [0, 1]$ . Especially, we have  $\tilde{\phi}_\alpha(0) = \tilde{\phi}_\alpha$  and  $\tilde{\phi}_\alpha(1) = \tilde{\phi}_\alpha^*$ . Now we take the derivative of (3.26) with respect to  $\tau$ . By straightforward calculation, we obtain

$$(3.27) \quad \frac{d\tilde{\phi}_\alpha(\tau)}{d\tau} = \sum_{d=1}^3 \left[ (w_d^* - w_d) \tilde{\phi}_{\alpha - e_d}(\tau) + \frac{1}{2} (\eta^* - \eta) \tilde{\phi}_{\alpha - 2e_d}(\tau) \right], \quad |\alpha| \leq M.$$

Considering the initial value  $\tilde{\phi}_\alpha(0) = \tilde{\phi}_\alpha$ , we claim that the solution of this ODE system is

$$(3.28) \quad \tilde{\phi}_\alpha(\tau) = \sum_{k=0}^{|\alpha|} \tilde{\phi}_\alpha^{(k)} \tau^k, \quad |\alpha| \leq M,$$

where  $\tilde{\phi}_\alpha^{(k)}$  is defined in (3.24). The verification of this claim is simply a direct substitution of (3.24) into (3.27), and the details are omitted. Setting  $\tau = 1$  in (3.28) and using  $\tilde{\phi}_\alpha(1) = \tilde{\phi}_\alpha^*$ , one completes the proof of (3.23).  $\square$

Theorem 3.1 provides an algorithm to obtain  $\tilde{f}_\alpha^{[\mathbf{u}, \theta]}$  from  $\tilde{f}_\alpha$ . In detail, we let

$$(3.29) \quad \mathbf{w} = \bar{\mathbf{u}}, \quad \eta = \bar{\theta}, \quad \mathbf{w}^* = \mathbf{u}, \quad \eta^* = \theta.$$

Then by (3.23) and (3.24), it can be seen that  $\tilde{f}_\alpha^{[\mathbf{u}, \theta]}$  can be represented by a linear combination of  $\tilde{f}_\beta$  with  $|\beta| \leq |\alpha|$ . Similarly, if we let  $\mathbf{w} = \mathbf{u}$ ,  $\eta = \theta$ ,  $\mathbf{w}^* = \bar{\mathbf{u}}$ ,  $\eta^* = \bar{\theta}$ , the same technique can help us to find  $\tilde{Q}_\alpha^*$  defined by

$$(3.30) \quad \tilde{Q}_\alpha^* = \frac{1}{\alpha!} \bar{\theta}^{\frac{|\alpha|}{2}} \int_{\mathbb{R}^3} H_\alpha \left( \frac{\mathbf{v} - \bar{\mathbf{u}}}{\sqrt{\bar{\theta}}} \right) Q^*[f, f](\mathbf{v}) \, d\mathbf{v},$$

where  $Q^*[f, f]$  is defined in (3.20). Thus our new collision model can be written as

$$(3.31) \quad Q^*[f, f](\mathbf{v}) = \sum_{\alpha \in \mathbb{N}^3} \tilde{Q}_\alpha^* \mathcal{H}_\alpha(\mathbf{v}),$$

where the map from  $\tilde{f}_\alpha$  to  $\tilde{Q}_\alpha^*$  can be summarized as follows:

$$(3.32) \quad \tilde{f}_\alpha \xrightarrow{\text{Theorem 3.1}} \tilde{f}_\alpha^{[\mathbf{u}, \theta]} \longrightarrow (3.20) \xrightarrow{\text{Theorem 3.1}} \tilde{Q}_\alpha^*.$$

Before closing this section, we would like to mention that the choice of the constant  $\nu_{M_0}$  in (3.19) and (3.20) should probably be determined by further numerical studies. Currently, we adopt the choice in [9, 31] and set  $\nu_{M_0}$  to be the spectral radius of the operator  $\hat{L}_{M_0} : F_{M_0}(0, 1) \rightarrow F_{M_0}(0, 1)$ , whose definition is

$$(3.33) \quad \hat{L}_{M_0}[h](\hat{\mathbf{v}}) = \sum_{|\alpha| \leq M_0} \sum_{|\beta| \leq M_0} (A_\alpha^{0,\beta} + A_\alpha^{\beta,0}) \tilde{h}_\beta \hat{\mathcal{H}}_\alpha(\mathbf{v}),$$

which is in fact the linearization of the quadratic operator  $\hat{Q}^*$  restricted on  $F_{M_0}(0, 1)$ . We refer the readers to [9, 31] for more details.

*Remark 3.2.* The transformation introduced in Theorem 3.1 can be avoided if we choose  $\bar{\mathbf{u}} = \mathbf{u}$  and  $\bar{\theta} = \theta$ . Thus the parameters  $\bar{\mathbf{u}}$  and  $\bar{\theta}$  are spatially and temporally dependent. Such a method has been implemented in [10] for linearized collision operators, and the ansatz of the distribution functions is essentially the same as Grad's expansion [18]. Here we do not take this option because Grad's method seems to have a convergence issue for problems with high temperature contrast. For details, we refer the readers to a recent work [11].

**3.3. Hermite spectral method for the Boltzmann equation with approximate collision term.** Having derived the approximate collision operator  $Q^*[f, f]$  in the previous subsection, we are ready to write down the equations for the coefficients  $\tilde{f}_\alpha(t, \mathbf{x})$  in (2.7). By Galerkin's method, the equations are obtained by the following equalities:

$$(3.34) \quad \frac{m}{\alpha!} \bar{\theta}^{\frac{|\alpha|}{2}} \int_{\mathbb{R}^3} H_\alpha(\hat{\mathbf{v}}) \left[ \frac{\partial f_M}{\partial t} + \nabla_{\mathbf{x}} \cdot (\mathbf{v} f_M) \right] d\mathbf{v} = \frac{m}{\alpha!} \bar{\theta}^{\frac{|\alpha|}{2}} \int_{\mathbb{R}^3} H_\alpha(\hat{\mathbf{v}}) Q^*[f_M, f_M](\mathbf{v}) d\mathbf{v},$$

where  $|\alpha| \leq M$  and  $\hat{\mathbf{v}}$  is defined in (3.3). To deal with the convection term, we need the recursion relation of the basis function  $\mathcal{H}_\alpha$ :

$$(3.35) \quad v_j \mathcal{H}_\alpha(\mathbf{v}) = \alpha_j \mathcal{H}_{\alpha - e_j}(\mathbf{v}) + \bar{u}_j \mathcal{H}_\alpha(\mathbf{v}) + \bar{\theta} \mathcal{H}_{\alpha + e_j}(\mathbf{v}).$$

Thus by orthogonality of Hermite polynomials, we obtain the following evolution equations for  $\tilde{f}_\alpha$ :

$$(3.36) \quad \frac{\partial}{\partial t} \tilde{f}_\alpha + \sum_{j=1}^3 \frac{\partial}{\partial x_j} \left( (\alpha_j + 1) \tilde{f}_{\alpha + e_j} + \bar{u}_j \tilde{f}_\alpha + \bar{\theta} \tilde{f}_{\alpha - e_j} \right) = \tilde{Q}_\alpha^*, \quad |\alpha| \leq M,$$

where  $\tilde{f}_\beta$  is regarded as zero if  $\beta$  contains negative indices or  $|\beta| > M$ . We remind the readers again that the right-hand side of (3.36) is a function of all  $\tilde{f}_\alpha$  by (3.32), which shows that the computation of  $\tilde{Q}_\alpha^*$  includes two parts:

1. application of the algorithm implied in Theorem 3.1 (the first and third arrows in (3.32)), whose time complexity is  $O(M^4)$  (see (3.23) and (3.24));
2. evaluation of all the coefficients in (3.20) (the second arrow in (3.32)), whose time complexity is  $O(M_0^9 + M^3)$ .

Therefore, the total time complexity for computing all  $\tilde{Q}_\alpha^*$  with  $|\alpha| \leq M$  is  $O(M_0^9 + M^4)$ .

To complete the problem, we need to supplement (3.36) with initial and boundary conditions. Suppose the initial condition for the original Boltzmann equation is  $f(0, \mathbf{x}, \mathbf{v}) = f_0(\mathbf{x}, \mathbf{v})$ . Then a natural initial condition for (3.36) is

$$(3.37) \quad \tilde{f}_\alpha(0, \mathbf{x}) = \frac{m}{\alpha!} \bar{\theta}^{\frac{|\alpha|}{2}} \int_{\mathbb{R}^3} H_\alpha \left( \frac{\mathbf{v} - \bar{\mathbf{u}}}{\sqrt{\bar{\theta}}} \right) f_0(\mathbf{x}, \mathbf{v}) d\mathbf{v}.$$

The boundary condition, especially the solid wall boundary condition, is slightly more complicated, and we will discuss this topic in the next section.

**4. Boundary condition.** Due to the hyperbolic nature of the Boltzmann equation, on the boundary of the spatial domain, we need to specify the value of the distribution function with velocity pointing into the domain. In the simulation of microflows, the wall boundary condition is especially important. In this paper, we focus on a popular type of boundary condition proposed by Maxwell in [25], which is a linear combination of the specular reflection and the diffuse reflection. Such a boundary condition has been studied for very similar methods in [7, 8], which make our work much easier. Below, we are going to first review the Maxwell boundary condition, and then propose the boundary condition for the Hermite spectral method.

#### 4.1. The Maxwell boundary condition for the Boltzmann equation.

Suppose  $\mathbf{x}_0 \in \partial\Omega$ . Let  $\mathbf{n}_0$  be the outer unit normal vector of the spatial domain  $\Omega$  at  $\mathbf{x}_0$ . Consider the case in which  $\mathbf{x}_0$  is the contact point of the gas and the solid wall. At point  $\mathbf{x}_0$ , the solid wall has temperature  $T^w$ , and is moving at velocity  $\mathbf{u}^w$ . By these assumptions, the Maxwell boundary condition is described as follows:

$$(4.1) \quad f(t, \mathbf{x}_0, \mathbf{v}) = \omega f_{\mathcal{M}}^w(t, \mathbf{x}_0, \mathbf{v}) + (1 - \omega)f(t, \mathbf{x}_0, \mathbf{v}^*) \quad \text{if } (\mathbf{v} - \mathbf{u}^w) \cdot \mathbf{n}_0 < 0,$$

where  $\omega \in [0, 1]$  is the accommodation coefficient of the wall, and  $f_{\mathcal{M}}^w$  and  $\mathbf{v}^*$  are defined as

$$(4.2) \quad f_{\mathcal{M}}^w(t, \mathbf{x}_0, \mathbf{v}) = \rho^w \mathcal{M}_{\mathbf{u}^w, \theta^w}(\mathbf{v}), \quad \mathbf{v}^* = \mathbf{v} - 2[(\mathbf{v} - \mathbf{u}^w) \cdot \mathbf{n}_0]\mathbf{n}_0.$$

In (4.2),  $\rho^w$  should be determined by the condition that the normal mass flux on the boundary is zero, that is,

$$(4.3) \quad \int_{(\mathbf{v} - \mathbf{u}^w) \cdot \mathbf{n}_0 < 0} [(\mathbf{v} - \mathbf{u}^w) \cdot \mathbf{n}_0] [f_{\mathcal{M}}^w(t, \mathbf{x}_0, \mathbf{v}) - f(t, \mathbf{x}_0, \mathbf{v}^*)] d\mathbf{v} = 0.$$

The boundary condition for the Hermite spectral method should be an approximation of the above boundary condition.

**4.2. Boundary condition for the Hermite spectral method.** The most natural idea to find the boundary conditions for (3.36) is to integrate the Maxwell boundary condition (4.1) against Hermite polynomials. However, if all the Hermite polynomials of degree less than or equal to  $M$  are taken into account, the resulting number of boundary conditions will generally be larger than the number required by the hyperbolicity. In general, for a hyperbolic system, the number of boundary conditions at  $\mathbf{x}_0 \in \Omega$  should be equal to the number of characteristics pointing into the domain  $\Omega$ . Below we will first find all the characteristic speeds of the system.

For clarification purposes, we rewrite (3.36) in the matrix-vector form:

$$(4.4) \quad \frac{\partial \tilde{\mathbf{f}}}{\partial t} + \sum_{j=1}^3 \mathbf{A}_j \frac{\partial \tilde{\mathbf{f}}}{\partial x_j} = \mathbf{Q}(\tilde{\mathbf{f}}),$$

where  $\tilde{\mathbf{f}}$  is a column vector with all the unknowns  $\tilde{f}_\alpha$ ,  $|\alpha| \leq M$ , as its components, and  $\mathbf{A}_j$  and  $\mathbf{Q}$  are, respectively, defined by the convection and collision terms in (3.36).

We first consider the case  $\mathbf{n}_0 = (1, 0, 0)^T$ , in which it is only necessary to find all the eigenvalues of  $\mathbf{A}_1$ . The matrix  $\mathbf{A}_1$  is in fact a reducible matrix, which can be

observed if we divide  $\tilde{\mathbf{f}}$  into the following subvectors:

$$(4.5) \quad \tilde{\mathbf{f}}_{\alpha'} = \left( \tilde{f}_{0,\alpha'}, \tilde{f}_{1,\alpha'}, \dots, \tilde{f}_{M-|\alpha'|,\alpha'} \right)^T, \quad \alpha' = (\alpha'_1, \alpha'_2) \in \mathbb{N}^2, \quad |\alpha'| \leq M.$$

Here the notation  $\tilde{f}_{k,\alpha'}$  designates the coefficient  $\tilde{f}_\alpha$  with  $\alpha = (k, \alpha'_1, \alpha'_2)$ . Apparently, the vector  $\tilde{\mathbf{f}}$  can be formed by gluing up  $\tilde{\mathbf{f}}_{\alpha'}$  for all  $\alpha' \in \mathbb{N}^2$ . Thus by (3.36), one can find that  $\mathbf{A}_1$  has a block-diagonal structure, and each block has a tridiagonal form

$$(4.6) \quad \mathbf{A}_{1\alpha'} = \begin{pmatrix} \frac{\bar{u}_1}{\bar{\theta}} & 1 & & & \\ & \frac{\bar{u}_1}{\bar{\theta}} & 2 & & \\ & & \frac{\bar{u}_1}{\bar{\theta}} & 3 & \\ & & & \ddots & \ddots \\ & & & & \frac{\bar{u}_1}{\bar{\theta}} & M-|\alpha'| \\ & & & & & \frac{\bar{u}_1}{\bar{\theta}} \end{pmatrix}, \quad \alpha' \in \mathbb{N}^2, \quad |\alpha'| \leq M.$$

All the eigenvalues of the above matrix have been given in [4] as<sup>3</sup>

$$(4.7) \quad \lambda(\mathbf{A}_{1\alpha'}) = \left\{ \bar{u}_1 + c_0 \sqrt{\bar{\theta}}, \bar{u}_1 + c_1 \sqrt{\bar{\theta}}, \dots, \bar{u}_1 + c_{M-|\alpha'|} \sqrt{\bar{\theta}} \right\},$$

where  $c_0, \dots, c_{M-|\alpha'|}$  are all the roots of the one-dimensional Hermite polynomial of degree  $M+1-|\alpha'|$ . Consequently, all the eigenvalues of  $\mathbf{A}_1$  are given by

$$(4.8) \quad \lambda(\mathbf{A}_1) = \bigcup_{\substack{\alpha' \in \mathbb{N}^2 \\ |\alpha'| \leq M}} \lambda(\mathbf{A}_{1\alpha'}).$$

When  $\mathbf{n}_0 = (1, 0, 0)^T$ , the number of boundary conditions at  $\mathbf{x}_0$  should equal the number of eigenvalues less than  $u_1^w$ . In general, this number varies with  $\bar{u}_1$  and  $\bar{\theta}$ , which makes it difficult to discuss the boundary conditions in the general setting. As a workaround, we assume that  $\bar{\mathbf{u}}$  is chosen such that  $\bar{u}_1 = u_1^w$ . Then, by the symmetry of the Hermite polynomials, the number of boundary conditions to be specified at  $\mathbf{x}_0$  is

$$(4.9) \quad \sum_{\substack{\alpha' \in \mathbb{N}^2 \\ |\alpha'| \leq M}} \left\lceil \frac{M-|\alpha'|}{2} \right\rceil.$$

In [8], it is proven that the number (4.9) equals the number of indices in the following index set:

$$(4.10) \quad \mathcal{A} = \{\alpha \in \mathbb{N}^3 \mid |\alpha| \leq M, \alpha_1 \text{ is odd}\}.$$

Therefore, as stated in the beginning of this section, all the boundary conditions can be formulated by

$$(4.11) \quad \begin{aligned} & \frac{m}{\alpha!} \bar{\theta}^{\frac{|\alpha|}{2}} \int_{(\mathbf{v}-\mathbf{u}^w) \cdot \mathbf{n}_0 < 0} H_\alpha \left( \frac{\mathbf{v}-\bar{\mathbf{u}}}{\sqrt{\bar{\theta}}} \right) f_M(t, \mathbf{x}_0, \mathbf{v}) \, d\mathbf{v} \\ &= \frac{m}{\alpha!} \bar{\theta}^{\frac{|\alpha|}{2}} \int_{(\mathbf{v}-\mathbf{u}^w) \cdot \mathbf{n}_0 < 0} H_\alpha \left( \frac{\mathbf{v}-\bar{\mathbf{u}}}{\sqrt{\bar{\theta}}} \right) [\omega f_M^w(t, \mathbf{x}_0, \mathbf{v}) + (1-\omega) f_M(t, \mathbf{x}_0, \mathbf{v}^*)] \, d\mathbf{v} \end{aligned}$$

<sup>3</sup>In [4], such a matrix is denoted as  $\mathbf{M}(\bar{u}_1, \bar{\theta})$ . It is shown in [4] that this matrix is similar to a diagonal matrix called  $\mathbf{A}(\mathbf{w})$ , whose diagonal entries are exactly the numbers given in (4.7).

for all  $\alpha \in \mathcal{A}$ . Here the function  $f$  in (4.3) should be changed to  $f_M$  when defining the “wall Maxwellian”  $f_M^w$ . The boundary conditions given by (4.11) also agree with Grad’s idea of using “odd moments” to ensure the continuity of the boundary conditions with respect to the accommodation coefficient  $\omega$ . We refer the readers to [18] for more details.

For a general normal vector  $\mathbf{n}_0$ , the above method still applies. We need to assume  $\bar{\mathbf{u}} \cdot \mathbf{n}_0 = \mathbf{u}^w \cdot \mathbf{n}_0$ , and replace  $H_\alpha((\mathbf{v} - \bar{\mathbf{u}})/\sqrt{\bar{\theta}})$  by  $H_\alpha(\mathbf{R}(\mathbf{v} - \bar{\mathbf{u}})/\sqrt{\bar{\theta}})$  in (4.11), where  $\mathbf{R}$  is a rotation matrix satisfying  $\mathbf{R}\mathbf{n}_0 = (1, 0, 0)^T$ . The remaining task is just to evaluate the integrals in (4.11). For the case  $\mathbf{n}_0 = (1, 0, 0)^T$ , this has been done in [8]. The results are

$$(4.12) \quad \tilde{f}_\alpha = \frac{2\omega}{2-\omega} \left[ \sqrt{\frac{2\pi}{\theta^w}} \hat{J}_{\alpha_1} J_{\alpha_2}(u_2^w - u_2) J_{\alpha_3}(u_3^w - u_3) \sum_{k=0}^{\lfloor M/2 \rfloor} S(1, 2k) \bar{\theta}^{1/2-k} \tilde{f}_{2ke_1} + \sum_{k=0}^{K(\alpha)} S(\alpha_1, 2k) \bar{\theta}^{\alpha_1/2-k} \tilde{f}_{\alpha+(2k-\alpha_1)e_1} \right], \quad \alpha \in \mathcal{A},$$

where  $K(\alpha) = \lfloor (M - \alpha_2 - \alpha_3)/2 \rfloor$ , and  $J_r(\cdot)$  and  $\hat{J}_r(\cdot)$  are recursively defined by

$$\begin{aligned} J_{-1}(u) &= 0, & J_0(u) &= 1, & J_r(u) &= \frac{1}{r}[(\theta^w - \bar{\theta})J_{r-2}(u) + uJ_{r-1}(u)], & r \geq 1, \\ S_0 &= 0, & S_1 &= \sqrt{\frac{\theta^w}{2\pi}}, & S_r &= -\frac{r-2}{r(r-1)}\bar{\theta}S_{r-2}, & r \geq 2, \\ \hat{J}_{-1} &= 0, & \hat{J}_0 &= 1/2, & \hat{J}_r &= \frac{1}{r}(\theta^w - \bar{\theta})\hat{J}_{r-2} - S_r, & r \geq 1. \end{aligned}$$

To define  $S(\cdot, \cdot)$ , we first introduce  $K(\cdot, \cdot)$  by

$$(4.13) \quad K(r, s) = \begin{cases} \frac{(-1)^{(r+s-1)/2} \sqrt{2\pi} (s-1)!!}{r 2^{(r-1)/2} (\frac{r-1}{2})!}, & r \text{ is odd and } s \text{ is even,} \\ 0, & \text{otherwise,} \end{cases}$$

which makes it convenient to define  $S(\cdot, \cdot)$ :

$$(4.14) \quad S(r, s) = \begin{cases} 1/2, & r = s = 0, \\ K(1, s-1), & r = 0 \text{ and } s \neq 0, \\ K(r, 0), & r \neq 0 \text{ and } s = 0, \\ K(r, s) + S(r-1, s-1)s/r, & \text{otherwise.} \end{cases}$$

It can be verified that when  $\alpha = e_1 = (1, 0, 0)$ , the boundary condition (4.12) can be simplified as  $\tilde{f}_{e_1} = 0$ , which indicates that the mass flux on the boundary is zero. In this paper, such a special case ( $\mathbf{n}_0 = (1, 0, 0)^T$ ) is sufficient for our numerical experiments. General discussions on the implementation of boundary conditions will be left for future work.

**5. Numerical algorithms and experiments.** Numerical algorithms to solve the system (4.4) with the boundary condition (4.12) are briefly introduced in this

section. Two spatially one-dimensional problems and a spatially two-dimensional problem, with the convenient settings

$$(5.1) \quad \frac{\partial \tilde{\mathbf{f}}}{\partial x_2} = \frac{\partial \tilde{\mathbf{f}}}{\partial x_3} \equiv 0 \quad \text{and} \quad \frac{\partial \tilde{\mathbf{f}}}{\partial x_1} \equiv 0,$$

respectively, are then given to illustrate the effectiveness of the proposed solver. Here  $\tilde{\mathbf{f}}$  is still the vector of coefficients for a three-dimensional distribution function.

**5.1. Numerical algorithm.** Suppose the spatial domain  $\Omega \subset \mathbb{R}^N$  is discretized by a uniform grid with cell size  $\Delta x$  and cell centers  $\mathbf{x}_j = (x_{j_1}, \dots, x_{j_N})$ ,  $j \in \mathbb{Z}^N$ . Using  $\tilde{\mathbf{f}}_j^n$  to approximate the average of  $\tilde{\mathbf{f}}$  over the  $j$ th grid cell  $[x_{j_1-1/2}, x_{j_1+1/2}] \times \dots \times [x_{j_N-1/2}, x_{j_N+1/2}]$  at time  $t^n$ , the system (4.4) can be solved by Euler's method with time step size  $\Delta t$  as follows:

$$(5.2) \quad \tilde{\mathbf{f}}_j^{n+1} = \tilde{\mathbf{f}}_j^n - \frac{\Delta t}{\Delta x} \sum_{d=1}^N \left[ \mathbf{F}_{j+\frac{1}{2}e_d}^n - \mathbf{F}_{j-\frac{1}{2}e_d}^n \right] + \Delta t \mathbf{Q}(\tilde{\mathbf{f}}_j^n),$$

where the finite volume method is employed for spatial discretization, and  $\mathbf{F}_{j+\frac{1}{2}e_d}^n$  is the numerical flux at the boundary between the cells with center  $\mathbf{x}_j$  and  $\mathbf{x}_{j+e_d}$ . In the present experiments, the Harten, Lax, and van Leer flux [19], given by

$$(5.3) \quad \mathbf{F}_{j+\frac{1}{2}e_d}^n = \begin{cases} \mathbf{A}_d \tilde{\mathbf{f}}_{j+\frac{1}{2}e_d}^{n,L}, & \lambda_d^L \geq 0, \\ \frac{\lambda_d^R \mathbf{A}_d \tilde{\mathbf{f}}_{j+\frac{1}{2}e_d}^{n,L} - \lambda_d^L \mathbf{A}_d \tilde{\mathbf{f}}_{j+\frac{1}{2}e_d}^{n,R} + \lambda_d^R \lambda_d^L (\tilde{\mathbf{f}}_{j+\frac{1}{2}e_d}^{n,R} - \tilde{\mathbf{f}}_{j+\frac{1}{2}e_d}^{n,L})}{\lambda_d^R - \lambda_d^L}, & \lambda_d^L < 0 < \lambda_d^R, \\ \mathbf{A}_d \tilde{\mathbf{f}}_{j+\frac{1}{2}e_d}^{n,R}, & \lambda_d^R \leq 0, \end{cases}$$

is adopted. Here  $\lambda_d^L = \bar{u}_d - C_{M+1}\sqrt{\bar{\theta}}$  and  $\lambda_d^R = \bar{u}_d + C_{M+1}\sqrt{\bar{\theta}}$ , where  $C_{M+1}$  is the maximal root of the Hermite polynomial of degree  $M+1$ . In our experiments,  $\bar{\mathbf{u}}$  will be set to be 0, and thus only the middle case of (5.3) is active. The scheme (5.2) can be improved straightforwardly to higher-order temporal schemes by Runge-Kutta methods. In order to get second-order spatial accuracy, the approximate solutions on the cell boundary  $\tilde{\mathbf{f}}_{j-\frac{1}{2}e_d}^{n,R}$  and  $\tilde{\mathbf{f}}_{j+\frac{1}{2}e_d}^{n,L}$  are computed by the linear reconstruction

$$(5.4) \quad \tilde{\mathbf{f}}_{j-\frac{1}{2}e_d}^{n,R} = \tilde{\mathbf{f}}_j^n - \frac{1}{2}\Delta x \mathbf{g}_d^n, \quad \tilde{\mathbf{f}}_{j+\frac{1}{2}e_d}^{n,L} = \tilde{\mathbf{f}}_j^n + \frac{1}{2}\Delta x \mathbf{g}_d^n$$

with  $\mathbf{g}_d^n = \frac{1}{2}(\tilde{\mathbf{f}}_{j+e_d}^n - \tilde{\mathbf{f}}_{j-e_d}^n)/\Delta x$ .

In our numerical experiments, we are interested in the steady state of microflows. However, due to the stability restriction of the explicit time-stepping scheme, the time step size should be chosen to satisfy the CFL condition

$$(5.5) \quad \Delta t \sum_{d=1}^N \frac{|\bar{u}_d| + C_{M+1}\sqrt{\bar{\theta}}}{\Delta x} < 1,$$

which indicates a long time simulation would be taken to achieve the steady state.

Such a method will be used in our two-dimensional examples to be shown in section 5.3. For one-dimensional steady-state problems, several additional techniques can be taken into account to accelerate the simulation by giving up the time accuracy of the solution. A simple way is to revise the computation of (5.2) on the whole spatial domain from the Jacobi-type iteration into a cell-by-cell symmetric Gauss–Seidel (SGS) iteration as shown in [21]. The SGS iteration is in general several times faster than the explicit time-stepping scheme, although for both methods, the total number of iterations is expected to grow linearly as the grid number increases.

Further acceleration of the steady-state computation can be obtained by using the multigrid technique, which has been explored in [20, 22]. The same framework of the nonlinear multigrid method as proposed in [20] is used in our simulation, except that the single level iteration is replaced by the above SGS iteration. By noting that  $\bar{u}$  and  $\bar{\theta}$  are constants,<sup>4</sup> the implementation is in fact much easier than that in [20].

**5.2. One-dimensional numerical experiments.** Numerical experiments of the planar Couette flow and the Fourier flow are carried out below. Numerical solutions of the quadratic collision term (3.20) as well as its linearization (3.33) are provided. In all simulations, a uniform grid with 256 cells is used for spatial discretization, and the gas of argon, which has molecular mass  $m = 6.63 \times 10^{-26}$  kg and molecular diameter  $d_{\text{ref}} = 4.17 \times 10^{-10}$  m at the reference temperature  $T_{\text{ref}} = 273.15$  K, is considered. The Maxwellian with density  $\rho = 9.282 \times 10^{-6}$  kg · m<sup>-3</sup>, velocity  $\mathbf{u} = 0$  m/s, and temperature  $T = 273.15$  K is adopted to set the initial value of the simulation. In order to match the reference results produced by the DSMC method [3], the viscosity coefficient  $\mu$  used in the collision term (3.20) is set to be

$$(5.6) \quad \mu = \frac{60(\eta - 1)^2 \sqrt{mk_B T_{\text{ref}}/\pi}}{(\eta - 2)(3\eta - 5)d_{\text{ref}}^2} \cdot \left( \frac{T}{T_{\text{ref}}} \right)^{\frac{1}{2}(\eta+3)/(\eta-1)},$$

where the Boltzmann constant  $k_B = 1.380658 \times 10^{-23}$  m<sup>2</sup> · kg · s<sup>-2</sup> · K<sup>-1</sup>, and the index  $\eta$  is set to be 10.

**5.2.1. The planar Couette flow.** Consider the gas between two infinite parallel plates, which have the temperature  $T^w = 273.15$  K, and move in the opposite direction along the plate with a speed of 119.25 m/s. Both plates are assumed to be completely diffusive, which indicates the accommodation coefficient  $\omega = 1$  in the boundary condition. Driven by the motion of the plates, the flow will reach a steady state as time tends to infinity. Numerically, we let the computational domain be  $[-D/2, D/2]$ , where  $D$  is the distance between the two plates. Four choices of the distance, i.e.,  $D = 0.092456$  m,  $0.018491$  m,  $0.003698$  m, and  $0.00074$  m, corresponding to the dimensionless Knudsen numbers  $Kn = 0.1, 0.5, 2.5$ , and  $12.5$ , respectively, are investigated. Additionally, since the problem setting is symmetric, it can be expected that the average speed equals zero. Therefore we set  $\bar{u} = 0$ . As for  $\bar{\theta}$ , we expect that the temperature of the flow does not differ too much from the wall temperature, which yields our choice  $\bar{\theta} = \frac{k_B}{m} T_{\text{ref}}$ .

(1)  $D = 0.092456$  m,  $Kn = 0.1$ : Numerical results for the quadratic collision term (3.20) with  $M_0 = 5$ , as well as the DSMC solutions, are listed in Figure 1. Only half of the domain is plotted, by noting that the density, the temperature, and the shear stress are even functions, and the heat flux is an odd function. Fast convergence of

<sup>4</sup>In [20], basis functions vary spatially. The idea has been sketched in section 1 and the implementation is more difficult due to the nonlinearity.



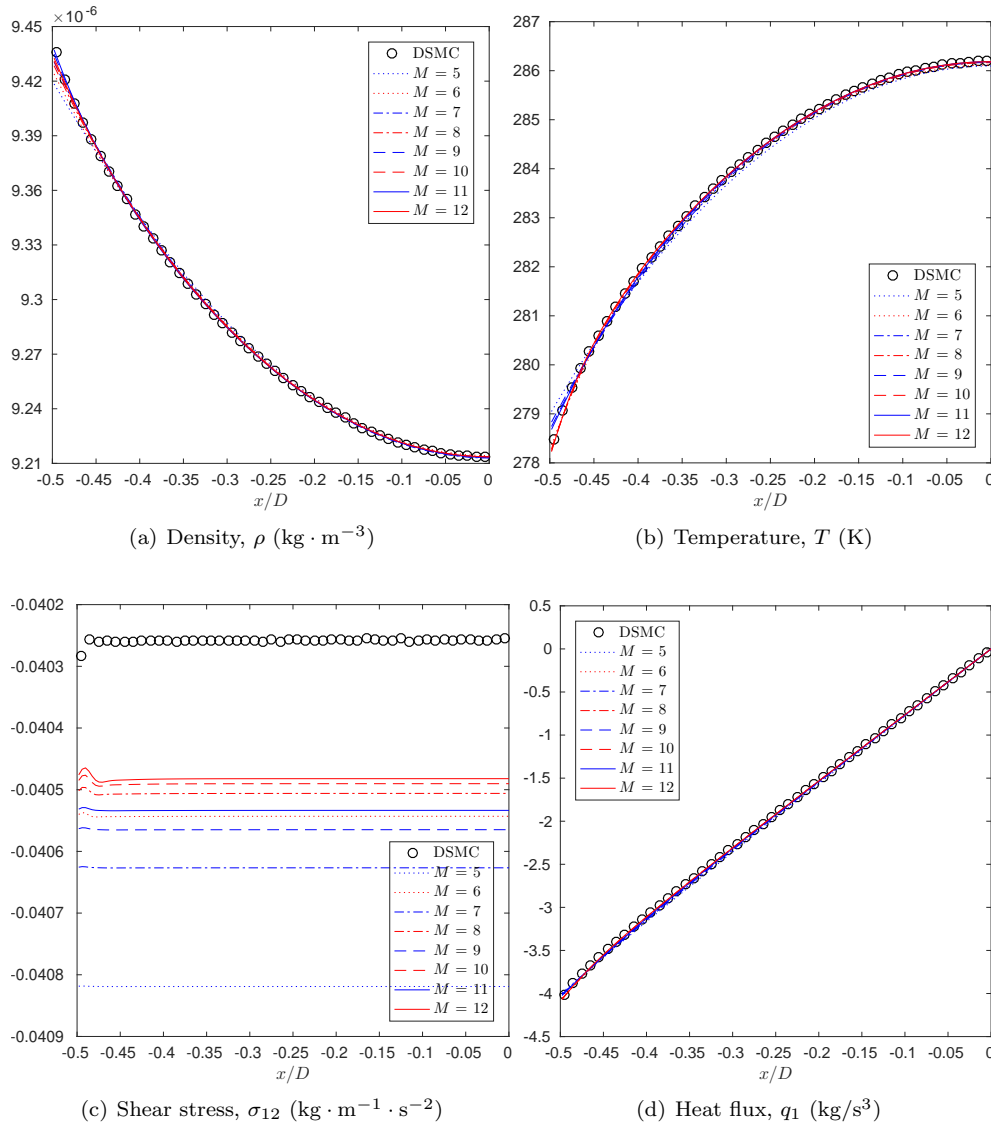


FIG. 1. Solution of the Couette flow for the quadratic collision term (3.20) with  $M_0 = 5$  and  $D = 0.092456$  m ( $Kn = 0.1$ ).

these quantities is observed as  $M$  increases. All results coincide very well with the DSMC results. Note that the actual relative error of shear stress  $\sigma_{12}$  is less than 1.5% even for the worst case  $M = 5$ , although an evident deviation can be seen from the figure. It turns out that a small  $M$ , e.g.,  $M = 5$ , with  $M_0 = 5$  for the quadratic collision term (3.20) is enough to give satisfactory results in this case. In fact, even for the linearized collision term (3.33) with  $M_0 = 5$ , numerical results also agree well with the results shown in Figure 1, except that a slight deviation can be observed for temperature. The comparison of temperature profiles between the quadratic collision term (3.20) and its linearization (3.33) can be found in Figure 2, from which one can see that the quadratic form provides a more accurate description of the fluid states.

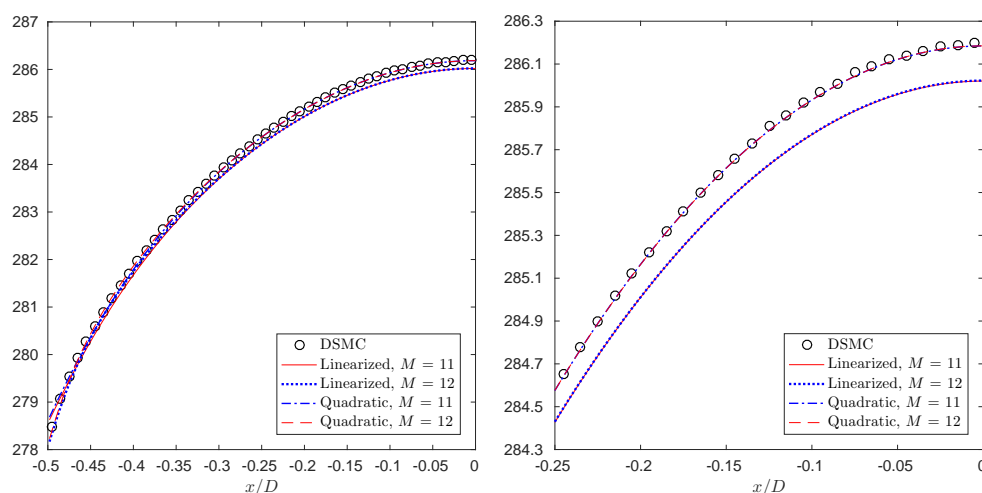


FIG. 2. Comparison of temperature (K) profiles (left) and its zoom (right) between the quadratic collision term (3.20) and its linearization (3.33) with  $M_0 = 5$  and  $D = 0.092456$  m ( $Kn = 0.1$ ).

(2)  $D = 0.018491$  m,  $Kn = 0.5$ : As the Knudsen number gets larger, a larger  $M$  is necessary to be considered. Numerical results for the quadratic collision term (3.20) with  $M_0 = 5$ , as well as the DSMC solutions, are shown in Figure 3. Again, only half of the domain is displayed. In this case, significant deviation can be observed between solutions with small  $M$  and the DSMC solutions. And the solutions behave differently for odd and even  $M$ , as exhibited many times in the literature (see, e.g., [10, 8]). In spite of this, the convergence can still be obtained for all plotted quantities, and they match the DSMC solutions better as  $M$  increases. Nevertheless, the quadratic collision term (3.20) with  $M_0 = 5$  still seems to be sufficient for Knudsen number 0.5, as long as  $M$  is sufficiently large.

Numerical results for the linearized collision term (3.33) with  $M_0 = 10$ , which is expected better than the same collision term with  $M_0 = 5$ , are presented in Figure 4 for comparison. Although convergence of these quantities is also observed with respect to  $M$ , the results are not as good as those obtained by the quadratic collision term with  $M_0 = 5$  and the same  $M$ . More precisely, there is a significant gap between the possible limiting temperature and the reference temperature given by the DSMC method. This indicates that the linearized collision term is indeed inadequate for problems with such a Knudsen number.

(3)  $D = 0.003698$  m,  $Kn = 2.5$ : This example tests our collision model for the flow in the transitional regime. Since the Knudsen number is even larger, we consider only the quadratic collision term (3.20) with  $M_0 = 10$ . The comparison between our results and the DSMC solutions are provided in Figure 5. It shows that the Hermite spectral method still provides high-quality solutions for lower-order moments such as density, temperature, and shear stress. Precisely speaking, for density, the relative deviation between the DSMC solution and our solution is lower than 0.03%, and the relative deviation of temperature and shear stress is less than 0.05% and 0.9%, respectively. For the heat flux, our solution agrees well with DSMC results when the flow is away from the boundary, while obvious discrepancy can be observed near the boundary, where the relative deviation is close to 9%.

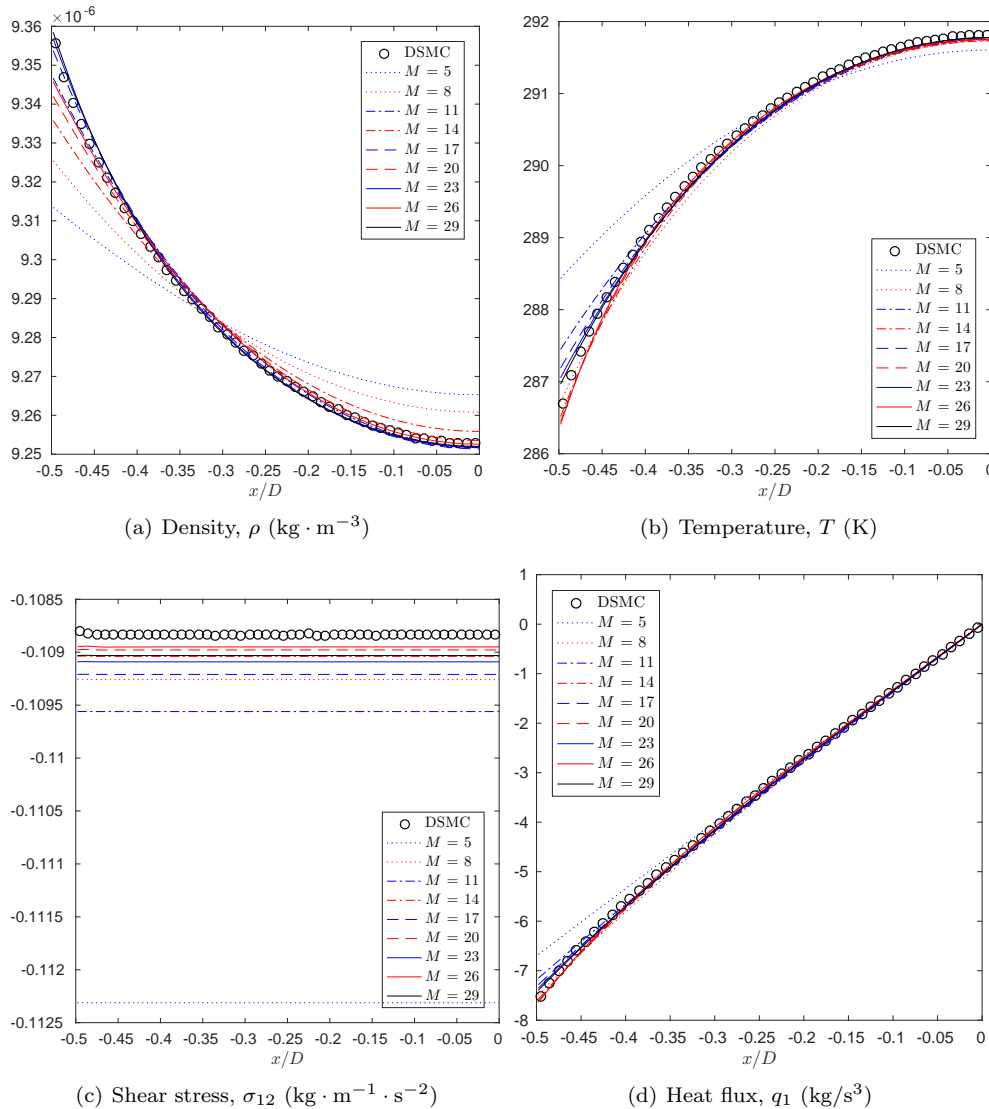


FIG. 3. Solution of the Couette flow for the quadratic collision term (3.20) with  $M_0 = 5$  and  $D = 0.018491$  m ( $Kn = 0.5$ ).

(4)  $D = 0.00074$  m,  $Kn = 12.5$ : For such a high Knudsen number, the flow is in the free molecular regime. The strong nonequilibrium requires an accurate modeling of the collision term to precisely capture the flow structure. Again we only present the results for the quadratic collision term (3.20) with  $M_0 = 10$ . Our numerical results and DSMC solutions are shown in Figure 6, where we see that our solutions are comparable to the DSMC solution for the density, temperature, and shear stress with  $M = 65$ . For the density, the relative deviation between our solution and the DSMC solution is 0.02%, and for temperature and shear stress, the relative errors are less than 0.3% and 1%, respectively. However, the structure of heat flux is not well captured. The relative deviation is around 30%. This may be because when the Knudsen number is large, a sharp discontinuity exists in the distribution function,

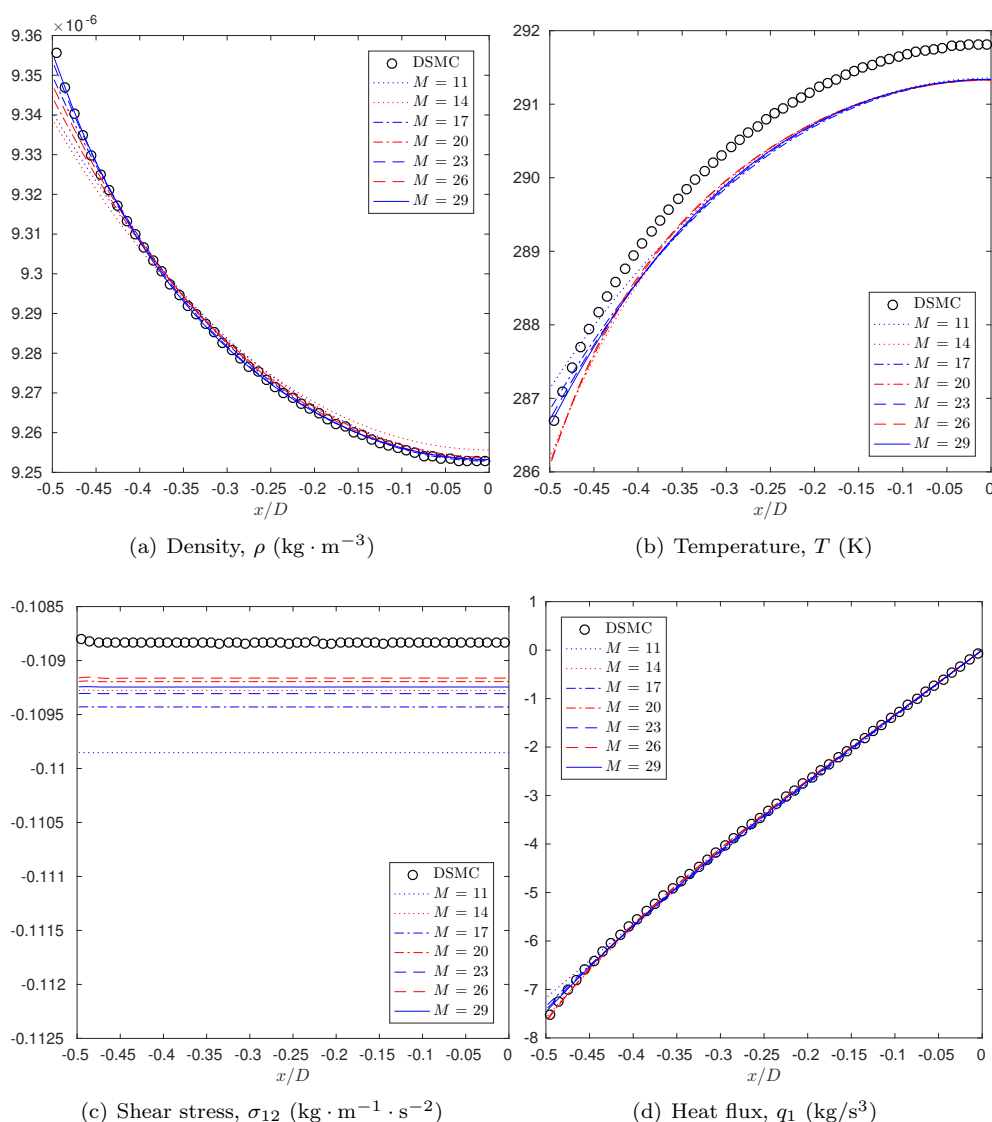


FIG. 4. Solution of the Couette flow for the linearized collision term (3.33) with  $M_0 = 10$  and  $D = 0.018491$  m ( $Kn = 0.5$ ).

which causes the Gibbs phenomenon when the distribution function is approximated using the spectral method. Thus, the spectral Galerkin method becomes inefficient. A similar observation is also presented in [30].

**5.2.2. The Fourier flow.** The second benchmark problem is the Fourier flow which also studies the motion of the gas between two infinite parallel plates. In contrast to the planar Couette flow, both plates are stationary, while their temperature is different. Specifically, the left plate has the temperature  $T_l^w = 273.15$  K and the right plate has the temperature  $T_r^w = 4T_l^w = 1092.6$  K. In this situation, the gas also reaches a steady state as time tends to infinity. To simulate it, we adopt  $\bar{\mathbf{u}} = 0$  since the steady state flow is static, and  $\bar{\theta}$  is chosen as  $\frac{k_B}{m} T_r^w$  to well capture the tail of the

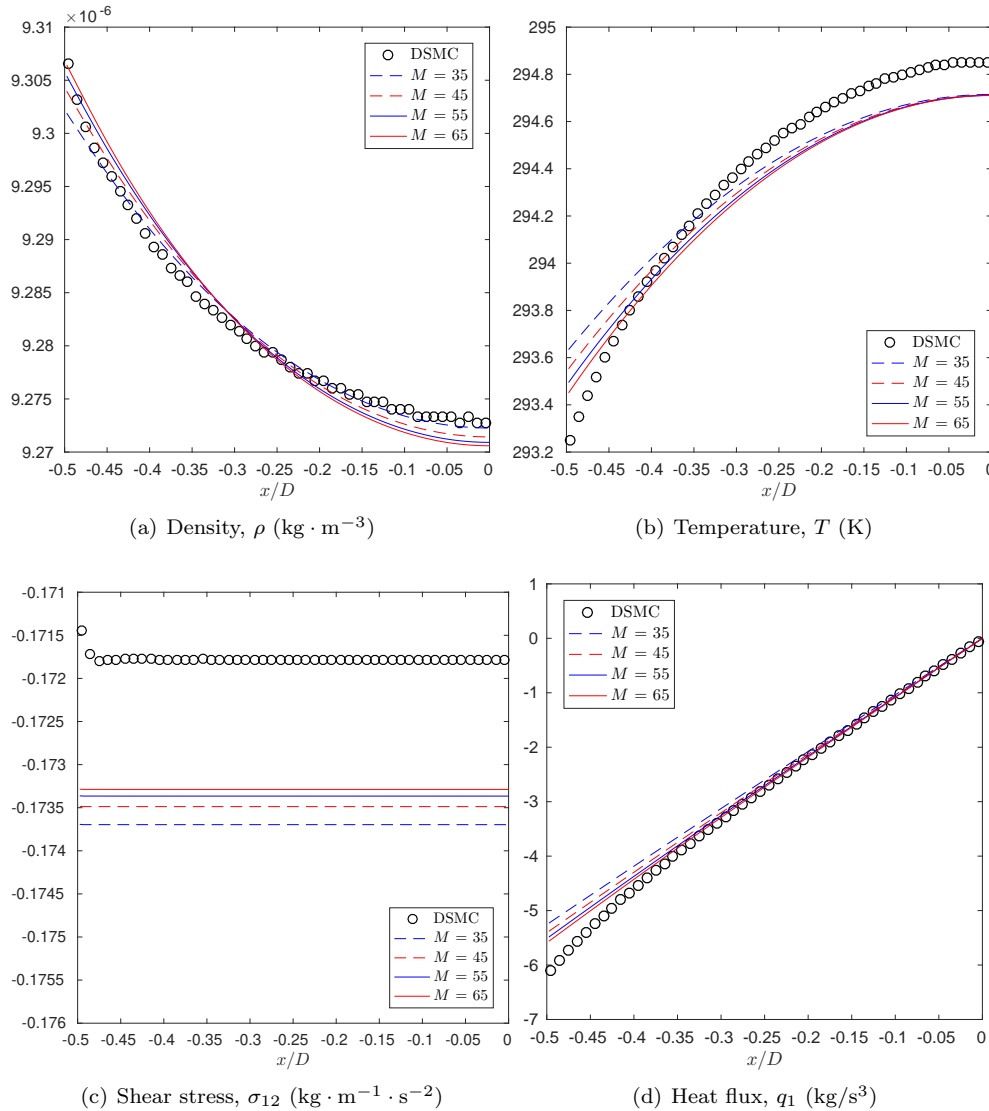


FIG. 5. Solution of the Couette flow for the quadratic collision term (3.20) with  $M_0 = 10$  and  $D = 0.003698$  m ( $Kn = 2.5$ ).

Maxwellian generated by particles reflected from the hot plate. The accommodation coefficient in the boundary condition is set to be  $\omega = 1$ , and the computational domain is still  $[-D/2, D/2]$  with  $D$  being the distance between two plates. Four distances  $D = 0.092456$  m,  $0.018491$  m,  $D = 0.003698$  m and  $0.00074$  m with the corresponding dimensionless Knudsen number  $Kn = 0.1$ ,  $0.5$ ,  $2.5$ , and  $12.5$ , respectively, are considered. Only results for the quadratic collision term (3.20) are presented.

(1)  $D = 0.092456$  m,  $Kn = 0.1$ : Numerical results for density  $\rho$ , temperature  $T$ , normal stress  $\sigma_{11}$ , and heat flux  $q_1$  with  $M_0 = 5$ , together with the DSMC solutions, are shown in Figure 7. Our results coincide very well with the DSMC solutions for density and temperature, while a small deviation for normal stress  $\sigma_{11}$  and heat flux  $q_1$  can be observed. Note that for heat flux  $q_1$ , the relative deviation between the

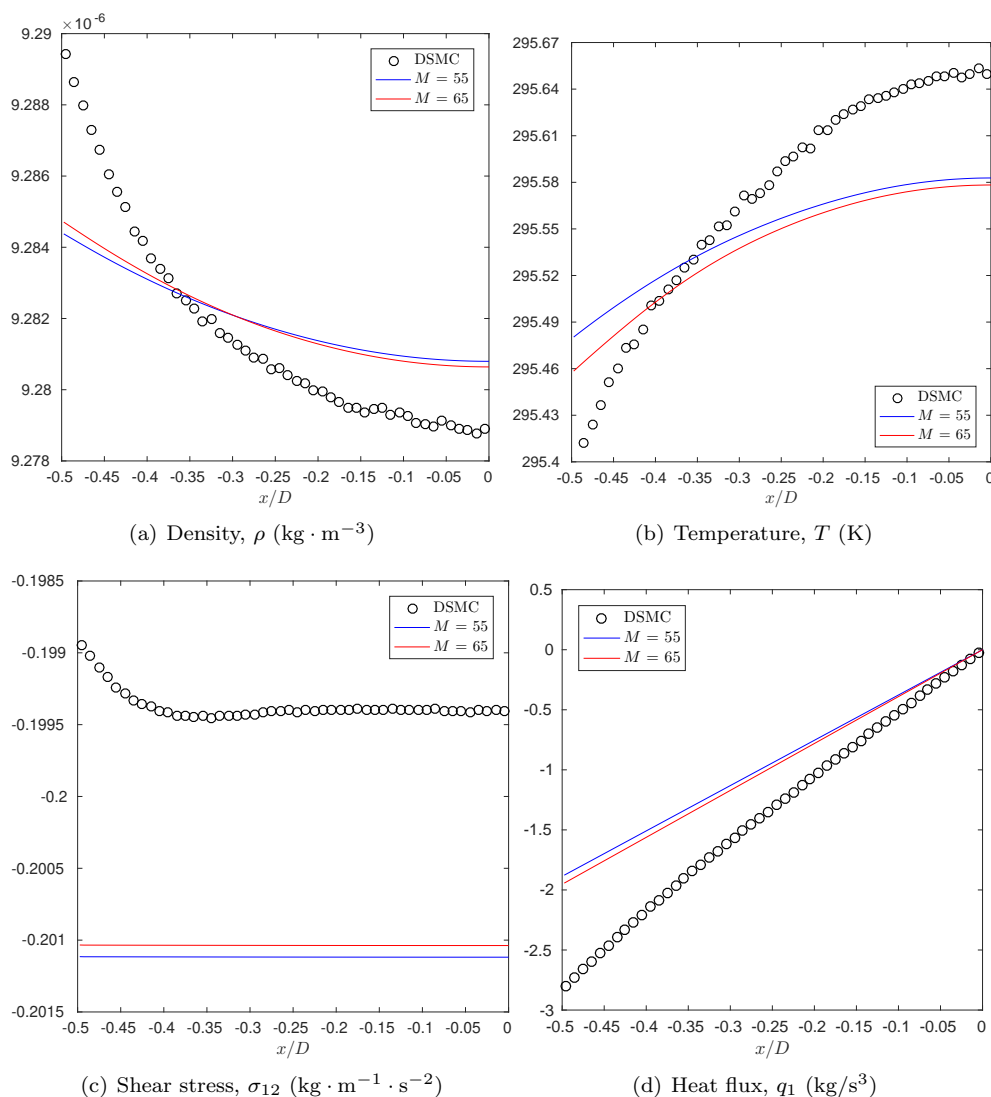


FIG. 6. Solution of the Couette flow for the quadratic collision term (3.20) with  $M_0 = 10$  and  $D = 0.00074$  m ( $Kn = 12.5$ ).

DSMC solution and our solution is less than 3% for all  $M$ . It is worth mentioning that  $q_1$  should be a constant in the steady-state solution, while the DSMC method provides a slanting profile. Such a result suggests the possible numerical error in the DSMC method, although we have run the DSMC code more than six days. It is left to the future work to determine what this constant should be.

Nevertheless, the deviation between our results and the DSMC solutions can be reduced by increasing  $M_0$  in the collision term. As an example, we plot the results of normal stress  $\sigma_{11}$  and heat flux  $q_1$  with  $M_0 = 10$  in Figure 8. Remarkable improvement can be observed.

(2)  $D = 0.018491$  m,  $Kn = 0.5$ : Numerical results with  $M_0 = 5$  and  $M_0 = 10$ , are presented in Figures 9 and 10, respectively. For this larger Knudsen number, evident deviations for all plotted quantities, in comparison to the DSMC results, can

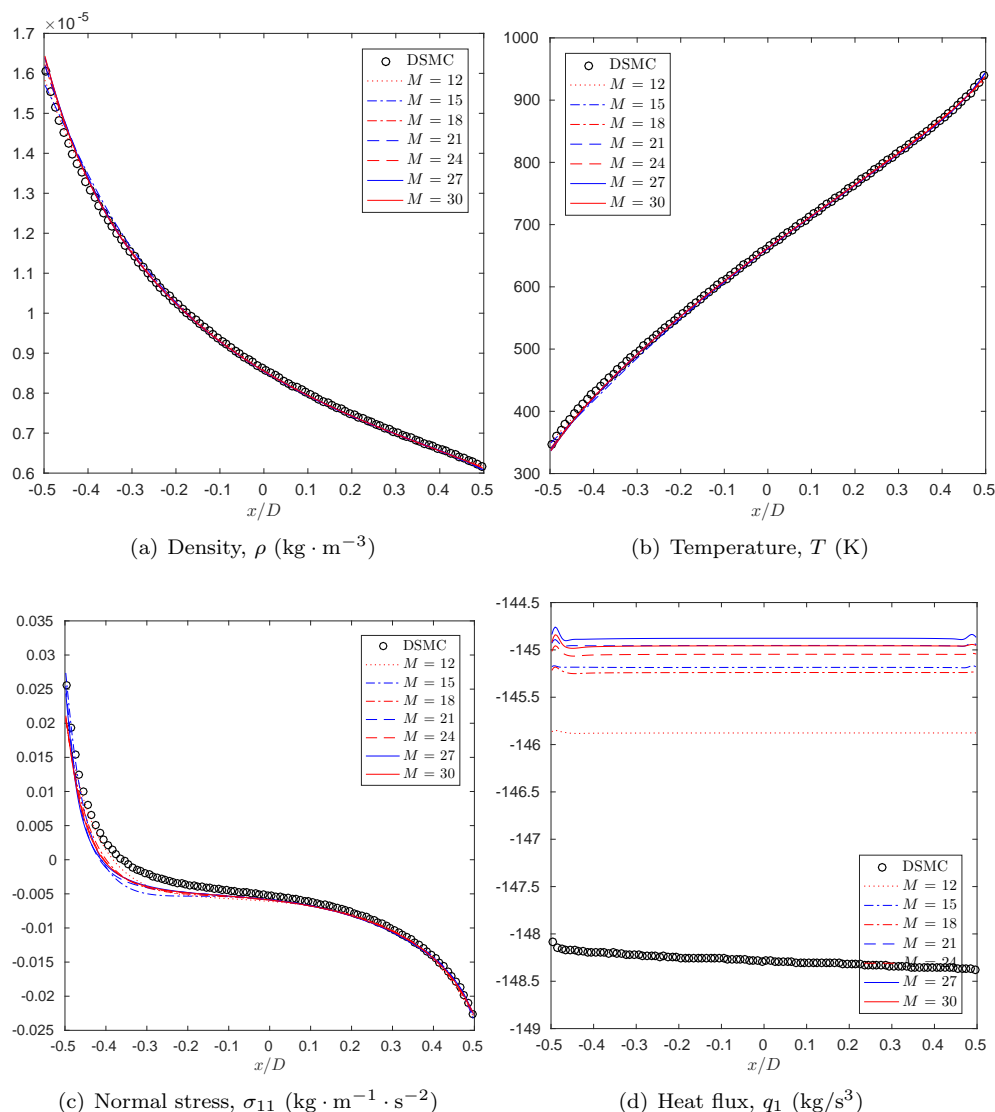


FIG. 7. Solution of the Fourier flow for the quadratic collision term (3.20) with  $M_0 = 5$  and  $D = 0.092456$  m ( $Kn = 0.1$ ).

be observed even for a large  $M$  in the case  $M_0 = 5$ . This indicates  $M_0 = 5$  is not enough for the simulation in this case.

As shown in Figure 10, the results with  $M_0 = 10$  again show considerable improvement, especially for  $M$  which is odd and larger than 20. For these  $M$ , all plotted quantities match the DSMC solutions quite well. It can also be observed that convergence of all plotted quantities with an even  $M$  is much slower, especially in the region near the left plate. The underlying reason remains to be further studied.

(3)  $D = 0.003698$  m,  $Kn = 2.5$ : Numerical solutions for  $M_0 = 10$  in this case are given in Figure 11, which shows the results for  $M = 35, 45, 55, 65$ . Despite a large Knudsen number, for all quantities, the profiles for different  $M$  are very close to each other, and they all agree well with DSMC solutions. The maximum relative deviation

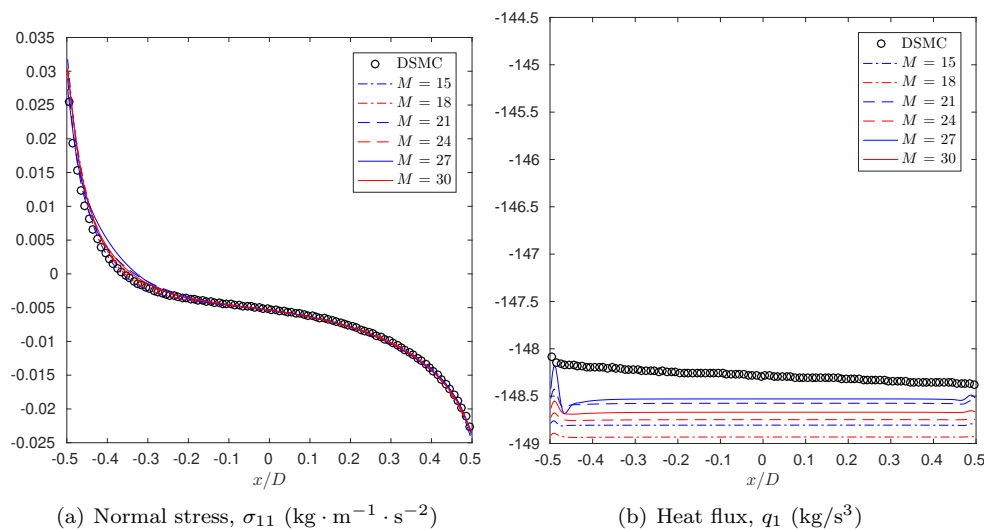


FIG. 8. Normal stress and heat flux of the Fourier flow for the quadratic collision term (3.20) with  $M_0 = 10$  and  $D = 0.092456$  m ( $Kn = 0.1$ ).

for all these quantities is less than 0.2%, which again shows the applicability of the Hermite spectral method for transitional flows.

(4)  $D = 0.00074$  m,  $Kn = 12.5$ : Numerical results for density  $\rho$ , temperature  $T$ , normal stress  $\sigma_{11}$ , and heat flux  $q_1$  with  $M_0 = 10$ , together with the DSMC solutions, are shown in Figure 12. It seems that our solutions are comparable with the DSMC solution for all these four quantities with  $M = 65$ . The relative deviation between our solution and the DSMC solution for the density and the temperature is 1.2% and 1.5%, respectively. But for the normal stress, the relative deviation is up to 15%. The relative deviation for heat flux is still quite small as to 0.1%. This may indicate the inadequacy of  $M_0$  in this simulation.

**5.3. Two-dimensional numerical experiments.** As a preliminary study of our method in the multidimensional case, we consider the two-dimensional lid-driven cavity flow which has been studied in [24, 32, 10]. In this case, the argon gas ( $m = 6.63 \times 10^{-26}$  kg) is confined in a square cavity with side length  $L = 1.25 \times 10^{-6}$  m. The temperature of the cavity walls is  $T = T_{\text{ref}} = 273$  K. The viscosity coefficient  $\mu$  is set to be

$$\mu = \mu_{\text{ref}} \left( \frac{T}{T_{\text{ref}}} \right)^{\frac{1}{2}(\eta+3)/(\eta-1)},$$

where the reference viscosity is  $\mu_{\text{ref}} = 2.117 \times 10^{-5}$  kg/(m · s), and the value of  $\eta$  is 7.45. Initially, the gas is in a uniform equilibrium with velocity  $\mathbf{u} = 0$  m/s, and temperature  $T = 273$  K, and the following two initial densities are considered:

1.  $\rho = 0.891$  kg/m<sup>3</sup>, corresponding to Knudsen number  $Kn = 0.1$ ;
2.  $\rho = 0.0891$  kg/m<sup>3</sup>, corresponding to Knudsen number  $Kn = 1.0$ .

The gas flow is driven by the top lid of the cavity, which moves to the right at a constant speed  $\mathbf{u}_w = (50, 0, 0)$  m/s. We expect that the steady state can be achieved after sufficiently long time. The simulation is carried out on a  $100 \times 100$  grid by explicit time stepping until  $5.24 \times 10^{-8}$  s. Since we are simulating the fluid in a closed cavity, whose average speed should be zero, we choose  $\bar{\mathbf{u}} = 0$ . The parameter  $\bar{\theta}$  is



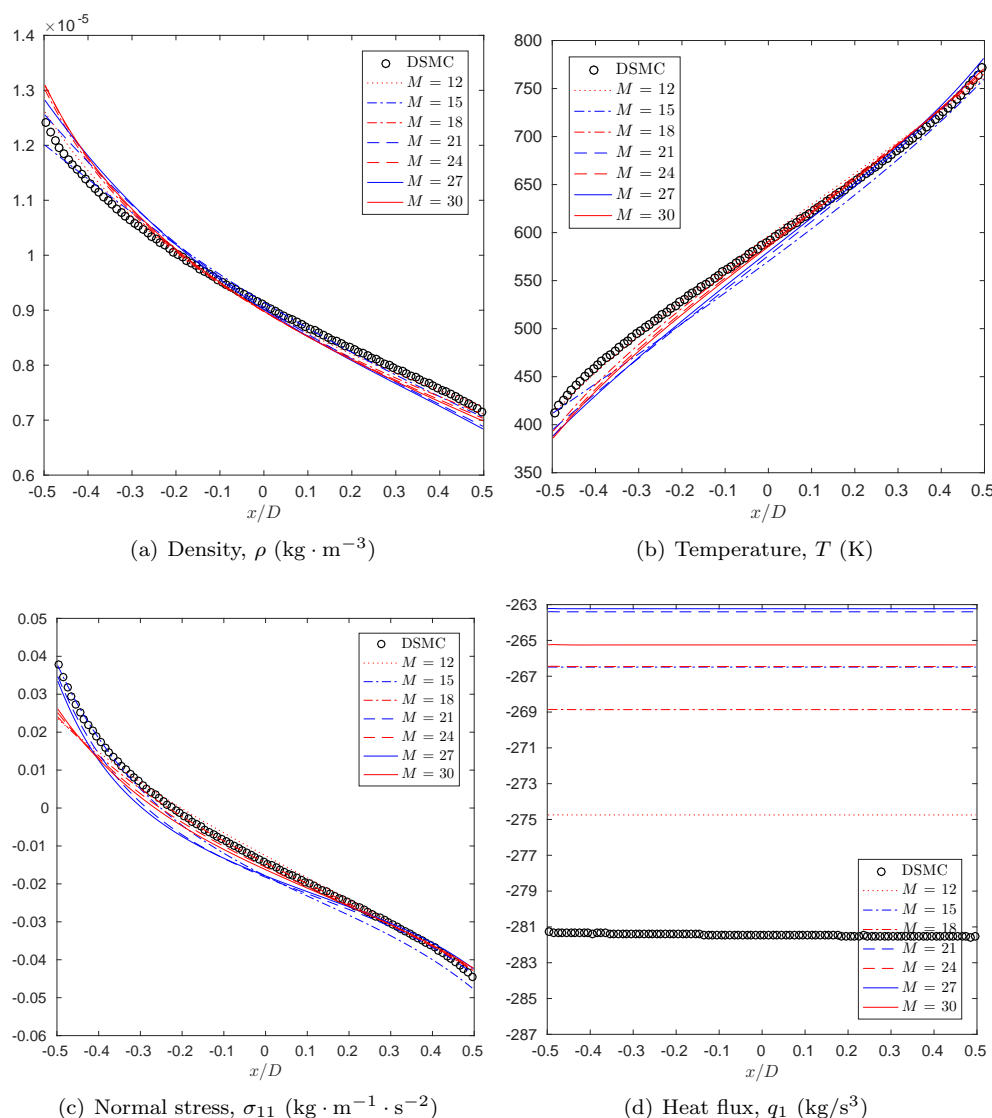


FIG. 9. Solution of the Fourier flow for the quadratic collision term (3.20) with  $M_0 = 5$  and  $D = 0.018491$  m ( $Kn = 0.5$ ).

chosen as  $\frac{k_B}{m} T_{\text{ref}}$  since we expect only a small variation of the temperature in this numerical test. For both Knudsen numbers, we use  $M_0 = 10$  in our simulation.

The simulation is run on the CPU model Intel Xeon E5-2680 v4 @ 2.40 GHz, and 28 threads are used in the simulation. Details of the simulations are given in Table 1. Here the total CPU time is obtained by the C function `clock()`, whose result is the sum of CPU time for all threads. Inspired by the tables presented in [14], we also provide the CPU time for each time step, each spatial grid, and each degree of freedom for easier comparison.

The results are again compared with DSMC results [24], which are provided in Figure 13 and 14. In general, two results agree well with each other, while some discrepancy can be found on the boundary of the domain. Such a discrepancy is

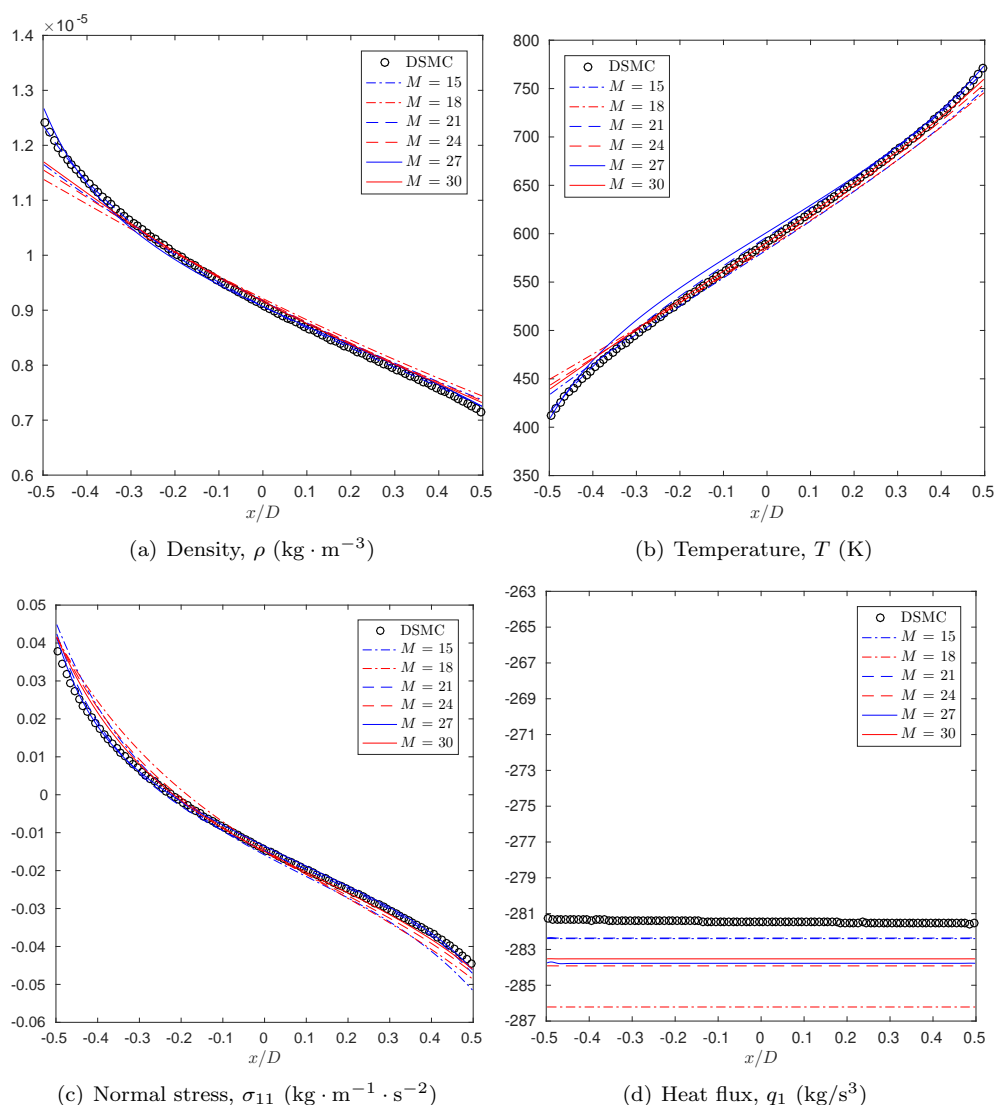


FIG. 10. Solution of the Fourier flow for the quadratic collision term (3.20) with  $M_0 = 10$  and  $D = 0.018491$  m ( $Kn = 0.5$ ).

probably related to the Gibbs phenomenon in the spectral method, since the distribution function on the boundary of the domain is generally discontinuous. Possible improvement includes using filters [1] or other boundary conditions [29], which will be part of our future work.

**6. Conclusion.** Based on the Hermite spectral method, we have developed a numerical solver for the Boltzmann equation with an approximate collision operator proposed in [31]. The approximate collision operator is derived from the original quadratic collision operator, but the quadratic form is preserved only for the first few moments. Our numerical simulation shows that a small number of degrees of freedom for the quadratic part can already provide much better results than the linear models, which makes it possible to design numerical methods that can well

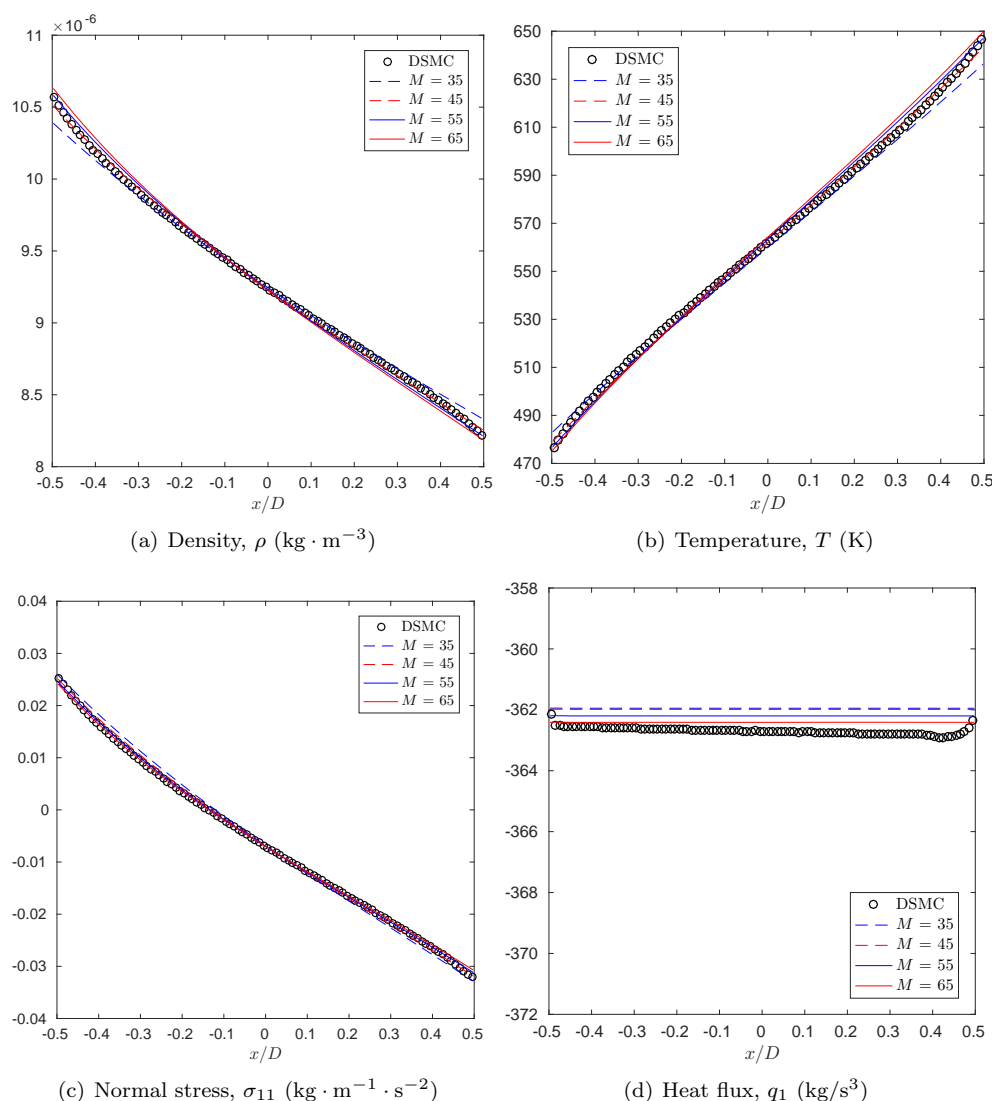


FIG. 11. Solution of the Fourier flow for the quadratic collision term (3.20) with  $M_0 = 10$  and  $D = 0.003698$  m ( $Kn = 2.5$ ).

balance the workload and the accuracy. Our major contribution to the algorithm is a special implementation of the collision operator. As is mentioned in section 1, the implementation of such a special collision operator in the spatially inhomogeneous case is not as straightforward as for the spatially homogeneous and normalized equation considered in [31]. By introducing a fast algorithm to change basis functions, we eventually obtain a numerical scheme with time complexity  $O(M_0^9 + M^4)$ .

Such a numerical cost makes the algorithm highly promising when applied to more complicated problems. Based on this framework, one can improve the efficiency of spatial discretization by a discontinuous Galerkin method, and consider using a GPU to distribute the computational load on all spatial grid cells (cf. [23]). Besides, research works on more multidimensional problems and polyatomic gases are ongoing.

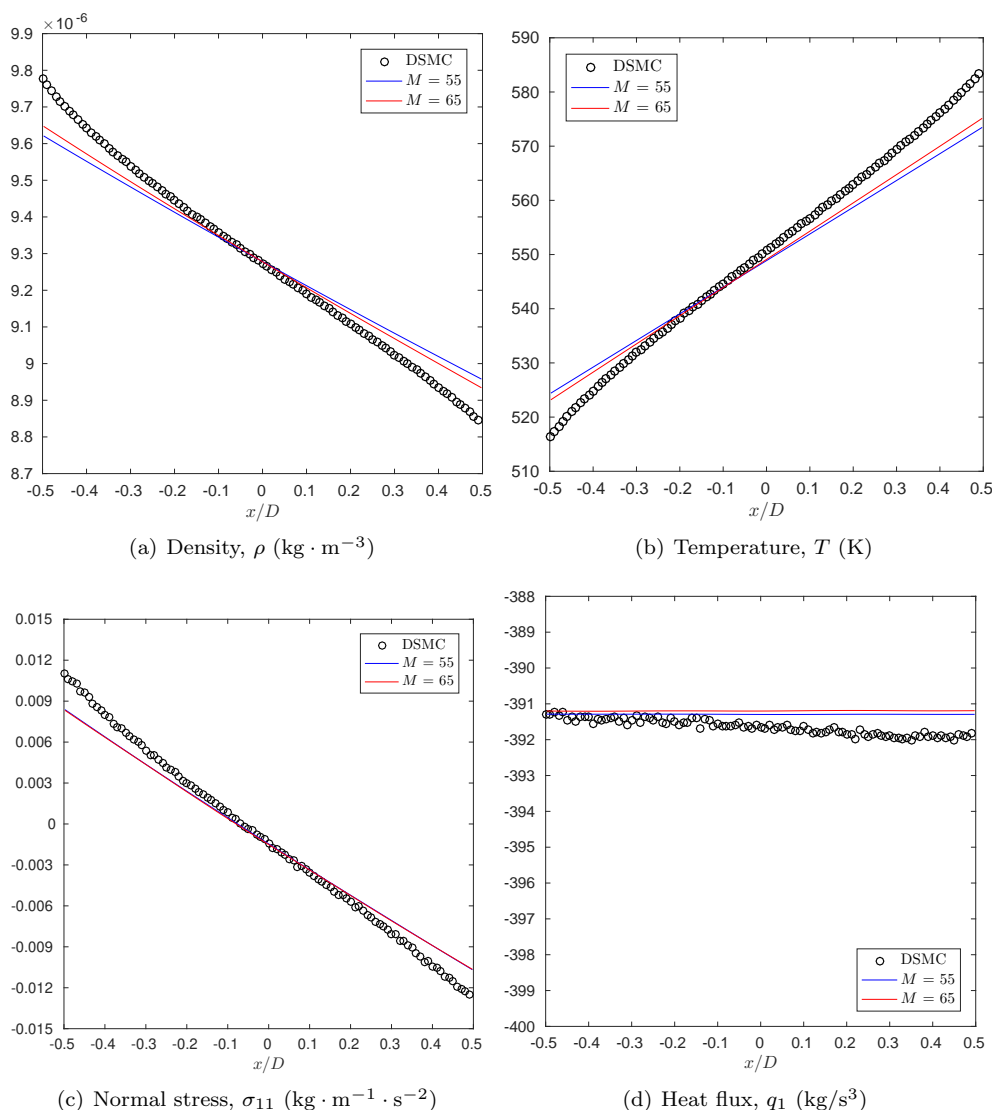


FIG. 12. Solution of the Fourier flow for the quadratic collision term (3.20) with  $M_0 = 10$  and  $D = 0.00074$  m ( $Kn = 12.5$ ).

TABLE 1  
Runtime data for the lid-driven cavity flow simulations.

Test case	$Kn = 0.1$	$Kn = 1.0$
$M$	25	35
Number of coefficients ( $N_M$ )	3276	8436
Time step ( $\Delta t$ )	$2.64 \times 10^{-12}$ s	$2.19 \times 10^{-12}$ s
Number of time steps ( $N_s$ )	19829	23968
Total CPU time ( $T_{\text{total}}$ )	$6.93 \times 10^6$ s	$1.07 \times 10^7$ s
CPU time per time step ( $T_s = T_{\text{total}}/N_s$ )	$3.50 \times 10^2$ s	$4.46 \times 10^2$ s
CPU time per grid ( $T_g = T_s/100^2$ )	$3.50 \times 10^{-2}$ s	$4.46 \times 10^{-2}$ s
CPU time per degree of freedom ( $T_d = T_g/N_M$ )	$1.07 \times 10^{-5}$ s	$5.30 \times 10^{-6}$ s

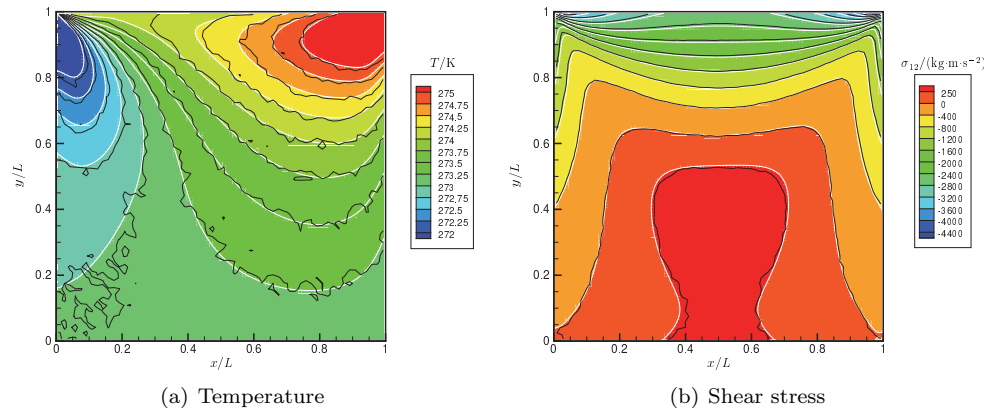


FIG. 13. Numerical results for  $Kn = 0.1$ . White contours: Hermite spectral method. Black contours: DSMC.

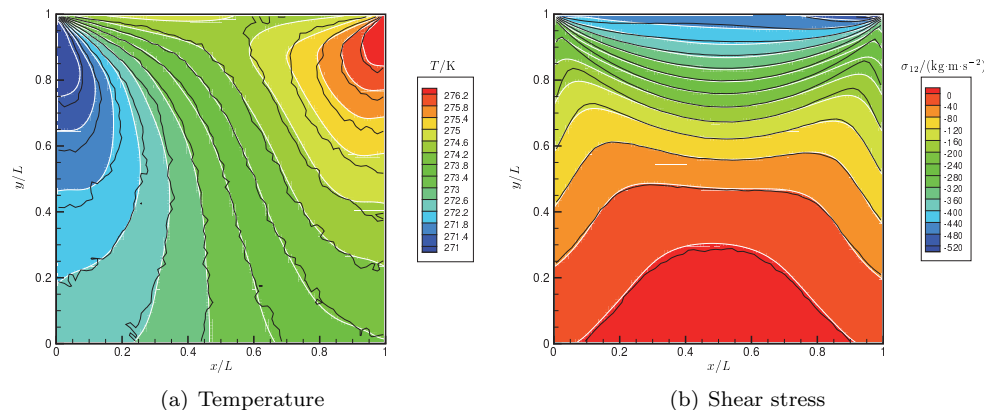


FIG. 14. Numerical results for  $Kn = 1.0$ . White contours: Hermite spectral method. Black contours: DSMC.

# REFERENCES

- [1] J. AGUIRRE AND J. RIVAS, *A spectral viscosity method based on Hermite functions for nonlinear conservation laws*, SIAM J. Numer. Anal., 46 (2008), pp. 1060–1078.
- [2] G. A. BIRD, *Approach to translational equilibrium in a rigid sphere gas*, Phys. Fluids, 6 (1963), pp. 1518–1519.
- [3] G. A. BIRD, *Molecular Gas Dynamics and the Direct Simulation of Gas Flows*, Clarendon Press, Oxford, 1994.
- [4] Z. CAI, Y. FAN, AND R. LI, *A framework on moment model reduction for kinetic equation*, SIAM J. Appl. Math., 75 (2015), pp. 2001–2023.
- [5] Z. CAI, Y. FAN, R. LI, AND Z. QIAO, *Dimension-reduced hyperbolic moment method for the Boltzmann equation with BGK-type collision*, Commun. Comput. Phys., 15 (2014), pp. 1368–1406.
- [6] Z. CAI AND R. LI, *Numerical regularized moment method of arbitrary order for Boltzmann-BGK equation*, SIAM J. Sci. Comput., 32 (2010), pp. 2875–2907.
- [7] Z. CAI, R. LI, AND Z. QIAO, *NRxx simulation of microflows with Shakhov model*, SIAM J. Sci. Comput., 34 (2012), pp. A339–A369.
- [8] Z. CAI, R. LI, AND Z. QIAO, *Globally hyperbolic regularized moment method with applications to microflow simulation*, Comput. & Fluids, 81 (2013), pp. 95–109.

- [9] Z. CAI AND M. TORRILHON, *Approximation of the linearized Boltzmann collision operator for hard-sphere and inverse-power-law models*, J. Comput. Phys., 295 (2015), pp. 617–643.
- [10] Z. CAI AND M. TORRILHON, *Numerical simulation of microflows using moment methods with linearized collision operator*, J. Sci. Comput., 74 (2018), pp. 336–374.
- [11] Z. CAI AND M. TORRILHON, *On the Holway-Weiss debate: Convergence of the Grad-moment-expansion in kinetic gas theory*, Phys. Fluids, 31 (2019), 126105.
- [12] S. CHAPMAN AND T. G. COWLING, *The Mathematical Theory of Non-uniform Gases*, 3rd ed., Cambridge University Press, Cambridge, 1990.
- [13] S. CHEN, K. XU, AND Q. CAI, *A comparison and unification of ellipsoidal statistical and Shakhov BGK models*, Adv. Appl. Math. Mech., 7 (2015), pp. 245–266.
- [14] G. DIMARCO, R. LOUBÉRE, J. NARSKI, AND T. REY, *An efficient numerical method for solving the Boltzmann equation in multidimensions*, J. Comput. Phys., 353 (2018), pp. 46–81.
- [15] I. M. GAMBA, J. R. HAACK, C. D. HAUCK, AND J. HU, *A fast spectral method for the Boltzmann collision operator with general collision kernels*, SIAM J. Sci. Comput., 39 (2017), pp. B658–B674.
- [16] I. M. GAMBA AND S. RJASANOW, *Galerkin-Petrov approach for the Boltzmann equation*, J. Comput. Phys., 366 (2018), pp. 341–365.
- [17] D. GOLDSTEIN, B. STURTEVANT, AND J. E. BROADWELL, *Investigations of the motion of discrete-velocity gases*, Progr. Astronaut. Aeronaut., 117 (1989), pp. 100–117.
- [18] H. GRAD, *On the kinetic theory of rarefied gases*, Comm. Pure Appl. Math., 2 (1949), pp. 331–407.
- [19] A. HARTEN, P. D. LAX, AND B. VAN LEER, *On upstream differencing and Godunov-type schemes for hyperbolic conservation laws*, SIAM Rev., 25 (1983), pp. 35–61.
- [20] Z. HU AND R. LI, *A nonlinear multigrid steady-state solver for 1D microflow*, Comput. Fluids, 103 (2014), pp. 193–203.
- [21] Z. HU, R. LI, AND Z. QIAO, *Acceleration for microflow simulations of high-order moment models by using lower-order model correction*, J. Comput. Phys., 327 (2016), pp. 225–244.
- [22] Z. HU, R. LI, AND Z. QIAO, *Extended hydrodynamic models and multigrid solver of a silicon diode simulation*, Commun. Comput. Phys., 20 (2016), pp. 551–582.
- [23] S. JAISWAL, A. A. ALEXEENKO, AND J. HU, *A discontinuous Galerkin fast spectral method for the full Boltzmann equation with general collision kernels*, J. Comput. Phys., 378 (2019), pp. 178–208.
- [24] B. JOHN, X.-J. GU, AND D. R. EMERSON, *Investigation of heat and mass transfer in a lid-driven cavity under nonequilibrium flow conditions*, Numer. Heat Transf., 58 (2010), pp. 287–303.
- [25] J. C. MAXWELL, *On stresses in rarefied gases arising from inequalities of temperature*, Proc. R. Soc. Lond., 27 (1878), pp. 304–308.
- [26] C. MOUHOT AND L. PARESCHI, *Fast algorithms for computing the Boltzmann collision operator*, Math. Comp., 75 (2006), pp. 1833–1852.
- [27] A. V. PANFEROV AND A. G. HEINTZ, *A new consistent discrete-velocity model for the Boltzmann equation*, Math. Methods Appl. Sci., 25 (2002), pp. 571–593.
- [28] L. PARESCHI AND B. PERTHAME, *A Fourier spectral method for homogeneous Boltzmann equations*, Transp. Theor. Stat., 25 (1996), pp. 369–382.
- [29] N. SARNA AND M. TORRILHON, *On stable wall boundary conditions for the Hermite discretization of the linearised Boltzmann equation*, J. Stat. Phys., 170 (2018), pp. 101–126.
- [30] W. SU, S. LINDSAY, H. LIU, AND L. WU, *Comparative study of the discrete velocity and lattice Boltzmann methods for rarefied gas flows through irregular channels*, Phys. Rev. E (3), 96 (2017), 023309.
- [31] Y. WANG AND Z. CAI, *Approximation of the Boltzmann collision operator based on Hermite spectral method*, J. Comput. Phys., 397 (2019), 108815, <https://doi.org/10.1016/j.jcp.2019.07.014>.
- [32] L. WU, J. M. REESE, AND Y. ZHANG, *Solving the Boltzmann equation deterministically by the fast spectral method: Application to gas microflows*, J. Fluid Mech., 746 (2014), pp. 53–84.