REVIEW   **Open Access**

Particle-based hybrid and multiscale methods for nonequilibrium gas flows

Jun Zhang[1], Benzi John[2], Marcel Pfeiffer[3], Fei Fei[4] and Dongsheng Wen[1,5*]

* Correspondence: d.wen@buaa.
edu.cn
[1]School of Aeronautic Science and
Engineering, Beihang University,
Beijing 100191, China
[5]School of Chemical and Process
Engineering, University of Leeds,
Leeds LS2 9JT, UK
Full list of author information is
available at the end of the article

Abstract

Over the past half century, a variety of computational fluid dynamics (CFD) methods and the direct simulation Monte Carlo (DSMC) method have been widely and successfully applied to the simulation of gas flows for the continuum and rarefied regime, respectively. However, they both encounter difficulties when dealing with multiscale gas flows in modern engineering problems, where the whole system is on the macroscopic scale but the nonequilibrium effects play an important role. In this paper, we review two particle-based strategies developed for the simulation of multiscale nonequilibrium gas flows, i.e., DSMC-CFD hybrid methods and multiscale particle methods. The principles, advantages, disadvantages, and applications for each method are described. The latest progress in the modelling of multiscale gas flows including the unified multiscale particle method proposed by the authors is presented.

Keywords: Nonequilibrium gas flow, Multiscale simulation, Hybrid method, DSMC, Fokker-Planck equation, Boltzmann equation, Unified statistical particle method

1 Introduction

While the traditional computational fluid dynamics (CFD) methods based on Navier-Stokes-Fourier (NSF) equations have been widely used for the simulation of gas flows, they encounter difficulties when applying to nonequilibrium gases, with too less molecular collisions in the flow timescales to ensure local thermodynamic equilibrium. The Knudsen number (Kn), which is defined as the ratio of the molecular mean free path to the characteristic length scale of the system, is usually used as a dimensionless parameter to evaluate whether molecular-level interactions affect continuum-level transport of momentum and energy. Correspondingly, different flow regimes can be identified based on Kn number, i.e. continuum regime ($Kn < 0.001$), slip regime ($0.001 < Kn < 0.1$), transition regime ($0.1 < Kn < 10$), and free molecular regime ($Kn > 10$).

For the continuum and slip regime, it is generally accepted that the NSF equations with linear constitutive relations are trustable to describe gas behaviors. It should be noted that slip boundary conditions need to be implemented for the slip regime. On the contrary, when the gas flow is in the transition or free molecular regime, the NSF equations with linear constitutive models break down, and the Boltzmann equation based on the kinetic theory of gases is proved to be an appropriate counterpart instead. It is well-known that solving the Boltzmann equation is cumbersome due to the complexity of the collision integral term. An alternative approach is the particle-based

direct simulation Monte Carlo (DSMC) method [1, 2], which tracks a large number of representative molecules, assuming that molecular motions and inter-molecular collisions are uncoupled during small time intervals. Theoretically, the DSMC method can be applied to the simulation of gas flows in the whole regime. However, due to the requirements that the cell sizes and time steps in DSMC need to be smaller than the mean free path and mean collision time, respectively, the application of DSMC to the continuum regime is computationally too expensive and sometimes inaccessible.

In many engineering problems, gas flows are not in a single flow regime but involve a variety of regimes and intrinsically multiscale [3, 4]. For instance, spacecrafts reentry meets the full Knudsen number range continuously when descending, i.e. from free molecular flow in the outer atmosphere to the continuum regime close to the earth. Another example is micro/nano-devices developed rapidly in the past three decades. It is common that in these devices the length scale of one dimension is on the micro or nano scale, so the corresponding Knudsen number is quite large. However, the length scale of the other dimensions might be on the macro scale, making the whole system multiscale. Even within a mostly continuum flow, there may be local regions of rarefaction (or nonequilibrium effects), such as the region of the sharp leading edge in near-space hypersonic vehicles, expanding plumes from a rocket at high altitude, shock waves, and boundary layers.

Computational methods designed for one scale are not adequate to simulate these multiscale flows in terms of physical accuracy and computational efficiency. In order to bridge this gap, many efforts have been made in the past two decades to develop hybrid or multiscale methods. One straightforward strategy is to combine the best of both the CFD solvers for the continuum regime and the DSMC methods for the rarefied regime [5]. In the literature, three different hybrid frameworks have emerged for the simulation of multiscale gas flows: domain-decomposition method (DDM) [6–8], heterogeneous multiscale method (HMM) [9], and internal-flow multiscale method (IMM) [10]. Note that the key point for these kinds of hybrid methods is how to transfer information between micro (DSMC) and macro (CFD) solutions.

Besides DSMC-CFD coupling, another strategy for multiscale simulation is to develop a unified solver for the whole flow regimes. Significant progress has been made in this research direction, such as the unified gas-kinetic scheme (UGKS) proposed by Xu and Huang [11] and the discrete unified gas-kinetic scheme (DUGKS) proposed by Guo et al. [12]. For both continuum and rarefied regimes, these two methods compute the gaseous distribution functions through discrete molecular velocities. Alternatively, it is promising to develop particle-particle hybrid methods [13, 14], which employ the DSMC method for the rarefied regime and particle methods based on the BGK or the Fokker-Planck model for the continuum regime. Compared to the UGKS and DUGKS methods, particle-particle hybrid methods are more efficient for the simulation of high-speed gas flows, especially when complex physical and chemical effects are considered. The latest progress in this direction is the unified particle method by combining the particle convection and collision effects [15]. The advantage is that it can be used with large time steps for the continuum flow, and thus the computational efficiency is highly improved.

The remainder of this paper is organized as follows. In section 2, we first review the existing continuum breakdown criterion and give a brief introduction of the DSMC

method. In section 3 and 4, two main hybrid and multiscale methods, i.e. the DSMC-CFD coupling and multiscale particle methods are reviewed, respectively. Lastly, conclusions are presented in Section 5.

2 Continuum breakdown criterion and DSMC method

2.1 Continuum breakdown criterion

It is well known that the macroscopic NSF equations, which provide a continuum description of gas flows, begin to break down under rarefied conditions. A popular assessment of the possibility of continuum breakdown is based on the Knudsen number. However, the Knudsen number in its global formulation depends on a macroscopic length scale, which is somewhat arbitrary. More important, in many flows of interest, some parts of the flow domain might be in the continuum regime, while other parts are in the rarefied regime. The continuum regions can be simulated using continuum method, while the rarefied regions must be simulated based on molecular description. Therefore, it is critical for an efficient hybrid or multiscale simulation to identify a switching criterion (also called "continuum breakdown parameter") to determine where and when to switch between two types of simulation models.

An often-used continuum breakdown parameter is a local Knudsen number [16, 17], where the length scale is defined based on the local spatial gradients of hydrodynamic variables, i.e.,

$$\mathrm{Kn}_L = \frac{\lambda}{Q}\left|\frac{dQ}{dx}\right|, \tag{1}$$

where λ is the molecular mean free path and Q can be a typical flow quantity such as density, temperature or macroscopic velocity. Note that different choices of Q may result in significant different values of the local Knudsen number. Wang and Boyd [18] proposed to use the maximum value of the different choices as the continuum breakdown parameter, i.e.,

$$\mathrm{Kn}_{L,\,\max} = \max\left(\mathrm{Kn}_{L,\rho}, \mathrm{Kn}_{L,T}, \mathrm{Kn}_{L,u}\right), \tag{2}$$

where $\mathrm{Kn}_{L,\,\rho}$, $\mathrm{Kn}_{L,\,T}$, and $\mathrm{Kn}_{L,\,u}$ are obtained using density, temperature, and velocity as the flow quantity in Eq. (1), respectively. The results obtained by Wang and Boyd [18] demonstrated that the modified breakdown parameter given in Eq. (2) could predict the failure of the continuum approach accurately for the simple cone flow and fairly well for the more complex cylinder/flare flow.

Alternatively, some researchers have developed other formulations of continuum breakdown parameters by measuring the departure from hydrodynamic behavior. For example, Lockerby et al. [19] proposed a new criterion based on the difference between the hydrodynamic near-equilibrium fluxes (the Navier–Stokes stress and Fourier heat flux) and the actual values of stress and heat flux (computed from molecular simulations or approximated through higher order constitutive relations); Meng et al. [20] proposed a kinetic criterion based on the molecular distribution function to indirectly assess the errors introduced by the description using NSF equations. Their numerical tests demonstrated that the new criterion performs very well in both low-speed and high-speed flows.

In addition, there are a few other continuum breakdown parameters reported in the literature. For example, Chen et al. [21] proposed a parameter based on the entropy generation rate; Gorji and Jenny [14] recommended a parameter based on the local collision rate. It should be noted that the continuum breakdown criterion is still an open question.

2.2 DSMC method

When the continuum NSF equations break down, it is required to employ molecular description instead. One widely used simulation method for gas flows is the DSMC method. DSMC method employs a large number of representative molecules that move in the computational domain, interact with boundaries, and collide with each other. The macroscopic quantities are obtained by sampling molecular information at the computational cells. It was first applied to the simulation of high-speed gas flows in the context of aerospace engineering. One successful example is that the DSMC results obtained for the space shuttle in the transitional regime (including the ratio of drag to lift) [22] have excellent agreement with the flight experimental data. More important, the molecular velocity distribution function in the shock wave structure predicted by DSMC calculations [23] is consistent with the experimental results. The verification of the DSMC method by the experiments enhanced its status and significance. Afterwards, the application of DSMC method has been extended to investigate a variety of gas flows at the molecular level, such as micro flows [24], flow instability [25–28], and even turbulence [29].

The key point of the DSMC method is that the molecular motions and inter-molecular collisions are assumed to be uncoupled during small computational time steps. The molecular motions in DSMC are implemented deterministically. In one calculating time step, each molecule moves from its original position to a new position according to its velocity. If the new position of a molecule is out of the computational domain, it means that the interaction of it with the boundary needs to be considered. The collision time is first calculated based on the molecular velocity and the distance between the molecule and the boundary, and then the reflected velocity is determined depending on the reflection models such as specular, diffuse, and Maxwell reflection models. Afterwards, the molecule moves according to its reflected velocity in the remaining time.

The inter-molecular collisions are implemented as a probabilistic process, which is the critical difference between DSMC and other deterministic methods such as molecular dynamics (MD). The basic idea is to simulate an appropriate number of representative collisions in each cell, with randomly selected molecule collision pairs. Several collision modeling techniques have been proposed within the framework of DSMC. The most widely used model is the no-time-counter technique proposed by Bird [1]. The post-collision velocities of the molecules depend on the particular molecular models employed, and this plays an important role in the exact modeling of the real gas flows. The DSMC method allows the introduction of the phenomenological models capable of reflecting the essential features of the flow field, such as the variable hard sphere (VHS) and the variable soft sphere (VSS) models [1].

Theoretically, the DSMC method can be applied to simulate the whole regime of gas flows. However, for an accurate simulation using DSMC method, the time step needs

to be smaller than the molecular mean collision time, and the cell sizes for the selection of collision pairs need to be smaller than the molecular mean free path [30, 31]. These requirements make the application of DSMC to large-scale continuum flows tremendously expensive. Therefore, hybrid or multiscale simulation techniques are highly required for modelling multiscale gas flows in modern engineering applications.
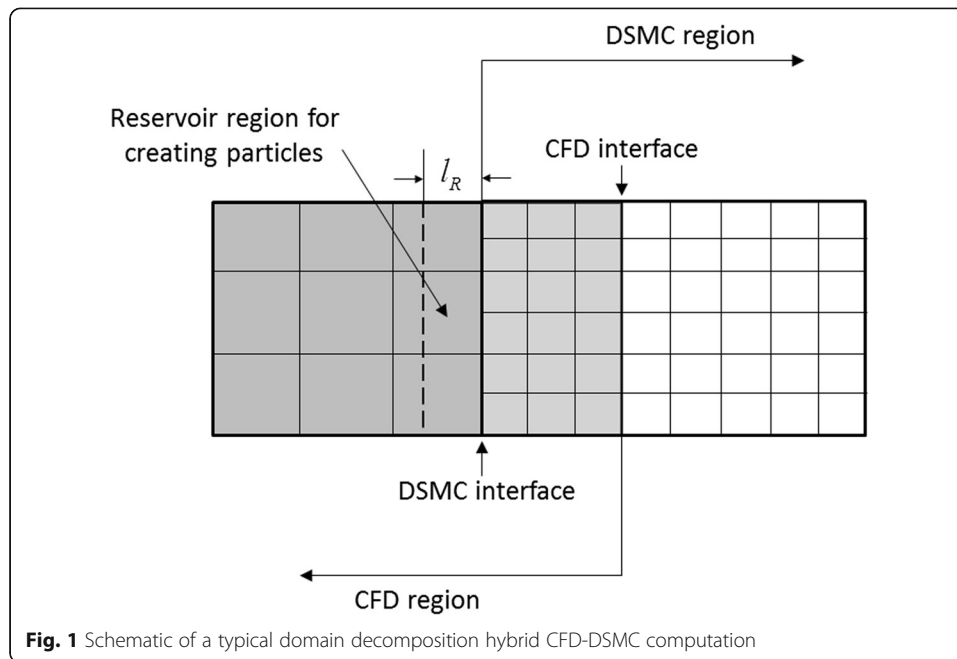
3 DSMC-CFD hybrid methods

Hybrid DSMC-CFD methods enable the combination of continuum and nonequilibrium models in gaseous flow problems. The selection of a particular hybrid method is dependent on the characteristic of the gas flows under consideration. The most popular hybrid method is the domain decomposition method (DDM), in which nonequilibrium regions (modelled by a molecular/particle method like the DSMC method) and equilibrium regions (modelled by a CFD method solving Navier-Stokes-Fourier equations) need to be demarcated. Information exchange is typically implemented via an overlap region, to enable a coupled hybrid solution. Besides DDM, two other approaches have emerged for multiscale coupling in the simulation of fluid flows, namely, heterogeneous multiscale method (HMM) and internal-flow multiscale method (IMM). The HMM and IMM are specifically tailored to suit particular classes of flow problems, and hence they are not general in nature like DDM. HMM is a point-wise coupling approach, suited particularly for complex fluid flows where there is a large degree of length scale separation. The molecular solver in HMM is solved on the continuum grid points to provide the missing information for the hydrodynamic solver. IMM is particularly tailored for high aspect ratio internal-flows, where the coupling is essentially enabled by the constraints applied by macroscopic conservation laws. These three multi-scale approaches are discussed in more details in the following.

3.1 Domain decomposition method

A schematic of the DDM coupling approach is shown in Fig. 1. The entire computational domain is divided into separate DSMC and CFD regions. In DDM, the computational efficiency is achieved by limiting the DSMC calculations only to the regions where it is really needed. The accuracy depends on various aspects like correct demarcation of the equilibrium and nonequilibrium regions, the appropriate selection of a coupling method, and a proper information exchange methodology between the DSMC and the CFD regions.

The choice of a coupling method is usually dictated by the flow characteristics of the application under consideration. For instance, this could be either based on a time-dependent explicit coupling or a time-independent implicit coupling formulation. A time-explicit coupling works best for high-speed flows, where compressible formulation is the method of choice for solving the hydrodynamic equations. This is because that the magnitude of the characteristic timescale in high-speed flows is relatively comparable to both the compressible Courant–Friedrichs–Lewy (CFL) time step and the DSMC time step. Additionally, high speeds imply that the statistical uncertainties of DSMC results are significantly smaller. Time-explicit coupling can be based on either flux-based [5] or state-based [7] variables.

Fig. 1 Schematic of a typical domain decomposition hybrid CFD-DSMC computation

On the other hand, in low-speed flows, incompressible formulations for solving hydrodynamic equations are typically coupled using either state (Dirichlet) properties or gradient (Neumann) variables. Low-speed flow applications are typically encountered in MEMS applications, wherein there is a wide disparity between the timescales of molecular and continuum models. For such cases, explicit integration at the molecular time step to the global time scale is computationally prohibitive, especially if the CFD sub-domain is large. Additionally, flux-based coupling is not suitable for low-speed flows because of the high statistical noise of the flux variables compared to the state variables. Therefore, implicit frameworks like the Schwarz alternating method [32, 33] provide a more appropriate choice for obtaining coupled solutions in such cases. Here, coupling is achieved by an iterative exchange of boundary conditions (state variables) across an overlap region, as shown in Fig. 1. Time-scale decoupling is achieved through iterations between steady state solutions of the continuum and molecular subdomains until convergence.

Irrespective of the choice of the coupling method in DDM as described above, the information exchange between the continuum and molecular regions needs careful consideration. Information exchange typically occurs over an overlap region in which both continuum and atomistic approaches are valid. The transfer from the molecular to the continuum subdomain is achieved by extracting macroscopic fields from molecular simulations. Due to the inherent statistical noise in the DSMC data, usually a correction is imposed to the boundary conditions so that discrepancies are kept to a minimum. Information exchange from the continuum to atomistic domain takes place with the aid of buffer cells (volume reservoirs) which are filled with particles.

3.2 Heterogeneous multiscale method
In contrast to DDM, where the domain is decomposed clearly into continuum and particle sub-domains, the grids of the continuum-fluid solver in HMM cover the entire

computational domain. Particle simulations (solvers) are dispersed at the nodes of the computational grids to provide the missing information for the hydrodynamic equations, as illustrated in Fig. 2. Concurrently, the particle solver is constrained by the local continuum solution at the collocated macro nodes. Thus, the particle solver enters as a refinement (point-wise coupling), rather than computing a specific part of the flow field [34, 35].

HMM coupling is typically employed when either the constitutive relations are not known or accurate wall boundary conditions cannot be formulated. Specifically, the particle simulations provide information on stresses and transport properties like viscosity or diffusion coefficients to the macroscopic solver, when constitutive relations are not known in the bulk of the flow. On the other hand, in the near wall regions, particle simulations can be used to provide accurate boundary conditions with respect to shear stress, heat flux or slip velocity. Length scale decoupling is essentially achieved by keeping the particle domains very small compared to the continuum cell sizes. HMM also achieves time scale decoupling by performing particle simulations only for very short time periods compared to the continuum time step sizes.

In the context of fluid dynamics, HMM has been mainly employed for pure liquid flows or complex fluid flows, using molecular dynamics (MD) as the molecular solver [34, 35]. The HMM point-wise coupling approach is ideal for problems where there is a large degree of spatial scale separation in the system, and therefore significant computational savings over a full molecular level simulation can be achieved. However, its application to gaseous flows has been comparatively limited. Recently, an HMM field-wise coupling approach (HMM-FWC) [9] has been shown appropriate for multi-scale gaseous flow problems where, typically, there are smaller degrees of length scale separations. In the HMM-FWC model, both the position and size of the micro elements are flexible and their locations are not restricted to the computational nodes. A two-way information exchange is then performed iteratively between the micro elements and continuum sub-regions to achieve a coupled solution.

3.3 Internal-flow multiscale method

IMM was originally developed as an extension of the HMM approach. This coupling approach has been specifically designed for the high aspect ratio internal micro-flows, where the length scale in the direction of flow is significantly larger than the length scales transverse to the flow direction. Similar to the HMM, the IMM covers a flow
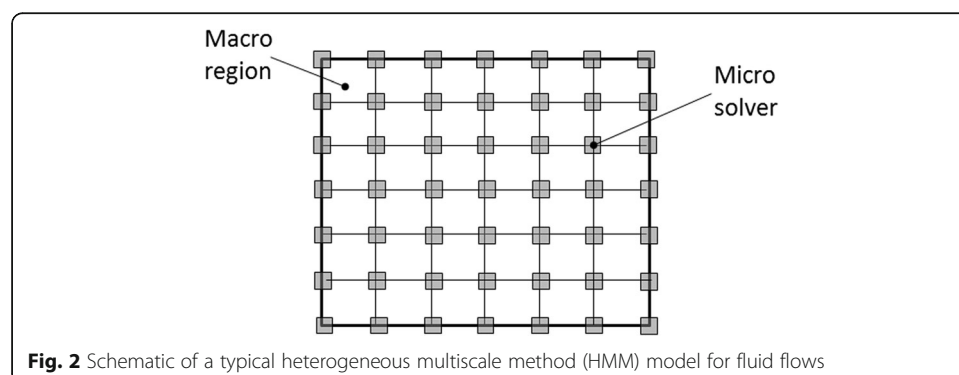

Fig. 2 Schematic of a typical heterogeneous multiscale method (HMM) model for fluid flows

domain with a distribution of sub-domains. However, instead of these sub-domains representing a point in the macro domain, as in the point-wise coupling, the sub-domains in IMM represent a two-dimensional cross-sectional 'slice' of the internal flow as shown in Fig. 3.

The IMM domain representation allows scale separation between molecular and hydrodynamic scales to be exploited in the flow direction. The full physics is considered in the transverse direction through molecular (DSMC) simulations of the sub-domains. The DSMC simulations are periodic and are carried out with the aid of a body force, which represents the effective pressure gradient at the sub-domain locations. The micro sub-domains then interact indirectly with each other through the constraints applied by the macroscopic conservation laws [36]. For gradually varying micro-channels with very low Reynolds number, the mass flow rate can be characterized in terms of the sum of its fundamental flow components like Couette and Poiseuille flow rates. The continuum formulation is thus based on mass conservation, and there is no additional requirement of any constitutive models, since this is indirectly provided by the micro sub-domains. Thus in an IMM simulation, typically, the micro simulations supplies information on the mass flow rate for the macro domain and the macro solver supplies the density and effective pressure gradient at the subdomain locations for the micro-solver. Information exchange between the macro domain and micro sub-domains occurs in an iterative manner, until convergence is obtained and mass conservation is collectively satisfied in all subdomains.

For transient flows which involve temporal scale separation, the unsteady IMM has been developed by Lockerby et al. [37] in which information exchange occurs at well-defined time intervals. IMM has been applied successfully to several multi-scale gaseous flow problems like Knudsen compressors [38] and slider bearing [10].

4 Multiscale particle methods

Another strategy for the multiscale simulation of nonequilibrium gas flows is to develop a unified particle method. Some researchers have made efforts in this direction to develop a particle-particle hybrid method, such as BGK-DSMC [13] and Fokker-Planck-DSMC [14] methods, where the particle simulation methods based on BGK or Fokker-Planck model are employed for the continuum regime, while DSMC method is used for the rarefied regime.
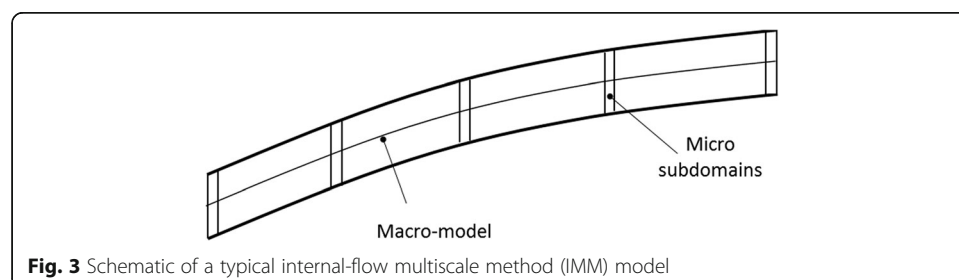

Fig. 3 Schematic of a typical internal-flow multiscale method (IMM) model

4.1 Particle Fokker-Planck method

The particle-based Fokker-Planck (FP) method was originally proposed by Jenny et al. [39]. The fundamental idea of this method is the approximation of the collision integral in the Boltzmann equation with the Fokker Planck model, i.e.,

$$\frac{\partial f}{\partial t} + v_i \frac{\partial f}{\partial x_i} = -\frac{\partial (A_i f)}{\partial v_i} + \frac{1}{2}\frac{\partial^2 (D_{ij} f)}{\partial v_i \partial v_j},$$ (3)

where $f(\mathbf{x}, \mathbf{v}, t)$ is the particle distribution function with the velocity $\mathbf{v}$, the position $\mathbf{x}$, and the time t, $\mathbf{A}$ is the drift coefficient, and $\mathbf{D}$ is the diffusion coefficient. Therefore, the binary collision process described by the Boltzmann collision integral, which is resolved within DSMC, is replaced by a continuous Markov process. Theoretically, the particle FP method is more efficient than DSMC at low Knudsen numbers, as the time steps can be larger than the mean collision time if the boundary conditions are implemented appropriately [40].

Note that in the particle FP method, FP equation is not solved directly but converted to an equivalent Langevin-type stochastic differential equation, i.e.,

$$dv_i = A_i dt + \sqrt{D_{ij}} dW_j,$$ (4)

where dW_j represents a standard Wiener process. An efficient time integration scheme for the Langevin equation as Eq. (4) was proposed by Jenny et al. [39]. In their original work, a linear drift coefficient $A_i = -c_i/\tau_{FP}$ and the diffusion coefficient $D_{ij} = 2k_B T\delta_{ij}/\tau_{FP}m$ were used, and they ensure the relaxation toward a Maxwellian distribution with the temperature T (k_B is the Boltzmann constant and m is the molecule mass). The drift coefficient is constructed from the thermal velocity $c_i = v_i - u_i$ with the macroscopic velocity u_i and the relaxation time $\tau_{FP} = 2\mu/p$ with the viscosity μ and the pressure p. The particle Fokker-Planck model has been applied to study a variety of gas flows, including Rayleigh-Bénard instability [41]. The drawback of the simple linear FP model is that the predicted Prandtl number has an incorrect value of 3/2 instead of the correct value of 2/3 for monoatomic gases [42].

A model to overcome this Prandtl number problem was the cubic-FP model proposed by Gorji et al. [43]. In this model, the linear drift term was replaced with the cubic term, i.e.,

$$A_i = -\frac{c_i}{\tau_{FP}} + \chi_{ij}c_j + \gamma_i\left(c_j c_j - \frac{3k_B T}{m}\right) + \Lambda\left(c_i c_j c_j - \frac{2q_i}{\rho}\right),$$ (5)

to achieve an exact relaxation of the stresses and heat fluxes concerning the Boltzmann collision operator under the assumption of Maxwell type molecules. The coefficients χ_{ij}, γ_i, and Λ form a 9×9 linear system [43, 44]. The cubic FP model has been successfully applied for a variety of gas flows [44–46]. Unfortunately, it was shown that the cubic-FP model has difficulties to reproduce the shock structure correctly compared with DSMC calculations [47]. To overcome this problem, a hybrid cubic-FP-DSMC method [14, 47] and an entropic FP model derived based on the Boltzmann collision integrals and maximum entropy distribution [48] has been proposed.

Another method to overcome the Prandtl problem with the FP model is described by Mathiaud and Mieussens [49]. In this model, the drift coefficient of the linear FP model is used, while the diffusion coefficient is replaced by a non-diagonal temperature tensor resulting from the ellipsoidal statistical model (ES-FP). An important advantage of the

ES-FP model is that it satisfies the H-theorem. Jiang et al. [50] demonstrated that the ES-FP model could predict good shock structure of gas flow over a flat plate with $Ma = 6$. Jun et al. [51] performed a comparative study between cubic-FP and ES-FP model for the simulation of gas flows. They found that the ES-FP model performed better than cubic-FP model in the shock region of the supersonic flow, while the cubic-FP model showed a better accuracy in the near continuum regime.

Recently, a multiscale temporal discretization FP (MTD-FP) method was proposed by Fei et al. [52]. The advantage of this method is that it can be used with significant larger time steps than that required in the cubic-FP method, and the predicted results in the continuum limit are consistent with the solutions predicted by the NSF equations. The previous studies have demonstrated its efficiency and accuracy in the simulation of simple flows, such as Couette flows, Poiseuille flows, and the Sod tube test case. It is desired to extend the application of this method to more complex 3D flows in the future.

4.2 Particle BGK method

The BGK approach simplifies the Boltzmann collision integral with a linear relaxation form,

$$\frac{\partial f}{\partial t} + v_i \frac{\partial f}{\partial x_i} = \frac{1}{\tau_{BGK}} (f^e - f), \tag{6}$$

here the particle distribution function f relaxes to the target equilibrium distribution function f^e with the relaxation time τ_{BGK}. The commonly used equilibrium distribution function is the Maxwell distribution, i.e.,

$$f^e = f^M = n \left(\frac{m}{2\pi k_B T} \right)^{\frac{3}{2}} \exp\left(\frac{-m\mathbf{c}^2}{2 k_B T} \right). \tag{7}$$

In one calculating time step Δt, the solution of Eq. (7) can be written as,

$$f(\Delta t) = (1 - \exp(-\Delta t / \tau_{BGK})) f^e + \exp(-\Delta t / \tau_{BGK}) f(0). \tag{8}$$

It means that the new distribution of particle velocities is a mixture of the initial distribution in the cell and a statistical approximation of the target equilibrium distribution. Using the BGK collision approximation, the corresponding particle method was proposed by Macrossan [53] and Gallis and Torczynski [54] independently. Specifically, the process of molecular movements in particle-BGK method are the same as that in the DSMC method, while the process of inter-molecular collisions in DSMC is replaced by a "redistribution phase" in which only a subset of the molecules in the cell, a fraction $1 - \exp(-\Delta t / \tau_{BGK})$ of the total, are assigned new velocities according to the local Maxwell distribution. The velocities of the remaining fraction of particles are unchanged.

A major drawback of this simple target Maxwell distribution is a wrong resulting Prandtl number of $Pr = 1$. To overcome this problem, different models have been proposed, such as the ellipsoidal statistical BGK (ESBGK) model [55] and Shakhov BGK (SBGK) model [56]. The ESBGK model replaces the Maxwell distribution with an anisotropic Gaussian distribution, which is the same as that used in the ES-FP method. It was validated in detail for a variety of gas flows and showed a very good agreement with DSMC results [57]. Recently, the particle ESBGK method has been extended to diatomic molecules [58] and included quantized vibrational energies [59].

Another target distribution function that is frequently used to correct the Prandtl number is the Shakhov distribution function [56]. The particle SBGK method was described in the reference [57]. It was shown that the SBGK method performs slightly better in the reproduction of correct shock structures in hypersonic flows compared with the ESBGK method. Nevertheless, it was also shown that the energy conservation scheme for the particle method becomes more complicated for the SBGK model and that the computational time increases compared to the ESBGK model. An overview and comparison of different sampling methods for each of the mentioned target distribution functions as well as the discussion of different energy and momentum conserving schemes are given in the reference [57].

As an example, the simulation results of the 70° blunted cone obtained using the ESBGK and DSMC methods with the particle code PICLas [60] are presented in Fig. 4. The inflow conditions are $v_\infty = 1502.6$ *m/s*, $T_\infty = 13.32$ *K* and $n_\infty = 3.715 \times 10^{20}$ *m*$^{-3}$, resulting in a Knudsen number of $Kn = 0.033$ and a Mach number of $Ma = 20$. As depicted in Fig. 5, the ESBGK results of the translational, rotational and vibrational temperatures agree well with the corresponding DSMC results.

4.3 Unified Stochastic Particle BGK method

In the particle-ESBGK and particle-SBGK method, the simulated particles move freely first, and then their velocity relaxes to the target distribution with a probability. Similar to DSMC method, the decoupling of the molecular movements and collisions inevitably introduces large numerical errors of transport coefficients if $\Delta t / \tau_{BGK} > 1$. This limits the application of the particle BGK method to continuum flows. To reduce the numerical dissipation of the ESBGK and SBGK method, and obtain an asymptotic-preserving property in the NS limit, a Unified Stochastic Particle-BGK (USP-BGK) method [15] has been developed recently. In the USP-BGK method, the BGK equation is rewritten as
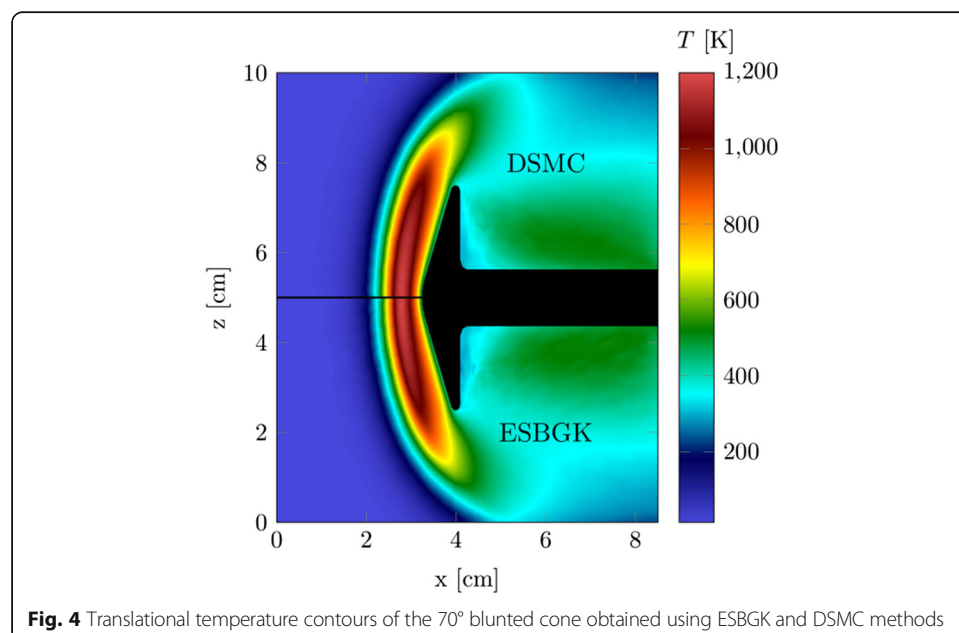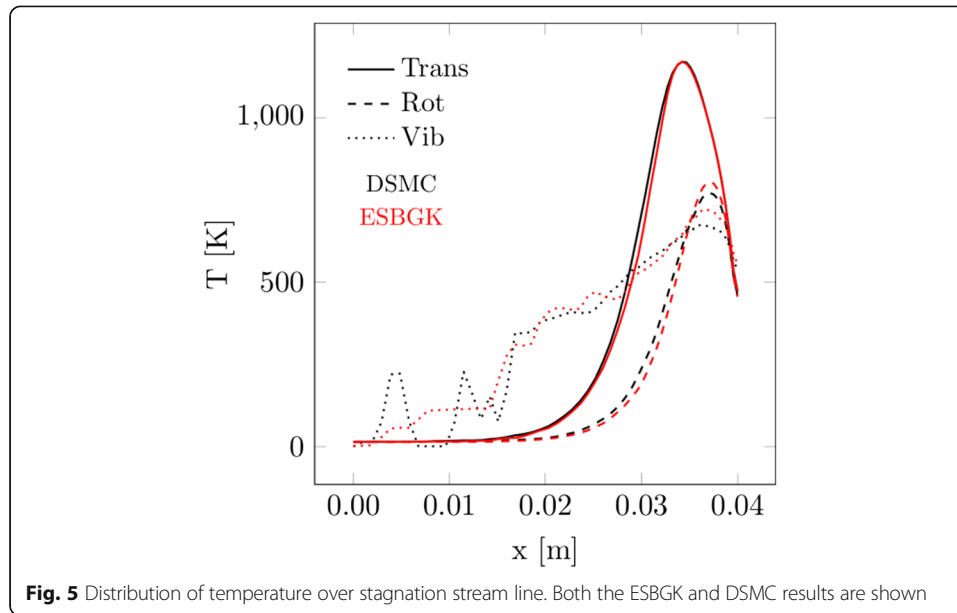

Fig. 4 Translational temperature contours of the 70° blunted cone obtained using ESBGK and DSMC methods
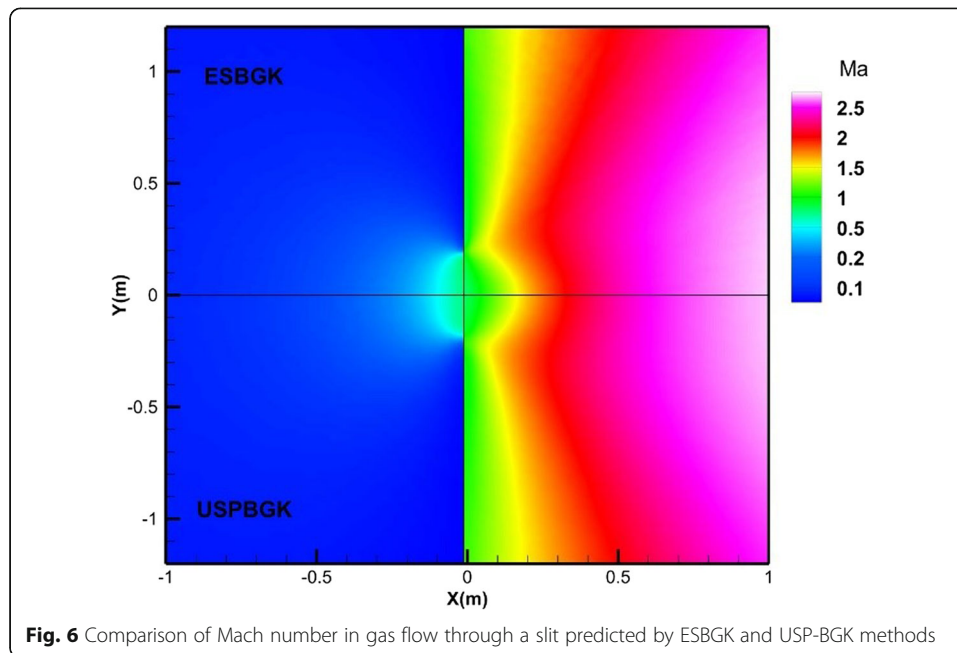
Fig. 5 Distribution of temperature over stagnation stream line. Both the ESBGK and DSMC results are shown

$$\underbrace{\frac{\partial f}{\partial t} + v_i \frac{\partial f}{\partial x_i} - \frac{P_E}{\tau_{BGK}} \left(f^e - f^{Grad} \right)}_{convection} = \underbrace{\frac{1}{\tau_{BGK}} (f^e - f) - \frac{P_E}{\tau_{BGK}} \left(f^e - f^{Grad} \right)}_{relaxation} \qquad (9)$$

Compared to the original BGK equation, a NS collision term with 13 moments Grad's distribution function f^{Grad} is added on both sides. The parameter P_E is in the range $0 < P_E < 1$, and has a value of 0 or 1 when the gas flow is in the limit of the free molecular regime or the continuum regime, respectively. Similar to the ESBGK and SBGK method, the left-hand side of Eq. (9) is implemented with particle convection, and the right-hand side of Eq. (9) is implemented with the relaxation of particle velocities. It should be noted that the particle convection in the USP-BGK method is coupled with a NS collision term. For rarefied regime ($P_E \rightarrow 0$), the modified BGK equation (Eq. (9)) is identical to those computed in the ESBGK and SBGK method. In contrast, for continuum regime ($P_E \rightarrow 1$), the right-hand side of Eq. (9) tends to zero as the first order expansions of f and f^{Grad} are the same, so the velocity distribution is determined solely by the left-hand side of Eq. (9). Therefore, the accurate momentum and energy transport in the continuum limit can be obtained in the USP-BGK method.

In the USP-BGK method, the treatments of the convection and the relaxation processes are similar to those in other Particle-BGK methods. The only difference is that in the convection process, the distribution function should be reconstructed before particle movements due to the additional NS collision term. The numerical details can be found in the reference [15].

The multiscale gas flow through a slit, as shown in Fig. 6, is simulated using USP-BGK and ESBGK methods. The computational domain is consisted of two chambers, and the gas flow is generated by the pressure gradient across the slit. The left and right chambers contact with two reservoirs. The temperature and number density of the left reservoir are 273 K and $10^{20} m^{-3}$, respectively, and vacuum is assumed for the right reservoir. The height of the computational domain is 20 times of the slit width. It can be seen from Fig. 6 that the

Fig. 6 Comparison of Mach number in gas flow through a slit predicted by ESBGK and USP-BGK methods

Mach number predicted by both USP-BGK and ESBGK methods agree well with each other. Note that in this case the USP-BGK method is performed with a much larger time step ($3.0\tau_{BGK}$) than that in ESBGK method ($0.2\tau_{BGK}$), so the computation using USP-BGK method is more efficient than ESBGK method.

5 Summary

In order to solve multiscale gas flows encountered in engineering problems, it is desirable to develop multiscale modelling methods instead of using the traditional computational methods designed for one scale. In this paper, we review two typical hybrid and multiscale methods, i.e. DSMC-CFD hybrid methods and multiscale particle methods. The principles and recent progresses of these methods have been described in detail.

DSMC-CFD hybrid method combines the best of worlds of both the CFD solvers (for the continuum regime) and the DSMC methods (for the rarefied regime). Depending on the characteristic of the multiscale gas flows under consideration, different hybrid frameworks could be applied like DDM, HMM, and IMM. The advantages and disadvantages of these three coupling methods are discussed. DDM is general in nature and valid for a large class of problems. It is important to carefully consider an appropriate continuum breakdown parameter and choose a coupling strategy based on the physics of the flow, in order to get maximum computational efficiency and accuracy. On the other hand, HMM and IMM types of coupling work best specifically for gas flows where there is a high degree of length scale separation.

Alternatively, some researchers have made efforts to develop a particle-particle hybrid method. Earlier researches on the hybrid BGK-DSMC and Fokker-Planck-DSMC methods have demonstrated their potential for the simulation of multiscale gas flows. Recently, the unified statistical particle BGK (USP-BGK) method has been proposed by combining the molecular movement and collision effects, which enables this method to simulate the small-scale nonequilibrium and large-scale continuum effects in a unified computational framework. Further research on this is expected in the future.

Availability of data and materials
All data generated or analyzed during this study are included in this published article.

Authors' contributions
The contribution of the authors to the work is equivalent. All authors read and approved the final manuscript.

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

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]School of Aeronautic Science and Engineering, Beihang University, Beijing 100191, China. [2]Scientific Computing Department, STFC Daresbury Laboratory, Warrington, Cheshire WA4 4AD, UK. [3]Institute of Space Systems, University of Stuttgart, Pfaffenwaldring 29, D-70569 Stuttgart, Germany. [4]School of Aerospace Engineering, Huazhong University of Science and Technology, Wuhan 430074, China. [5]School of Chemical and Process Engineering, University of Leeds, Leeds LS2 9JT, UK.

References
1. Bird GA (1994) Molecular gas dynamics and direct simulation of gas flows. Clarendon, Oxford
2. Shen C (2005) Rarefied gas dynamics: fundamentals, simulations and micro flows. Springer, Berlin
3. Ivanov MS, Gimelshein SF (1998) Computational hypersonic rarefied flows. Annu Rev Fluid Mech 30:469–505
4. Titov EV, Zeifman MI, Levin DA (2005) Application of the kinetic and continuum techniques to the multi-scale flows in MEMS devices. AIAA paper 2005–1399
5. Hash DB, Hassan HA (1995) A hybrid DSMC/Navier-Stokes solver. AIAA paper 1995–0410
6. Sun QH, Boyd ID, Candler GV (2004) A hybrid continuum/particle approach for modeling subsonic, rarefield gas flow. J Comput Phys 194:256–277
7. Schwartzentruber TE, Scalabrin LC, Boyd ID (2007) A modular particle-continuum numerical method for hypersonic non-equilibrium gas flows. J Comput Phys 225:1159–1174
8. Xu X, Wang X, Zhang M, Zhang J, Tan J (2018) A parallelized hybrid N–S/DSMC-IP approach based on adaptive structured/unstructured overlapping grids for hypersonic transitional flows. J Comput Phys:409–433
9. Docherty SY, Borg MK, Lockerby DA, Reese JM (2016) Coupling heterogeneous continuum-particle fields to simulate non-isothermal microscale gas flows. Int J Heat Mass Transf 98:712–727
10. John B, Lockerby DA, Patronis A, Emerson DR (2018) Simulation of the head-disk interface gap using a hybrid multi-scale method. Microfluid Nanofluid 22:106
11. Xu K, Huang J-C (2010) A unified gas-kinetic scheme for continuum and rarefied flows. J Comput Phys 229:7747–7764
12. Guo Z, Xu K, Wang R (2013) Discrete unified gas kinetic scheme for all Knudsen number flows: low-speed isothermal case. Phys Rev E 88:033305
13. Kumar R, Titov EV, Levin DA (2013) Development of a particle-particle hybrid scheme to simulate multiscale transitional flows. AIAA J 51:200–217
14. Gorji MH, Jenny P (2015) Fokker-Planck-DSMC algorithm for simulations of rarefied gas flows. J Comput Phys 287:110–129
15. Fei F, Zhang J, Li J, Liu Z (2018) A Unified Stochastic Particle Bhatnagar-Gross-Krook Method for Multiscale Gas Flows, arXiv preprint arXiv:1808.03801, 2018
16. Bird GA (1970) Breakdown of translational and rotational equilibrium in gaseous expansions. AIAA J 8:1998–2003
17. Boyd ID, Chen G, Candler GV (1995) Predicting failure of the continuum fluid equations in transitional hypersonic flows. Phys Fluids 7:210–219
18. Wang WL, Boyd ID (2003) Predicting continuum breakdown in hypersonic viscous flows. Phys Fluids 15:91–100
19. Lockerby DA, Reese JM, Struchtrup H (2009) Switching criteria for hybrid rarefied gas flow solvers. Proc R Soc A-Math Phys Eng Sci 465:1581–1598
20. Meng J, Dongari N, Reese JM, Zhang Y (2014) Breakdown parameter for kinetic modeling of multiscale gas flows. Phys Rev E 89:063305
21. Chen P-H, Boyd I, Camberos J (2003) Assessment of entropy generation rate as a predictor of continuum breakdown, AIAA paper 2003–3783
22. Bird G (1990) Application of the direct simulation Monte Carlo method to the full shuttle geometry, AIAA paper 1990–1692

23. Pham-Van-Diep G, Erwin D, Muntz EP (1989) Nonequilibrium molecular-motion in a hypersonic shock-wave. Science 245:624–626
24. Sun QH, Boyd ID (2002) A direct simulation method for subsonic, microscale gas flows. J Comput Phys 179:400–425
25. Stefanov S, Roussinov V, Cercignani C (2002) Rayleigh-Benard flow of a rarefied gas and its attractors. I. Convection regime. Phys Fluids 14:2255–2269
26. Zhang J, Fan J (2009) Monte Carlo simulation of thermal fluctuations below the onset of Rayleigh-Bénard convection. Phys Rev E 79:056302
27. Zhang J, Fan J, Fei F (2010) Effects of convection and solid wall on the diffusion in microscale convection flows. Physics of Fluids 22:122005
28. Manela A, Zhang J (2012) The effect of compressibility on the stability of wall-bounded Kolmogorov flow. J Fluid Mech 694:29–49
29. Gallis MA, Bitter NP, Koehler TP, Torczynski JR, Plimpton SJ, Papadakis G (2017) Molecular-level simulations of turbulence and its decay. Phys Rev Lett 118:064501
30. Alexander FJ, Garcia AL, Alder BJ (1998) Cell size dependence of transport coefficients in stochastic particle algorithms. Phys Fluids 10:1540–1542
31. Garcia AL, Wagner W (2000) Time step truncation error in direct simulation Monte Carlo. Phys Fluids 12:2621–2633
32. Wijesinghe HS, Hadjiconstantinou NG (2004) Discussion of hybrid atomistic-continuum methods for multiscale hydrodynamics. Int. J. Multiscale Comput. Eng. 2:189–202
33. John B, Damodaran M (2009) Computation of head-disk interface gap micro flowfields using DSMC and continuum-atomistic hybrid methods. Int J Numer Methods Fluids 61:1273–1298
34. Weinan E, Engquist B, Huang Z (2003) Heterogeneous multiscale method: a general methodology for multiscale modeling. Phys Rev B 67:092101
35. Ren WQ, Weinan E (2005) Heterogeneous multiscale method for the modeling of complex fluids and micro-fluidics. J Comput Phys 204:1–26
36. Borg MK, Lockerby DA, Reese JM (2015) A hybrid molecular–continuum method for unsteady compressible multiscale flows. J Fluid Mech 768:388–414
37. Lockerby DA, Duque-Daza CA, Borg MK, Reese JM (2013) Time-step coupling for hybrid simulations of multiscale flows. J Comput Phys 237:344–365
38. Patronis A, Lockerby DA (2014) Multiscale simulation of non-isothermal microchannel gas flows. J Comput Phys 270:532–543
39. Jenny P, Torrilhon M, Heinz S (2010) A solution algorithm for the fluid dynamic equations based on a stochastic model for molecular motion. J Comput Phys 229:1077–1098
40. Önskog T, Zhang J (2015) An accurate treatment of diffuse reflection boundary conditions for a stochastic particle Fokker–Planck algorithm with large time steps. Physica A: Statistical Mechanics and its Applications 440:139–146
41. Zhang J, Onskog T (2017) Langevin equation elucidates the mechanism of the Rayleigh-Benard instability by coupling molecular motions and macroscopic fluctuations. Phys Rev E 96:043104
42. Zhang J, Zeng DD, Fan J (2014) Analysis of transport properties determined by Langevin dynamics using green-Kubo formulae. Physica A: Statistical Mechanics and its Applications 411:104–112
43. Gorji MH, Torrilhon M, Jenny P (2011) Fokker-Planck model for computational studies of monatomic rarefied gas flows. J Fluid Mech 680:574–601
44. Gorji MH, Jenny P (2014) An efficient particle Fokker-Planck algorithm for rarefied gas flows. J Comput Phys 262:325–343
45. Pfeiffer M, Gorji MH (2017) Adaptive particle–cell algorithm for Fokker–Planck based rarefied gas flow simulations. Comput Phys Commun 213:1–8
46. Gorji MH, Jenny P (2013) A Fokker–Planck based kinetic model for diatomic rarefied gas flows. Phys Fluids 25:062002
47. Jun E, Gorji MH, Grabe M, Hannemann K (2018) Assessment of the cubic Fokker–Planck–DSMC hybrid method for hypersonic rarefied flows past a cylinder. Comput Fluids 168:1–13
48. Gorji MH, Torillhon M (2016) A Fokker-Planck model of hard sphere gases based on H-theorem. AIP Conference Proceedings 1786(1):090001
49. Mathiaud J, Mieussens L (2016) A Fokker-Planck model of the Boltzmann equation with correct Prandtl number. J Stat Phys 162:397–414
50. Jiang Y, Gao Z, Lee C-H (2018) Particle simulation of nonequilibrium gas flows based on ellipsoidal statistical Fokker–Planck model. Comput Fluids 170:106–120
51. Jun E, Pfeiffer M, Mieussens L, Gorji MH (2019) Comparative study between cubic and ellipsoidal Fokker–Planck kinetic models. AIAA J:1–10
52. Fei F, Liu ZH, Zhang J, Zheng CG (2017) A particle Fokker-Planck algorithm with multiscale temporal discretization for rarefied and continuum gas flows. Commun Comput Phys 22:338–374
53. Macrossan MN (2001) Nu-DSMC: a fast simulation method for rarefied flow. J Comput Phys 173:600–619
54. Gallis MA, Torczynski JR (2011) Investigation of the ellipsoidal-statistical Bhatnagar-gross-Krook kinetic model applied to gas-phase transport of heat and tangential momentum between parallel walls. Phys Fluids 23:030601
55. Holway LH Jr (1966) New statistical models for kinetic theory: methods of construction. The physics of fluids 9:1658–1673
56. Shakhov E (1968) Generalization of the Krook kinetic relaxation equation. Fluid Dynamics 3:95–96
57. Pfeiffer M (2018) Particle-based fluid dynamics: comparison of different Bhatnagar-gross-Krook models and the direct simulation Monte Carlo method for hypersonic flows. Phys Fluids 30:106106
58. Tumuklu O, Li Z, Levin DA (2016) Particle ellipsoidal statistical Bhatnagar-gross-Krook approach for simulation of hypersonic shocks. AIAA J 54:3701–3716
59. Pfeiffer M (2018) Extending the particle ellipsoidal statistical Bhatnagar-gross-Krook method to Di-atomic molecules including quantized vibrational energies. Phys Fluids 30:116103
60. Munz C-D, Auweter-Kurtz M, Fasoulas S, Mirza A, Ortwein P, Pfeiffer M, Stindl T (2014) Coupled particle-in-cell and direct simulation Monte Carlo method for simulating reactive plasma flows. Comptes Rendus Mécanique 342:662–670