RESEARCH

A hybrid Gaussian mixture/DSMC approach to study the Fourier thermal problem

Shahin Mohammad Nejad[1] · Frank A. Peters[1] · Silvia V. Nedea[1] · Arjan J. H. Frijns[1] · David M. J. Smeulders[1]

Abstract

In rarefied gas dynamics scattering kernels deserve special attention since they contain all the essential information about the effects of physical and chemical properties of the gas–solid surface interface on the gas scattering process. However, to study the impact of the gas–surface interactions on the large-scale behavior of fluid flows, these scattering kernels need to be integrated in larger-scale models like Direct Simulation Monte Carlo (DSMC). In this work, the Gaussian mixture (GM) model, an unsupervised machine learning approach, is utilized to establish a scattering kernel for monoatomic (Ar) and diatomic (H_2) gases directly from Molecular Dynamics (MD) simulations data. The GM scattering kernel is coupled to a pure DSMC solver to study isothermal and non-isothermal rarefied gas flows in a system with two parallel walls. To fully examine the coupling mechanism between the GM scattering kernel and the DSMC approach, a one-to-one correspondence between MD and DSMC particles is considered here. Benchmarked by MD results, the performance of the GM-DSMC is assessed against the Cercignani–Lampis–Lord (CLL) kernel incorporated into DSMC simulation (CLL-DSMC). The comparison of various physical and stochastic parameters shows the better performance of the GM-DSMC approach. Especially for the diatomic system, the GM-DSMC outperforms the CLL-DSMC approach. The fundamental superiority of the GM-DSMC approach confirms its potential as a multi-scale simulation approach for accurately measuring flow field properties in systems with highly nonequilibrium conditions.

Keywords Rarefied gas dynamics · Molecular dynamics · Gas-surface scattering models · Gaussian mixture approach · Direct simulation Monte Carlo

1 Introduction

Rarefied gas dynamics is an active research topic in numerous cutting-edge engineering applications ranging from aerospace to biological applications(Karniadakis et al. 2006). At the design stage of such applications, numerical modeling is required to predict flow field properties and thermal energy transport between components to guarantee their optimal performance and lifetime. Depending upon the degree of rarefaction in the system, determined by the Knudsen number (Kn), different modeling approaches can be used to study the rarefied gas dynamics. In the case of low to moderate degree of rarefaction (Kn < 0.1) the continuum approach,

based on the Navier–Stokes equations, is the most common simulation approach. On the other hand, for a highly rarefied gas (Kn ≥ 0.1), particle-based simulation techniques, such as Molecular Dynamics (MD) (Allen and Tildesley 2017), Direct Simulation Monte Carlo (DSMC) (Bird and Brady 1994), and lattice Boltzmann method (LBM) (Chen and Doolen 1998), are typically employed to determine flow field properties. Among these techniques, DSMC is the most commonly used one that has been successfully applied in a broad range of engineering applications (Karniadakis et al. 2006). Nevertheless, prescribing accurate boundary conditions at the gas–surface interface is a vital parameter to achieving reliable results in DSMC. In fact, going toward a higher degree of rarefaction, the complex physical interactions at the interface become even more dominant than the gas–gas interactions happening at the bulk of the fluid (Shen 2006; Zhang et al. 2012). Therefore, a detailed understanding of such interfacial nanoscale phenomena is of extreme importance in rarefied gas dynamics.

✉ Arjan J. H. Frijns
a.j.h.frijns@tue.nl

1   Department of Mechanical Engineering,
    Eindhoven University of Technology, PO Box 513,
    5600 MB Eindhoven, The Netherlands

The diffuse reflection scattering model was the first boundary model used to describe the scattering process in rarefied gases (Shen 2006). This model assumes that the reflected gas molecules are fully accommodated with the adjacent solid surface, and their outgoing velocity distribution is determined by the Maxwellian distribution based on the wall temperature. However, later experimental studies (Devienne et al. 1965; Gregory and Peters 1986) revealed that the complete diffuse reflection assumption is not always valid. For example, at a very clean or high-temperature surface or under ultra vacuum conditions, gas molecules can experience near specular reflection. Consequently, to establish more comprehensive boundary models, researchers proposed various empirical scattering kernels that could anticipate both specular and diffuse reflections in specific rarefied gas flow applications (Maxwell 1878; Epstein 1967; Cercignani and Lampis 1971; Lord 1989, 1991; Yamanishi et al. 1999; Yamamoto et al. 2006; Yakunchikov et al. 2012; Hossein Gorji and Jenny 2014; Frezzotti and Gibelli 2008). The performance of these scattering kernels normally depends on several parameters, known as accommodation coefficients (ACs). In these scattering kernels, ACs are applied to quantify the accommodating level of different kinetic energy modes of gas molecules on a neighboring solid surface. The Cercignani-Lampis-Lord (CLL) scattering kernel is one of the most employed scattering kernels that can be utilized to describe the scattering process for both monoatomic and diatomic gas molecules (Lord 1991). Despite the acceptable performance of the classical scattering kernels in the systems near equilibrium conditions, it was shown that these models are incapable of fully capturing the complex physical phenomena, such as a considerable temperature jump, happening in systems in a highly non-equilibrium situation (Yamamoto et al. 2007; Liang et al. 2013). This shortcoming, alongside the lack of generality of such scattering kernels, raises the need for more elaborated gas scattering kernels.

As the most accurate particle-based simulation approach, an MD simulation is a valuable tool for studying interfacial nanoscale phenomena. In an MD simulation, the interactions between individual gas molecules with neighboring solid molecules are modeled deterministically. This fact makes MD computationally very expensive. Therefore, its application is usually restricted to the nanoscale level. Nevertheless, an MD simulation is exploited in various approaches for investigating interfacial physics. Most commonly, MD is used to compute different ACs between a gas-solid pair. These ACs are then fed into one of the previously discussed empirical scattering kernels that are served as boundary conditions for high scale simulation approaches such as DSMC (Watvisave et al. 2015).

Other researchers developed hybrid simulation schemes combining the classical MD and DSMC approaches to study rarefied gas flow at mesoscale level (Nedea et al. 2005; Yamamoto et al. 2006; Liang and Ye 2014; Watvisave et al. 2015; Longshaw et al. 2020). In these schemes, the simulation domain is decomposed into smaller regions. To benefit the precision of MD and the speed of DSMC at the same time, the first approach is applied in the vicinity of the solid surfaces, while the latter one is utilized in the bulk of the domain. Although these hybrid schemes are relatively faster than pure MD, since a large number of solid atoms must still be simulated in the MD part, these schemes are computationally expensive. In addition, there are considerable differences in the order of magnitude of time and space between MD and DSMC. Therefore, while applying these hybrid schemes, special attention must be taken into account to ensure particle and energy conservation at the coupling interfaces.

Another category of wall scattering kernels is nonparametric scattering kernels (Liao et al. 2018; Andric et al. 2019; Liu et al. 2021). Unlike the classical empirical scattering kernels, no intermediate calibration based on ACs is needed, and MD data are used directly to construct these scattering kernels. These scattering kernels generally show better performance in the highly non-equilibrium situation compared to the classical scattering kernels (Andric et al. 2019). As a first step to constructing these scattering kernels, single gas molecules with independent thermal velocities are beamed onto a specific solid surface in an MD simulation. Then, the discrete gas molecule trajectories, characterized by the system's initial condition, such as gas molecular velocity or surface conditions, are gathered in a database. Based on the relation between the precollisional and postcollisional molecular velocity vectors in the MD dataset, a conditional probability density function is derived that can be used for sampling the reflected gas molecular velocity depending on its initial state. Liu et al. (2021) developed such a scattering kernel and coupled it to a DSMC code to model hypersonic flow over a rounded wedge. Comparing the simulation results with another case study, in which the CLL scattering kernel was used as the boundary model, showed that the DSMC code based on their proposed model performs better in describing different flow properties. Nevertheless, in all these scattering kernels, the gas–gas interactions that can affect the reflected gas molecules properties in the early transition regime ($0.1 < Kn < 1$) are ignored, and these wall models cannot deal with adsorption-related problems. Besides, the usual nonparametric approximations of high-dimensional multimultivariate data can be a complex task requiring advanced learning methods (Sung 2004). To avoid this problem, before the main fitting step, usually various techniques are employed to reduce the dimensionality of the dataset (Andric et al. 2019).

Machine learning is another promising technique that can be used to establish a gas scattering kernel directly based on the collisional data obtained from MD simulations (Liao et al. 2018; Mohammad Nejad et al. 2021, 2022; Wang et al. 2021; Wu et al. 2022). As an example, in our previous works (Mohammad Nejad et al. 2021, 2022), the Gaussian mixture (GM) approach, an unsupervised machine learning approach, was employed to construct a scattering kernel for monoatomic gases (Ar, He) interacting with the Au surface and diatomic gases (H_2, N_2) interacting with the Ni surface, respectively. In these studies, a two parallel walls MD setup was used as the reference system to include the impact of both gas–wall and gas–gas interactions on the postcollisional behavior of gas molecules. The physical and stochastic properties predicted by the GM scattering kernel were assessed against the CLL model results and the original MD data in the Fourier thermal and the combined Fourier-Couette flow problems. It was illustrated that the GM model outperforms the CLL model in both benchmark systems. The main advantage of the GM approach-driven scattering kernel over the empirical scattering kernels is that its performance does not rely on a finite number of parameters. In fact, we can adjust the number of fitting parameters to get the best performance from the scattering kernel. Wang et al. (2021) used a data-based scattering kernel model for gas-solid interactions. The main advantage of our GM scattering approach over this approach is that in the case of the GM scattering model the MD results have been used as a whole and the velocity components are not separated from each other. Therefore, the interplay between different velocity components or energy modes (in the case of diatomic molecules) will be fully taken into account in the case of our GM approach. In addition, unlike nonparametric approaches whose extension to higher dimensions is limited (Sung 2004), the GM scattering kernel can handle high-dimensional data sets rather straightforwardly. To the best of our knowledge, the scattering kernel based on the GM approach has not been coupled to a DSMC solver before to study rarefied gas flow systems.

Knowing the vital importance of the applied scattering kernels in rarefied gas modeling, the main objective in this work is to develop a robust gas scattering model for higher scale simulation techniques. To fulfill this goal, we investigate the capability of the GM-driven scattering kernels to be used as a boundary model in a DSMC solver. Besides, we develop a dedicated mechanism which can be used to efficiently couple the GM scattering kernel to a DSMC solver. Initially, based on a two parallel walls system, the interactions of Ar gas with an Au surface and H_2 gas with a Ni surface are studied using MD simulations. Two benchmark problems are considered: an equilibrium gas system confined between isothermal walls and a Fourier thermal problem. For the Ar–Au system, the pre- and postcollisional

translational velocities of Ar molecules are used for training the GM model. Whereas, for the H_2–Ni system, both translational and rotational velocities are employed for the training purpose. Such consideration guarantees the model ability to anticipate the possible energy transfer between the translational and rotational modes at non-equilibrium conditions. Implementing the GM model on the MD collisional data in each case study, a conditional multivariate probability distribution is derived that can be used to generate the postcollisional velocities of gas molecules based on their precollisional states. The GM approach-driven boundary models for the Ar–Au and H_2–Ni are incorporated in pure DSMC simulations based on a one-to-one mapping between the corresponding MD and DSMC simulation setups. The DSMC simulations coupled to the GM scattering kernel (GM-DSMC) results are evaluated against the DSMC simulations coupled to the CLL scattering kernel (CLL-DSMC) and the reference MD solutions. The evaluation is performed based on different physical and stochastic criteria.

2 Methods

In this section, the most relevant features of MD, the applied gas-surface interaction models, and DSMC related to the hybrid GM-DSMC approach are addressed. Further details of MD, gas-surface interactions models, and DSMC can be found in Allen and Tildesley (2017), in Liao et al. (2018), Mohammad Nejad et al. (2021), Hossein Gorji and Jenny (2014), Sung (2004), and Mohammad Nejad et al. (2022), and in Bird and Brady (1994), respectively.

2.1 MD simulation

The exact particle trajectories are calculated in an MD simulation based on the interaction potentials and Newton's second law of motion. Thus, MD is considered as the most accurate method for modeling the scattering process and is used to provide benchmark solutions in this work. Our MD setup to study the 1-D Fourier thermal problem in a nanochannel consists of two infinite parallel walls at a distance, L_y, apart from each other and of gas molecules confined between these two walls (see Fig. 1). Each wall is constructed with five layers of FCC planes. In the case of the Ar–Au system, each wall has a cross-sectional area of 10 nm × 10 nm, while for the H_2–Ni system, the cross-sectional area is 10.8 nm × 10.8 nm. In each wall, the outermost layer is constrained to prevent the translational motion of the wall in the normal direction. The distance between the two walls is fixed at L_y = 20 nm for the Ar–Au system, and at L_y = 30 nm for H_2–Ni system.

For the Ar–Au system, the Knudsen number in the bulk and the gas reduced density (η) are Kn = 0.33 and η = 0.008,
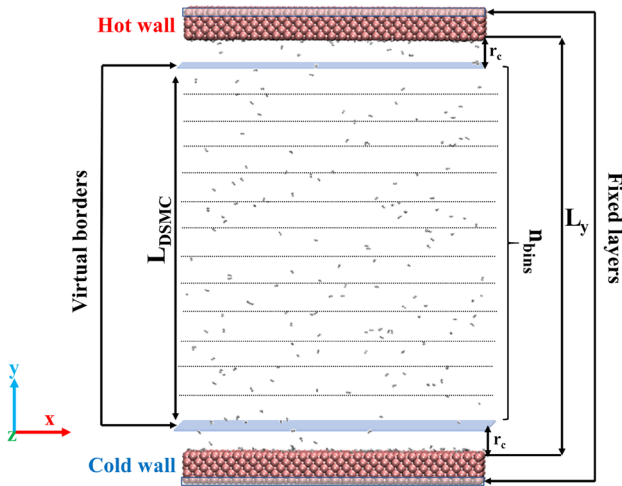
Fig. 1 Schematic representation of the simulation model

Table 1 Lennard–Jones potential parameters

Atom pair	ϵ (eV)	σ (Å)
Ar–Au	$1.14e^{-2}$	3.819
Ar–Ar	$1.22e^{-2}$	3.35
H–H	$9.29e^{-4}$	1.421
H–Ni	$1.19e^{-2}$	2.016

respectively. The reduced density is defined as $\eta = \pi n a^3 / 6$, where n is the number density and a is the particle diameter (Frezzotti 1999). On the other hand, for the H_2–Ni system $Kn = 0.37$ and $\eta = 0.003$. In addition, H_2 molecules are considered as rigid rotors with a fixed bond length of 0.7414 Å(Atkins and De Paula 2011).

In both studied gas-solid pairs, periodic boundary conditions are considered along x and z directions. The interactions between the solid Au and Ni atoms located in the walls are models using the corresponding embedded atom model (EAM) potentials developed by (Sheng et al. 2011) and Foiles et al. (1986), respectively. For the Ar–Au system, the non-bonded gas–gas and gas-wall interactions are modeled by the Lennard–Jones (LJ) 12-6 potential. However, in the case of the H_2–Ni system, the non-bonded interactions are modeled using COMPASS force field (Sun 1998), in which an LJ 9-6 function is applied to describe the interactions. All the gas–gas and gas-wall interatomic potential parameters utilized in this work are presented in Table 1.

The cutoff distances for gas–gas interactions are set at 2.5σ. Considering the gas-wall interactions, the cutoff distances (r_c in Fig. 1) is set at 12 Å and 10 Å for the Ar–Au and H_2–Ni systems, respectively. In each MD simulation, after deposition of the target number of gas molecules between the solid walls, energy minimization is carried out

by iteratively rearranging atom positions to eliminate the possible overlapping of the neighboring atoms. Afterward, each plate is connected to a Nose-Hoover (NVT) thermostat to maintain its temperature at the desired level. On the other hand, gas molecules are modeled in the microcanonical ensemble (NVE), and their temperature can change via the collision with other atoms in the box. In order to speed up the equilibration process, the velocity components of gas molecules are initially sampled from a Gaussian distribution with a mean value of 0.0 and a standard deviation of $\sqrt{\frac{k_B T_a}{2m_g}}$, where T_a is the average value of the bottom and top plates temperatures, and m_g is the mass of the gas molecule. Each MD setup is equilibrated for 3 ns with a time step of 1 and 0.5 fs for the Ar–Au and H_2–Ni systems, respectively. After the complete thermalization of the system, the production run is started, which is proceeded for 25 ns for each MD system. All MD simulations are carried out using LAMMPS (Plimpton 1995) package. The overall computational time for extracting the required MD data is around 13 h running on a computer with 16 cores.

2.2 Gas-surface interaction models

In rarefied gas flow simulations, gas-surface interaction models are used as boundary conditions. Such interactions usually are described in terms of scattering kernel, $R(v|v')$, representing the probability density that an impinging gas molecule with velocity v' is reflected with velocity v. In a general form, the boundary condition for the impinging gas molecular velocity distribution, $f(v')$, can be expressed as Shen (2006),

$$v_n f(v) = \int_{v'_n < 0} |v'_n| R(v|v') f(v') dv', \quad v_n > 0 \qquad (1)$$

where v'_n and v_n are the normal components of the impinging and reflected gas molecular velocity, respectively. In the case of a diatomic gas molecule, in addition to the translational velocity components of the center of mass (v', v), the rotational velocity vectors (ω', ω) need to be taken into account, as well. Therefore, for a diatomic gas molecules the scattering kernel is presented as $R(v, \omega|v', \omega')$, and the probability density is denoted by $f(v, \omega)$.

2.2.1 CLL scattering model

Among all the proposed empirical scattering kernels, the CLL model is used in this work. We choose this particular model since it is one of the most reliable scattering models for monoatomic and diatomic gases and implementing it in DSMC is rather straightforward (Liang et al. 2013;

Hossein Gorji and Jenny 2014; Watvisave et al. 2015). For diatomic gas molecules, the CLL scattering kernel is given as (Lord 1991):

$$R^{CLL}(\boldsymbol{v}, \boldsymbol{\omega}|\boldsymbol{v}', \boldsymbol{\omega}') = \frac{2v_n}{\pi^2 \alpha_t (2-\alpha_t) \alpha_n \alpha_{rot}} \exp\left[-\frac{[\boldsymbol{v_t} - \sqrt{1-\alpha_t}\boldsymbol{v_t'}]^2}{\alpha_t(2-\alpha_t)}\right]$$

$$\times \exp\left[-\frac{v_n^2 + (1-\alpha_n)v_n'^2}{\alpha_n}\right] I_0 \left[\frac{2(\sqrt{1-\alpha_n})v_n v_n'}{\alpha_n}\right]$$

$$\times \exp\left[-\frac{(\boldsymbol{\omega} - \sqrt{1-\alpha_{rot}}\boldsymbol{\omega}')^2}{\alpha_{rot}}\right]$$

(2)

where α_t, α_n, and α_{rot} are the accommodation coefficients corresponding to the tangential momentum, normal translational kinetic energy, and rotational energy, respectively. $\boldsymbol{v_t}$ represents the tangential velocity vector and I_0 is the modified Bessel function of the first kind and zeroth order. The translational ($\boldsymbol{v}'$, $\boldsymbol{v}$) and rotational ($\boldsymbol{\omega}'$, $\boldsymbol{\omega}$) velocities are normalized by $\sqrt{\frac{2k_B T_w}{m_g}}$ and $\sqrt{\frac{2k_B T_w}{I}}$, respectively. Here, T_w describes the wall temperature and I is the mass moment of inertia of the diatomic gas molecule. It is noteworthy to mention that monoatomic gases do not possess rotational energy. Therefore, the last part in Eq. (2), referring to the rotational velocities ($\boldsymbol{\omega}'$, $\boldsymbol{\omega}$), needs to be eliminated in the case of the Ar–Au system. As a result, only α_t, α_n are needed in the CLL model for the Ar–Au system. The accommodation coefficients needed in Eq. (2) are computed using the approach proposed by Spijker et al. (2010), in which the correlation between the corresponding impinging and outgoing kinetic properties (e.g., the normal translational energy or rotational energy) is applied to derive the relevant accommodation coefficients (e.g., α_n or α_{rot}).

The impinging and outgoing kinetic properties, known as the collisional data, that are required to compute the accommodation coefficients and training of the GM scattering model (discussed in the following part), are recorded at the virtual borders located at the distances r_c away from the walls (see Fig. 1). The number of data points in different gathered collisional datasets varies between 100,000 and 150,000.

2.2.2 GM scattering model

The Gaussian mixture model, a well-known unsupervised machine learning approach, can be exploited to derive a formalism describing gas-surface interactions directly based on MD simulation results. As it has also been addressed in the previous section, the main input required to train the GM model is the MD collisional data. From each MD collisional dataset, 75% is used for the training and the remaining 25% is considered for the verification purpose. The training of the

GM scattering kernel for monoatomic and diatomic gases have been extensively discussed in our previous works (Mohammad Nejad et al. 2021, 2022). However, for the sake of clarity, we briefly revisit the main findings in this section.

In the case of the Ar–Au system, having only the translational degrees of freedom, the collisional data is a 6-dimensional dataset including the impinging and outgoing translational velocities of the center of mass (COM) of Ar atoms (v_x', v_y', v_z', v_x, v_y, v_z). On the other hand, for the H_2–Ni system, we must also account for the rotational degrees of freedom. Therefore, the impinging and outgoing rotational velocities are added to the training data. Here, the final training dataset is a 10-dimensional matrix (v_x', v_y', v_z', ω_1', ω_2', v_x, v_y, v_z, ω_1, ω_2). Another parameter that directly influences the performance of the GM model and the computational cost of the training process is the number of applied Gaussian functions (K) in the model. This parameter must be specified adequately by the user to avoid the probable overfitting or underfitting. Here, a sensitivity analysis is performed to ascertain the optimal K. The details of it can be found in "Appendix". From this analysis, $K = 100$ and $K = 500$ are assigned as the number of Gaussian for the Ar–Au and H_2–Ni systems, respectively. Using these numbers of the Gaussian functions, the training of the GM model on a regular laptop computer takes around 3 and 40 min for the Ar–Au and H_2–Ni systems. The GM model manifests its best performance when all the components of the training data are normally distributed (Mohammad Nejad et al. 2021). In the case of both gas-solid pairs considered in this work, except for the normal velocities (v_y', v_y) following the Rayleigh distributions, the other components in the training data follow a Gaussian distribution. To obtain the same kind of distribution for the normal velocity components, initially, for each normal velocity pair (v_y', v_y), an implicit pair ($-v_y'$, $-v_y$) was added to the dataset. Afterward, implementing the expression given in Eq. (3) results in the Gaussian distributions for the normal velocity components (Liao et al. 2018).

$$T(\theta) = \sqrt{2\beta_I}\, erf^{-1}\left[1 - 2\exp\left(-\frac{\theta^2}{2\beta_I}\right)\right], \beta_I = \frac{k_B T_I}{m_g} \quad (3)$$

where θ represents an arbitrary velocity component in one of the new created velocity sets, which are (v_y', $-v_y'$) and (v_y, $-v_y$) for the impinging and outgoing velocities, respectively. T_I refers to the temperature of impinging gas molecules and can be calculated based on the average translational kinetic energy of the gas molecules. Applying such a transformation doubles the size of the data in the normal direction. Thus, to be consistent with the number of data points related to the

other velocity components, half of the resulted new normal velocity components are added to the final training data.

After performing those mentioned above preprocessing on the MD collisional data (X), including the impinging (x_I) and the outgoing (x_O) velocity components, the GM model is employed to estimate the joint probability density of the collisional data. GM employs the superposition of multiple multivariate Gaussians to describe the probability density of the collisional data as:

$$P(x_O, x_I) = \sum_{i=1}^{K} \rho_i p_i(x_O, x_I)$$
$$p_i(x_O, x_I) = N(X|\vec{\mu}_i, \Sigma_i)$$
(4)

where $\rho_i, i = 1, 2, ..., K$ are the mixture component weights with the constraint that $\sum_{i=1}^{K} \rho_i = 1$, $\vec{\mu}_i$ is the mean vector, and Σ_i is the covariance matrix. These model parameters are determined using the expectation-maximization (EM) optimization algorithm(Dempster et al. 1977). Each component, p_i, of the mixture model is a multivariate Gaussian function given as:

$$N(X|\vec{\mu}_i, \Sigma_i) = \frac{1}{(2\pi)^{M/2}|\Sigma_i|^{1/2}} exp\left[-\frac{1}{2}(X - \vec{\mu}_i)'\Sigma_i^{-1}(X - \vec{\mu}_i)\right]$$
(5)

where M refers to the dimensionality of the dataset. To properly incorporate the GM model results into a DSMC simulation a conditional PDF, $R^{GM}(x_O|x_I)$, is required. To attain such Probability Density Function (PDF), in the first step, the obtained mean vector ($\vec{\mu}_i$) and covariance matrix ($\sum_i$) for each individual i, multivariate Gaussian component ($\forall i$ in$\{1 \cdots K\}$) are partitioned as follows:

$$\vec{\mu}_i = \begin{bmatrix} \mu_{iO} \\ \mu_{iI} \end{bmatrix}, \Sigma_i = \begin{bmatrix} \Sigma_{iOO} \Sigma_{iOI} \\ \Sigma_{iIO} \Sigma_{iII} \end{bmatrix}$$
(6)

where all the sub-matrices presented in $\sum_i$ are square matrices with the size of (3×3) and (5×5) for the Ar–Au and H_2–Ni systems, respectively. The GM scattering kernel is of the following form:

$$R^{GM}(X_O|X_I) = \sum_{i=1}^{K} \tilde{\rho}_i(X_I) p_i(X_O|X_I)$$
(7)

where the new set of weights $\tilde{\rho}_i(X_I)$ can be computed as:

$$\tilde{\rho}_i(X_I) = \frac{\rho_i p_i(X_I)}{\sum_{i=1}^{K} \rho_i p_i(X_I)}$$
(8)

The marginal $p_i(X_I)$ and the conditional $p_i(X_O|X_I)$ distributions demonstrated in Eqs. (8) and (7) are evaluated as (Williams and Rasmussen 2006)

$$p_i(X_I) = \int p_i(X_O, X_I) dX_O = N(X_I|\mu_{iI}, \Sigma_{iII})$$
(9)

and

$$p_i(X_O|X_I) = \frac{p_i(X_O, X_I)}{p_m(X_I)} = N(X_O|\mu_{iO|I}, \Sigma_{iO|I})$$
$$\mu_{iO|I} = \mu_{iO} + \Sigma_{iOI}\Sigma_{iII}^{-1}(x_I - \mu_{iI})$$
$$\Sigma_{iO|I} = \Sigma_{iOO} - \Sigma_{iOI}\Sigma_{iII}^{-1}\Sigma_{iIO}$$
(10)

From each MD collisional dataset, 75 % is applied for the training of the GM model, and the remaining data is utilized for validation purpose.

Finally, it should be pointed out that all the predicted velocity components by the GM model, in accordance with the initial distributions of the training data, follow Gaussian distributions. Therefore, to compare the GM model predictions with the original MD data, the normal velocity components need to be transferred back into the Rayleigh distribution using the following expression (Liao et al. 2018):

$$Z(\theta) = \sqrt{-2\beta_O \ln\left[\frac{1}{2} - \frac{1}{2} erf(\frac{\theta}{\sqrt{2\beta_O}})\right]}, \beta_O = \frac{k_B T_O}{m_g}$$
(11)

where T_O refers to the temperature of reflected gas molecules from the surface and can be computed based on the average translational kinetic energy of the gas reflected gas molecules.

2.3 DSMC simulation

DSMC simulation, a stochastic approach suitable for rarefied gases ($Kn > 0.1$), is considered the most popular particle-based simulation approach that has been successfully applied in a wide range of technological flow applications (Karniadakis et al. 2006).

The DSMC method models gases using discrete particles which normally represent a large number of real molecules. A probabilistic collision method is used to solve the Boltzmann equation (Bird and Brady 1994). The DSMC particles are initially distributed randomly in the simulation box. In this work, a one-to-one correspondence between the DSMC particles and MD molecules is established. In addition to particles' locations, their initial velocities are assigned randomly from a Maxwellian distribution at the given temperature. For H_2 particles, their initial rotational energy are generated based on equipartition theorem according to the prescribed temperature. The DSMC domain is subdivided into n_{bins} in the y-direction with dimensions smaller than

the particles' mean free path (see Fig. 1). Within each cell, the macroscopic flow properties, such as density or temperature, are calculated through sampling the particle statistics over a number of independent simulation trials, presented by time averaging. Particles can freely travel across the DSMC domain and collide with other particles. One key difference between the MD and DSMC is indeed the essence of the interparticle collisions in these approaches. In an MD simulation, the collisions are deterministic, described by interatomic potential functions. Integrating these functions numerically, one can derive the molecules' exact postcollisional velocity and position. On the contrary, in a DSMC simulation, the interparticle collisions are stochastic, defined by simplified molecular interaction models. In this work, the variable hard sphere (VHS) model (Bird and Brady 1994) is employed to describe the gas–gas collisions in the studied systems. The viscosity index, δ, for Ar and H_2 gases is set to 0.81 and 0.67, respectively (Bird and Brady 1994). Besides, in the case of H_2 molecules the Borgnakke-Larsen (BL) model (Borgnakke and Larsen 1975) is applied to deal with the possible energy exchange between the translational and rotational energy modes. The rotational collision number (Bird and Brady 1994), Z_r, required in the BL model computed using the empirical relation ($Z_r = 10480/T_a^\delta$) proposed by Boyd et al. (1994) for H_2 gas. Assuming stochastic nature for the intermolecular collisions is one of the substantial advantages of DSMC over MD simulation that makes it significantly less intensive from a computational point of view.

In general, DSMC is a promising tool for obtaining accurate gas transport properties outside the gas-solid surface interaction layer, defined by the cutoff distance of the interatomic potential (Liang and Ye 2014). In fact, near surface effects, such as the presence of the adsorption layer caused by the attractive part of the interaction potential, can not be captured by the common DSMC tools. To overcome this shortcoming, as it was suggested in Liang and Ye (2014), the DSMC domain length, L_{DSMC}, is shortened a bit and covers only the distance considered between the virtual borders in the MD simulation as $L_{DSMC} = L_y - 2r_c$ (see Fig. 1). As a result, the exact number of DSMC particles, N_{DSMC}, is a bit less than the number of particles used in the MD simulation since the gas particles adsorbed on the surface are not included in the DSMC simulation. This issue will be elaborated on more in the following section.

In DSMC, at the gas-solid interface, the velocity vectors for the reflected particles are typically determined by sampling from a specific distribution function known as a scattering kernel. Here, the CLL and GM scattering kernels are employed at the virtual borders (see Fig. 1). Periodic boundary conditions are applied in the lateral directions (x,z).

Another important issue is that in the case of H_2 molecules the rotational velocity vectors ($\boldsymbol{\omega}'$, $\boldsymbol{\omega}$) are used for the training of the GM model. However, in the currently available DSMC solvers particles are treated as the point-particles and they do not explicitly model rotational velocities but rather accounts for a scalar value for the rotational energy. To deal with this issue, as it was suggested by Yamamoto et al. (2006), the impinging rotational energy is equally decomposed into two rotational velocities in the rotational directions ($\omega_1' = \omega_2' = \sqrt{E_{rot-I}/I_g}$).

In this work, DSMC simulations are carried out using dsmcFoam+ (White et al. 2018) solver, a DSMC solver implemented within the Open- FOAM (Weller et al. 1998) software framework. More details about integrating the GM scattering kernel into the dsmcFoam+ solver can be found in the following section.

2.4 Integration of the GM scattering kernel into the DSMC solver

This section elaborates further upon integrating of the GM scattering kernel with the applied DSMC solver in the case of H_2–Ni system. The integration procedure for the Ar–Au system is not discussed here as it is similar except that it excludes the angular velocities.

First of all, considering $\boldsymbol{x_I} = (v_x', v_y', v_z', \omega_1', \omega_2')$ as the impinging, and $\boldsymbol{x_O} = (v_x, v_y, v_z, \omega_1, \omega_2)$ as the outgoing velocity components obtained from the MD simulation the corresponding translational temperatures (T_I, T_O) are computed. As addressed in Eq. (3), T_I is applied to transform the distribution of the impinging normal velocity component from Rayleigh to Gaussian distribution. On the other hand, T_O is utilized to transform the distribution of the outgoing normal velocity component produced by the GM scattering kernel from Gaussian to Rayleigh distribution (see Eq. 11).

After training the GM model using the the MD collisional data $X = (\boldsymbol{x_I}, \boldsymbol{x_O})$ at a specific wall, the model parameters ($\{\rho_i, \vec{\mu}_i, \Sigma_i\} \forall i \text{ in} \{1 \cdots K\}$) are employed to calculate the following constants for each multivariate Gaussian function i:

$$S_{A,i} = \frac{1}{(2\pi)^{M/4}|\Sigma_{iII}|^{1/2}} \tag{12}$$

$$S_{B,i} = \Sigma_{iII}^{-1} \tag{13}$$

$$S_{C,i} = \Sigma_{iOI}\Sigma_{iII}^{-1} \tag{14}$$

$$S_{D,i} = Chol[\Sigma_{iO|I} + \epsilon I] \tag{15}$$

$S_{A,i}$ and $S_{B,i}$ are used in Eq. (9) to compute $p_i(X_I)$. S_{Ci} and $S_{D,i}$ are applied to sample from a multivariate Gaussian with

arguments $\boldsymbol{\mu}_{iO|I}$ and $\Sigma_{iO|I}$. The procedure proposed in Williams and Rasmussen (2006) is followed for sampling purpose. Assuming $\Sigma_{iO|I}$ is a positive definitive matrix, the Cholesky decomposition (*Chol*) is used to decompose the covariance matrix into the product of a lower triangular matrix and its transpose. In practice, adding a small multiple of the identity matrix I to the covariance matrix may be required for numerical stability. Here, $\epsilon = 0.0001$ is used as a small perturbation. Generating data based on the trained GM scattering kernel is one of the most crucial steps in the GM-DSMC model. As it is shown in Fig. 2, the sampling

procedure starts with computing the probability, p_i, of a given transformed impinging velocity, $X^*_{I,j}$ belonging to the Gaussian component i (block C1 in Fig. 2). Doing such for all K components of the GM model, the accumulated weight, α, is computed. Using p_is and α, the new set of weights, $\tilde{\rho}_i$, is computed in block C2. These new weights, alongside selecting a random number, R_1, uniformly distributed between 0 and 1, are applied to choose the specific component i from the mixture model to generate the final sample (block C4 in Fig. 2). The arguments of the chosen
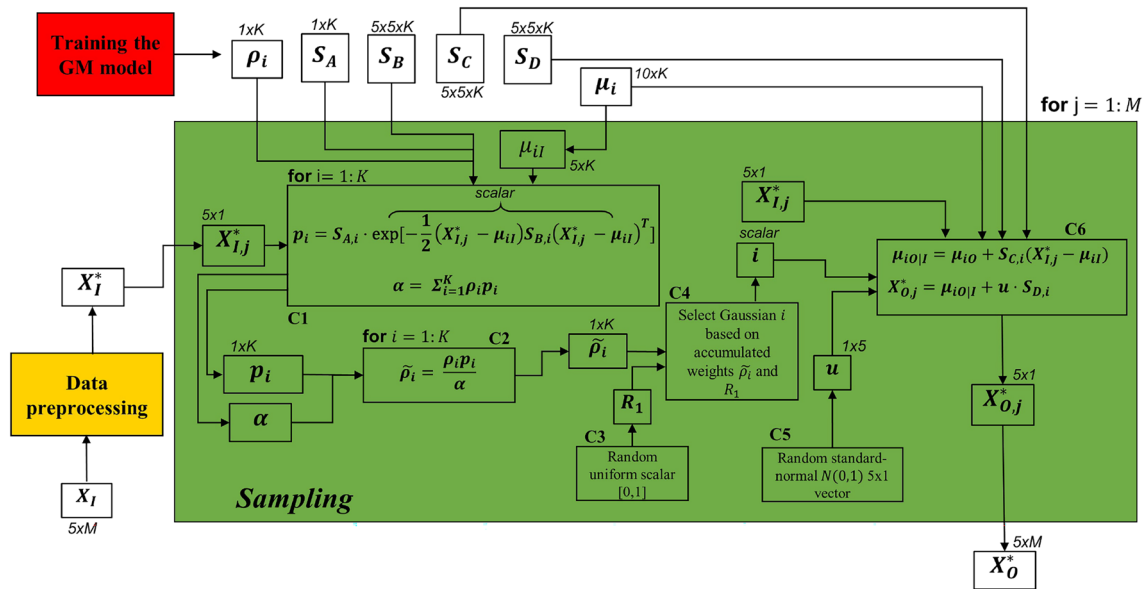
Fig. 2 The steps (C1:C6) followed to generate M samples from a GM model with K components; X^*_I and X^*_O are the transformed impinging and outgoing velocity vectors
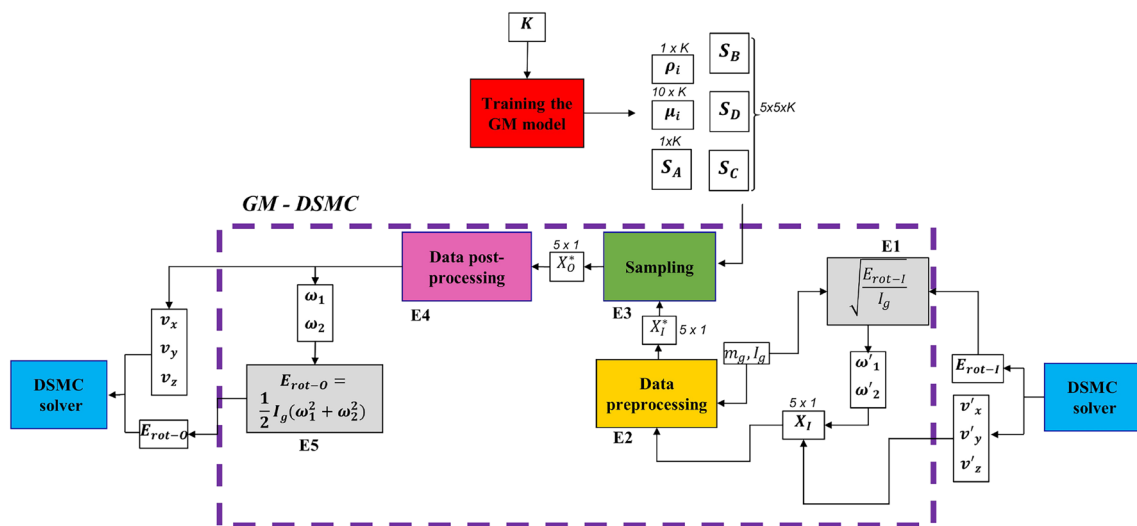

Fig. 3 The scheme showing the integration of the GM scattering kernel into the DSMC solver

Gaussian component (μ_{iI}, μ_{iO}, S_{Ci}, $S_{D,i}$) are used to generate the final sample (block C6 in Fig. 2) based on a random vector, u, drawn from a normal distribution with the mean value of 0 and variance of 1 (block C5 in Fig. 2). This procedure is repeated M times to generate the corresponding outgoing velocity of all the impinging velocities.

The scheme followed to couple the GM scattering kernel into the DSMC solver is shown in Fig. 3. A DSMC particle approaching to the system boundary has the translational velocity (v'_x, v'_y, v'_z) and the rotational energy E_{rot-I}. Since the GM mode is trained based on the rotational velocities, in the first step (block E1 in Fig. 3) E_{rot-I} is applied to derive the corresponding rotational velocities (ω'_1, ω'_2). In the next step (block E2 in Fig. 3), the data preprocessing is carried out, which includes transforming the normal velocity distribution from the Rayleigh into the Gaussian distribution, using Eq. (3), and normalizing the translational and rotational velocity vectors. The preprocessed impinging velocity, X^*_I, is used to generate the outgoing velocity vector, X^*_O, in block E3. Afterward, X^*_O is converted back into the initial units, and the normal velocity distribution is also transformed back into the Rayleigh distribution using Eq. (11) (block E4 in Fig. 3). Finally, the rotational velocities are used to compute the corresponding outgoing rotational energy value, E_{rot-O}, that alongside the outgoing translational velocity components (v'_x, v'_y, v'_z) are assigned to the reflected DSMC particle.

3 Results and discussion

The performance of the GM and CLL scattering kernels employed as boundary conditions in DSMC simulations are investigated using the MD results as the reference solutions. For each gas-solid pair, two benchmarking case studies are considered: (i) isothermal Fourier problem, in which the temperature of the bottom and top walls are set to 300 K; (ii) non-isothermal Fourier problem, in which the temperature of the bottom wall is maintained at 300 K, while that of the top wall is fixed at 500 K. For each case study after performing the MD simulations, the required ACs in the CLL scattering kernel are computed. These ACs for the Ar–Au and H$_2$–Ni systems are listed in Tables 2 and 3, respectively. Different physical and

Table 2 Tangential momentum (α_x) and normal energy (α_n) accommodation coefficients of the Fourier thermal problem for the Ar–Au system obtained from MD simulations

System	Wall	α_x	α_n
Isothermal	Bottom/top	0.876	0.910
Non-isothermal	Bottom	0.886	0.910
	Top	0.775	0.856

Table 3 Tangential momentum (α_x), normal energy (α_n), and rotational energy (α_{rot}) accommodation coefficients of the Fourier thermal problem for the H$_2$–Ni system obtained from MD simulations

System	Wall	α_x	α_n	α_{rot}
Isothermal	Bottom/top	0.951	0.758	0.656
Non-isothermal	Bottom	0.930	0.724	0.596
	Top	0.922	0.730	0.600

statistical criteria are employed to assess the performance of the applied scattering kernels, including the number density and temperature profiles between two walls, as well as the predicted velocity distributions at the gas-wall interfaces. In the case of the predicted number density and temperature by DSMC, the accuracy of the simulation results is determined by measuring the deviations of the DSMC results from the pure MD results. Considering x_{y-DSMC} as the predicted number density or temperature in a spacial bin by the DSMC approach coupled to the y scattering kernel, the deviations are computed by $\sqrt{(x_{y-DSMC} - x_{MD})^2}/x_{MD}$, where x_{MD} refers to the corresponding property obtained from the MD simulation.

3.1 Isothermal Ar–Au system

Figure 4 shows the density and temperature profiles of the isothermal Fourier problem for Ar–Au system obtained from the DSMC simulations combined with GM and CLL scattering models alongside the pure MD results. As indicated in the previous section, gas–wall interaction zones need to be excluded to have a fair comparison between the MD and DSMC results. The reason is that DSMC cannot predict gas transport properties in these areas. In all the case studies in this work, looking at the number density profile obtained from the MD simulations, the number of the adsorbed gas molecules on the surfaces, N_{ads}, are excluded from the total number of the MD particles, N_{MD}, resulted in $N_{DSMC} = N_{MD} - N_{ads}$. As an example, in the present case study $N_{MD} = 800$. Based on the number density profile (see Fig. 4a), the total number of Ar atoms adsorbed at the bottom and top walls are about $N_{ads} = 225$, which leads to $N_{DSMC} = 575$.

From Fig. 4 it is deduced that the DSMC results using both the CLL and GM scattering models are in good agreement with the original MD results. To be more specific, the average deviations of the predicted number densities by the GM-DSMC and CLL-DSMC are 0.4% and 0.5%, respectively. While for the temperature, the deviations are 0.2% on average.
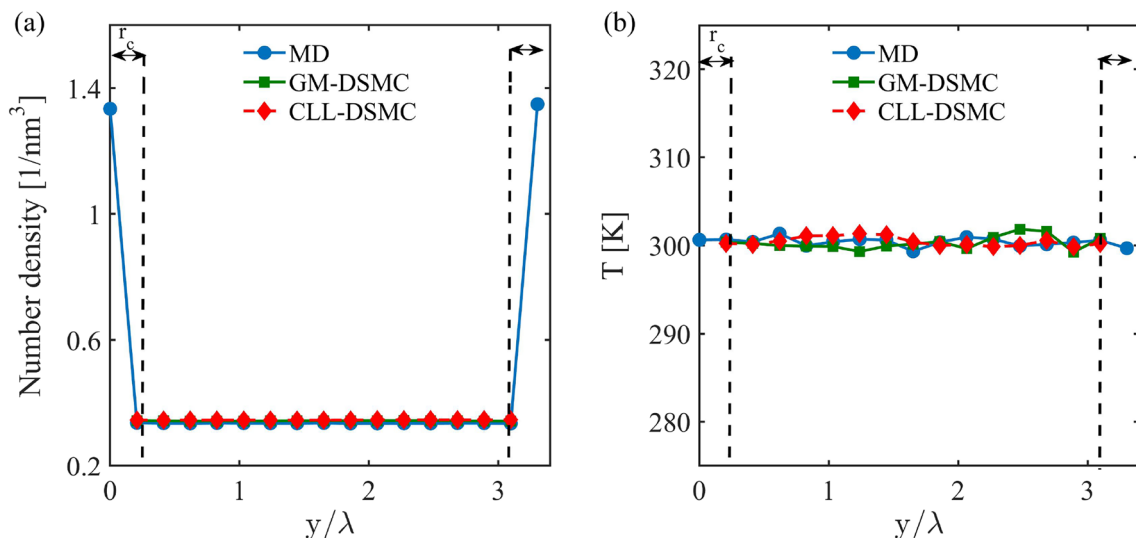
Fig. 4 Variation of the macroscopic quantities of the isothermal Fourier problem for Ar–Au system obtained from the pure MD simulation and DSMC simulations combined with the GM and CLL scattering models. **a** Number density, **b** temperature; $r_c = 12$ Å

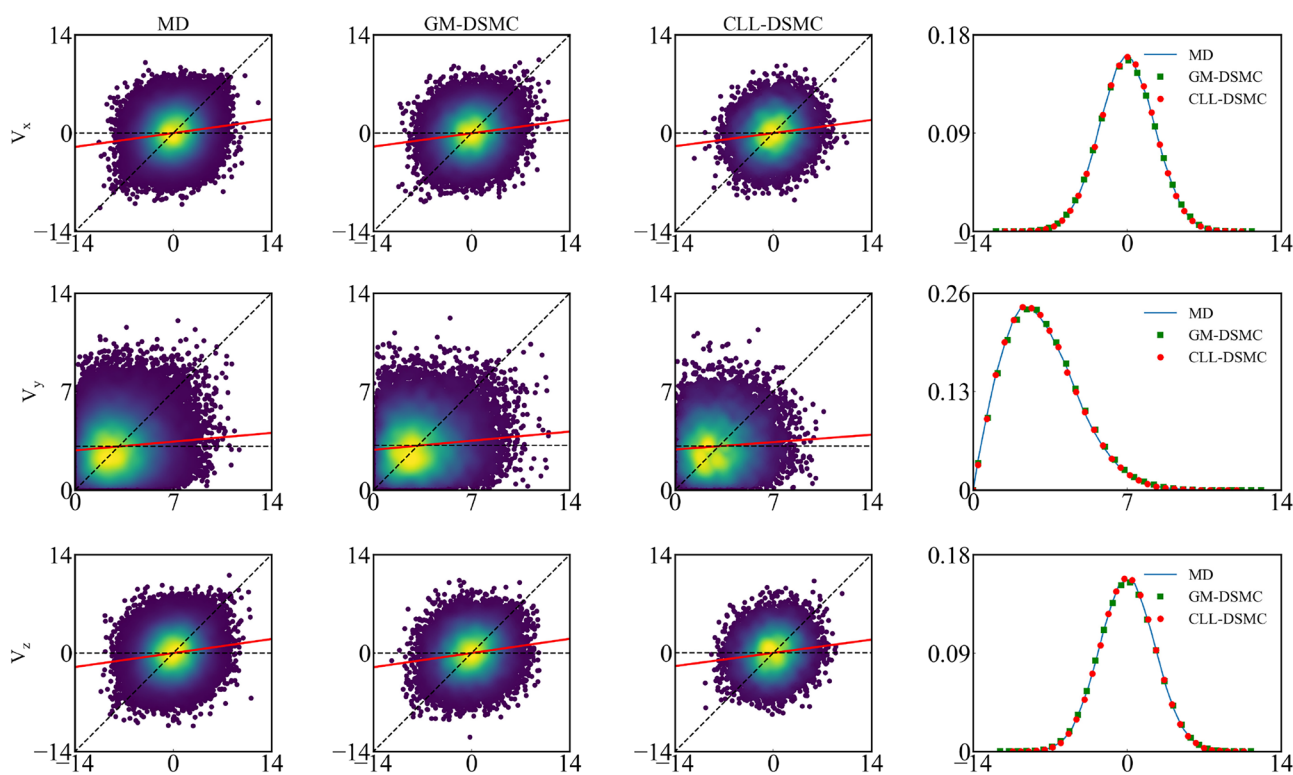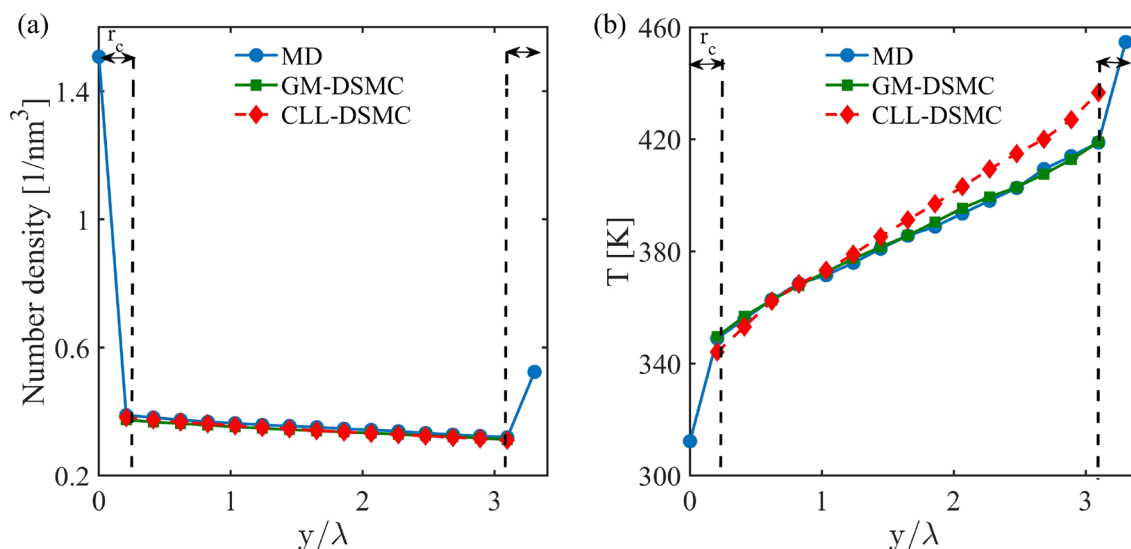**Fig. 5** Velocity correlations of impinging (horizontal-axis) and reflected (vertical-axis) velocity components in [Å/ps] of the isothermal Fourier problem for Ar–Au system at the bottom wall. The dashed horizontal and diagonal lines indicate fully diffusive and specular conditions, respectively. Red lines indicate the least-square linear fit of the data, its slope infers: 1-AC. In the last column the corresponding PDF for the reflecting particles are shown (color figure online)

Fig. 6 Variation of the macroscopic quantities of the non-isothermal Fourier problem for Ar–Au system obtained from the pure MD simulation and DSMC simulations combined with the GM and CLL scattering models. **a** Number density, **b** temperature, $r_c = 12$ Å (color figure online)

The correlation between the impinging and outgoing velocity components, as well as the PDF of the outgoing velocity components obtained from the applied simulation approaches for the isothermal Ar–Au system, are depicted in Fig. 5. It is seen that in this case study, which resembles the fully equilibrium condition in the system, the velocity distributions predicted by the GM-DSMC and CLL-DSMC approaches match well the MD data.

3.2 Non-isothermal Ar–Au system

In Fig. 6, the variation of the local number density and temperature observed in the MD simulation and the predicted trends by the GM-DSMC and CLL-DSMC of the non-isothermal Fourier problem for Ar–Au case study are presented. In general, having a higher temperature at the top wall causes less number of atoms to be adsorbed at the surfaces in this case (i.e., $N_{ads} = 222$, $N_{DSMC} = 578$). Regarding the number density (see Fig. 6a), similar to the previous case study, the predicted results by DSMC incorporating both scattering models are consistent with the MD data. Here, the average deviations of the predicted results are 0.6% for both GM-DSMC and CLL-DSMC approaches. In the case of the temperature profile (see Fig. 6b), the trend predicted by the GM-DSMC in the bulk of the domain matches well the pure MD results. Here, the deviations are 0.2%. On the other hand, in most parts of the simulation domain, the results predicted by the CLL-DSMC method deviate from the MD results, observing the highest deviation close to the top wall, which is 4%.

Another observation is the noticeable temperature jump between the consecutive bins located at the beginning and end of the simulation domain. This observation is in line with the previously observed temperature jump adjacent to the solid surface induced by the strong gas-wall interactions within the potential cutoff distance (Markvoort et al. 2005).

The scattering results obtained from the MD, GM-DSMC, and CLL-DSMC approaches at the bottom and top walls are presented in Figs. 7 and 8, respectively. It is seen that the velocity clouds predicted by the GM-DSMC and CLL-DSMC approaches are very similar to the MD results. Looking at the PDFs of the outgoing velocities in the *x* and *z*-directions at both walls, some discrepancies between the CLL-DSMC and MD results around the peak of the graph are seen. Nevertheless, the results from the GM-DSMC approach are always in good agreement with the MD results.

3.3 Isothermal H$_2$–Ni system

Variations of the number density and temperature in the case of the isothermal Fourier problem for the H$_2$–Ni system are shown in Fig. 9. Comparing the number density profile obtained from the MD simulation with the one for the isothermal Ar–Au system (see Fig. 4a), presence of the weaker adsorption layer is seen in the current case study. This is manifested through the relatively smaller difference between the number densities of the consecutive bins near the walls in the current case than the isothermal Ar–Au system. It is seen that for the isothermal H$_2$–Ni system the measured number densities in all bins are in the same order of magnitudes, while for the isothermal Ar–Au system the number densities near the walls are one order of magnitude higher than the ones in the bulk. This outcome is caused by significantly higher mass of the Ar atom compared to the
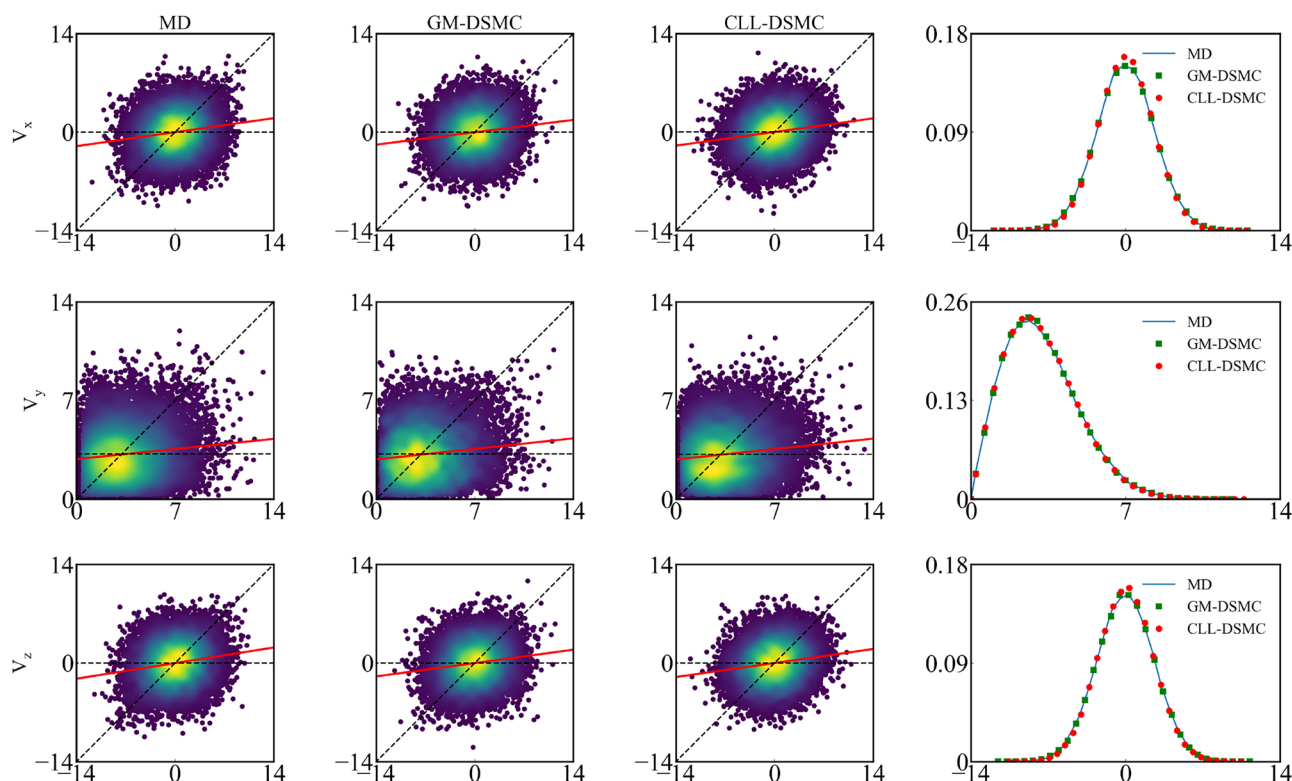
Fig. 7 Velocity correlations of impinging (horizontal-axis) and reflected (vertical-axis) velocity components in [Å/ps] of the non-isothermal Fourier problem for Ar–Au system at the bottom wall. The dashed horizontal and diagonal lines indicate fully diffusive and specular conditions, respectively. Red lines indicate the least-square linear fit of the data, its slope infers: 1-AC. In the last column the corresponding PDF for the reflecting particles are shown (color figure online)

H_2 ($m_{Ar} \approx 20m_{H2}$). Regarding the number of particles used in the MD and DSMC simulations of the H_2–Ni system, $N_{MD} = 900$ particles were used in the MD simulation. It was observed that around 100 particles were trapped in the gas-surface interaction zones during the MD simulation. Therefore, $N_{DSMC} = 800$ particles were considered in the DSMC simulation.

Going back to the predicted trends of the number density and temperature obtained based on different approaches in this case study, the results of GM-DSMC and CLL-DSMC agree with MD data. Here, the deviations of the predicted number densities by the GM-DSMC and CLL-DSMC are 0.6% and 0.4%, respectively. On the other hand, the temperature results of the DSMC simulations coupled with the GM and CLL scattering models on average deviate around 0.2% from the MD results.

The correlation plots and PDFs for different translations velocity components and energy modes of the isothermal Fourier problem for H_2–Ni system are presented in Fig. 10. It is shown that in this case study, there is a good agreement between the correlation plots and PDFs of the partial translational velocity components, and the rotational energy mode. However, E_{tr} and E_{tot} correlation clouds of the pure MD and GM-DSMC are narrower than the CLL-DSMC approach, which since $E_{tot} = E_{tr} + E_{rot}$ the mismatch in E_{tr} induced the mismatch in the results for E_{tot}.

3.4 Non-isothermal H_2–Ni system

Figure 11 shows the number density and temperature profile in the case of the non-isothermal Fourier problem for the H_2–Ni system. First of all, from the MD simulation it is realized that $N_{ads} = 94$ H_2 are adsorbed on the surfaces. Therefore, $N_{DSMC} = 806$ particles are used in the DSMC simulation. Regarding the number density variation (see Fig. 11a), it is observed that within the bulk of the simulation domain, the GM-DSMC approach results match well with the MD data. Here, the deviations of the GM-DSMC results are around 0.6% on average. However, unlike all the previously investigated ones, a notable discrepancy between the density profiles obtained from the MD and CLL-DSMC is observed in the current case study. In this case, the highest deviation is measured near the top wall, which is 8%. In Fig. 11b, it is depicted that the predicted temperature profiles based on both GM-DSMC and CLL-DSMC approaches deviate from the reference MD results. Nevertheless, the GM-DSMC
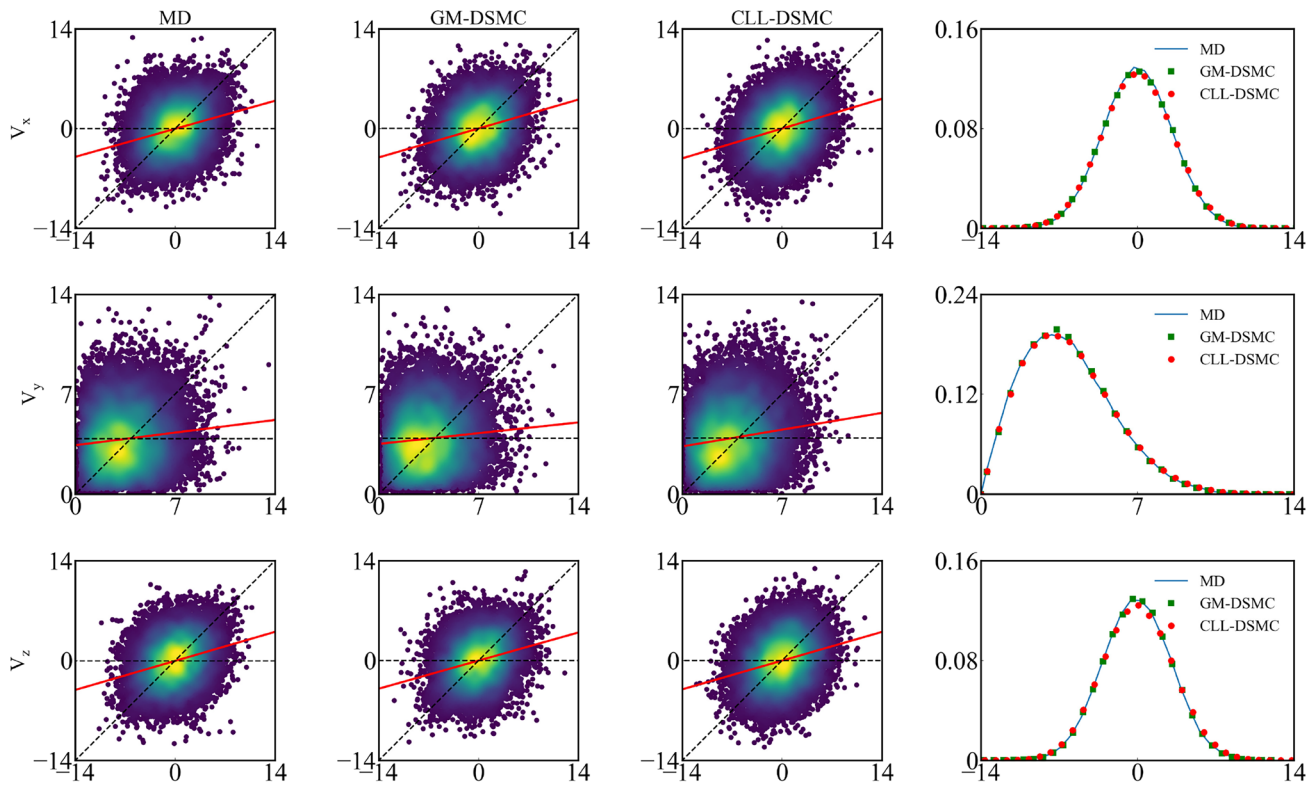
Fig. 8 Velocity correlations of impinging (horizontal-axis) and reflected (vertical-axis) velocity components in [Å/ps] of the non-isothermal Fourier problem for Ar–Au system at the top wall. The dashed horizontal and diagonal lines indicate fully diffusive and specular conditions, respectively. Red lines indicate the least-square linear fit of the data, its slope infers: 1-AC. In the last column the corresponding PDF for the reflecting particles are shown (color figure online)
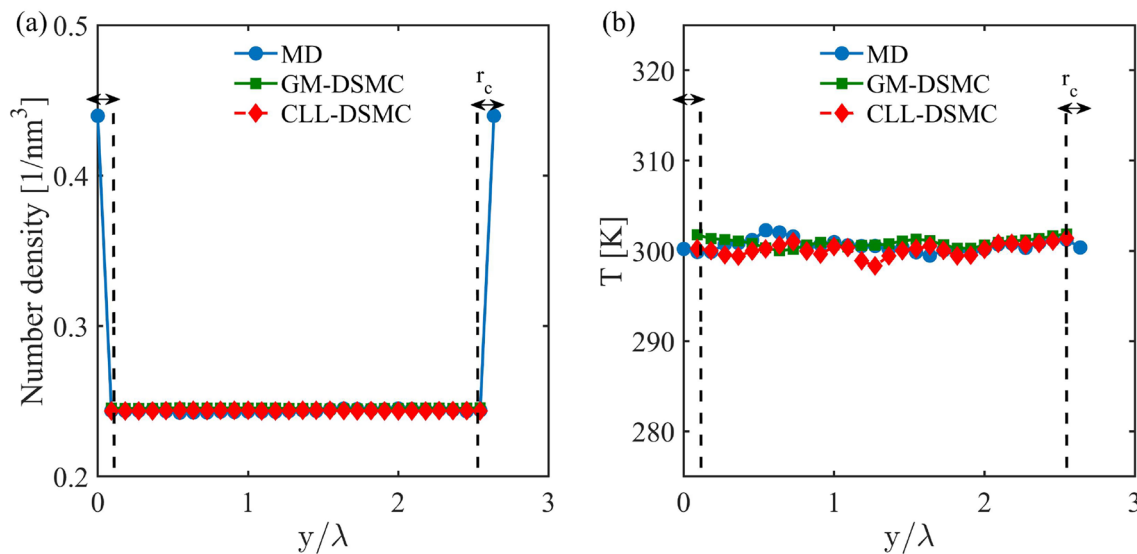
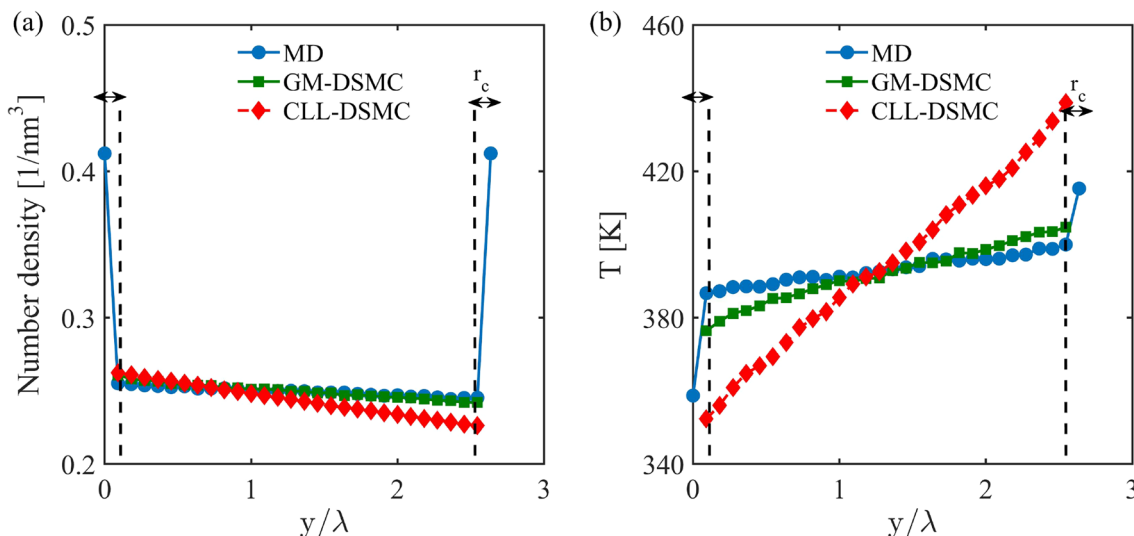**Fig. 9** Variation of the macroscopic quantities of the isothermal Fourier problem for the H_2–Ni system obtained from the pure MD simulation and DSMC simulations combined with the GM and CLL scattering models. **a** Number density, **b** temperature. $r_c = 10$ Å
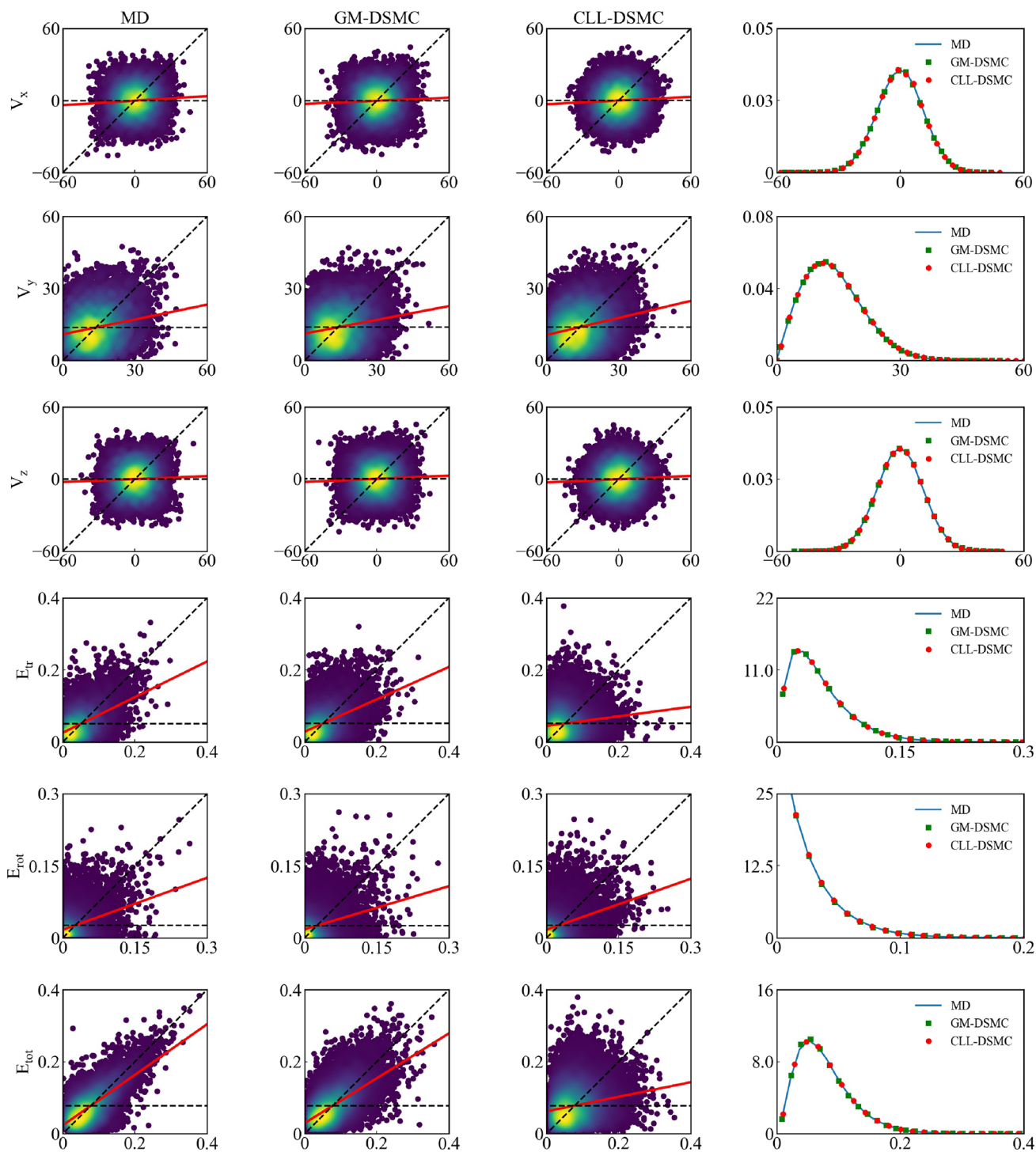
Fig. 10 Correlations between incoming (horizontal-axis) and outgoing (vertical-axis) translational velocity components in [Å/ps] and energy modes in [eV] of the isothermal Fourier problem for the H_2–Ni system at the bottom wall. The dashed horizontal and diagonal lines demonstrate fully diffusive and specular reflection, respectively.

Solid red lines demonstrate the least-square linear fit of the kinetic data, its slope infers: 1-AC. In the last column the corresponding PDF of translational velocity components and energy modes for the reflecting particles are presented (color figure online)

Fig. 11 Variation of the macroscopic quantities of the non-isothermal Fourier problem for H_2–Ni system obtained from the pure MD simulation and DSMC simulations combined with the GM and CLL scattering models. **a** Number density, **b** temperature. $r_c = 10$ Å

approach still outperforms the CLL-DSMC approach. Herein, the highest deviations based on the GM-DSMC approach are measured at the first and last bins, which are 2% and 1%, respectively. On the other hand, using the CLL-DSMC approach, the deviations on the same bins are 9% and 10%, respectively.

The scattering results at the bottom wall of the non-isothermal Fourier problem for the H_2–Ni system are shown in Fig. 12. It is seen that the correlations plots of the translation velocity components obtained from the GM-DSMC and CLL-DSMC are in good agreement with the MD results. However, while the correlation graphs for V_x and V_z components obtained from the CLL-DSMC resemble perfect ellipsoid, the MD results look more skewed around the diagonal line. This observation, which can be perfectly captured by the GM-DSMC approach, indicates that many gas molecules with high velocity experience almost specular reflection. In addition, the PDFs of the outgoing velocity components predicted by the GM-DSMC approach are a perfect match, while the CLL-DSMC predictions deviate from the MD results around the peak value of the PDF plots. Looking at the scattering results related to the different energy modes, except for the rotational energy mode, in the other energy modes, the results from the CLL-DSMC deviate from the MD results. Nevertheless, the GM-DSMC results are in better agreement with the MD results. Except for the PDF of rotational energy mode, the peak values of all PDFs predicted by the CLL-DSMC approach are higher than MD and GM-DSMC results. A higher peak implies a lower temperature. Therefore, the CLL-DSMC overestimates the degree of H_2 accommodation to the surface. This issue is also seen in the correlation graphs for E_{tr} and E_{tot}, in which

the slopes of the red lines, indicating 1-AC, are smaller for CLL-DSMC in comparison with the MD and GM-DSMC results.

The relatively significant mismatch between the reference MD and the CLL-DSMC results in this case study, shown in Fig. 11b and the last column of Fig. 12, is caused by the small molecular weight of H_2 molecules and the small size of the channel. In the MD simulation, the light H_2 molecules move very fast between the two plates causing the temperature in different parts of the bulk to converge towards the average value of the bottom and top plates temperatures ($T_a = 400K$). Based on the incoming and outgoing collisional data recorded at each wall, we can compute the incoming and outgoing gas temperatures adjacent to the walls. For the non-isothermal H_2–Ni system based on the MD results, the incoming and outgoing temperatures at the bottom wall are $T_{in-MD-b} = 403$ K and $T_{out-MD-b} = 378$ K, respectively. On the top wall, the incoming and outgoing gas temperatures are $T_{in-MD-t} = 385$ K and $T_{out-MD-t} = 411$ K. These temperatures are close to the MD values near the walls shown in Fig. 11b. Our hybrid GM-DSMC model can deal with this, while the CLL scattering model can not anticipate such behavior. Next, the performance of the CLL model also highly depends on the values of applied ACs. As an example, Spijker et al. (2010) computed ACs for the Ar-Pt system based on the classical and correlation approaches. The obtained ACs from the correlation approach were slightly higher than the classical approach. In their study, they showed a significant difference in the CLL scattering model results, based on the computed ACs. In this case study, the obtained ACs (see Table 3), especially the tangential momentum AC, are high, indicating a large degree
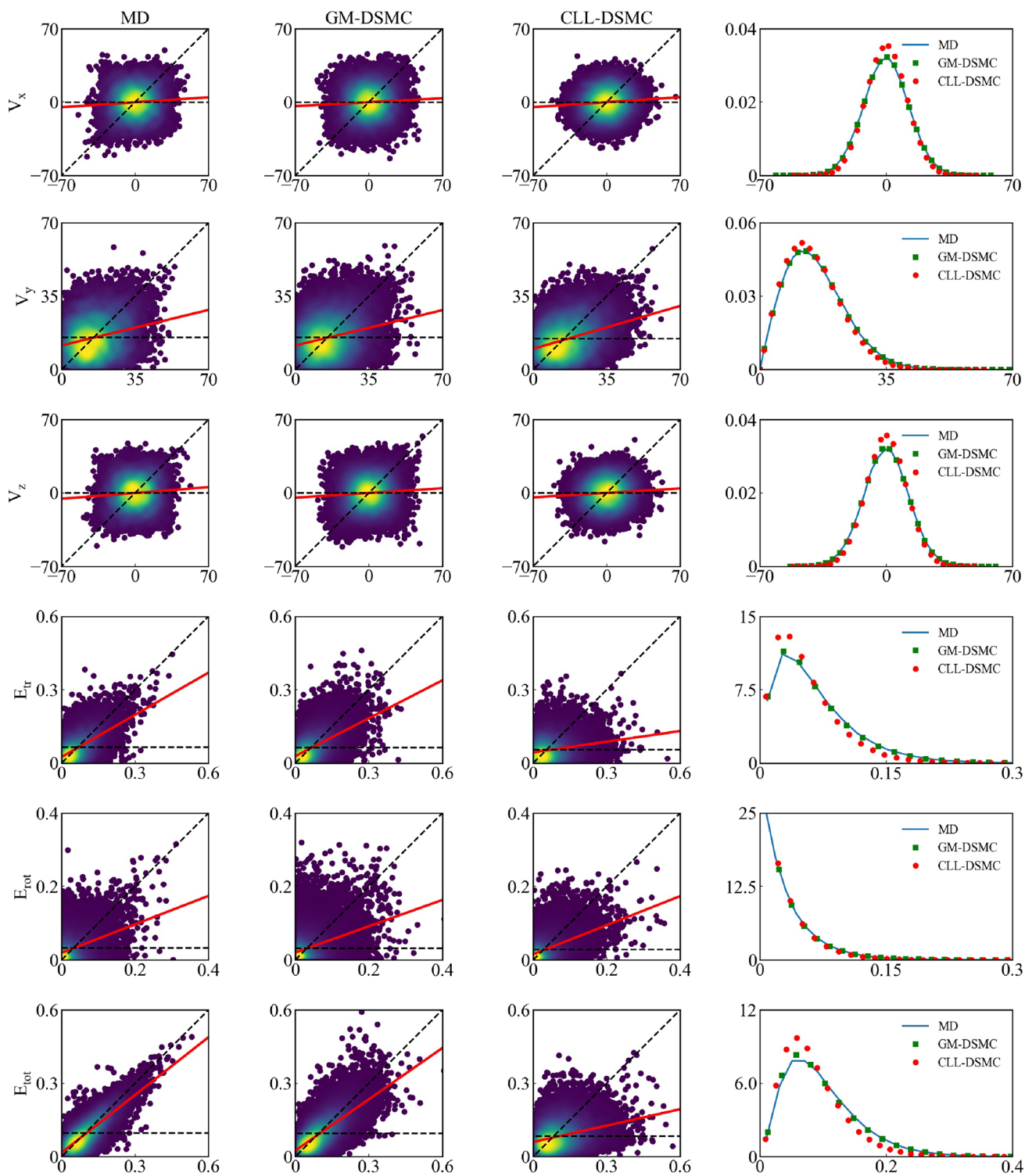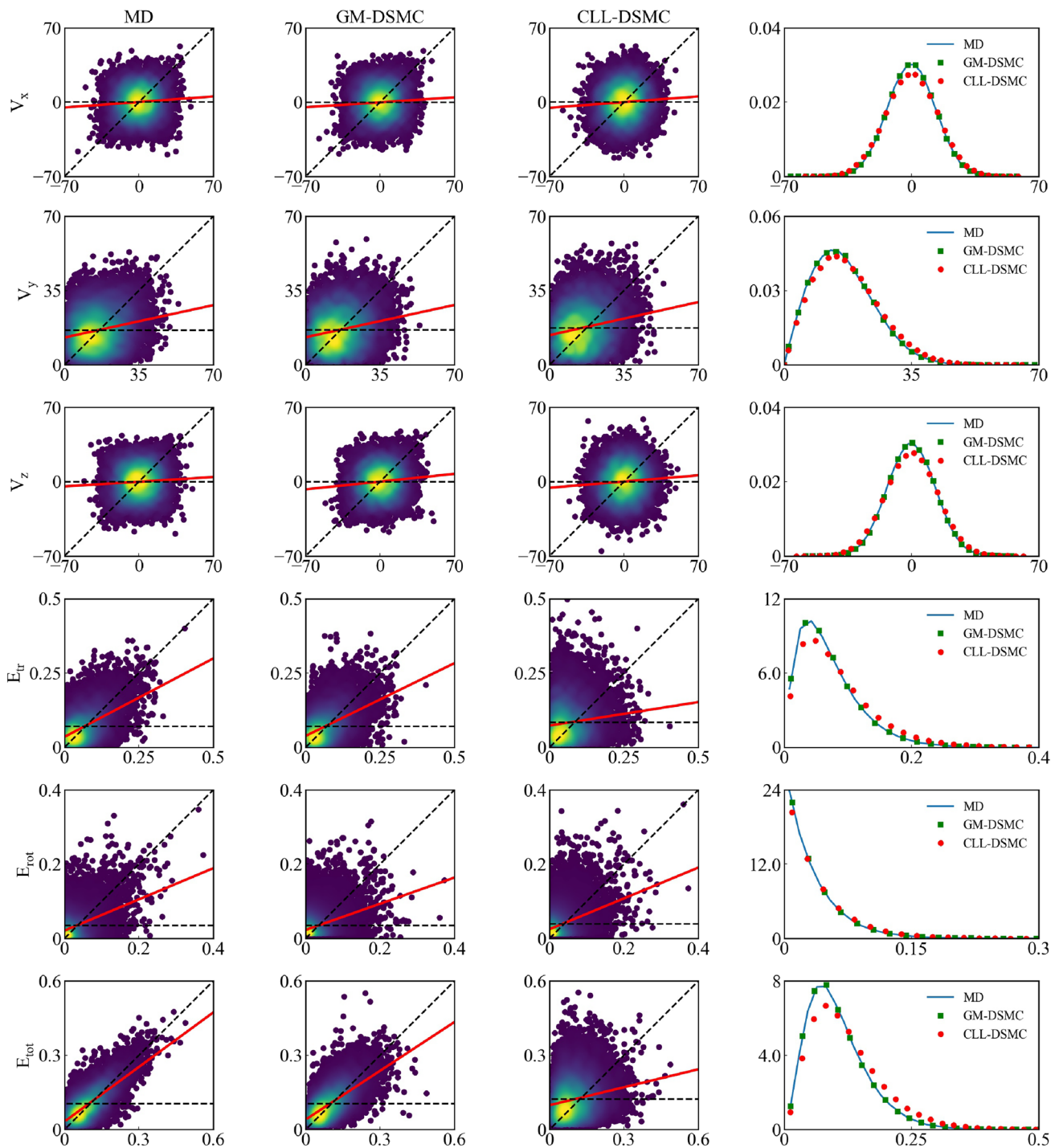
Fig. 12 Correlations between incoming (horizontal-axis) and outgoing (vertical-axis) translational velocity components in [Å/ps] and energy modes in [eV] of the non-isothermal Fourier problem for H_2–Ni system at the bottom wall. The dashed horizontal and diagonal lines demonstrate fully diffusive and specular reflection, respectively. Solid red lines demonstrate the least-square linear fit of the kinetic data, its slope infers: 1-AC. In the last column the corresponding PDF of translational velocity components and energy modes for the reflecting particles are presented (color figure online)

Fig. 13 Correlations between incoming (horizontal-axis) and outgoing (vertical-axis) translational velocity components in [Å/ps] and energy modes in [eV] of the non-isothermal Fourier problem for the H_2–Ni system at the top wall. The dashed horizontal and diagonal lines demonstrate fully diffusive and specular reflection, respectively. Solid red lines demonstrate the least-square linear fit of the kinetic data, its slope infers: 1-AC. In the last column the corresponding PDF of translational velocity components and energy modes for the reflecting particles are presented (color figure online)

of accommodating the gas molecules to the neighboring surfaces. To entirely study the impact of obtained ACs on the gas temperature, using the algorithm proposed in

Hossein Gorji and Jenny (2014), the incoming rotational and translational MD velocities captured at the bottom wall are utilized to generate post-collisional velocities according to

the CLL scattering model. Based on the obtained outgoing velocities, the outgoing gas temperature is $T_{out-CLL-b} = 314$ K. Looking to Fig. 11b, the temperature in the very first bin at the left side of temperature profile related to the CLL-DSMC case study is 359 K. This temperature is very close to the average of incoming and outgoing temperatures computed here ($T_{in-MD-b} = 403$ K, $T_{out-CLL-b} = 314$ K).

The scattering results at the top wall of the non-isothermal H_2–Ni system are shown in Fig. 13. Comparing the results obtained from the MD, GM-DSMC, and CLL-DSMC, similar trends to the bottom wall can also be seen here. For E_{rot}, the results of GM-DSMC and CLL-DSMC are consistent with the MD data. For the other energy modes and translational velocity components, the GM-DSMC results are always in better agreement with MD results than the CLL-DSMC results. Looking at the PDFs, except for E_{rot}, in the other components, the peak value of the PDF predicted by the CLL-DSMC is lower than MD and GM-DSMC results. It means CLL-DSMC predicts a higher temperature for reflected gases. This observation, alongside underestimating the outgoing temperature at the bottom wall, indicates that the CLL-DSMC is not able to predict the actual temperature jump happening at the gas-solid interfaces. Such a conclusion can also be derived from Fig. 11b, in which the predicted temperatures by the CLL-DSMC near the cold and hot walls are respectively lower and higher than the reference MD results.

4 Conclusions

DSMC simulations, as the most common particle-based simulation techniques, can be applied to derive precise solutions of gas transport properties outside the gas-wall interaction layer based on the accurate velocity distribution function of the reflected gas molecules provided by MD simulations. In this work, MD simulations on monoatomic and diatomic gases-surface interactions under thermal equilibrium and non-equilibrium conditions are carried out. Using the MD collisional data, including the pre and postscattered molecular gas velocities from each case study, an unsupervised machine learning approach, called the GM model, is employed to construct a gas-surface boundary model. The GM boundary model is rewritten in the form of a conditional multivariate PDF that can reproduce the postscattered gas molecular velocities based on their initial imposed velocities and therefore can be used as boundary condition in a DSMC simulation.

This new conditional GM scattering kernel is successfully incorporated in a pure DSMC simulation as a boundary condition to study isothermal and non-isothermal Fourier problems. In each case study, the

performance of the proposed model (GM-DSMC) is assessed against the DSMC simulations based on the CLL scattering kernel (CLL-DSMC) using full MD results as the reference solution. Comparing different physical (e.g., temperature field) and stochastic (e.g., velocity and energy distributions) parameters obtained from the applied simulation approaches confirm the superiority of the GM-DSMC approach. The hybrid GM-DSMC approach demonstrated superior results, particularly in the diatomic case (H_2-Ni), surpassing the outcomes obtained with the CLL-DSMC approach. The performance was closely aligned with the full MD results, highlighting the significant potential of this new approach.

These results also pave the way to develop a multiscale simulation scheme, which combining an accurate generalized scattering model based on MD data with DSMC simulation, can accurately measure different flow field properties by efficiently including the microscopic non-continuum phenomena. To achieve such a scattering kernel, an expanded training data set including other physical parameters of the system, such as the wall temperatures and gas density is required, and this is in our plan for future studies.

Appendix: Sensitivity analysis to determine the number of Gaussian functions (K) used in the GM model

In the case of the GM model a sensitivity analysis is required to choose the optimal number of Gaussian functions. In this work, for both gas-solid pairs the isothermal Fourier thermal problem ($T_b = T_t = 300$ K) is chosen as the benchmark system. The number of Gaussians, K, is varied in the range of $K = 1-1000$.

For the Ar–Au system, including only the translational degrees of freedom, the translational kinetic energy AC (α_{tr}) driven from the GM model is compared with the corresponding value obtained from MD simulation. On the other hand, for the H_2–Ni system the ACs for different gas molecule's energy modes consist of the translational energy AC (α_{tr}), the rotational energy AC (α_{rot}), and the total energy AC (α_{tot}) driven from the GM model are compared with the reference MD results. The results of sensitivity analysis for both gas-solid pairs are shown in Fig. 14. In the case of the Ar–Au system, for $K \geqslant 50$ already the difference between α_{tr} based on the GM model and the MD data is less than 5%. Therefore, K = 100 is utilized as the number of Gaussians for training the GM model based on the results from the Ar–Au system. For the H_2–Ni system for $K \geqslant 500$ the computed ACs become stable and no significant change in their values are observed. Therefore, $K = 500$ is employed
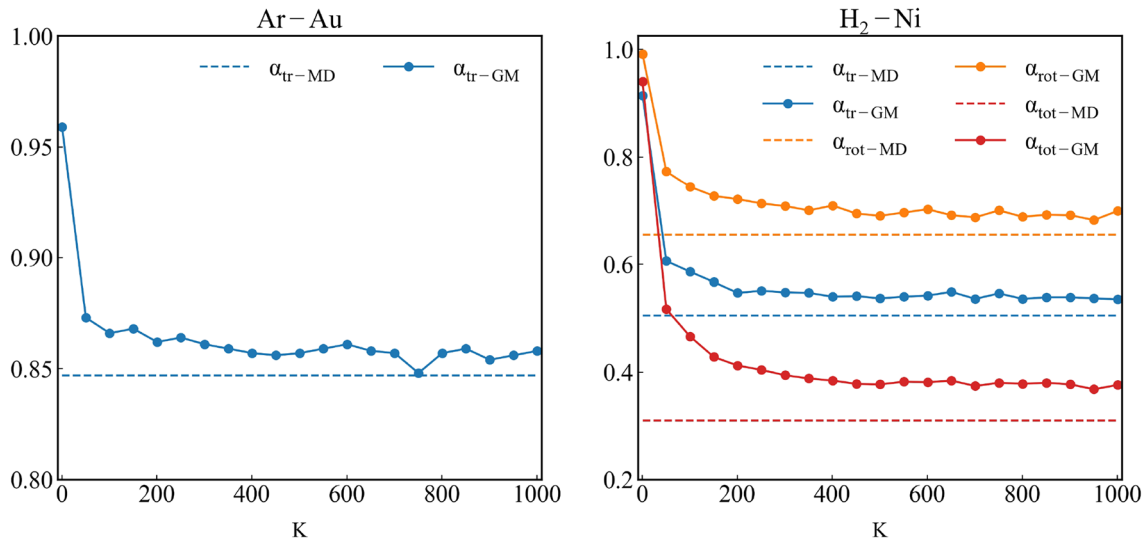
Fig. 14 Accommodation coefficients related to the translational energy (α_{tr}), rotational energy (α_{rot}), and total energy (α_{tot}) computed based on MD simulations and the GM scattering model using different number of Gaussian functions K

as the number of Gaussian functions in the GM model in the case of H_2–Ni system.

Author contributions SM and FP carried out the required simulations. SM wrote the main manuscript. AF and SN were the main supervisors. SM, FP, AF, and SN contributed in the analysis and interpretation of the obtained results from the simulations. All authors reviewed the final manuscript.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

Allen MP, Tildesley DJ (2017) Computer simulation of liquids. Oxford University Press, Oxford

Andric N, Meyer DW, Jenny P (2019) Data-based modeling of gas–surface interaction in rarefied gas flow simulations. Phys Fluids 31(6):067109

Atkins P, De Paula J (2011) Physical chemistry for the life sciences. Oxford University Press, Oxford

Bird GA, Brady J (1994) Molecular gas dynamics and the direct simulation of gas flows, vol 5. Clarendon press, Oxford

Borgnakke C, Larsen PS (1975) Statistical collision model for Monte Carlo simulation of polyatomic gas mixture. J Comput Phys 18(4):405–420

Boyd I, Beattie D, Cappelli M (1994) Numerical and experimental investigations of low-density supersonic jets of hydrogen. J Fluid Mech 280:41–67

Cercignani C, Lampis M (1971) Kinetic models for gas–surface interactions. Transp Theory Stat Phys 1(2):101–114

Chen S, Doolen GD (1998) Lattice Boltzmann method for fluid flows. Annu Rev Fluid Mech 30(1):329–364

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the em algorithm. J R Stat Soc Ser B (Methodol) 39(1):1–22

Devienne F, Souquet J, Roustan J (1965) Study of the scattering of high energy molecules by various surfaces. Rarefied Gas Dyn 2(2):584

Epstein M (1967) A model of the wall boundary condition in kinetic theory. AIAA J 5(10):1797–1800

Foiles S, Baskes M, Daw MS (1986) Embedded-atom-method functions for the fcc metals Cu, Ag, Au, Ni, Pd, Pt, and their alloys. Phys Rev B 33(12):7983

Frezzotti A (1999) Monte Carlo simulation of the heat flow in a dense hard sphere gas. Eur J Mech B Fluids 18(1):103–119

Frezzotti A, Gibelli L (2008) A kinetic model for fluid–wall interaction. Proc Inst Mech Eng Part C J Mech Eng Sci 222(5):787–795

Gregory JC, Peters PN (1986) A measurement of the angular distribution of 5 eV atomic oxygen scattered off a solid surface in earth orbit. In: International symposium on rarefied gas dynamics

Hossein Gorji M, Jenny P (2014) A gas–surface interaction kernel for diatomic rarefied gas flows based on the Cercignani–Lampis–Lord model. Phys Fluids 26(12):122004

Karniadakis G, Beskok A, Aluru N (2006) Microflows and nanoflows: fundamentals and simulation, vol 29. Springer, New York

Liang T, Ye W (2014) An efficient hybrid DSMC/MD algorithm for accurate modeling of micro gas flows. Commun Comput Phys 15(1):246–264

Liang T, Li Q, Ye W (2013) Performance evaluation of Maxwell and Cercignani–Lampis gas–wall interaction models in the modeling of thermally driven rarefied gas transport. Phys Rev E 88(1):013009

Liao M, To Q-D, Léonard C, Monchiet V (2018) Non-parametric wall model and methods of identifying boundary conditions for moments in gas flow equations. Phys Fluids 30(3):032008

Liao M, To Q-D, Léonard C, Yang W (2018) Prediction of thermal conductance and friction coefficients at a solid–gas interface from statistical learning of collisions. Phys Rev E 98(4):042104

Liu W, Zhang J, Jiang Y, Chen L, Lee C-H (2021) DSMC study of hypersonic rarefied flow using the Cercignani–Lampis–Lord model and a molecular-dynamics-based scattering database. Phys Fluids 33(7):072003

Longshaw S, Pillai R, Gibelli L, Emerson D, Lockerby D (2020) Coupling molecular dynamics and direct simulation Monte Carlo using a general and high-performance code coupling library. Comput Fluids 213:104726

Lord R (1989) Application of the Cercignani–Lampis scattering kernel to direct simulation Monte Carlo calculations. In: Rarefied gas dynamics: 17th international symposium on rarefied gas dynamics, pp 1427–1433

Lord R (1991) Some extensions to the Cercignani–Lampis gas–surface scattering kernel. Phys Fluids A 3(4):706–710

Markvoort AJ, Hilbers P, Nedea S (2005) Molecular dynamics study of the influence of wall–gas interactions on heat flow in nanochannels. Phys Rev E 71(6):066702

Maxwell JC (1878) III. On stresses in rarefied gases arising from inequalities of temperature. Proc R Soc Lond 27(185–189):304–308

Mohammad Nejad S, Iype E, Nedea S, Frijns A, Smeulders D (2021) Modeling rarefied gas–solid surface interactions for Couette flow with different wall temperatures using an unsupervised machine learning technique. Phys Rev E 104(1):015309

Mohammad Nejad S, Gaastra-Nedea S, Frijns A, Smeulders D (2022) Development of a scattering model for diatomic gas–solid surface interactions by an unsupervised machine learning approach. Phys Fluids 34:117122

Nedea SV, Frijns AJH, Van Steenhoven AA, Markvoort AJ, Hilbers PAJ (2005) Hybrid method coupling molecular dynamics and Monte Carlo simulations to study the properties of gases in microchannels and nanochannels. Phys Rev E 72(1):016705

Plimpton S (1995) Fast parallel algorithms for short-range molecular dynamics. J Comput Phys 117(1):1–19

Shen C (2006) Rarefied gas dynamics: fundamentals, simulations and micro flows. Springer, Berlin

Sheng H, Kramer M, Cadien A, Fujita T, Chen M (2011) Highly optimized embedded-atom-method potentials for fourteen fcc metals. Phys Rev B 83(13):134118

Spijker P, Markvoort AJ, Nedea SV, Hilbers PAJ (2010) Computation of accommodation coefficients and the use of velocity correlation profiles in molecular dynamics simulations. Phys Rev E 81(1):011203

Sun H (1998) Compass: an ab initio force-field optimized for condensed-phase applications overview with details on alkane and benzene compounds. J Phys Chem B 102(38):7338–7364

Sung HG (2004) Gaussian mixture regression and classification. Rice University

Wang Z, Song C, Qin F, Luo X (2021) Establishing a data-based scattering kernel model for gas–solid interaction by molecular dynamics simulation. J Fluid Mech 928:A34

Watvisave DS, Puranik BP, Bhandarkar UV (2015) A hybrid MD-DSMC coupling method to investigate flow characteristics of micro-devices. J Comput Phys 302:603–617

Weller HG, Tabor G, Jasak H, Fureby C (1998) A tensorial approach to computational continuum mechanics using object-oriented techniques. Compu Phys 12(6):620–631

White C, Borg MK, Scanlon TJ, Longshaw SM, John B, Emerson DR, Reese JM (2018) dsmcfoam+: an openfoam based direct simulation Monte Carlo solver. Comput Phys Commun 224:22–43

Williams CK, Rasmussen CE (2006) Gaussian processes for machine learning, vol 2. MIT Press, Cambridge

Wu H, Chen W, Jiang Z (2022) Gaussian mixture models for diatomic gas–surface interactions under thermal non-equilibrium conditions. Phys Fluids 34(8):082007

Yakunchikov A, Kovalev V, Utyuzhnikov S (2012) Analysis of gas–surface scattering models based on computational molecular dynamics. Chem Phys Lett 554:225–230

Yamamoto K, Takeuchi H, Hyakutake T (2006) Characteristics of reflected gas molecules at a solid surface. Phys Fluids 18(4):046103

Yamamoto K, Takeuchi H, Hyakutake T (2007) Scattering properties and scattering kernel based on the molecular dynamics analysis of gas–wall interaction. Phys Fluids 19(8):087102

Yamanishi N, Matsumoto Y, Shobatake K (1999) Multistage gas–surface interaction model for the direct simulation Monte Carlo method. Phys Fluids 11(11):3540–3552

Zhang W-M, Meng G, Wei X (2012) A review on slip models for gas microflows. Microfluid Nanofluid 13(6):845–882