A HIGHER ORDER MOMENT PRESERVING REDUCTION SCHEME FOR THE STOCHASTIC WEIGHTED PARTICLE METHOD *

SONAM LAMA†, JOHN ZWECK†, AND MATTHEW GOECKNER‡

Abstract. The Stochastic Weighted Particle Method (SWPM) is a Monte Carlo technique developed by Rjasanow and Wagner that generalizes Bird's Direct Simulation Monte Carlo (DSMC) method for solving the Boltzmann equation. To reduce computational cost due to the gradual increase in the number of stochastic particles in the SWPM, Rjasanow and Wagner proposed several particle reduction schemes designed to preserve specified moments of the velocity distribution. Here, we introduce an improved particle reduction scheme that preserves all moments of the velocity distribution up to the second order, as well as the raw and central heat flux both within each group of particles to be reduced and for the entire system. Furthermore, we demonstrate that with the new reduction scheme the scalar fourth-order moment can be computed more accurately at a reduced computational cost.

Key words. Boltzmann equation, stochastic weighted particle method, deterministic particle reduction, higher order moments

AMS subject classifications. 65C05, 65Z05, 76P05, 82C08

1. Introduction.

A fundamental problem in the computational modeling of rarefied gases and plasmas is to determine the velocity probability density function (pdf) of each particle species. The evolution of these velocity pdfs is governed by the Boltzmann equation, which models particle transport and collision processes [3, 4, 5]. Both deterministic and stochastic particle methods are used to solve the Boltzmann equation. Although deterministic methods avoid the uncertainties inherent in stochastic approaches, the cost of computing the Boltzmann collision operator can still be prohibitively high, especially in the low probability tails of the pdf. However, recent theoretical advances in combination with increased computational power have led to the introduction of several promising deterministic spectral methods [7, 8, 9, 10, 15, 16, 24]. For example, Gamba and Rjasanow recently proposed a Petrov-Gelerkin method whose computational efficiency is comparable to that of stochastic methods [10]. Despite the recent reduction of their computational cost, deterministic methods are not as flexible as stochastic methods for the modeling of the diverse range of collision, transport, and boundary surface phenomena, and particle gain and loss mechanisms that occur in experimental settings [11, 13].

Accurate modeling of the low probability tails of the velocity distribution is also of interest to experimentalists. For example, reaction rates in plasmas are determined by the overlap between the electron velocity pdf and the electron-impact cross sections of the various species. Therefore, accurate calculation of the low probability tails of the electron velocity pdf is critical. If the plasma is in thermal equilibrium, the electron velocity pdf can often be assumed to be Maxwellian. However, experimental results demonstrate that the Maxwellian assumption is often invalid [1, 6, 21, 22, 23], especially for pulsed plasmas where the velocity pdf may depend strongly on both spatial position and on time [17]. Consequently, there is still a pressing need for

*Submitted to the editors DATE.

†Department of Mathematics, The University of Texas at Dallas, Richardson, TX (sonam.lama@utdallas.edu, zweck@utdallas.edu).

‡Department of Physics, The University of Texas at Dallas, Richardson, TX (goeckner@utdallas.edu).

improved stochastic particle methods that have both greater computational efficiency and higher accuracy, especially in the higher order moments and in the low probability tails of the distributions.

Unlike deterministic methods, particle methods are not based on solving the Boltzmann equation directly. Rather they simulate a real system using stochastic particles, each of which represents a group of physical particles that are in close proximity in phase space. Collisions between stochastic particles are designed so as to approximate the collision processes modeled by the Boltzmann equation. Modern particle methods are based on the Direct Simulation Monte Carlo (DSMC) method which was developed by Bird [2]. A convergence proof for this method was given by Wagner [26]. The DSMC method has many computational advantages over deterministic methods. However, the computational cost of accurately computing the low probability tails is still very high. To resolve the low probability tails with relatively low computational cost, Rjasanow and Wagner introduced a generalization of the DSMC method which they called the Stochastic Weighted Particle Method (SWPM).

One of the challenges for the SWPM is that the number of stochastic particles gradually increases over the course of the simulation. To reduce the computational cost, Rjasanow and Wagner proposed to use a particle reduction scheme in combination with a clustering technique [18, 20]. With these methods, the particles are partitioned into groups such that the particles are close together, and each group is replaced by a small number of particles. A reduction scheme that does not require clustering was proposed by Vikhansky and Kraft [25]. Their reduction scheme redistributes the statistical weights of the particles so as to conserve the mass, momentum and energy of the ensemble. They argue that the efficiency of a particle reduction scheme that relies on clustering primarily depends on the computational cost of the clustering algorithm. For the clustering algorithm used for the results in this paper, the computational cost scales linearly with the number of computational particles. In this context, it is important to note that the convergence theorem for the SWPM obtained by Rjasanow and Wagner [20] requires that the maximum diameter of the groups of particles converges to zero as the initial number of computational particles increases. As a result, there is a theoretical advantage to employing a clustering technique.

The reduction schemes proposed by Rjasanow and Wagner were designed to preserve a specified set of moments of the distribution. The particle reduction scheme of Rjasanow and Wagner that preserves the most moments is a deterministic reduction scheme that preserves the total weight, momentum, energy and central heat flux within each group [18, 20]. The total weight corresponds to the fraction of physical particles represented by the group. Although this reduction scheme preserves the central heat flux of each group, it does not preserve the raw heat flux, and consequently neither the raw nor the central heat flux are preserved for the entire system.

In this paper, we improve upon the reduction scheme of Rjasanow and Wagner by conserving all of the moments up to the second order (i.e. the full pressure and momentum flux tensors), as well as both the raw and central heat flux, which are third order moments. Conservation of all these moments within each group automatically guarantees that they are conserved for the entire system.

We performed two series of simulation studies to evaluate the degree to which our new deterministic particle reduction scheme improves upon that of Rjasanow and Wagner's deterministic reduction schemes. First, we present results which confirm that the existing scheme of Rjasanow and Wagner does not conserve the raw heat flux within each group, while our new scheme conserves both raw and central heat

flux of each group. Second, we study the convergence rate of the SWPM with the new and existing reduction schemes. In particular, we will present results showing the rate at which the scalar fourth order moment converges to its true value as the number of stochastic particles increases. We compare the results of our new reduction scheme with the existing deterministic schemes to show that our scheme requires a fewer initial number of computational particles and less computational time for the convergence of the scalar fourth order moment compared to the existing reduction schemes.

In section 2, we review the stochastic weighted particle method, and in section 3, we discuss the reduction schemes of Rjasanow and Wagner and introduce our new reduction scheme. In section 4, we briefly show that the assumptions in Wagner's convergence theorem hold for our new particle reduction scheme. In section 5, we present our numerical results, and finally in section 6 we make some conclusions.

2. The stochastic weighted particle method. In this section, we review the stochastic weighted particle method for the spatially homogeneous Boltzmann equation. The stochastic weighted particle method is a particle method [19] that improves upon Bird's DSMC method [2] by decreasing the uncertainty in the computation of rare events. In Bird's method, each stochastic particle represents the same number of physical particles, and the number of stochastic particles is kept constant throughout the simulation. With the SWPM, the number of physical particles represented by a single stochastic particle varies over the course of the simulation. Each stochastic particle represents a group of physical particles that are in close proximity in phase space. Each stochastic particle is characterized by its velocity and weight. The weight quantifies the proportion of physical particles represented by the given stochastic particle. The SWPM is based on a generalized version of the collision process used in the DSMC method in which only the physical particles corresponding to some portion of the weights of the colliding stochastic particles undergo collisions. For each stochastic collision, this results in the creation of two new stochastic particles whose velocities are given by the post-collision velocities and whose weights quantify the proportion of physical particles involved in the collision process [18, 20]. The weights of the original pair of colliding stochastic particles are reduced so as to keep the total weight of the system constant. As the number of stochastic particles increases due to collisions, the number of high velocity particles in the low probability tails of the velocity pdf also increases. By periodically applying a clustering technique and a particle reduction scheme, the proportion of particles in the center of the distribution is reduced. The combined effect of these processes is to increase the fraction of stochastic particles occupying the low probability tails of the velocity pdf, which decreases the statistical uncertainty in the tails.

We consider the spatially homogeneous Boltzmann equation for a single species of particles with unit mass. This equation, which describes the evolution of the velocity probability density function (pdf), f, due to collisions, is given by

$$\frac{\partial f}{\partial t}(\mathbf{v},t) \;=\; \int_{\mathbb{R}^3}\int_{S^2} B(\mathbf{v},\mathbf{w},\mathbf{\Theta})\left[\left(f(\mathbf{v}',t)\,f(\mathbf{w}',t)\;-\;f(\mathbf{v},t)\,f(\mathbf{w},t)\right]\right. d\mathbf{\Theta}\,d\mathbf{w}, \tag{2.1}$$

with an initial condition of the form

$$f_0(\mathbf{v}) \;=\; f(\mathbf{v},0). \tag{2.2}$$

Here, S^2 denotes the unit sphere, B is the collision kernel, t is time, $\mathbf{v}$ and $\mathbf{w}$ are the pre-collision velocities, and $\mathbf{v}'$ and $\mathbf{w}'$ are the post-collision velocities. For simplicity,

for the results in this paper we consider isotropic Maxwell type interactions, for which

$$B(\mathbf{v}, \mathbf{w}, \mathbf{\Theta}) \;=\; \frac{1}{4\pi}. \tag{2.3}$$

Assuming that the collisions are elastic, the post-collision velocities are given in terms of the pre-collision velocities and the direction vector, $\mathbf{\Theta}$, by

$$\mathbf{v}' \;=\; \frac{1}{2}\left[\mathbf{v} + \mathbf{w} - \mathbf{\Theta}\,|\mathbf{w} - \mathbf{v}|\,\right] \quad \text{and} \quad \mathbf{w}' \;=\; \frac{1}{2}\left[\mathbf{v} + \mathbf{w} + \mathbf{\Theta}\,|\mathbf{w} - \mathbf{v}|\,\right]. \tag{2.4}$$

The state of the ith stochastic particle is given by $(\mathbf{v}_i, g_i)$, where $\mathbf{v}_i$ and g_i are the velocity and weight, respectively. The state of the entire stochastic system is

$$z \;=\; \{(g_1, \mathbf{v}_1), (g_2, \mathbf{v}_2), \ldots, (g_m, \mathbf{v}_m)\}, \tag{2.5}$$

where m is the current number of stochastic particles. To model a collision between the stochastic particles indexed by i and j, we introduce the weight transfer function, $\gamma_{\text{coll}}(z; i, j)$. This function encodes the proportion of physical particles represented by the stochastic particles indexed by i and j that undergo collisions when the state of the system is z. The weight transfer function cannot exceed the minimum of the weights of the colliding particles,

$$0 \;\le\; \gamma_{\text{coll}}(z; i, j) \;\le\; \min(g_i, g_j). \tag{2.6}$$

During a collision between the stochastic particles indexed by i and j, only the fraction of physical particles in the system represented by the weight $\gamma_{\text{coll}}(z; i, j)$ undergo collisions. This process is modeled by adding one or two new stochastic particles to the system. For the results in this paper, we use $\gamma_{\text{coll}}(z; i, j) = \frac{1}{2}\min(g_i, g_j)$, which always results in two new stochastic particles. In this case, the state, $[J_{\text{coll}}(z; i, j, \mathbf{\Theta})]_k$, of the k-th stochastic particle after a collision between particles i and j is given by [18, 20],

$$[J_{\text{coll}}(z; i, j, \mathbf{\Theta})]_k \;=\; \begin{cases} (\mathbf{v}_k, g_k), & \text{if } k \le m,\ k \notin \{i, j\}, \\ (\mathbf{v}_i, g_i - \gamma_{\text{coll}}(z; i, j)), & \text{if } k = i, \\ (\mathbf{v}_j, g_j - \gamma_{\text{coll}}(z; i, j)), & \text{if } k = j, \\ (\mathbf{v}_i', \gamma_{\text{coll}}(z; i, j)), & \text{if } k = m+1, \\ (\mathbf{v}_j', \gamma_{\text{coll}}(z; i, j)), & \text{if } k = m+2, \end{cases} \tag{2.7}$$

resulting in a new system state,

$$z \;=\; \{(g_1, \mathbf{v}_1), (g_2, \mathbf{v}_2), \ldots, (g_{m+1}, \mathbf{v}_{m+1}), (g_{m+2}, \mathbf{v}_{m+2})\}. \tag{2.8}$$

After the collision, the fraction of physical particles corresponding to the weight $\gamma_{\text{coll}}(z; i, j)$ are assigned the post-collision velocities, and the remaining fraction of particles is unchanged. The two new stochastic particles are indexed by $m+1$ and $m+2$. To keep the total weight constant this weight is subtracted from the weights of the colliding stochastic particles, indexed by i and j. For elastic collisions given by (2.4), this stochastic collision process conserves the total weight, momentum and energy.

To correctly model the evolution of the velocity pdf, we must relate the collision frequency for the stochastic system to that of the physical system. The total collision frequency in the physical system is given by

$$\nu \;=\; \int_{\mathbb{R}^3}\int_{\mathbb{R}^3}\int_{S^2} B(\mathbf{v}, \mathbf{w}, \mathbf{\Theta})\, f(\mathbf{v}, t)\, f(\mathbf{w}, t)\, d\mathbf{\Theta}\, d\mathbf{w}\, d\mathbf{v}. \tag{2.9}$$

If we let $\nu_{g_i g_j}$ denote the frequency of collisions between the physical particles that correspond to stochastic particles with states $(\mathbf{v}_i, g_i)$ and $(\mathbf{v}_j, g_j)$, then by (2.9) we obtain

$$\nu_{g_i g_j} \;=\; g_i\, g_j \int_{S^2} B(\mathbf{v}_i, \mathbf{v}_j, \boldsymbol{\Theta})\, d\boldsymbol{\Theta}. \tag{2.10}$$

Furthermore, if we let $\widetilde{\nu}_{ij}$ denote the frequency of collisions between particles i and j in the stochastic system, then by the definition of the weight transfer function, we have that

$$\widetilde{\nu}_{ij}\, \gamma_{\text{coll}}(z; i, j) \;=\; \nu_{g_i g_j}. \tag{2.11}$$

Therefore by (2.10), we obtain

$$\widetilde{\nu}_{ij} \;=\; \frac{g_i\, g_j}{\gamma_{\text{coll}}(z; i, j)} \int_{S^2} B(\mathbf{v}_i, \mathbf{v}_j, \boldsymbol{\Theta})\, d\boldsymbol{\Theta}, \tag{2.12}$$

and so, by (2.9) and (2.12), the total collision frequency in the stochastic system is given by

$$\widetilde{\nu}(z) \;=\; \frac{1}{2} \sum_{i=1}^{m} \sum_{\substack{j=1 \\ j \neq i}}^{m} \frac{g_i\ g_j}{\gamma_{\text{coll}}(z; i, j)} \int_{\boldsymbol{\Theta} \in \mathbf{S^2}} B(\mathbf{v}_i, \mathbf{v}_j, \boldsymbol{\Theta})\, d\boldsymbol{\Theta}. \tag{2.13}$$

Using this frequency, we can obtain the waiting time between stochastic collisions. Since it is memoryless, this waiting time is a Poisson process which follows an exponential distribution. Therefore, the probability that a collision did not occur by time, t, is given by the survival function, $P(s > t) = e^{-\widetilde{\nu}(z)t}$. Since the survival function has an uniform distribution on $[0, 1]$, the time between collisions is given by $\Delta t = -\ln(r)/\widetilde{\nu}(z)$, where r is a random number uniformly distributed on $[0, 1]$. Once the time interval between collisions is calculated, the time counter is updated.

The probability, $p(z; i, j)$, of a collision between the stochastic particles i and j is given by the ratio of the frequency $\widetilde{\nu}_{ij}$ of collisions between stochastic particles i and j given in (2.12) and the total collision frequency given in (2.13), that is,

$$p(z; k, l) \;=\; \frac{\frac{g_k\ g_l}{\gamma_{\text{coll}}(z; k, l)} \displaystyle\int_{\boldsymbol{\Theta} \in \mathbf{S^2}} B(\mathbf{v}_k, \mathbf{v}_l, \boldsymbol{\Theta})\, d\boldsymbol{\Theta}}{\displaystyle\sum_{i=1}^{m} \sum_{\substack{j=1 \\ j \neq i}}^{m} \frac{g_i\ g_j}{\gamma_{\text{coll}}(z; i, j)} \int_{\boldsymbol{\Theta} \in \mathbf{S^2}} B(\mathbf{v}_i, \mathbf{v}_j, \boldsymbol{\Theta})\, d\boldsymbol{\Theta}}. \tag{2.14}$$

Once a pair of colliding stochastic particles, k and l, has been randomly selected, the direction vector $\boldsymbol{\Theta}$ is chosen using the probability density function

$$\eta(\boldsymbol{\Theta}) = \frac{B(\mathbf{v}_k, \mathbf{v}_l, \boldsymbol{\Theta})}{\displaystyle\int_{\widetilde{\boldsymbol{\Theta}} \in S^2} B(\mathbf{v}_k, \mathbf{v}_l, \widetilde{\boldsymbol{\Theta}})\ d\widetilde{\boldsymbol{\Theta}}}. \tag{2.15}$$

In the case of the constant collision kernel given by (2.3), the probability in (2.14) further simplifies to

$$p(z; k, l) \;=\; \frac{\frac{g_k\ g_l}{\gamma_{\text{coll}}(z; k, l)}}{\displaystyle\sum_{i=1}^{m} \sum_{\substack{j=1 \\ j \neq i}}^{m} \frac{g_i\ g_j}{\gamma_{\text{coll}}(z; i, j)}}, \tag{2.16}$$

and $\eta(\mathbf{\Theta}) = \frac{1}{4\,\pi}$. After the colliding pair of stochastic particles and the direction vector have been chosen, the velocities and weights of the colliding particles are updated using (2.7).

The computation of the collision frequency in (2.13) and collision probability in (2.14) can be computationally expensive since in general these quantities need to be updated after each collision. This issue also arises for the DSMC method when the collision kernel is not constant. To overcome this computational issue, the technique of null collisions was developed by Koura [12] for the DSMC method. With this technique, an equal maximum collision frequency is assigned to all pairs of particles, which leads to an equal probability of collision for all pairs. Consequently the colliding pair can be selected at random from a uniform distribution. Once a pair is chosen, we decide whether the collision is an actual one or a null collision based on the probability given by the ratio between the actual collision frequency and the assigned equal collision frequency. Rjasanow and Wagner generalized the technique of the null collisions to the SWPM [18, 20].

3. A reduction scheme conserving total weight, momentum, pressure tensor and heat flux. As we explained in section 1, one of the challenges for the SWPM is that the number of stochastic particles gradually increases. For computational feasibility, it is necessary to periodically reduce the number of particles. There are two steps in the reduction process. First, the stochastic particles need to be clustered into groups, and then each group of particles needs to be replaced by a small number of particles.

A number of clustering techniques have been proposed by Rjasanow, Wagner and their collaborators [14, 18, 20]. One of these techniques is based on partitioning particles into two groups with a cutting plane whose normal vector is in the direction of the eigenvector corresponding to the largest eigenvalue of the covariance matrix of the particles [18, 20]. This partitioning method is performed iteratively on each of the partitioned groups using the group's covariance matrix. The iteration continues until the product of the total weight and the standard deviation of the particle speeds within each group is minimized, which results in a roughly uniform number of stochastic particles in each group. We use this clustering method for the results in this paper. Rjasanow and Wagner also proposed several stochastic and deterministic particle reduction schemes to replace each group by a group with a small number of particles. These schemes are based on conserving a specific set of moments of the distribution within each group. The details of these reduction schemes can be found in [18, 20].

We are interested in deterministic reduction schemes that conserve as many moments as possible, so that the structure of the velocity pdf is preserved. In this paper, we propose a particle reduction scheme that conserves all the moments of the velocity pdf up to second order, given in Table 1, together with the raw and central heat flux, which are the most physically relevant third order moments. The raw heat flux vector, $\mathbf{h}$, is computed relative to the origin, while the central heat flux, $\mathbf{q}$, is relative to the drift velocity, $\mathbf{V}$. They are given by

$$\mathbf{h} \;=\; \frac{1}{2}\sum_{i=1}^{m} g_i\,\mathbf{v}_i\,|\mathbf{v}_i|^2 \quad \text{and} \quad \mathbf{q} \;=\; \frac{1}{2}\sum_{i=1}^{m} g_i\,(\mathbf{v}_i-\mathbf{V})\,|\mathbf{v}_i-\mathbf{V}|^2. \tag{3.1}$$

In the following discussion, when we refer to third order moments we simply mean the raw and central heat flux.

The reduction scheme of Rjasanow and Wagner that preserves the most moments and is closest to our scheme is the one that preserves the total weight, momentum,

TABLE 1
Moments of the velocity pdf

Moment order	Raw Moment	Symbol	Central Moment	Symbol
Zero	Total Weight	ϱ	—	—
First	Momentum	$\varrho\,\mathbf{V}$	—	—
Second	Momentum Flux Tensor	Π	Pressure Tensor	P

energy, and central heat flux [18, 20]. With this scheme, although the central heat flux is conserved within each group, the momentum flux tensor and the pressure tensor are not. Only the total energy, which is the trace of the momentum flux tensor, is conserved. As a consequence, the raw heat flux for each group is not conserved, and therefore, the raw and central heat flux of the entire system are also not conserved.

To conserve both the raw and central moments of a group, it is necessary and sufficient to conserve either of these moments and all of the lower order moments. Because of the additive property of raw moments, if a raw moment is conserved within each group then it must also be conserved for the entire system. In particular, since the total weight and momentum are raw moments, conservation of ϱ and $\varrho\,\mathbf{V}$ within each group ensures that these two moments are conserved for the entire system. Therefore, if we could conserve the total weight, momentum, pressure tensor and central heat flux within each group, then all of the raw and central moments up to the second order together with the raw and central heat flux would be conserved for the entire system.

We formalize this idea as follows. Let m be the number of stochastic particles in the system, and suppose that the particles have been partitioned into $\widehat{n}$ groups with m_l stochastic particles in the l-th group, G_l. Let $g_{l,i}$ and $\mathbf{v}_{l,i}$ denote the weight and velocity of the i-th particle in the l-th group. Then, the total weight, ϱ_l, momentum, $\varrho_l\,\mathbf{V}_l$, momentum flux tensor, Π_l, and raw heat flux, $\mathbf{h}_l$, for the l-th group are given by

$$\begin{aligned} \varrho_l &= \sum_{i=1}^{m_l} g_{l,i}, & \varrho_l\,\mathbf{V}_l &= \sum_{i=1}^{m_l} g_{l,i}\,\mathbf{v}_{l,i}, \\ \Pi_l &= \sum_{i=1}^{m_l} g_{l,i}\,\mathbf{v}_{l,i}\,\mathbf{v}_{l,i}^T, & \mathbf{h}_l &= \frac{1}{2}\sum_{i=1}^{m_l} g_{l,i}\,\mathbf{v}_{l,i}\,|\mathbf{v}_{l,i}|^2, \end{aligned} \tag{3.2}$$

where $\mathbf{V}_l$ is the drift velocity of the l-th group. The pressure tensor, P_l, and the central heat flux, $\mathbf{q}_l$, of the l-th group are given by

$$P_l = \sum_{i=1}^{m_l} g_{l,i}\,(\mathbf{v}_{l,i}-\mathbf{V}_l)\,(\mathbf{v}_{l,i}-\mathbf{V}_l)^T \quad \text{and} \quad \mathbf{q}_l = \frac{1}{2}\sum_{i=1}^{m_l} g_{l,i}\,(\mathbf{v}_{l,i}-\mathbf{V}_l)\,\big|\mathbf{v}_{l,i}-\mathbf{V}_l\big|^2. \tag{3.3}$$

The energy, E_l, and temperature, T_l, are given by

$$E_l = \sum_{i=1}^{m_l} g_{l,i}\,|\mathbf{v}_{l,i}|^2, \quad \text{and} \quad 3\,\varrho_l\,T_l = \sum_{i=1}^{m_l} g_{l,i}\,|\mathbf{v}_{l,i}-\mathbf{V}_l|^2, \tag{3.4}$$

where the quantity on the right hand side of the formula for T_l is the trace of the pressure tensor. The energy is given in terms of the temperature by

$$E_l = \varrho_l\,|\mathbf{V}_l|^2 + 3\,\varrho_l\,T_l. \tag{3.5}$$

The raw moments of the entire system are given by

$$\varrho \;=\; \sum_{l=1}^{\hat{n}} \varrho_l, \quad \varrho\,\mathbf{V} \;=\; \sum_{l=1}^{\hat{n}} \varrho_l\,\mathbf{V}_l, \quad \Pi \;=\; \sum_{l=1}^{\hat{n}} \Pi_l, \quad \text{and} \quad \mathbf{h} \;=\; \sum_{l=1}^{\hat{n}} \mathbf{h}_l. \tag{3.6}$$

Here ϱ is the total weight, $\mathbf{V}$ is the drift velocity, Π is the momentum flux tensor and $\mathbf{h}$ is the raw heat flux of the entire system.

The relationship between the raw and central second order moments is given by

$$P_l \;=\; \Pi_l - \varrho_l\,\mathbf{V}_l\,\mathbf{V}_l^T. \tag{3.7}$$

Using this relationship, we observe that for a reduction scheme to preserve both of the second order moments, P_l and Π_l, it is sufficient to conserve the total weight, ϱ_l, the momentum, $\varrho_l\,\mathbf{V}_l$, and either P_l or Π_l. Since the raw moments are additive (see (3.6)), conservation of the total weight, momentum and momentum flux tensor within each group leads to the conservation of the these moments for the entire system. Using (3.7) for the entire system, we conclude that the pressure tensor for the entire system is also conserved. Therefore, a reduction scheme that conserves the total weight, momentum and either of the second order moments for each group leads to the conservation of the moments up to second order for the entire system.

Similarly, the relationship between the raw and central third order moment can be determined using (3.2), (3.4), and (3.7) giving the equation,

$$\mathbf{q}_l \;= \mathbf{h}_l - P_l\,\mathbf{V}_l - \frac{1}{2}\,\varrho_l\,\mathbf{V}_l\,\big|\mathbf{V}_l\big|^2 - \frac{3}{2}\,\varrho_l\,T_l\,\mathbf{V}_l, \tag{3.8}$$

which relates the raw and central heat flux to each other via the lower order moments. As above, to conserve both the raw and central moments of a group up to third order it is sufficient to conserve the total weight, momentum, pressure tensor, and either of the third order moments of the group. Similarly, using the additivity property of the raw moments, and the relationships between the moments given by (3.7) and (3.8) for the entire system, the pressure tensor and central heat flux of the system are also conserved together with all the raw moments. This verifies our claim that to conserve the raw and central moments of the system up to the third order during a reduction process, it is sufficient to conserve the total weight, momentum, pressure tensor and central heat flux of each group.

Next, we present a novel particle reduction scheme that conserves the total weight, momentum, pressure tensor, and central heat flux in a group. First, we outline the idea behind the conservation of these moments. Before describing this scheme, we briefly recall that for the reduction scheme that preserves the total weight and momentum of a group, we simply replace all the stochastic particles in the group by a single stochastic particle with the given weight and momentum [20]. The next higher order moments are the momentum flux tensor and the pressure tensor. Since, the pressure tensor is a 3×3 real symmetric positive semi-definite matrix, it can be diagonalized using an orthonormal basis of normalized eigenvectors, with the non-negative eigenvalues as its diagonal entries. This simplifies the problem, as we only have to conserve the diagonal entries of the pressure tensor. To conserve the pressure tensor in this new orthonormal basis, each group can be replaced by a group with between one and three pairs of particles. The number of pairs of particles depends on the number of nonzero eigenvalues. Specifically, we choose to assign an equal portion of the total weight to each pair of particles. For each pair, the velocity of one of the particles relative

to the drift velocity is chosen to be in the direction of an eigenvector with nonzero eigenvalue, while the other particle moves in the opposite direction. The magnitudes of these velocity pairs relative to the drift velocity of the group are equal, and are chosen to ensure the conservation of each of the diagonal entries of the pressure tensor in the new basis, which leads to the conservation of the pressure tensor. The total weight and momentum of the group are conserved as a consequence of the choices we made.

To additionally conserve the raw and central heat flux, we utilize a degree of freedom in the choice of weights and in the magnitudes of the velocities relative to the drift velocity. We choose the sum of the weights of the particles for each pair to be an equal portion of the total weight. For each pair, the weights of the two particles and the magnitudes of their velocities relative to the drift velocity are not required to be equal. These quantities are determined by solving the conditions required to conserve the weight, momentum, pressure tensor, and central heat flux in the new basis. Once the post reduction velocities are determined, the transformation of these velocities to the standard basis leads to the conservation of the moments in the standard basis.

The following theorem summarizes our new particle reduction scheme for the conservation of the total weight, momentum, pressure tensor, and central heat flux of a group.

THEOREM 3.1. *Let G_l be a group of stochastic particles. Suppose that the pressure tensor, P_l, has k non-zero eigenvalues, $\lambda_1, \ldots, \lambda_k$, for some $k \in \{1, 2, 3\}$, and an associated orthonormal set of eigenvectors, $\mathbf{\Theta}_1, \ldots, \mathbf{\Theta}_k$, with the direction of $\mathbf{\Theta}_i$ chosen so that $\widehat{q}_{l,i} = \mathbf{\Theta}_i^T \mathbf{q}_l > 0$. Let $\widetilde{G}_l$ be the reduced group of $2k$ stochastic particles whose weights and velocities, $(\widetilde{\mathbf{v}}_i, \widetilde{g}_i)$, for $i = 1, \ldots, 2k$ are given by*

$$
\begin{aligned}
\widetilde{\mathbf{v}}_i &= \mathbf{V}_l + \gamma_i \sqrt{\frac{k\,\lambda_i}{\varrho_l}} \mathbf{\Theta}_i, \qquad & \widetilde{g}_i &= \frac{\varrho_l}{k} \frac{1}{1+\gamma_i^2}, \\
\widetilde{\mathbf{v}}_{i+k} &= \mathbf{V}_l - \frac{1}{\gamma_i} \sqrt{\frac{k\,\lambda_i}{\varrho_l}} \mathbf{\Theta}_i, \qquad & \widetilde{g}_{i+k} &= \frac{\varrho_l}{k} \frac{\gamma_i^2}{1+\gamma_i^2}, \quad \textit{for} \quad i = 1, \ldots, k
\end{aligned}
\tag{3.9}
$$

where,

$$
\gamma_i = \frac{\sqrt{\varrho_l}\,\widehat{q}_{l,i}}{\sqrt{k}\,\lambda_i^{\frac{3}{2}}} + \sqrt{1 + \frac{\varrho_l\,\widehat{q}_{l,i}^2}{k\,\lambda_i^3}}, \quad \textit{for} \quad i = 1, \ldots, k. \tag{3.10}
$$

Then, $\widetilde{G}_l$ preserves the total weight, momentum, pressure tensor, and central heat flux of G_l, which leads to the preservation of all the moments up to the second order as well as the raw and central heat flux of G_l.

Proof. We consider the case where the eigenvalues of the pressure tensor are all nonzero. We let the reduced group, $\widetilde{G}_l$, consist of three pairs of particles, with each pair of the form,

$$
\widetilde{\mathbf{v}}_i = \mathbf{V}_l + \alpha_i \mathbf{\Theta}_i, \quad \widetilde{\mathbf{v}}_{i+3} = \mathbf{V}_l - \alpha_{i+3} \mathbf{\Theta}_i, \quad \text{and} \quad \widetilde{g}_i + \widetilde{g}_{i+3} = \frac{\varrho_l}{3}, \quad i \in \{1, 2, 3\}, \tag{3.11}
$$

for some $\alpha_i \in \mathbb{R}$ and $\mathbf{\Theta}_i \in S^2$. We derive the conditions on the unknown parameters, $\widetilde{g}_i$, α_i, and $\mathbf{\Theta}_i$, so as to conserve the total weight, momentum, pressure tensor and heat flux.

By construction, the total weight is conserved,

$$\sum_{i=1}^{6} \widetilde{g}_i \;=\; \varrho_l. \tag{3.12}$$

Similarly, if we impose the condition

$$\widetilde{g}_i\, \alpha_i \;=\; \widetilde{g}_{i+3}\, \alpha_{i+3}, \quad \text{for} \quad i \in \{1,2,3\}, \tag{3.13}$$

we find that the momentum of the group is conserved, since

$$\sum_{i=1}^{6} \widetilde{g}_i\, \widetilde{\mathbf{v}}_i \;=\; \sum_{i=1}^{3} \widetilde{g}_i(\mathbf{V}_l + \alpha_i\, \boldsymbol{\Theta}_i) + \widetilde{g}_{i+3}(\mathbf{V}_l - \alpha_{i+3}\, \boldsymbol{\Theta}_i) = \varrho_l\, \mathbf{V}_l. \tag{3.14}$$

Next, to conserve the pressure tensor, P_l, we use the fact that it is a 3×3 real symmetric matrix with positive eigenvalues. Therefore, there is a diagonal matrix $D = \operatorname{diag}[\lambda_1, \lambda_2, \lambda_3]$ and an orthonormal matrix $Q = [\boldsymbol{\Theta_1}, \boldsymbol{\Theta_2}, \boldsymbol{\Theta_3}]$ such that

$$D \;=\; Q^T\, P_l\, Q. \tag{3.15}$$

That is, each $\{\lambda_i, \boldsymbol{\Theta}_i\}$ is an eigenpair of the matrix P_l. The condition, $\widetilde{P}_l \;=\; P_l$, that the reduction scheme preserves the pressure tensor is therefore equivalent to the condition

$$\begin{aligned} D \;&= Q^T \Big[\sum_{i=1}^{6} \widetilde{g}_i \left(\widetilde{\mathbf{v}}_i - \mathbf{V}_l\right)\left(\widetilde{\mathbf{v}}_i - \mathbf{V}_l\right)^T \Big] Q \\ &= \sum_{i=1}^{3} \left(\widetilde{g}_i\, {\alpha_i}^2 + \widetilde{g}_{i+3}\, \alpha_{i+3}^2\right) \left(Q^T\, \boldsymbol{\Theta}_i\right) \left(Q^T\, \boldsymbol{\Theta}_i\right)^T. \end{aligned} \tag{3.16}$$

Therefore, to conserve the pressure tensor, we require that

$$\widetilde{g}_i\, {\alpha_i}^2 + \widetilde{g}_{i+3}\, \alpha_{i+3}^2 \;=\; \lambda_i, \quad \text{for} \quad i \in \{1,2,3\}. \tag{3.17}$$

In the basis of eigenvectors, the central heat flux, $\widehat{\mathbf{q}}_l$, is given by

$$\widehat{\mathbf{q}}_l \;=\; Q^T\, \mathbf{q}_l. \tag{3.18}$$

As in the statement of the theorem, we choose the direction of $\boldsymbol{\Theta}_i$ so that the i-th component, $\widehat{q}_{l,i} \;=\; \boldsymbol{\Theta}_i^T\, \mathbf{q}_l$, of $\widehat{\mathbf{q}}_l$ is positive. To conserve the central heat flux in the new basis, we have

$$\widehat{\mathbf{q}}_l \;=\; \frac{1}{2} \sum_{i=1}^{3} \left(\widetilde{g}_i\, \alpha_i^3 - \widetilde{g}_{i+3}\, \alpha_{i+3}^3\right) Q^T \boldsymbol{\Theta}_i \;=\; \frac{1}{2} \sum_{i=1}^{3} \left(\widetilde{g}_i\, \alpha_i^3 - \widetilde{g}_{i+3}\, \alpha_{i+3}^3\right) \mathbf{e}_i, \tag{3.19}$$

and we obtain

$$\widehat{q}_{l,i} \;=\; \frac{1}{2} \big[\widetilde{g}_i\, \alpha_i^3 - \widetilde{g}_{i+3}\, \alpha_{i+3}^3\big]. \tag{3.20}$$

Next, to solve for α_i and $\widetilde{g}_i$, we apply a technique used by Rjasanow and Wagner [18, 20] . We introduce a new parameter, γ_i, and express α_i as

$$\alpha_i \;=\; \gamma_i \sqrt{\frac{3\, \lambda_i}{\varrho_l}}. \tag{3.21}$$

Substituting (3.13) into (3.17) we obtain

$$\lambda_i \;=\; \frac{\widetilde{g}_i}{\widetilde{g}_{i+3}}\,\alpha_i^2\left[\widetilde{g}_{i+3}+\widetilde{g}_i\right]. \tag{3.22}$$

Substituting the expression for α_i given by (3.21), and using (3.11) for the sum of weights, we obtain

$$\frac{\widetilde{g}_i}{\widetilde{g}_{i+3}}\,\gamma_i^2 \;=\; 1. \tag{3.23}$$

Using this relationship in (3.11), we obtain

$$\widetilde{g}_i \;=\; \frac{\varrho_l}{3}\,\frac{1}{1+\gamma_i^2},\quad \alpha_i \;=\; \gamma_i\sqrt{\frac{3\,\lambda_i}{\varrho_l}},\quad \text{and}\quad \widetilde{g}_{i+3} \;=\; \frac{\varrho_l}{3}\,\frac{\gamma_i^2}{1+\gamma_i^2},\quad \alpha_{i+3} \;=\; \frac{1}{\gamma_i}\sqrt{\frac{3\,\lambda_i}{\varrho_l}}. \tag{3.24}$$

To determine γ_i, we substitute (3.24) into (3.20) to obtain

$$\widetilde{g}_i\,\alpha_i^3-\widetilde{g}_{i+3}\,\alpha_{i+3}^3 \;=\; \sqrt{\frac{3}{\varrho_l}}\,\frac{\lambda_i^{\frac{3}{2}}}{\gamma_i}\,\big(\gamma_i^2-1\big). \tag{3.25}$$

Now by (3.20), $\widehat{q}_{l,i} \;=\; \frac{1}{2}\big[\widetilde{g}_i\,\alpha_i^3-\widetilde{g}_{i+3}\,\alpha_{i+3}^3\big] > 0$. Therefore $\gamma_i > 1$ and

$$\widehat{q}_{l,i} \;=\; \frac{1}{2}\sqrt{\frac{3}{\varrho_l}}\,\frac{\lambda_i^{\frac{3}{2}}}{\gamma_i}\,\big(\gamma_i^2-1\big) \;\;\Longrightarrow\;\; \gamma_i^2-2\,\frac{\sqrt{\varrho_l}\,\widetilde{q}_i}{\sqrt{3}\,\lambda_i^{\frac{3}{2}}}\,\gamma_i-1 \;=\; 0. \tag{3.26}$$

Solving for the positive root, we obtain

$$\gamma_i \;=\; \frac{\sqrt{\varrho_l}\,\widehat{q}_{l,i}}{\sqrt{3}\,\lambda_i^{\frac{3}{2}}}+\sqrt{1+\frac{\varrho_l\,\widehat{q}_{l,i}^2}{3\,\lambda_i^3}}. \tag{3.27}$$

Therefore, the post reduction particles are given by (3.9) and (3.10) as required.

If the pressure tensor has at least one zero eigenvalue, the moments can be conserved with fewer than six particles. The reason is that there is no need to introduce particles whose heat flux is in the direction of the eigenvectors corresponding to the zero eigenvalues. In this situation, the result follows similarly to the calculations above. ∎

4. Theoretical convergence of SWPM with the new reduction scheme. In this section, we show that our new reduction scheme satisfies the assumptions in Wagner's convergence theorem for the SWPM [20, Thm. (3.22)]. This theorem provides a collection of assumptions which guarantee that the sequence of empirical measures of the Markov process produced by the SWPM converges to the weak solution of the Boltzmann equation as $n \to \infty$. These assumptions on the reduction scheme are given by [20, eq. (3.162)], and [20, eq. (3.164)]. According to Rjasanow and Wagner, assumption [20, eq. (3.162)] assures that the reduction is sufficiently precise, and [20, eq. (3.164)] restricts the increase in energy during reduction. Since the energy is conserved in our new reduction scheme, the second assumption related to the energy is satisfied. For assumption [20, eq. (3.162)], the arguments given by Rjasanaw and Wagner for their deterministic reduction schemes also apply to our new deterministic reduction scheme. Therefore, for this assumption to hold for our new

reduction scheme, it is sufficient to show that the inequality given by [20, eq. (3.273)] holds. This inequality states that

$$\left| \int_{Z_l} \Phi(\widetilde{z}_l)\, p_{\mathrm{red}}(z_l; d\widetilde{z}_l) - \Phi(z_l) \right| \;\leq\; ||\varphi||_L \left[\sum_{i=1}^{m_l} g_i \left| \mathbf{V}_l - \mathbf{v}_i \right| + \varrho_l \sqrt{3\, T_l} \right]. \tag{4.1}$$

Here,

$$z_l \;=\; \{(g_1, \mathbf{v}_1), (g_2, \mathbf{v}_2), \ldots, (g_{m_l}, \mathbf{v}_{m_l})\} \tag{4.2}$$

is the state of a group G_l prior to reduction, $\widetilde{z}_l$ is the post-reduction state, and $p_{\mathrm{red}}(z_l; d\widetilde{z}_l)$ is a measure that gives the probability that the post-reduction state lie in the volume element, $d\widetilde{z}_l$. The function Φ, which approximates the velocity pdf, is given by

$$\Phi(z_l) \;=\; \sum_{i=1}^{m_l} g_i\, \varphi(\mathbf{v}_i), \tag{4.3}$$

for the particles in the group G_l. Here φ is an arbitrary test function. The norm for the test function, $||\varphi||_L$, is defined as

$$||\varphi||_L \;=\; \max\left\{ ||\varphi||_\infty, \sup_{\mathbf{v} \neq \mathbf{w} \in \mathbb{R}^3} \frac{|\varphi(\mathbf{v}) - \varphi(\mathbf{w})|}{|\mathbf{v} - \mathbf{w}|} \right\}. \tag{4.4}$$

In the inequality (4.1), $\displaystyle\int_{Z_l} \Phi(\widetilde{z}_l)\, p_{\mathrm{red}}(z_l; d\widetilde{z}_l)$ gives the expectation of Φ for the reduced system.

For our new deterministic reduction scheme, in the spatially homogeneous case, for each group only one state is possible after reduction. Therefore, in the case where all three eigenvalues of the pressure tensor are positive, $p_{\mathrm{red}}(z_l; d\widetilde{z}_l) = \delta_{J_{\mathrm{red}(z_l)}}(d\widetilde{z}_l)$, where $[J_{\mathrm{red}(z_l)}]_i = (\widetilde{\mathbf{v}}_i(z_l), \widetilde{g}_i(z_l))$, for $i = 1, \ldots, 6$, is the post reduction state given by Theorem 3.1. Therefore,

$$\int_{Z_l} \Phi(\widetilde{z}_l)\, p_{\mathrm{red}}(z_l; d\widetilde{z}_l) \;=\; \Phi(J_{\mathrm{red}}(\widetilde{z}_l)) \;=\; \sum_{j=1}^{6} \widetilde{g}_j\, \varphi(\widetilde{\mathbf{v}}_j). \tag{4.5}$$

Since $\displaystyle\sum_{j=1}^{6} \widetilde{g}_j = \varrho_l$, and applying the triangle inequality, we obtain

$$\begin{aligned}
\left| \int_{Z_l} \Phi(\widetilde{z}_l)\, p_{\mathrm{red}}(z_l; d\widetilde{z}_l) - \Phi(z) \right| \;&=\; \left| \sum_{j=1}^{6} \widetilde{g}_j\, \varphi(\widetilde{\mathbf{v}}_j) - \sum_{i=1}^{m_l} g_i\, \varphi(\mathbf{v}_i) \right| \\
&\leq\; \sum_{j=1}^{6} \left| \frac{\widetilde{g}_j}{\varrho_l} \sum_{i=1}^{m_l} g_i\, \varphi(\widetilde{\mathbf{v}}_j) - \frac{\widetilde{g}_j}{\varrho_l} \sum_{i=1}^{m_l} g_i\, \varphi(\mathbf{v}_i) \right| \\
&\leq\; \sum_{j=1}^{6} \frac{\widetilde{g}_j}{\varrho_l} \sum_{i=1}^{m_l} g_i \left| \varphi(\widetilde{\mathbf{v}}_j) - \varphi(\mathbf{v}_i) \right| \\
&\leq\; ||\varphi||_L \sum_{j=1}^{6} \frac{\widetilde{g}_j}{\varrho_l} \sum_{i=1}^{m_l} g_i \left| \widetilde{\mathbf{v}}_j - \mathbf{v}_i \right|,
\end{aligned} \tag{4.6}$$

where the final inequality follows from (4.4). Furthermore, using (3.11), the triangle inequality, and the fact that $\sum_{i=1}^{m_l} g_i = \varrho_l$, we obtain

$$
\begin{aligned}
&\sum_{j=1}^{6} \frac{\widetilde{g}_j}{\varrho_l} \sum_{i=1}^{m_l} g_i \left| \widetilde{\mathbf{v}}_j - \mathbf{v}_i \right| \\
&\leq \sum_{i=1}^{m_l} g_i \left| \mathbf{V}_l - \mathbf{v}_i \right| + \sum_{j=1}^{6} \widetilde{g}_j \, \alpha_j \\
&= \sum_{i=1}^{m_l} g_i \left| \mathbf{V}_l - \mathbf{v}_i \right| + \left(\sum_{j=1}^{6} \widetilde{g}_j^2 \, \alpha_j^2 + 2 \sum_{j=1}^{6} \sum_{k>j} \widetilde{g}_j \, \widetilde{g}_k \, \alpha_j \, \alpha_k \right)^{\frac{1}{2}} \\
&\leq \sum_{i=1}^{m_l} g_i \left| \mathbf{V}_l - \mathbf{v}_i \right| + \left(\sum_{j=1}^{6} \widetilde{g}_j^2 \, \alpha_j^2 + \sum_{j=1}^{6} \sum_{k>j} \widetilde{g}_j \, \widetilde{g}_k \, (\alpha_j^2 + \alpha_k^2) \right)^{\frac{1}{2}} \\
&= \sum_{i=1}^{m_l} g_i \left| \mathbf{V}_l - \mathbf{v}_i \right| + \left(\varrho_l \sum_{j=1}^{6} \widetilde{g}_j \, \alpha_j^2 \right)^{\frac{1}{2}} \\
&= \sum_{i=1}^{m_l} g_i \left| \mathbf{V}_l - \mathbf{v}_i \right| + \varrho_l \sqrt{3 \, T_l}.
\end{aligned}
\tag{4.7}
$$

Here, the final equality is obtained from (3.4) and (3.11). Therefore, the desired inequality (4.1) holds for our new reduction scheme, and Wagner's convergence theorem applies in this context.

5. Numerical results. In this section, we discuss our numerical results. The algorithm was implemented in C++ and all simulations were performed on a desktop machine with a 3.6 GHz single processor. We verified that the total times for the particle collisions and for the clustering and particle reductions both scale linearly with the initial number of computational particles, m_0. The time taken to simulate the clustering and particle reductions was approximately four times larger than the time taken to simulate the particle collisions. However, as we will show in Table 3, for the results in Figures 1 to 3 below, the total computational time is only about 30 seconds for $N = 100$ ensembles with $m_0 = 10,240$ particles per ensemble.

First, to numerically verify the conclusions of Theorem 3.1, we study the sum over all the groups of the reduction errors for the raw and central heat flux. For this study, we consider an initial Maxwellian distribution with temperature, $T = 1$, and drift velocity, $\mathbf{V} = \langle 0, 0, 0 \rangle$. We used a single ensemble to obtain these results, and the initial number of computational particles is chosen to be $m_0 = 10{,}240$. Once the number of computational particles reaches $4m_0$, we reduce it to $\widetilde{m} \approx \frac{m_0}{4}$, which was the strategy that produced the largest errors for the deterministic reduction schemes of Rjasanow and Wagner [18]. We chose this strategy to demonstrate that our method performs well even under this condition.

In Table 2, we compare three reduction schemes. All three schemes conserve the total weight and momentum, and in addition to these moments, the reduction schemes conserve the moments associated to their names. The first two schemes, energy conservation, and energy and central heat flux conservation (Ct. HF), are the reduction schemes of Rjasanow and Wagner, and the third one is our reduction scheme which conserves the pressure tensor (PT) and central heat flux (Ct. HF). To

TABLE 2
Maximum average relative errors in the central and raw heat flux for the three reduction schemes.

Reduction Scheme	Central Heat Flux Error	Raw Heat Flux Error
Energy	1	0.01743
Energy and Central Heat Flux (Ct. HF)	3.00844e-15	0.015072
Pressure Tensor (PT) and Central Heat Flux (Ct. HF)	2.12406e-15	6.81888e-16

compare these schemes, we compute the relative 2-norm errors for each of the third order moments of each group and take their average over all the groups, that is we let

$$\mathcal{E} \;=\; \frac{1}{\#\text{Grps}} \sum_{l=1}^{\#\text{Grps}} \frac{||\mathfrak{m}_{\text{After},l} - \mathfrak{m}_{\text{Before},l}||_2}{||\mathfrak{m}_{\text{Before},l}||_2}. \tag{5.1}$$

We obtained these average relative errors for the first ten reductions, and show the maximum of these errors in Table 2. The errors for the pressure tensor and central heat flux scheme are smaller than 10^{-14}, which is negligible. However, in the third column of the table, we observe that for the energy and central heat flux scheme the raw heat flux error is about 2×10^{13} times larger than that for the pressure tensor and central heat flux scheme. These results support the theory in section 3 that the energy and central heat flux scheme only conserves the central heat flux in each group, and does not conserve the raw heat flux, while the pressure tensor and central heat flux scheme conserves both. Furthermore, the energy scheme has the largest error for both third order moments. Since this scheme replaces a group by two particles with equal weights and opposite velocities relative to the drift velocity, the central heat flux of the group after reduction is zero. This observation explains why the relative error in the central heat flux is 1 for the energy reduction scheme.

In [20], Rjasanow and Wagner observed that the higher order moments of a distribution are conserved statistically when averaged over a large number of ensembles, even if the reduction scheme only conserves the lower order moments. However, they found that the existing deterministic reduction schemes require a larger initial number of computational particles for the convergence of the scalar fourth order moment than for the lower moments. To examine this, for each reduction scheme we studied the convergence of (1,1)-component of the momentum flux tensor, $\Pi_{1,1}$, the second component of the raw heat flux, $\mathbf{h}_2$, and the scalar fourth order moment,

$$s \;=\; \sum_{i=0}^{m} g_i |\mathbf{v}_i|^4, \tag{5.2}$$

as we increase m_0. For this study, we chose the initial condition to be a mixture of Maxwellian distributions, since for this pdf there is an analytical formula for the given moments as a function of time, t [20]. The initial distribution is given by

$$f_0(\mathbf{v}) \;=\; \alpha\, M_{\mathbf{V}_1,T_1}(\mathbf{v}) + (1-\alpha)\, M_{\mathbf{V}_2,T_2}(\mathbf{v}), \tag{5.3}$$

where $M_{\mathbf{V}_1,T_1}(\mathbf{v})$ and $M_{\mathbf{V}_2,T_2}(\mathbf{v})$ are Maxwellian distributions with drift velocities $\mathbf{V}_1$ and $\mathbf{V}_2$, and temperatures T_1 and T_2, respectively. We chose $\alpha = 0.5$, $\mathbf{V}_1 = \langle -2, 2, 0\rangle$, $\mathbf{V}_2 = \langle 2, 0, 0\rangle$, and $T_1 = T_2 = 1$. We performed two sets of simulations in which we studied the short term (transient) behavior of s in the time interval $[0, 3]$. For this study, we calculated the relative error of the moments and the half-width of

the 99.9% confidence interval as a function of time in the interval $[0, 3]$. The relative error for a moment $\mathfrak{m}$ is given by

$$E = \frac{|\mathfrak{m}_{\text{anal}} - \overline{\mathfrak{m}}|}{|\mathfrak{m}_{\text{anal}}|}, \tag{5.4}$$

where $\overline{\mathfrak{m}} = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N} \mathfrak{m}_i$ is the average of the simulated moments over the N ensembles. Similarly, the half-width of the relative confidence interval is given by

$$CI = \frac{z_{(1-\frac{\alpha}{2})}}{|\mathfrak{m}_{\text{anal}}|}\sqrt{\frac{\sigma^2}{N}}, \tag{5.5}$$

where $\sigma^2 = \dfrac{\sum_{i=1}^{N}\left(\mathfrak{m}_i - \overline{\mathfrak{m}}\right)^2}{N-1}$ is the variance of the simulated moments and $z_{(1-\frac{\alpha}{2})}$ is the z-score for the confidence interval with $\alpha = 10^{-3}$. A statistical simulation computes a moment accurately if $E < CI$, that is if there is a high probability that $\mathfrak{m}_{\text{anal}}$ lies in the confidence interval centered at $\overline{\mathfrak{m}}$, and that this confidence interval is relatively narrow.

In the first set of simulations, we used $N = 500$ ensembles and various initial numbers of computational particles, m_0. In the left column of Figure 1, we show the relative error, E, in (5.4) and confidence interval, CI, in (5.5) at time $t = 3$ for the $(1,1)$-component of the momentum, $\Pi_{1,1}$, (top), the second component of the raw heat flux, $\mathbf{h}_2$, (middle), and the scalar fourth-order moment, s, (bottom). The percentage relative error, for the energy, the energy and central heat flux, and the pressure tensor and central heat flux reduction schemes are shown using the symbols in the legends. The half-width of the relative confidence intervals are shown using the corresponding vertical lines. These quantities are plotted for the different values of m_0, which is displayed using a logarithmic scale. For each value of m_0, we have offset the results for the three reduction schemes from each other to aid comprehension.

First, we observe that for each moment, the confidence intervals primarily depend on m_0 rather than on the reduction scheme. Furthermore, with one slight exception, the errors for $\Pi_{1,1}$ and $\mathbf{h}_2$ are within the confidence intervals, even for a small number of computational particles. For $\Pi_{1,1}$, this result is to be expected since all three reductions schemes are designed to conserve momentum. However, as we saw in Table 2, the energy and energy and central heat flux reduction schemes do not preserve the raw heat flux. Therefore, the accuracy of the computation of $\mathbf{h}_2$ with these two schemes is simply due to statistical averaging over the 500 ensembles. Significantly, in most cases the errors for $\Pi_{1,1}$ and $\mathbf{h}_2$ are smaller with the pressure tensor and central heat flux scheme than with the other two reduction schemes.

The main advantage to be gained from using the new pressure tensor and heat flux reduction scheme can be seen in the results for the scalar fourth-order moment, s (see the bottom left panel of Figure 1). With our method, the error in s lies within the confidence interval for $m_0 \geq 1{,}024$. However, with the other two methods the errors are larger than the width of the confidence interval, even for $m_0 = 10{,}240$. Therefore, the energy conservation, and energy and central heat flux conservation reduction schemes require more than 10 times the initial number of computational particles as the pressure tensor and heat flux conservation scheme to approximate the scalar fourth-order moment with the same degree of accuracy. As we see in Table 3, this requires at least seventeen times the computational time.

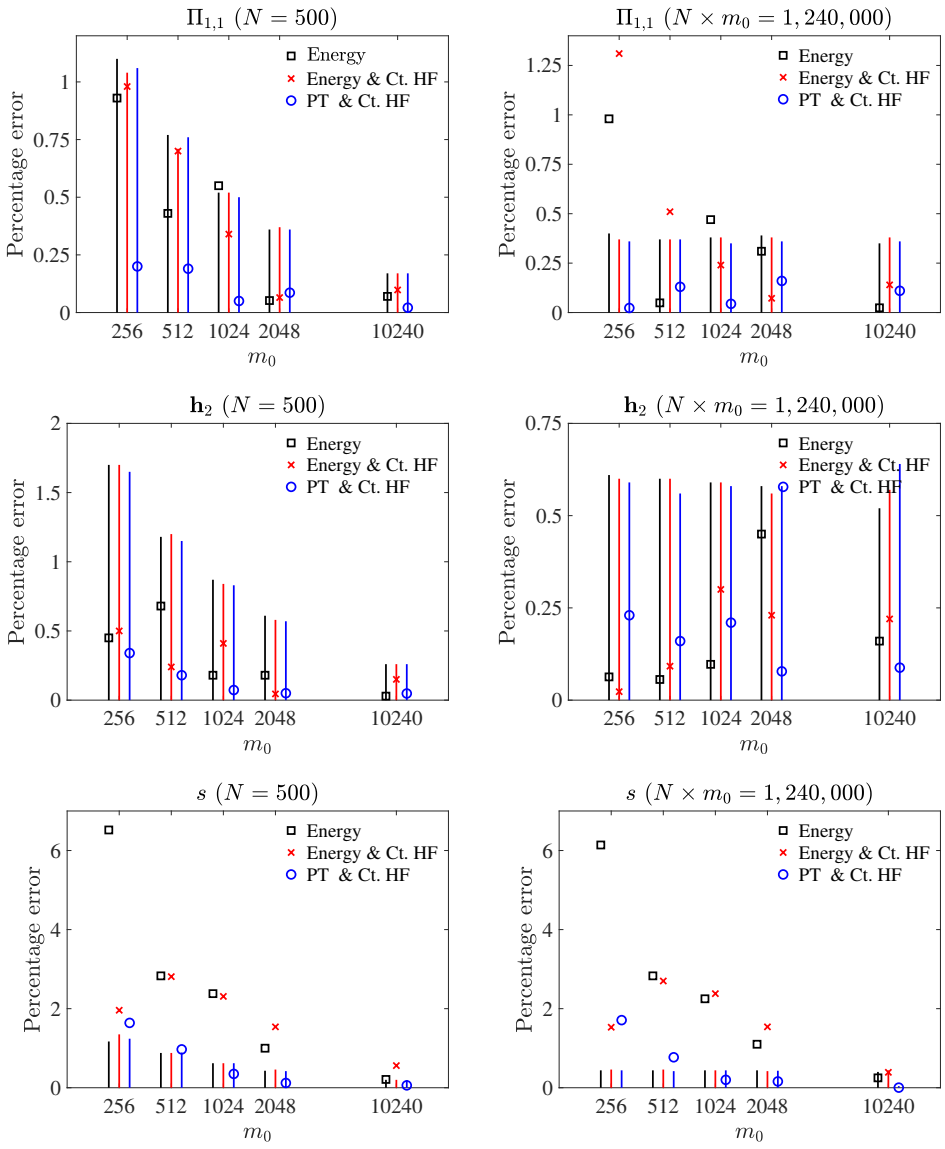

FIG. 1. *Percentage relative error, E, in (5.4) at time $t = 3$ for selected moments, $\mathfrak{m}$, of the velocity pdf. We compare the performance of the SWPM with the three different reduction schemes shown in the legends. We show results for the $(1,1)$-component of the momentum, $\Pi_{1,1}$, (top row), the second component of the raw heat flux, $\mathbf{h}_2$, (middle row), and the scalar fourth-order moment, s, (bottom row). We plot the errors using symbols and the half confidence intervals with vertical lines, so that $E < CI$ when the symbol lies on the line. In the left column, we plot E as a function of the number of particles, m_0, per ensemble for $N = 500$ ensembles. In the right column, we plot E as a function of m_0 when N is chosen so that $N \times m_0 = 1,240,000$.*

To further examine how accurately the three reductions schemes compute the scalar fourth-order moment, in Figure 2 we plot the evolution of s as a function of time, together with 99.9% confidence intervals. The numerical results are shown with red-dashed lines and the true values are shown with solid blue lines. The results for

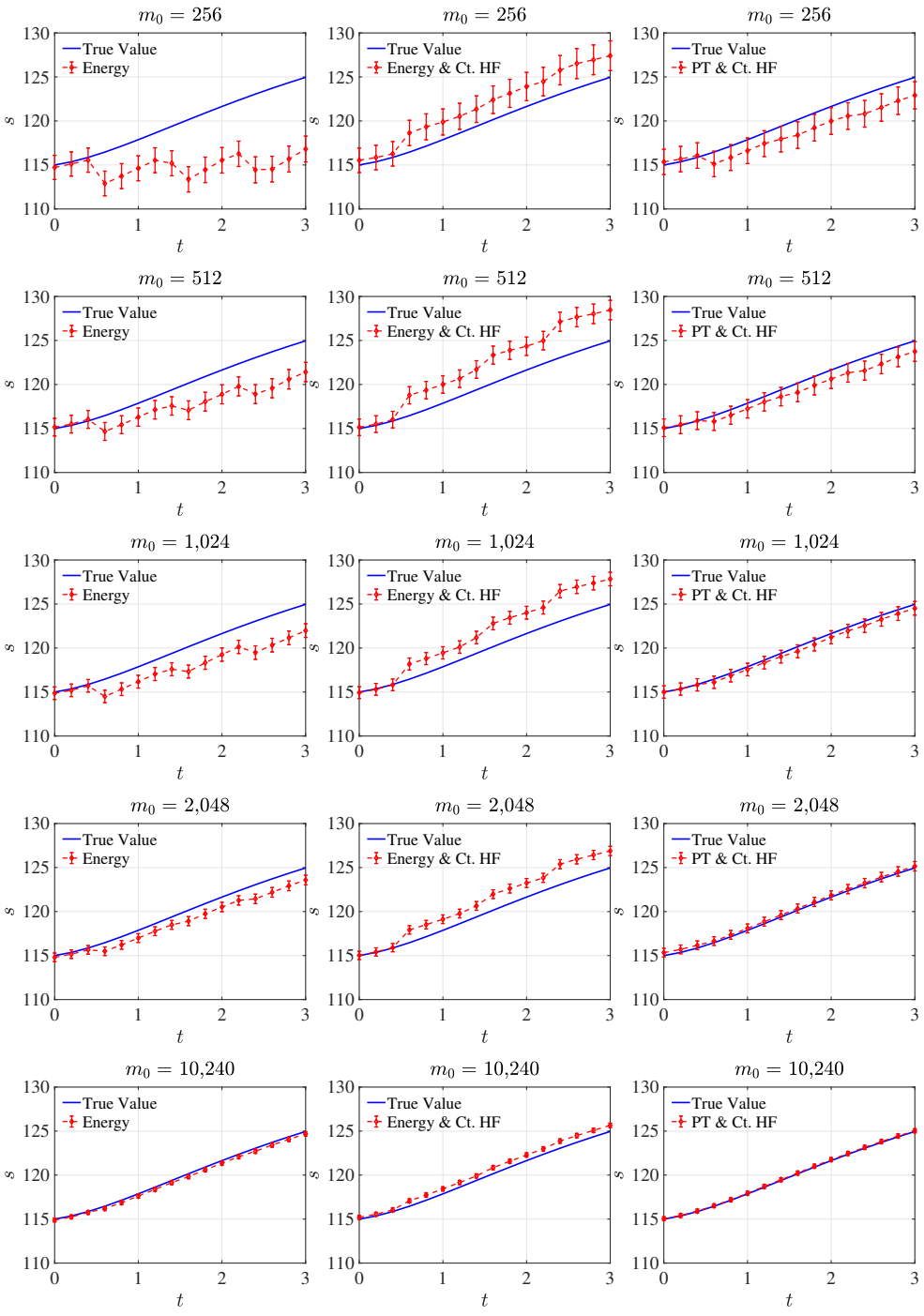

FIG. 2. *Evolution of the scalar fourth order moment, s, as a function of time together with 99.9% confidence intervals. In the different rows we show the results for different initial numbers of computational particles, m_0, per ensemble. We used $N = 500$ ensembles in all the panels. We show the results for the energy scheme (left column), energy and central heat flux scheme (middle column), and pressure tensor and central heat flux scheme (right column).*

TABLE 3

Total computational time for the simulation results shown in Figure 1. These results were obtained with the pressure tensor and central heat flux scheme. The computational times for the two reduction schemes of Rjasanow and Wagner were similar ($\approx \pm 10\%$).

	$N = 500$	$N \times m_0 = 1,240,000$
m_0	t (sec)	t (sec)
256	2.61	20.85
512	6.06	22.02
1,024	13.76	24.93
2,048	31.26	31.26
10,240	205.56	39.05

the energy and the energy and central heat flux reduction schemes are shown in the left and middle columns. The numerical results are visually close to the true values only for $m_0 = 10,240$ (bottom left and middle panels). On the other hand, with the pressure tensor and central heat flux scheme (right column), the numerical results are reasonably accurate across the entire time range for $m_0 = 1,240$. In particular, examining each column in turn, we see that the convergence of s is significantly faster for the pressure tensor and central heat flux conservation scheme, than for the other two reduction schemes.

To summarize our conclusions so far, the new reduction scheme provides improved accuracy at a significantly reduced computational cost. To provide additional evidence for this conclusion, we performed a second set of simulations where we fixed the total number of computational particles, $m_0 \times N$, to be 1,024,000. This value was kept constant to obtain approximately equal sized confidence intervals for all the simulations. The results for these simulations are shown in the right column of Figure 1 and in Figure 3. Comparing the errors and the half width of the confidence intervals for $\Pi_{1,1}$ in Figure 1 (top right panel), we observe that our scheme is accurate even with $m_0 = 256$, while the other two schemes require a larger initial number of computational particles. On the other hand, for $\mathbf{h}_2$ the results obtained with all three reduction schemes are acceptable for all the values of m_0. In the case of the fourth order moment, for both reduction schemes of Rjasanow and Wagner, the error lies within the confidence interval only for $m_0 = 10{,}240$. On the other hand, the errors for our scheme lies within the confidence intervals for $m_0 \geq 1,024$. In the right column of Figure 1, for each moment, the confidence intervals only depend on the total number of computational particles, $m_0 \times N$, and not on m_0 or the choice of reduction scheme. Furthermore, in Table 3 we see that as m_0 decreases the computational time decreases. As a consequence, for each moment, for the same level of accuracy the computational time for our reduction scheme is significantly less than that for the other two reduction schemes. We also observe this phenomenon in Figure 3. For example, we observe the same degree of accuracy in the fourth order moment for our reduction scheme with $m_0 = 1{,}024$ and $N = 1{,}000$ (the right panel in the third row) as for the other two methods with $m_0 = 10{,}240$ and $N = 100$ (the left and middle panels in the last row). However, the computational time of 43 seconds for the two reduction schemes of Rjasanow and Wagner is reduced by 42% to 25 seconds for our scheme.

6. Conclusions. We have confirmed that the reduction scheme of Rjasanow and Wagner that conserves total weight, momentum, energy and central heat flux of a group does not conserve the raw heat flux in each group. Consequently, the raw and central heat flux of the entire system are not conserved. We resolved this problem by

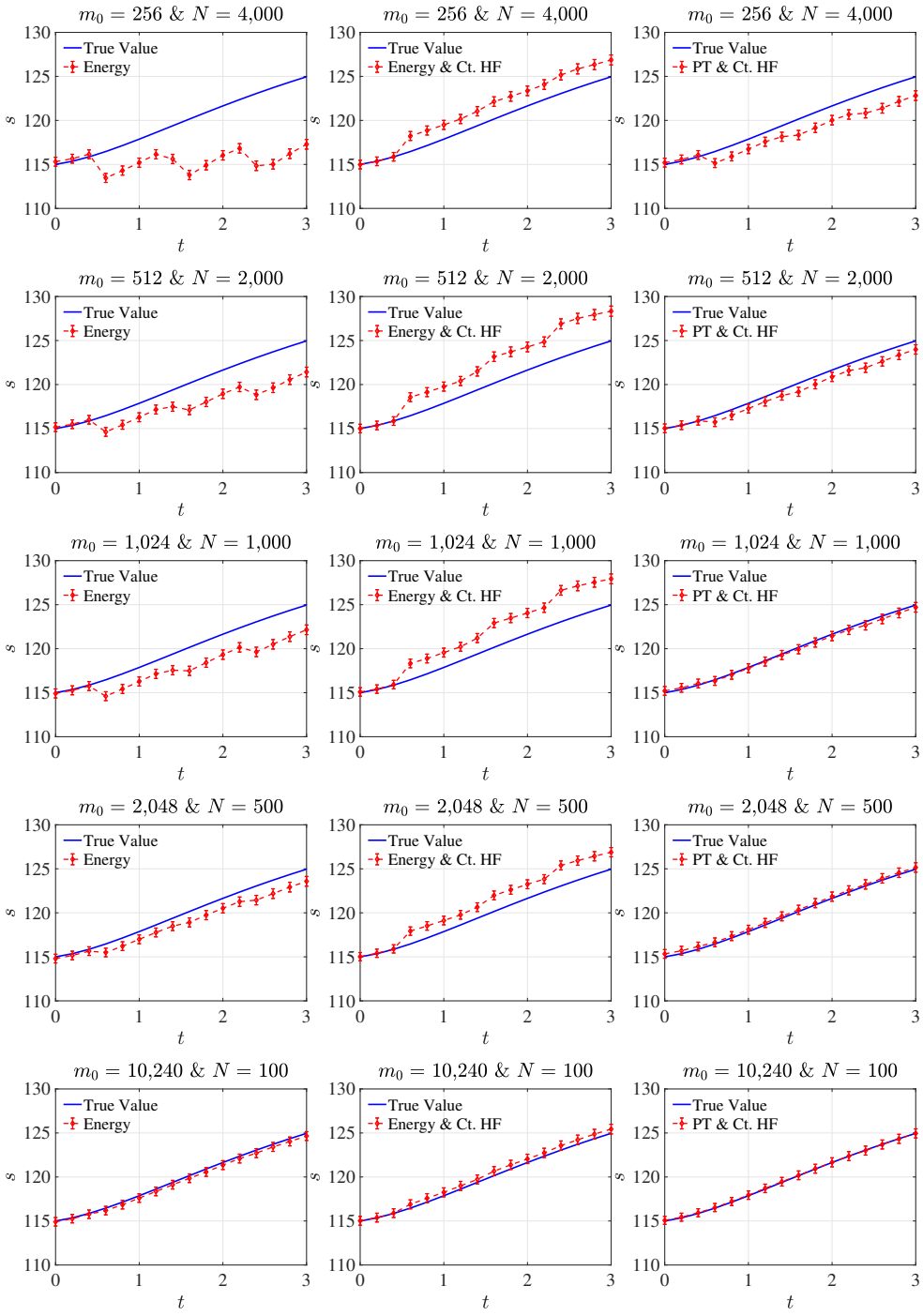

FIG. 3. *Evolution of the scalar fourth order moment, s, as a function of time, together with 99.9% confidence intervals, for different initial numbers of computational particles, m_0, per ensemble and different number of ensembles, N. For these results the total number of computational particles, $m_0 \times N$, was kept constant. We show the results for the energy scheme (left column), energy and central heat flux scheme (middle column), and pressure tensor and central heat flux scheme (right column).*

devising a new reduction scheme that conserves the total weight, momentum, pressure tensor and heat flux within each group. Conservation of these moments within a group results in the conservation of all of the moments up to the second order, and both raw and central heat flux among the third order moments of a group. This further leads to the preservation of these moments for the entire system.

To examine the accuracy of our new reduction scheme, we performed simulation studies to analyze the convergence of $\Pi_{1,1}$, $\mathbf{h}_2$, and the scalar fourth order moment for the existing and new reduction schemes. The new reduction scheme leads to the convergence of these moments, particularly the scalar fourth order moment, with significantly less computational cost compared to the existing deterministic reduction schemes. This shows that the preservation of additional moments in the new reduction scheme conserves the higher moments with better accuracy, and also minimizes the reduction error.

Although the conservation of higher-order moments reduces the systematic error introduced by the reduction process, the clustering technique must also be carefully designed in order to accurately and efficiently compute the low-probability tails of the velocity pdf. Specifically, since the tails occupy a proportionately large volume of phase space, we need to ensure that particles in the tails that are assigned to the same group are sufficiently close together. In a forthcoming article we will use the proof of the convergence theorem for the SWPM [20] to develop such a clustering algorithm. In combination with the reduction scheme introduced in this paper, we will demonstrate that this leads to a more efficient method for the computation of tail functionals.

REFERENCES

[1] J. Allen, *On the applicability of the Druyvesteyn method of measuring electron energy distributions*, Journal of Physics D: Applied Physics, 11 (1978), p. L35.

[2] G. A. Bird and J. Brady, *Molecular gas dynamics and the direct simulation of gas flows*, vol. 5, Clarendon press Oxford, 1994.

[3] J. A. Bittencourt, *Fundamentals of plasma physics*, Springer Science & Business Media, 2013.

[4] L. Boltzmann, *Weitere studien über das wärmegleichgewicht unter gasmolekülen*, in Kinetische Theorie II, Springer, 1970, pp. 115–225.

[5] C. Cercignani, *The Boltzmann equation*, in The Boltzmann equation and its applications, Springer, 1988, pp. 40–103.

[6] J. V. DiCarlo and M. J. Kushner, *Solving the spatially dependent Boltzmann's equation for the electron-velocity distribution using flux corrected transport*, Journal of Applied Physics, 66 (1989), pp. 5763–5774.

[7] I. Gamba, V. Panferov, and C. Villani, *Upper Maxwellian bounds for the spatially homogeneous Boltzmann equation*, Archive for Rational Mechanics and Analysis, 194 (2009), pp. 253–282.

[8] I. M. Gamba and J. R. Haack, *A conservative spectral method for the Boltzmann equation with anisotropic scattering and the grazing collisions limit*, Journal of Computational Physics, 270 (2014), pp. 40–57.

[9] I. M. Gamba, J. R. Haack, C. D. Hauck, and J. Hu, *A fast spectral method for the Boltzmann collision operator with general collision kernels*, SIAM Journal on Scientific Computing, 39 (2017), pp. B658–B674.

[10] I. M. Gamba and S. Rjasanow, *Galerkin–Petrov approach for the Boltzmann equation*, Journal of Computational Physics, 366 (2018), pp. 341–365.

[11] J. Johannes, T. Bartel, G. A. Hebner, J. Woodworth, and D. J. Economou, *Direct simulation Monte Carlo of inductively coupled plasma and comparison with experiments*, Journal of the Electrochemical Society, 144 (1997), pp. 2448–2455.

[12] K. Koura, *Null-collision technique in the direct-simulation Monte Carlo method*, The Physics of fluids, 29 (1986), pp. 3509–3511.

[13] M. J. Kushner, *Hybrid modelling of low temperature plasmas for fundamental investigations*

and equipment design, Journal of Physics D: Applied Physics, 42 (2009), p. 194013.
[14] I. Matheis and W. Wagner, *Convergence of the stochastic weighted particle method for the Boltzmann equation*, SIAM Journal on Scientific Computing, 24 (2003), pp. 1589–1609.
[15] C. Mouhot and L. Pareschi, *Fast algorithm for computing the Boltzmann collision operator*, Math. Comp. , 75 (2006), pp. 1833–1852.
[16] L. Pareschi and B. Perthame, *A Fourier spectral method for homogeneous Boltzmann equations*, Transport Theory Statist., 25 (2002), pp. 369–382.
[17] J. Poulose, M. Goeckner, S. Shannon, D. Coumou, and L. Overzet, *Driving frequency fluctuations in pulsed capacitively coupled plasmas*, The European Physical Journal D, 71 (2017), p. 242.
[18] S. Rjasanow, T. Schreiber, and W. Wagner, *Reduction of the number of particles in the stochastic weighted particle method for the Boltzmann equation*, Journal of Computational Physics, 145 (1998), pp. 382–405.
[19] S. Rjasanow and W. Wagner, *A stochastic weighted particle method for the Boltzmann equation*, Journal of Computational Physics, 124 (1996), pp. 243–253.
[20] S. Rjasanow and W. Wagner, *Stochastic numerics for the Boltzmann equation*, Springer, 2005.
[21] T. Sheridan, M. Goeckner, and J. Goree, *Electron velocity distribution functions in a sputtering magnetron discharge for the E× B direction*, Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films, 16 (1998), pp. 2173–2176.
[22] C. Sozzi, E. De La Luna, D. Farina, J. Fessey, L. Figini, S. Garavaglia, G. Grossetti, S. Nowak, P. Platania, A. Simonetto, et al., *Measurements of electron velocity distribution function*, in AIP Conference Proceedings, vol. 988, AIP, 2008, pp. 73–80.
[23] W. Tan, *Langmuir probe measurement of electron temperature in a Druyvesteyn electron plasma*, Journal of Physics D: Applied Physics, 6 (1973), p. 1206.
[24] M.-B. Tran, *Nonlinear approximation theory for the homogeneous Boltzmann equation*, arXiv preprint arXiv:1305.1667, (2013).
[25] A. Vikhansky and M. Kraft, *Conservative method for the reduction of the number of particles in the Monte Carlo simulation method for kinetic equations*, Journal of Computational Physics, 203 (2005), pp. 371–378.
[26] W. Wagner, *A convergence proof for Bird's direct simulation Monte Carlo method for the Boltzmann equation*, Journal of Statistical Physics, 66 (1992), pp. 1011–1044.