GENERALIZED UGK SCHEME IN THE DIFFUSIVE LIMIT

NICOLAS CROUSEILLES[*], JULIEN MATHIAUD[†], AND LUC MIEUSSENS[‡]

Abstract. The unified gas kinetic scheme (UGKS) was initially designed to address multiscale challenges in rarefied gas dynamics and then extended to radiative transfert theory, as described by BGK like relaxation models. In this work, we extend its application to linear kinetic models with non isotropic scattering collision operators, as well as Fokker-Planck models . These problems typically exhibit a fully diffusive nature in the optically thick limit (corresponding to a small Knudsen number). It still leads to an asymptotic preserving (AP) property not only in this diffusive regime but also in the free transport limit. A series of numerical experiments confirm the effectiveness of the approach.

Key word. Transport equations, diffusion limit, asymptotic preserving schemes, BGK, Fokker-Planck, scattering.

Introduction.

Kinetic models play a fundamental role in describing particle system dynamics across various domains, such as rarefied gas dynamics (RGD), neutron transport, semiconductor physics, and radiative transfer.

From a numerical perspective, solving accurately these models presents significant challenges. The computational cost is driven by the necessity to capture the smallest microscopic scales, which constrain spatial discretization and, consequently, time steps for stability. The Knudsen number, denoted by ε, represents the ratio of the mean free path of particles to a macroscopic length scale and characterizes the transition between collision-dominated and free-streaming regimes. As ε tends to zero, a global diffusive behavior emerges at the macroscopic level. However, standard numerical schemes for moment models do not necessarily recover the correct macroscopic diffusion equation in this limit so that asymptotic-preserving (AP) numerical methods have been developed, ensuring stability and consistency across different regimes [8, 7, 9, 10, 2, 11, 13, 1, 3, 4, 6, 12, 16].

A notable AP approach is the Unified Gas Kinetic Scheme (UGKS), originally introduced by Xu and Huang [19] in the context of rarefied gas dynamics. UGKS leverages a finite volume framework where numerical fluxes incorporate information from the collision operator, allowing it to bridge different regimes. Instead, UGKS makes use of the relaxation form of the BGK collision operator to derive an approximation of the interface flux by using a Duhamel representation of the solution. In other words, this representation allows for an approximation of the solution of the generalized Riemann problem at the interface that accounts for collisions. Since its conception, UGKS has been extended to more complex gas flows [14, 20] and applied to linear models with a diffusion limit [16, 17]. A key advantage of UGKS is that it does not require decomposition of the distribution function (like the micro-macro or the odd-even decomposition), thus avoiding issues related to boundary conditions, and it operates without staggered grids, facilitating multi-dimensional implementations.

By construction, the extension of UGKS to non relaxation operators is not obvious. Among these collision operators, we mention the original Boltzmann operator and Fokker-Planck models of Rarefied Gas Dynamics, and non isotropic scattering and Fokker-Planck models in neutron transport and radiative transfer, for instance. A first attempt [17] was to use the relaxation technique of Filbet-Jin ([5]): the UGKS is applied to a asymptotically consistent BGK operator, while the deviation to the original collision operator is used as a source terme. However, this was proved to induce a non physical statbility constraint on the collision kernel, at least for the diffusion regime. Another related extension was proposed by Liu et al. for rarefied gas Dynamics ([15]), where the BGK model is still used for the UGKS flux, while a the collision operator is replaced by a con-

---

[*]Univ Rennes, INRIA (MINGuS), CNRS, IRMAR, UMR 6625, Rennes, France (nicolas.crouseilles@inria.fr).

[†]Univ Rennes, CNRS, IRMAR, UMR 6625, Rennes, France (julien.mathiaud@univ-rennes.fr).

[‡]University of Bordeaux, Bordeaux INP, CNRS, IMB, UMR 5251, Talence, France (luc.mieussens@math.u-bordeaux.fr).

vex combination between the Boltzmann operator and the BGK one. Here, we propose a related approach that looks more general, in which we make use of the eigenvalue of the pseudo-inverse of the collision operator to derive a new relaxation form. Combined with the UGKS approach, and various simplifications, this allows to capture the correct diffusion limit without any stability restriction, for both non isotropic scattering collision operator and the Fokker-Planck operator as well. We demonstrate that UGKS offers a robust framework for AP numerical schemes in the context of the diffusive limit. By leveraging kinetic-inspired flux approximations, UGKS enables a seamless transition between kinetic and macroscopic scales, making it a promising approach for modeling various physical phenomena.

The rest of this work is organized as follows: first, the continuous and semi-discrete (in velocity) models are presented, and their diffusion limit is discussed. In the second section, the new UGKS framework is presented and the AP property is verified. Section 3 proposes some possible extensions and links with the UGKS scheme are discussed. Then, in Section 4, some numerical results are presented to illustrate the capabilities of the new approach.

1. Continuous models. In this section, we present models at continuous level and discretized level: some properties and the difusion limit are given.

1.1. Linear collision operators with continuous velocities. We consider the equation on the particle density function $f(t, x, v)$ which depends on time $t \geq 0$, position $x \in [0, L] \subset \mathbb{R}$ ($L > 0$) and velocity $v \in [-c, c] \subset \mathbb{R}$ (with c a typical velocity of the phenomena):

$$(1.1) \qquad \eta \partial_t f + v \partial_x f = \frac{\sigma}{\varepsilon} \mathcal{D} f,$$

ε and η being two parameters that will vary according to limit one considers. σ is the characteristic collision frequency. The collision operator $\mathcal{D}$ can be given by:

$$(1.2) \qquad \mathcal{D} f(v) := \int_{-c}^{c} k(v, v')(f(v') - f(v)) \, dv',$$

or

$$(1.3) \qquad \mathcal{D} f(v) = \partial_v \Big((c^2 - v^2) \partial_v f \Big).$$

The collisional kernels in (1.2) and (1.3) respectively correspond to a simple scattering model (including the BGK model for $k = 1$) and a Fokker-Planck-like model projected in the v direction. The function $k : (v, v') \rightarrow k(v, v')$ is the probability of passing from state v to state v' and is symmetric so that $\langle k(v, \cdot), 1 \rangle = 1$, $\forall v \in [-c, c]$, with $\langle f, g \rangle = \int_{-c}^{c} f(v)g(v)dv$ for two v-dependent functions f, g.

It is well known (see [13, 2, 6, 10, 7]) that in the diffusion limit, ie $\eta = \varepsilon$ tends to zero, the function $f(t, x, v)$ matches the Chapman-Enskog expansion

$$(1.4) \qquad f(t, x, v) = \rho(t, x) + \frac{\varepsilon}{\sigma} \mathcal{D}^+(v) \partial_x \rho(t, x) + \mathcal{O}(\varepsilon^2),$$

where $\mathcal{D}^+$ denotes the pseudo-inverse of $\mathcal{D}$ and the density $\rho(t, x) = \int_{-c}^{c} f(t, x, v)dv$ satisfies the following diffusion equation:

$$(1.5) \qquad \partial_t \rho + \partial_x(\kappa \partial_x \rho) = 0,$$

where $\kappa = \langle v, \mathcal{D}^+ v \rangle$. The operator $\mathcal{D}^+$ is defined for functions with zero mean by the following condition: for any function ϕ such that $\langle \phi \rangle = 0$, the function $\psi = \mathcal{D}^+ \phi$ is the unique solution of $\mathcal{D}\psi = \phi$ such that $\langle \psi, 1 \rangle = 0$.

1.2. Continuous Problem with Discrete Velocities.

We denote $F_j(t,x) = F(t,x,v_j)$ for $j = 1, \ldots, 2N$, where v_j is a symmetric regular velocity grid with step size $\Delta v = \dfrac{2c}{2N} = \dfrac{c}{N}$, with $v_j = -c + \dfrac{\Delta v}{2} + (j-1)\Delta v$ that ranges from $-c$ to c, and $F(t,x) \in \mathbb{R}^{2N}$ is the solution of the following equation:

$$(1.6) \qquad \partial_t F_j + \frac{1}{\eta} v_j \partial_x F_j = \frac{\sigma}{\eta \varepsilon} \langle e_j, DF \rangle,$$

where $D \in \mathcal{M}_{2N,2N}(\mathbb{R})$ is a matrix discretizing a collision operator $\mathcal{D}$ (diffusion (1.3) or scattering (1.2)), e_j is the j-th vector of the canonical basis of $\mathbb{R}^{2N}$, and for $U, W \in \mathbb{R}^{2N}$ we denote $\langle U, W \rangle = \sum_{k=1}^{2N} U_k W_k$ (let us remark the same notation as in the continuous case is used). Let also introduce the notation $V \in \mathbb{R}^{2N}$ which is the vector of discrete velocities

$$(1.7) \qquad V = \begin{pmatrix} v_1 & \ldots & v_{2N} \end{pmatrix}^T \in \mathbb{R}^{2N},$$

which satisfies $\sum_{k=1}^{2N} v_k = 0$.

From now on we choose $\mathbf{c} = \mathbf{1}$ so that $\mathbf{\Delta v} = \frac{1}{\mathbf{N}}$.

First of all, we recall some useful properties for the matrix D that are common to discretization of operators like (1.3) and (1.2) (see subsections 4.3 and 4.4 for examples of such discretization matrices). We assume the matrix D to be symmetric, negative, and its kernel is Ker $D = \text{Span}(\mathbf{1})$, with

$$(1.8) \qquad \mathbf{1} = \begin{pmatrix} 1 & \ldots & 1 \end{pmatrix} \in \mathbb{R}^{2N}.$$

It implies that $\forall i, \ \sum_{j=1}^{2N} D_{i,j} = \sum_{j=1}^{2N} D_{j,i} = 0$.

Moreover, we introduce the following notations:
- $\lambda_k, k = 0, \ldots, N'$ the eigenvalues of D, with $N' \le 2N - 1$ and $\lambda_0 = 0$.
- P_0 the orthogonal projector onto Ker D: $P_0 = \frac{1}{2N} \begin{pmatrix} 1 & \ldots & 1 \\ \vdots & & \vdots \\ 1 & \ldots & 1 \end{pmatrix}$, associated with the eigenvalue $\lambda_0 = 0$,
- $P_1, \ldots, P_{N'}$ the orthogonal projectors onto the other eigenspaces ($N' \le 2N - 1$) associated with eigenvalues $\lambda_1, \ldots, \lambda_{N'}$.

We also assume that the matrix D satisfies:
- A positiveness condition:

$$(1.9) \qquad D_{i,j} \ge 0 \text{ for } i \ne j$$

.
- D is somehow "irreducible": there exists $\delta_0 > 0$ such that $\forall \delta$ such that $0 \le \delta \le \delta_0$, $I + \delta D$ is an irreducible bistochastic matrix: it implies that there is a sequence $k_0, \ldots, k_n$ of integers between 1 and $2N$ that contains all the integers between 1 and $2N$ such that 2 consecutive terms are different ($k_j \ne k_{j+1}$) and $D_{k_j, k_{j+1}} > 0$

Then, we recall some useful properties in the following proposition that can be proved using standard linear algebra tools for symmetric matrices.

PROPOSITION 1.1. *Let consider a matrix $D \in \mathcal{M}_{2N,2N}(\mathbb{R})$ symmetric, negative, and such that Ker $D = Span(\mathbf{1})$ with $\mathbf{1}$ defined in (1.8). Then, we have*

- the orthogonal projectors P_k associated with the eigenvalues λ_k of D satisfy

$$\sum_{k=0}^{N'} P_k = I, \qquad \sum_{k=0}^{N'} \lambda_k P_k = \sum_{k=1}^{N'} \lambda_k P_k = D, \quad and \quad P_k P_j = 0 \ if \ k \neq j, \ N' \leq 2N - 1.$$

- the pseudo-inverse D^+ satisfies

$$D^+ := \sum_{k=1}^{N'} \frac{1}{\lambda_k} P_k \quad and \quad D^+ D = D D^+ = I - P_0.$$

- there exists a unique $U \in \mathbb{R}^{2N}$ such that $\langle U, \mathbf{1} \rangle = 0$ and $DU = V$, being given by (1.7). We define $\lambda_\star$ as:

(1.10) $$\lambda_\star := \frac{\langle V, V \rangle}{\langle D^+ V, V \rangle} = \frac{\langle DU, DU \rangle}{\langle U, DU \rangle}.$$

the pseudo-eigenvalue. We have $U = V/\lambda_\star$ when $DV = \lambda V$.

REMARK 1.1. *Let us perform some remarks regarding Proposition 1.1 The pseudo-eigenvalue $\lambda_\star$ is always negative since D is negative. For the BGK operator, we have $\lambda_\star = -1$ whereas for the Fokker-Planck operator, we have $\lambda_\star = -2$, V being an eigenvector in both cases. Besides, last item leads to the second principle on the semi-discrete equation (1.6).*

1.3. Generic Properties of the discrete model and its Diffusive limit. In view of deriving a suitable Duhamel formula, we introduce a $\lambda_\star$-relaxation term in (1.6) to get the equivalent following form (let recall that $\lambda_\star$ is negative) : in other words, we rewrite D as $D = (D - \lambda_\star I) + \lambda_\star I$ so that the discrete equation is

(1.11) $$\partial_t F_j + \frac{1}{\eta} v_j \partial_x F_j = \langle e_j, \frac{\sigma}{\eta \varepsilon} [(D - \lambda_\star I)F + \lambda_\star F] \rangle.$$

Then, from (1.11), the relaxation term $\lambda_\star F$ is used to write a Duhamel-type formula

$$F_j(t,x) = \exp\left(\frac{\lambda_\star \sigma t}{\eta \varepsilon}\right) F_j\left(0, x - \frac{v_j t}{\eta}\right)$$
$$+ \langle e_j, \int_0^t \frac{\sigma}{\eta \varepsilon} \exp\left(\frac{\lambda_\star \sigma}{\eta \varepsilon}(t - t')\right)(D - \lambda_\star I)F\left(t', x + \frac{v_j(t' - t)}{\eta}\right) dt' \rangle.$$

We then perform the change of variable $u = \frac{\sigma(t - t')}{\eta \varepsilon}$ in the integral term to get

$$F_j(t,x) = \exp\left(\frac{\lambda_\star \sigma t}{\eta \varepsilon}\right) F_j\left(0, x - \frac{v_j t}{\eta}\right)$$

(1.12) $$+ \langle e_j, \int_0^{\frac{\sigma t}{\eta \varepsilon}} \exp\left(\lambda_\star u\right)(D - \lambda_\star I)F\left(t - \frac{\eta \varepsilon u}{\sigma}, x - \frac{v_j \varepsilon u}{\sigma}\right) du \rangle.$$

From the latter expression, we can formally derive a Chapman-Enskog expansion for F_j in the diffusion limit thanks to the second principle:

PROPOSITION 1.2 (Second principle). *The following equivalences hold:*
1. $DF = 0$,
2. $F = \rho \mathbf{1}$, with $\rho = \sum_{j=1}^{2N} F_j \Delta v$.

3. $\langle DF, \ln(F) \rangle = 0$.

Proof. Using the properties of D, the two first assertions are equivalent. The first assertion clearly leads to the third one. It remains to prove that last assertion implies the second one. To do so, let us expand the scalar product: $\langle DF, \ln(F) \rangle = \langle F, D\ln(F) \rangle$

$$\langle DF, \ln(F) \rangle = \sum_{1 \leq i,j \leq 2N} F_i D_{i,j} \ln(F_j)$$

Since $D\mathbf{1} = 0$ and D is symmetric, one gets that $\forall i, D_{i,i} = -\sum_{j \neq i} D_{i,j} = -\sum_{j \neq i} D_{j,i}$ so that

$$\langle DF, \ln(F) \rangle = \sum_{1 \leq i \neq j \leq 2N} F_i D_{i,j} \ln(F_j) - \sum_{1 \leq i \neq j \leq 2N} F_i D_{i,j} \ln(F_i)$$

$$= \sum_{1 \leq i \neq j \leq 2N} F_i D_{i,j} \ln(F_j/F_i)$$

$$= \frac{1}{2} \sum_{1 \leq i \neq j \leq 2N} F_i D_{i,j} \ln(F_j/F_i) + \frac{1}{2} \sum_{1 \leq i \neq j \leq 2N} F_j D_{i,j} \ln(F_i/F_j)$$

$$= \frac{1}{2} \sum_{1 \leq i \neq j \leq 2N} (F_i - F_j) D_{i,j} \ln(F_j/F_i)$$

If the scalar product $\langle DF, \ln(F) \rangle$ is equal to zero, we get that as soon as $D_{i,j} > 0$ $F_i = F_j$ since $(x, y) \to (y - x) \ln(x/y)$ is a negative function which is equal to zero only when $x = y$.

Since there is a sequel $k_0, ..., k_n$ of integers between 1 and $2N$ that contains all the integers between 1 and $2N$ such that 2 consecutive terms are different ($k_j \neq k_{j+1}$) and $D_{k_j,k_{j+1}} > 0$, all F_i are equal leading to the fact that $F = \rho \mathbf{1}$. $\qquad\square$

1.4. Diffusion Limit. We now study the diffusion limit $\varepsilon = \eta \to 0$ using Duhamel formula (1.12).

1.4.1. Zero-th Order . Passing to the limit $\varepsilon = \eta \to 0$ in (1.12), we obtain from the dominated convergence theorem (recalling that $\lambda_\star < 0$)

$$F_j(t,x) \to \langle e_j, (D - \lambda_\star I) F(t,x) \int_0^\infty \exp(\lambda_\star u) du \rangle$$

(1.13)
$$= -\frac{1}{\lambda_\star} \langle e_j, (D - \lambda_\star I) F(t,x) \rangle.$$

Hence, at the zero-th order, the last relation rewrites as $F(t,x) = \left(I - \frac{1}{\lambda_\star} D \right) F(t,x)$, so that

(1.14)
$$DF(t,x) = 0.$$

From the properties of D (see Proposition 1.2), we deduce that $F(t,x) = \rho(t,x)\mathbf{1}$ with $\rho(t,x) = \sum_{j=1}^{2N} F_j(t,x) \Delta v$.

1.4.2. First order. We now go to the next order by performing a Taylor expansion of F in (1.12):

$$F\left(t - \frac{\eta \varepsilon u}{\sigma}, x - \frac{v_j \varepsilon u}{\sigma} \right) = F(t,x) - \frac{\eta \varepsilon u}{\sigma} \frac{v_j}{\eta} \partial_x F(t,x) + O(\eta \varepsilon)$$

$$= F(t,x) - \frac{\varepsilon u v_j}{\sigma} \partial_x \rho(t,x) \mathbf{1} + O(\eta \varepsilon),$$

5

where we used the zero-th order approximation of F to set $\partial_x F(t,x) = \partial_x \rho(t,x)\mathbf{1} + \mathcal{O}(\varepsilon)$. Thus, the term in the integral in (1.12) becomes (recalling that $D\mathbf{1} = 0$):

$$\langle e_j, (D - \lambda_\star I)F\left(t - \frac{\eta \varepsilon u}{\sigma}, x - \frac{v_j \varepsilon u}{\sigma}\right)\rangle$$

$$=\langle e_j, (D - \lambda_\star I)F(t,x)\rangle - \frac{\varepsilon u}{\sigma}\partial_x \rho(t,x)\langle e_j, (D - \lambda_\star I)v_j \mathbf{1}\rangle + O(\varepsilon^2)$$

$$=\langle e_j, (D - \lambda_\star I)F(t,x)\rangle + \frac{\varepsilon u}{\sigma}\partial_x \rho(t,x)\lambda_\star v_j + O(\varepsilon^2).$$

Multiplying the latter expression by $\exp(\lambda_\star u)$ and integrating over $u \in [0, +\infty[$, equation (1.12) becomes:

$$\int_0^\infty \exp(\lambda_\star u)(D - \lambda_\star I)F\left(t - \frac{\eta \varepsilon u}{\sigma}, x - \frac{v_j \varepsilon u}{\sigma}\right) du =$$

$$\left(I - \frac{1}{\lambda_\star}D\right)F(t,x) + \frac{\varepsilon}{\lambda_\star \sigma}\partial_x \rho(t,x)V + O(\varepsilon^2),$$

where we used $\int_0^{+\infty} u\exp(\lambda_\star u)du = 1/\lambda_\star^2$ and $\int_0^{+\infty}\exp(\lambda_\star u)du = -1/\lambda_\star$. Consequently, since the first term in (1.12) tends to zero exponentially fast, (1.12) becomes, as $\varepsilon \to 0$

$$F(t,x) = \left(I - \frac{1}{\lambda_\star}D\right)F(t,x) + \frac{\varepsilon}{\lambda_\star \sigma}\partial_x \rho(t,x)V + O(\varepsilon^2),$$

or after simplifying $F(t,x)$ on both sides and multiplying by $\lambda_\star$

$$(1.15) \qquad\qquad DF(t,x) = \frac{\varepsilon}{\sigma}\partial_x \rho(t,x)V + O(\varepsilon^2).$$

Applying the pseudo-inverse D^+ (whose properties are recalled in Proposition 1.1) leads to

$$(1.16) \qquad\qquad F(t,x) = \rho(t,x)\mathbf{1} + \frac{\varepsilon}{\sigma}\partial_x \rho(t,x)D^+V + O(\varepsilon^2),$$

which is indeed the discrete counterpart of the desired result (1.4).

REMARK 1.2. *In* (1.12), *The Duhamel formula has been written on the time interval* $[0,t]$ *but it can also be written on the time interval* $[t_n, t]$. *In this case, it comes*

$$F_j(t,x) = \exp\left(\frac{\lambda_\star \sigma(t - t_n)}{\eta \varepsilon}\right)F_j\left(t_n, x - \frac{v_j(t - t_n)}{\eta}\right)$$

$$(1.17) \qquad + \langle e_j, \int_0^{\frac{\sigma(t - t_n)}{\eta \varepsilon}}\exp(\lambda_\star u)(D - \lambda_\star I)F\left(t - \frac{\eta \varepsilon u}{\sigma}, x - \frac{v_j \varepsilon u}{\sigma}\right) du\rangle.$$

2. Numerical scheme. In this section, we will describe a generalized UGKS for linear equations of the form (1.6). First of all, we introduce a uniform spatial mesh $x_i = i\Delta x$ is introduced and denote $x_{i+1/2} = (x_i + x_{i+1})/2$ the interface between two cells. The spatial interval being $x \in [0, L], L > 0$, the mesh step is defined as $\Delta x = L/N_x$, N_x being the number of cells. Moreover, we introduce the time discretization $t_n = n\Delta t, \Delta t > 0, n \in \mathbb{N}$.

To derive a UGK scheme, one of the main ingredient relies on a suitable interface value $F_j(t, x_{i+1/2})$, usually based on a Duhamel formula (see [19, 16]), that will serve in the finite volume formulation as a flux approximation of the space derivative in (1.6). Another ingredient is the space reconstruction of $F_j(t, x)$ in the integral term of Duhamel formula (1.17). After recalling the UGKS, we will present some spatial reconstructions and we will see that the general context considered here induces some difficulties.

6

2.1. UGKS. Our goal is to design a UGKS type numerical scheme for a general class of linear equations (1.1). Let recall the basics of UGKS framework (see [19, 16]). First, starting from (1.6), we define the averages of the density and distribution function on cell i at time t_n

$$\begin{pmatrix} \rho_i^n \\ F_{i,j}^n \end{pmatrix} = \frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} \begin{pmatrix} \rho(t_n, x) \\ F_j(t_n, x) \end{pmatrix} dx,$$

and the macroscopic and microscopic numerical fluxes across the interface $x_{i+1/2}$

$$\begin{pmatrix} \Phi_{i+1/2} \\ \phi_{i+1/2,j} \end{pmatrix} = \frac{1}{\eta \Delta t} \int_{t_n}^{t_{n+1}} \begin{pmatrix} \sum_{j=1}^{2N} v_j F_j(t, x_{i+1/2}) \Delta v \\ v_j F_j(t, x_{i+1/2}) \end{pmatrix} dt.$$

The finite volume formulations of the macroscopic and kinetic equations are thus

(2.1) $$\frac{\rho_i^{n+1} - \rho_i^n}{\Delta t} + \frac{1}{\Delta x}(\Phi_{i+1/2} - \Phi_{i-1/2}) = 0,$$

(2.2) $$\frac{F_{i,j}^{n+1} - F_{i,j}^n}{\Delta t} + \frac{1}{\Delta x}(\phi_{i+1/2,j} - \phi_{i-1/2,j}) = \frac{\sigma}{\eta \varepsilon} D F_{i,j}^{n+1},$$

where the collision term is implicit for stability reason.

As we can see, the interface value of $F_j(t, x_{i+1/2})$ plays an important role since it enables to compute the numerical fluxes of the finite volume method. As mentioned above, the main idea of UGKS relies on the suitable approximation of the interface value of $F_j(t, x_{i+1/2})$ obtained from a space approximation of a Duhamel formula. Indeed, the interface value is obtained from a suitable space approximation of the Duhamel formula of the original equation. Here, we will use the Duhamel formula (1.17) which includes a penalization procedure with the factor $\lambda_\star$ and we will discuss the space approximation that ensures consistency and good asymptotic behavior. Contrary to the BGK case studied in [19, 16] where the distribution function $F_j(t, x)$ and the density $\rho(t, x)$ have to be reconstructed, the general case considered here only involves $F_j(t, x)$. In the sequel, different reconstructions are discussed and we will see that some choices require the full knowledge of the spectral decomposition of the linear collision operator, which is not acceptable from a computational point of view. Indeed, our goal is to design a UGKS scheme which is computationally efficient, preserves the diffusion limit $\eta = \varepsilon \to 0$ and the collisionless limit η fixed and $\varepsilon \to \infty$.

2.2. Spatial Approximation: first attempt. A first (natural) reconstruction of F_j on the spatial mesh is:

(2.3) $$F_j(t, x) = F_{i+1/2,j}(t) + \frac{F_j(t, x_{i+1}) - F_j(t, x_i)}{\Delta x}(x - x_{i+1/2}), \quad x \in [x_i, x_{i+1}].$$

The reconstruction is inserted in (1.12) to get an approximation of the interface value $F_{i+1/2,j}(t) \approx F_j(t, x_{i+1/2})$. Using the reconstruction (2.3), the integral term in (1.12) evaluated at $x = x_{i+1/2}$ enables to get the following approximation

$$F\left(t - \frac{\eta \varepsilon(u - t_n)}{\sigma}, x_{i+1/2} - \frac{v_j \varepsilon u}{\sigma}\right) \approx F\left(t, x_{i+1/2} - \frac{v_j \varepsilon u}{\sigma}\right)$$

$$= F_{i+1/2}(t) + \frac{F_{i+1}(t) - F_i(t)}{\Delta x}\left(x_{i+1/2} - \frac{\varepsilon v_j u}{\sigma} - x_{i+1/2}\right)$$

(2.4) $$= F_{i+1/2}(t) - \frac{v_j \varepsilon u}{\sigma}\frac{F_{i+1}(t) - F_i(t)}{\Delta x}.$$

Let us remark that in the first approximation the shift in time has been neglected, as usual in the UGKS for the diffusion regime.

Regarding now the first term in (1.12), we consider a first order approximation in space based on the sign of v_j. We thus obtain

$$(2.5) \qquad F_j\left(t_n, x_{i+1/2} - \frac{v_j(t - t_n)}{\eta\varepsilon}\right) \approx F_{i,j}^n \mathbb{1}_{v_j > 0} + F_{i+1,j}^n \mathbb{1}_{v_j < 0},$$

where $\mathbb{1}_{v_j < 0}$ denotes the Heaviside function which is equal to one if $v_j < 0$ and zero else (same for $\mathbb{1}_{v_j > 0}$).

Thus, evaluating (1.17) at $x = x_{i+1/2}$ and using the previous approximations (2.4) and (2.5), we obtain the following approximation $F_{i+1/2,j}(t)$ of the interface value $F_j(t, x_{i+1/2})$:

$$F_{i+1/2,j}(t) := \exp\left(\frac{\lambda_\star \sigma(t - t_n)}{\eta\varepsilon}\right)\left(F_{i,j}^n \mathbb{1}_{v_j > 0} + F_{i+1,j}^n \mathbb{1}_{v_j < 0}\right)$$

$$(2.6) \qquad + \langle e_j, \int_0^{\frac{\sigma(t-t_n)}{\eta\varepsilon}} \exp\left(\lambda_\star u\right)\left(D - \lambda_\star I\right)\left[F_{i+1/2}(t) - \frac{\varepsilon v_j u}{\sigma}\frac{F_{i+1}(t) - F_i(t)}{\Delta x}\right] du\rangle.$$

This approximation can be inserted in (2.2) to get a first version of the scheme but we can observe the resulting scheme will couple the space and velocity indices so that the calculation of $F_{i,j}^{n+1}$ will be very costly. Moreover, as observed in (1.4), capturing the correct asymptotic behavior requires to introduce D^+ which is not the case with this first attempt. Thus, we will next try to work only with $F_{i+1/2}(t)$ and not with $F_i(t)$ and $F_{i+1}(t)$.

2.3. Spatial Approximation: second attempt. Thus, instead of the reconstruction (2.3), we propose the following reconstruction:

$$(2.7) \qquad F_j(t, x) = F_{i+1/2,j}(t) + \frac{F_{i+1,j}^n - F_{i,j}^n}{\Delta x}\left(x - x_{i+1/2}\right),$$

where we neglect the temporal variations of spatial gradients ($t_n \leq t \leq \Delta t$). Thus, evaluating (1.17) at $x = x_{i+1/2}$, using the reconstruction (2.7) for the integral term and (2.5) for the first term, we obtain:

$$F_{i+1/2,j}(t) := \exp\left(\frac{\lambda_\star \sigma(t - t_n)}{\eta\varepsilon}\right)\left(F_{i,j}^n \mathbb{1}_{v_j > 0} + F_{i+1,j}^n \mathbb{1}_{v_j < 0}\right)$$

$$(2.8) \qquad + \langle e_j, \int_0^{\frac{\sigma(t-t_n)}{\eta\varepsilon}} \exp\left(\lambda_\star u\right)\left(D - \lambda_\star I\right)\left[F_{i+1/2}(t) - \frac{\varepsilon v_j u}{\sigma}\frac{F_{i+1}^n - F_i^n}{\Delta x}\right] du\rangle.$$

From the relation (2.8), $F_{i+1/2}(t)$ can now be expressed as the solution of the following linear system (which is local in space):

$$(2.9) \qquad M(t)F_{i+1/2}(t) = S(t),$$

where the j-th component of the source term $S(t)$ is:

$$S_j(t) = \exp\left(\frac{\lambda_\star \sigma(t - t_n)}{\eta\varepsilon}\right)\left(F_{i,j}^n \mathbb{1}_{v_j > 0} + F_{i+1,j}^n \mathbb{1}_{v_j < 0}\right)$$

$$(2.10) \qquad + \langle v_j e_j, \int_0^{\frac{\sigma(t-t_n)}{\eta\varepsilon}} \exp\left(\lambda_\star u\right)\left(D - \lambda_\star I\right)\left[-\frac{\varepsilon u}{\sigma}\frac{F_{i+1}^n - F_i^n}{\Delta x}\right] du\rangle,$$

and the matrix $M(t)$ is given by

$$M(t) = I - \int_0^{\frac{\sigma(t-t_n)}{\eta\varepsilon}} \exp\left(\lambda_\star u\right)\left(D - \lambda_\star I\right)du$$

$$(2.11) \qquad = \exp\left(\frac{\lambda_\star \sigma(t - t_n)}{\eta\varepsilon}\right)I + \left(1 - \exp\left(\frac{\lambda_\star \sigma(t - t_n)}{\eta\varepsilon}\right)\right)\left(\frac{1}{\lambda_\star}D\right).$$

Some important properties of the matrix $M(t)$ are given in the following proposition.

PROPOSITION 2.1. *The matrix $M(t)$ defined by* (2.11) *is positive definite and thus invertible and its inverse is given by*

$$(2.12) \qquad M(t)^{-1} = \sum_{k=0}^{N'} \mathcal{A}_k^{-1} P_k,$$

where N' is the number of eigenvalues, P_k denotes the orthogonal projectors associated to the eigenvalues λ_k of D, and $\mathcal{A}_k \in \mathbb{R}$ are defined as follows

$$(2.13) \qquad \mathcal{A}_k = \exp\left(\frac{\lambda_\star \sigma(t - t_n)}{\eta \varepsilon}\right) + \left(\frac{\lambda_k}{\lambda_\star}\right)\left(1 - \exp\left(\frac{\lambda_\star \sigma(t - t_n)}{\eta \varepsilon}\right)\right).$$

Proof. One can observe that $M(t)$ is a convex combination of the two positive matrices I and $\frac{1}{\lambda_\star}D$ (since $\lambda_\star < 0$ and D is negative). Moreover, it is always positive definite since I is, $\left(\frac{1}{\lambda_\star}D\right)$ is positive, and the coefficient in front of I is always strictly positive, so the system will always be numerically invertible.

Concerning the inverse $M(t)^{-1}$, we recall the projector properties from Prop. 1.1: $\sum_{k=0}^{N'} P_k = I$ and $\sum_{k=1}^{N'} \lambda_k P_k = D$. Indeed, using these relations in (2.11), one has

$$M(t) = \exp\left(\frac{\lambda_\star \sigma(t - t_n)}{\eta \varepsilon}\right)\sum_{k=0}^{N'} P_k + \left(1 - \exp\left(\frac{\lambda_\star \sigma(t - t_n)}{\eta \varepsilon}\right)\right)\sum_{k=1}^{N'}\left(\frac{\lambda_k}{\lambda_\star}\right)P_k,$$

which reads: $M(t) = \sum_{k=0}^{N'} \mathcal{A}_k P_k$. The expression (2.12) of the inverse of $M(t)$ is deduced from this latter form. $\qquad \square$

Let us remark that the case $k = 0$ in (2.12) will play an important role since $\lambda_0 = 0$ and for $k = 0$, we have $\mathcal{A}_0 = \exp\left(\frac{\lambda_\star \sigma(t-t_n)}{\eta \varepsilon}\right)$.

2.3.1. Computation of the Interface Value. From the relation (2.9) satisfied by the interface value $F_{i+1/2}(t)$ and the expressions (2.10) of $S(t)$ and (2.12) of $M(t)^{-1}$, we will compute explicitly the interface value $F_{i+1/2}(t)$. In order to avoid the explicit calculation of the projectors P_k and thus to get a simple numerical scheme, some approximations will be performed.

First, let introduce some useful notations. The half densities associated with positive and negative velocities are

$$(2.14) \qquad \rho_i^{-,n} = \frac{1}{2N}\sum_{j=1}^{N} F_{i,j}^n = \Delta v \sum_{j=1}^{N} F_{i,j}^n \quad \text{and} \quad \rho_i^{+,n} = \frac{1}{2N}\sum_{j=N+1}^{2N} F_{i,j}^n.$$

Then, we introduce the coefficient

$$(2.15) \quad \mathcal{C}(t) = \lambda_\star \int_0^{\frac{\sigma(t-t_n)}{\eta \varepsilon}} \exp\left(\lambda_\star u\right) u\, du = \frac{1}{\lambda_\star}\left[1 + \left(\frac{\lambda_\star \sigma(t - t_n)}{\eta \varepsilon} - 1\right)\exp\left(\frac{\lambda_\star \sigma(t - t_n)}{\eta \varepsilon}\right)\right].$$

We now compute the interface value $F_{i+1/2}$ using the expressions (2.10) of $S(t)$ and (2.12) of $M(t)^{-1}$:

$$
\begin{aligned}
F_{i+1/2}(t) &= M(t)^{-1}S(t) \\
&= P_0 \left\{ F_{i,j}^n \mathbb{1}_{v_j>0} + F_{i+1,j}^n \mathbb{1}_{v_j<0} \right\}_{j=1,\ldots,2N} \\
&\quad - \frac{\varepsilon \mathcal{C}(t)}{\sigma} \frac{1}{2N} \langle (DV - \lambda_\star V), \frac{F_{i+1}^n - F_i^n}{\Delta x} \rangle \mathbf{1} \\
&\quad + \sum_{k=1}^{N'} \mathcal{A}_k^{-1} P_k \left\{ \exp\left(\frac{\lambda_\star \sigma(t - t_n)}{\eta \varepsilon} \right) \left(F_{i,j}^n \mathbb{1}_{v_j>0} + F_{i+1,j}^n \mathbb{1}_{v_j<0} \right) \right\}_{j=1,\ldots,2N} \\
&\quad + \sum_{k=1}^{N'} \mathcal{A}_k^{-1} P_k \left\{ \langle v_j e_j, \frac{\mathcal{C}(t)}{\lambda_\star} (D - \lambda_\star I) \left[-\frac{\varepsilon}{\sigma} \frac{F_{i+1}^n - F_i^n}{\Delta x} \right] \rangle \right\}_{j=1,\ldots,2N},
\end{aligned}
$$

(2.16)
$$
= \text{ⓐ} + \text{ⓑ} + \text{ⓒ} + \text{ⓓ},
$$

where we used the notations $\{w_j\}_{j=1,\ldots,2N}$ the vector $w \in \mathbb{R}^{2N}$ and we remind $\mathcal{A}_k$ is defined by (2.13). Let us now detail how we deal with terms ⓐ, ⓑ, ⓒ, ⓓ.

2.3.2. Computation of the terms ⓐ and ⓑ. Let consider in this part the first two terms ⓐ and ⓑ. First, from the definition (2.14) of the half densities, one has for ⓐ

(2.17)
$$
\text{ⓐ} = \left(\rho_i^{+,n} + \rho_{i+1}^{-,n} \right) \mathbf{1}.
$$

Second, we provide some details regarding ⓑ. Recalling $(D - \lambda_\star I)$ is symmetric and from the definition (2.15) of $\mathcal{C}$, one has

$$
\begin{aligned}
\text{ⓑ} &= -\lambda_\star \frac{\varepsilon}{\sigma} \frac{1}{2N} \langle (D - \lambda_\star I)V, \frac{\mathcal{C}(t)}{\lambda_\star} \frac{F_{i+1}^n - F_i^n}{\Delta x} \rangle \mathbf{1} \\
&= \lambda_\star \frac{1}{2N} \langle (D - \lambda_\star I)V, \int_0^{\frac{\sigma(t-t_n)}{\eta\varepsilon}} \exp(\lambda_\star u) \left[-\frac{\varepsilon u}{\sigma} \frac{F_{i+1}^n - F_i^n}{\Delta x} \right] du \rangle \mathbf{1} \\
&= \lambda_\star \frac{1}{2N} \langle V, (D - \lambda_\star I) \int_0^{\frac{\sigma(t-t_n)}{\eta\varepsilon}} \exp(\lambda_\star u) \left[-\frac{\varepsilon u}{\sigma} \frac{F_{i+1}^n - F_i^n}{\Delta x} \right] du \rangle \mathbf{1} \\
&= \lambda_\star \frac{1}{2N} \sum_{k=1}^{2N} \langle v_k e_k, \int_0^{\frac{\sigma(t-t_n)}{\eta\varepsilon}} \exp(\lambda_\star u)(D - \lambda_\star I) \left[-\frac{\varepsilon u}{\sigma} \frac{F_{i+1}^n - F_i^n}{\Delta x} \right] du \rangle \mathbf{1}.
\end{aligned}
$$

(2.18)

Let us remark that this term is independent of j so that it has no contribution in the diffusion limit.

2.3.3. Computation of the terms ⓒ and ⓓ. Now, we will consider the terms ⓒ and ⓓ for which some approximations will be performed to avoid the explicit calculation of the projectors P_k.

First let recall the expression of the term ⓒ

$$
\text{ⓒ} = \sum_{k=1}^{N'} \mathcal{A}_k^{-1} P_k \exp\left(\frac{\lambda_\star \sigma(t - t_n)}{\eta \varepsilon} \right) \left\{ F_{i,j}^n \mathbb{1}_{v_j>0} + F_{i+1,j}^n \mathbb{1}_{v_j<0} \right\}_{j=1,\ldots,2N}.
$$

On the one side, we observe that the term ⓒ decays exponentially fast and thus does not contribute

in the diffusion limit. On the other side, in the transport limit $\varepsilon \to +\infty$, one has

$$\sum_{k=1}^{N'} \mathcal{A}_k^{-1} P_k = \sum_{k=1}^{N'} \frac{1}{\exp\left(\frac{\lambda_\star \sigma(t-t_n)}{\eta\varepsilon}\right) + \frac{\lambda_k}{\lambda_\star}\left(1 - \exp\left(\frac{\lambda_\star \sigma(t-t_n)}{\eta\varepsilon}\right)\right)} P_k$$

$$(2.19) \qquad = \sum_{k=1}^{N'} P_k + O(1/\varepsilon) = (I - P_0) + O(1/\varepsilon).$$

Hence, from these asymptotic behaviors, and in view of constructing a method which does not require the knowledge of P_k, we propose the following approximation for ⓒ

$$(2.20) \qquad ⓒ \approx (I - P_0)\exp\left(\frac{\lambda_\star \sigma(t - t_n)}{\eta\varepsilon}\right)\left\{F_{i,j}^n \mathbb{1}_{v_j>0} + F_{i+1,j}^n \mathbb{1}_{v_j<0}\right\}_{j=1,\ldots,2N}.$$

Let now consider the term ⓓ and first, let recall its expression

$$(2.21) \qquad ⓓ = \sum_{k=1}^{N'} \mathcal{A}_k^{-1} P_k \left\{\left\langle v_j e_j, \frac{\mathcal{C}(t)}{\lambda_\star}(D - \lambda_\star I)\left[-\frac{\varepsilon}{\sigma}\frac{F_{i+1}^n - F_i^n}{\Delta x}\right]\right\rangle\right\}_{j=1,\ldots,2N}.$$

Then, we observe that in the diffusion limit, we have, for any arbitrary $q \in \mathbb{N}$:

$$(2.22) \qquad \sum_{k=1}^{N'} \frac{\mathcal{A}_k^{-1}}{\lambda_\star} P_k = \sum_{k=1}^{N'} \frac{1}{\lambda_k} P_k + O(\varepsilon^q) = D^+ + O(\varepsilon^q),$$

since from the definition (2.13) of $\mathcal{A}_k$, we deduce $\mathcal{A}_k^{-1}$ tends to $\lambda_\star/\lambda_k$ as $\varepsilon \to 0$ and from Prop. 1.1, we have $D = \sum_{k=1}^{N'} \lambda_k P_k$. However, little can be said about the term in braces in (2.21) except that $\lim_{\eta=\varepsilon\to 0} \mathcal{C}(t) = 1/\lambda_\star$. Moreover, using the approximation (2.22) imposes the term into braces to be orthogonal to $\mathbf{1}$ which is not the case due to the presence of $F_{i+1}^n - F_i^n$. Hence, the calculation of the term ⓓ would require the knowledge of the projectors P_k which we want to avoid since it may be very costly in the general case. Hence, in the next subsection, we will consider another reconstruction which will avoid the knowledge of the projectors P_k.

2.4. Spatial Approximation: third attempt. Due to the obstacle observed previously, we thus consider the following reconstruction for $F_j(t, x)$

$$(2.23) \qquad F_j(t, x) = F_{i+1/2,j}(t) + \frac{\rho_{i+1}^n - \rho_i^n}{\Delta x}(x - x_{i+1/2}).$$

As before, assessing (1.17) at $x = x_{i+1/2}$, using reconstruction (2.23) for the integral term and (2.5) for the first term, we get the following interface relation for $F_{i+1/2,j}(t)$

$$F_{i+1/2,j}(t) = \exp\left(\frac{\lambda_\star \sigma(t - t_n)}{\eta\varepsilon}\right)\left(F_{i,j}^n \mathbb{1}_{v_j>0} + F_{i+1,j}^n \mathbb{1}_{v_j<0}\right)$$

$$(2.24) \qquad + \left\langle e_j, \int_0^{\frac{\sigma(t-t_n)}{\eta\varepsilon}} \exp\left(\lambda_\star u\right)(D - \lambda_\star I)\left[F_{i+1/2}(t) - \frac{\varepsilon v_j u}{\sigma}\frac{\rho_{i+1}^n - \rho_i^n}{\Delta x}\mathbf{1}\right] du\right\rangle.$$

The same calculations as the ones done before can be performed since the expression of $M(t)^{-1}$ given by (2.12) is unchanged, but the expression of $S(t)$ in (2.10) is slightly modified since its

11

expression is now

$$S_j(t) = \exp\left(\frac{\lambda_\star \sigma(t-t_n)}{\eta\varepsilon}\right)\left(F^n_{i,j}\mathbb{1}_{v_j>0} + F^n_{i+1,j}\mathbb{1}_{v_j<0}\right)$$

(2.25)
$$+ \langle v_j e_j, \int_0^{\frac{\sigma(t-t_n)}{\eta\varepsilon}} \exp\left(\lambda_\star u\right)(D - \lambda_\star I)\left[-\frac{\varepsilon u}{\sigma}\frac{\rho^n_{i+1} - \rho^n_i}{\Delta x}\mathbf{1}\right] du\rangle.$$

So now $F_{i+1/2}$ satisfies $F_{i+1/2}(t) = M(t)^{-1}S(t)$ and decompose $F_{i+1/2}(t) = ⓐ + ⓑ + ⓒ + ⓓ$ as:

$$ⓐ = P_0\left\{F^n_{i,j}\mathbb{1}_{v_j>0} + F^n_{i+1,j}\mathbb{1}_{v_j<0}\right\}_{j=1,\ldots,2N}$$

$$ⓑ = -\frac{\varepsilon\mathcal{C}(t)}{\sigma}\frac{1}{2N}\langle(DV - \lambda_\star V), \frac{\rho^n_{i+1} - \rho^n_i}{\Delta x}\rangle\mathbf{1}$$

$$ⓒ = \sum_{k=1}^{N'}\mathcal{A}_k^{-1}P_k\exp\left(\frac{\lambda_\star\sigma(t-t_n)}{\eta\varepsilon}\right)\left\{F^n_{i,j}\mathbb{1}_{v_j>0} + F^n_{i+1,j}\mathbb{1}_{v_j<0}\right\}_{j=1,\ldots,2N}$$

$$ⓓ = \sum_{k=1}^{N'}\mathcal{A}_k^{-1}P_k\left\{\langle v_j e_j, \frac{\mathcal{C}(t)}{\lambda_\star}(D - \lambda_\star I)\left[-\frac{\varepsilon}{\sigma}\frac{\rho^n_{i+1} - \rho^n_i}{\Delta x}\mathbf{1}\right]\rangle\right\}_{j=1,\ldots,2N}.$$

As previously, we now deal with the decomposition. Since ⓐ and ⓒ do not depend on the gradient reconstruction, they are unchanged and respectively given by (2.17) and (2.20); we focus on ⓒ and ⓓ.

2.4.1. Calculation of the term ⓑ. Regarding the term ⓑ, we now have the following expression

$$ⓑ = -\frac{\varepsilon\mathcal{C}(t)}{\sigma}\frac{1}{2N}\langle(D - \lambda_\star I)V, \frac{\rho^n_{i+1} - \rho^n_i}{\Delta x}\rangle\mathbf{1}$$

$$= -\frac{\varepsilon\mathcal{C}(t)}{\sigma}\frac{1}{2N}\langle V, (D - \lambda_\star I)\frac{\rho^n_{i+1} - \rho^n_i}{\Delta x}\rangle\mathbf{1}$$

$$= \frac{\varepsilon\lambda_\star\mathcal{C}(t)}{\sigma}\frac{1}{2N}\langle V, \frac{\rho^n_{i+1} - \rho^n_i}{\Delta x}\rangle\mathbf{1} = \frac{\varepsilon\lambda_\star\mathcal{C}(t)}{\sigma}\frac{1}{2N}\sum_{k=1}^{2N}v_k\frac{\rho^n_{i+1} - \rho^n_i}{\Delta x}\mathbf{1} = \mathbf{0},$$

since the velocity grid satisfies $\sum_{k=1}^{2N}v_k = 0$ by assumption.

2.4.2. Calculation of the term ⓓ. Let now investigate the term ⓓ. Using the reconstruction (2.24), the term ⓓ becomes

$$ⓓ = \sum_{k=1}^{N'}\mathcal{A}_k^{-1}P_k\left\{\langle v_j e_j, \mathcal{C}(t)(D - \lambda_\star I)\left[-\frac{\varepsilon}{\sigma}\frac{\rho^n_{i+1} - \rho^n_i}{\Delta x}\mathbf{1}\right]\rangle\right\}_{j=1,\ldots,2N}$$

$$= D^+\left\{\langle v_j e_j, \mathcal{C}(t)(D - \lambda_\star I)\left[-\frac{\varepsilon}{\sigma}\frac{\rho^n_{i+1} - \rho^n_i}{\Delta x}\mathbf{1}\right]\rangle\right\}_{j=1,\ldots,2N} + \mathcal{O}(\varepsilon^q)$$

$$= \mathcal{C}(t)\lambda_\star D^+\left\{\langle v_j e_j, \frac{\varepsilon}{\sigma}\frac{\rho^n_{i+1} - \rho^n_i}{\Delta x}\mathbf{1}\rangle\right\}_{j=1,\ldots,2N} = \mathcal{C}(t)\lambda_\star D^+\left[\frac{\varepsilon}{\sigma}\frac{\rho^n_{i+1} - \rho^n_i}{\Delta x}\right]V$$

(2.26)
$$= \mathcal{C}(t)\lambda_\star\left[\frac{\varepsilon}{\sigma}\frac{\rho^n_{i+1} - \rho^n_i}{\Delta x}\right]U,$$

where we used $D\mathbf{1} = 0$ and $D^+V = U$. We can observe that with the reconstruction (2.24), the knowledge of the projectors P_k can be avoided but as we will see, the asymptotic behavior can still be recovered.

12

2.4.3. Computation of the Interface value. All in all, inserting the expressions (2.17), (2.20) and (2.26) of ⓐ, ⓒ and ⓓ (recalling that ⓑ = 0) in (2.16) leads to the following expression for the interface value:

$$F_{i+1/2}(t) = \left(\rho_i^{+,n} + \rho_{i+1}^{-,n}\right)\mathbf{1}$$

$$+ (I - P_0)\exp\left(\frac{\lambda_\star\sigma(t-t_n)}{\eta\varepsilon}\right)\left\{F_{i,j}^n\mathbf{1}_{v_j>0} + F_{i+1,j}^n\mathbf{1}_{v_j<0}\right\}_{j=1,\dots,2N}$$

$$+ \lambda_\star\mathcal{C}(t)\left[\frac{\varepsilon}{\sigma}\frac{\rho_{i+1}^n - \rho_i^n}{\Delta x}\right]U$$

$$= \exp\left(\frac{\lambda_\star\sigma(t-t_n)}{\eta\varepsilon}\right)\left\{F_{i,j}^n\mathbf{1}_{v_j>0} + F_{i+1,j}^n\mathbf{1}_{v_j<0}\right\}_{j=1,\dots,2N}$$

(2.27)
$$+ \left(1 - \exp\left(\frac{\lambda_\star\sigma(t-t_n)}{\eta\varepsilon}\right)\right)\left(\rho_i^{+,n} + \rho_{i+1}^{-,n}\right)\mathbf{1} + \lambda_\star\mathcal{C}(t)\left[\frac{\varepsilon}{\sigma}\frac{\rho_{i+1}^n - \rho_i^n}{\Delta x}\right]U.$$

This expression of $F_{i+1/2}(t)$ will be inserted in the finite volume scheme which will lead to a new UGKS described hereafter.

2.5. New UGKS. The complete scheme consists of considering (2.1)-(2.2) with the following definition for the flux $\phi_{i+1/2,j}$

$$\phi_{i+1/2,j} = \frac{1}{\eta\Delta t}\int_{t_n}^{t_{n+1}} v_j F_{i+1/2,j}(t)dt,$$

where the expression (2.27) is used for the interface value $F_{i+1/2,j}(t)$. Some calculation enables to get the following explicit expression for the flux:

$$\phi_{i+1/2,j} = \frac{1}{\eta\Delta t}\int_{t_n}^{t_n+\Delta t} v_j\left[\exp\left(\frac{\lambda_\star\sigma(t-t_n)}{\eta\varepsilon}\right)\left(F_{i,j}^n\mathbf{1}_{v_j>0} + F_{i+1,j}^n\mathbf{1}_{v_j<0}\right)\right.$$

$$\left.+ \left(1 - \exp\left(\frac{\lambda_\star\sigma(t-t_n)}{\eta\varepsilon}\right)\right)\left(\rho_i^{+,n} + \rho_{i+1}^{-,n}\right) + \lambda_\star\mathcal{C}(t)\frac{\varepsilon}{\sigma}\frac{\rho_{i+1}^n - \rho_i^n}{\Delta x}U_j\right]dt$$

$$= A(\Delta t, \sigma, \eta, \varepsilon)\left(F_{i,j}^n\mathbf{1}_{v_j>0} + F_{i+1,j}^n\mathbf{1}_{v_j<0}\right)v_j + C(\Delta t, \sigma, \eta, \varepsilon)\left(\rho_i^{+,n} + \rho_{i+1}^{-,n}\right)v_j$$

(2.28)
$$+ D(\Delta t, \sigma, \eta, \varepsilon)\frac{\rho_{i+1}^n - \rho_i^n}{\Delta x}\lambda_\star U_j v_j,$$

where the coefficients A,C,D are given by (using the notation $w = \lambda_\star\sigma\Delta t/(\eta\varepsilon)$):

(2.29)
$$A(\Delta t, \sigma, \eta, \varepsilon) = \frac{1}{\eta w}(e^w - 1)$$

(2.30)
$$C(\Delta t, \sigma, \eta, \varepsilon) = \frac{1}{\eta} - A(\Delta t, \sigma, \eta, \varepsilon)$$

(2.31)
$$D(\Delta t, \sigma, \eta, \varepsilon) = \frac{\varepsilon}{\sigma\lambda_\star}(C(\Delta t, \sigma, \eta, \varepsilon) - A(\Delta t, \sigma, \eta, \varepsilon)) + \frac{\varepsilon}{\eta\sigma\lambda_\star}e^w.$$

This exactly matches the scheme proposed by Mieussens in [16] for the BGK operator, if the interface value for the density used in [16] is set to $(\rho_i^n + \rho_{i+1}^n)/2$. Once the fluxes have been written, it remains to define the scheme satisfied by $F_i^{n+1} \in \mathbb{R}^{2N}$ from (2.2)

(2.32)
$$\left(I - \frac{\sigma\Delta t}{\varepsilon\eta}D\right)F_i^{n+1} = F_i^n - \frac{\Delta t}{\Delta x}\left(\phi_{i+1/2} - \phi_{i-1/2}\right).$$

13

Due to the implicit treatment of the collision operator D, a linear system has to be inverted for each spatial cell, which is fully aligned with other strategies for stiff kinetic problems.

Let us now consider the macroscopic flux $\Phi_{i+1/2}$ which is defined by

$$\Phi_{i+1/2} = \frac{1}{2N} \sum_{j=1}^{2N} \phi_{i+1/2,j},$$

where the (microscopic) flux $\phi_{i+1/2,j}$ is given by (2.28). Some simple calculations enable to get

$$\Phi_{i+1/2} = \mathrm{A}(\Delta t, \sigma, \eta, \varepsilon)(J_i^{+,n} + J_i^{-,n}) + \mathrm{D}(\Delta t, \sigma, \eta, \varepsilon) \frac{\lambda_\star}{2N} \frac{\rho_{i+1}^n - \rho_i^n}{\Delta x} \langle U, V \rangle,$$

with $J_i^{\pm,n}$ are given by

$$J_i^{-,n} = \frac{1}{2N} \sum_{j=1}^{N} v_j F_{i,j}^n \quad \text{and} \quad J_i^{+,n} = \frac{1}{2N} \sum_{j=N+1}^{2N} v_j F_{i,j}^n.$$

But from Prop. 1.1, one has $\lambda_\star \langle U, V \rangle = \langle V, V \rangle$ so that the macroscopic flux $\Phi_{i+1/2}$ finally becomes

$$(2.33) \qquad \Phi_{i+1/2} = \mathrm{A}(\Delta t, \sigma, \eta, \varepsilon)(J_i^{+,n} + J_i^{-,n}) + \frac{\mathrm{D}(\Delta t, \sigma, \eta, \varepsilon)}{2N} \frac{\rho_{i+1}^n - \rho_i^n}{\Delta x} \langle V, V \rangle,$$

and the macroscopic density is updated by as

$$(2.34) \qquad \frac{\rho_i^{n+1} - \rho_i^n}{\Delta t} + \frac{\Phi_{i+1/2} - \Phi_{i-1/2}}{\Delta x} = 0.$$

REMARK 2.1.
- *Let us remark that, contrary to the standard UGKS applied to the BGK equation, the update of ρ^{n+1} is not required to update the kinetic unknown F^{n+1} in (2.32). Indeed, in this version of UGKS, F^{n+1} is updated from (2.32) and then, ρ^{n+1} is computed from (2.34).*
- *Solving (2.32) requires a linear solver (Conjugate Gradient method in this paper) in each spatial cell if D has no clear properties. When D comes from BGK model, the trick is to update ρ through (2.34): the system becomes diagonal. When D is linked to FP model, one has to use Thomas algorithm for tridiagonal matrices for more efficiency and precision (the matrix is badly conditioned for Conjugate Gradient method).*

3. Asymptotic behavior. Here, we formally investigate the asymptotic behavior of the new UGKS (2.32)-(2.28)-(2.34)-(2.33) presented above. First, we study the free transport regime $\varepsilon \to +\infty$ and then the diffusion limit $\eta = \varepsilon \to 0$ is studied.

3.1. Free transport regime. The asymptotic behavior follows from the ones of the coefficients defined in (2.29), (2.30), (2.31).

PROPOSITION 3.1. *When $\varepsilon \to +\infty$ (and η constant), one has*
- *$\mathrm{A}(\Delta t, \sigma, \eta, \varepsilon)$ tends to $\frac{1}{\eta}$.*
- *$\mathrm{C}(\Delta t, \sigma, \eta, \varepsilon)$ tends to 0.*
- *$\mathrm{D}(\Delta t, \sigma, \eta, \varepsilon)$ tends to 0.*

As a consequence, the flux (2.28) verifies when $\varepsilon \to +\infty$

$$\phi_{i+1/2,j} \to \frac{v_j}{\eta} \left(F_{i,j}^n \mathbb{1}_{v_j>0} + F_{i+1,j}^n \mathbb{1}_{v_j<0} \right),$$

14

and the scheme becomes

$$F_{i,j}^{n+1} - F_{i,j}^n + \frac{v_j \Delta t}{\eta \Delta x} \left(F_{i,j}^n \mathbb{1}_{v_j > 0} + F_{i+1,j}^n \mathbb{1}_{v_j < 0} \right) = 0,$$

which is the standard first order upwind scheme for the free transport equation $\partial_t f + \frac{v}{\eta} \partial_x f = 0$.

3.2. Diffusion regime. We now investigate the diffusion limit $\eta = \varepsilon \to 0$.

PROPOSITION 3.2. *When $\eta = \varepsilon \to 0$, one has*
- *$A(\Delta t, \sigma, \eta, \varepsilon)$ tends to 0,*
- *$D(\Delta t, \sigma, \eta, \varepsilon)$ tends to $\frac{1}{\sigma \lambda_\star}$.*

As a consequence, the macroscopic flux $\Phi_{i+1/2}$ defined by (2.33) satisfies, as $\eta = \varepsilon \to 0$

$$(3.1) \qquad \Phi_{i+1/2} \to \frac{1}{2N} \frac{\rho_{i+1}^n - \rho_i^n}{\Delta x} \frac{\langle V, V \rangle}{\sigma \lambda_\star}.$$

Hence, the equation (2.34) becomes

$$\frac{\rho_i^{n+1} - \rho_i^n}{\Delta t} + \frac{\langle V, V \rangle/(2N)}{\sigma \lambda_\star} \frac{\rho_{i+1}^n - \rho_i^n + \rho_{i-1}^n}{\Delta x^2} = 0.$$

But from the definition (1.10) of $\lambda_\star$, we get $\langle V, V \rangle / \lambda_\star = \langle D^+ V, V \rangle$ which is consistent with the diffusion coefficient κ occurring in (1.5).

4. Extensions and remarks. We now present some extensions and remarks.

4.1. Implicit diffusion. The version where the limit scheme is implicit is obtained simply by modifying the flux as:

$$\phi_{i+1/2,j} = A(\Delta t, \sigma, \eta, \varepsilon) \left(F_{i,j}^n \mathbb{1}_{v_j > 0} + F_{i+1,j}^n \mathbb{1}_{v_j < 0} \right) v_j$$
$$(4.1) \qquad + C(\Delta t, \sigma, \eta, \varepsilon) \left(\rho_i^{+,n} + \rho_{i+1}^{-,n} \right) v_j + D(\Delta t, \sigma, \eta, \varepsilon) \frac{\rho_{i+1}^{n+1} - \rho_i^{n+1}}{\Delta x} \lambda_\star U_j v_j.$$

However, to compute ρ_i^{n+1}, the macroscopic fluxes $\Phi_{i+1/2,j}$ are now obtained by summing on j the microscopic fluxes (4.1). From the previous calculations, we get

$$\Phi_{i+1/2} = \frac{1}{2N} \sum_{j=1}^{2N} \phi_{i+1/2,j}$$
$$(4.2) \qquad = A(\Delta t, \sigma, \eta, \varepsilon)(J_i^{+,n} + J_{i+1}^{-,n}) + \frac{\langle V, V \rangle}{2N} D(\Delta t, \sigma, \eta, \varepsilon) \frac{\rho_{i+1}^{n+1} - \rho_i^{n+1}}{\Delta x},$$

so that the scheme on ρ finally becomes (with the notations $J_{i+1/2}^n = J_{i+1}^{-,n} + J_i^{+,n}$):

$$\frac{\rho_i^{n+1} - \rho_i^n}{\Delta t} = -A(\Delta t, \sigma, \eta, \varepsilon) \frac{J_{i+1/2}^n - J_{i-1/2}^n}{\Delta x} - \frac{\langle V, V \rangle}{2N} D(\Delta t, \sigma, \eta, \varepsilon) \frac{\rho_{i+1}^{n+1} - 2\rho_i^{n+1} + \rho_{i-1}^{n+1}}{\Delta x^2},$$

where A and D are given by (2.29) and (2.31).

REMARK 4.1. *Since we have an implicit diffusion in space, one needs first to solve ρ^{n+1} through a linear solver and then solve F^{n+1}.*

4.2. Some remarks for the BGK operator. In the BGK case, the computations simplify significantly. Indeed, first, all the eigenvalues $\lambda_k (k \geq 1)$ are equal to $\lambda_\star = -1$, and $\lambda_0 = 0$. Hence, the coefficients $\mathcal{A}_k$ defined in (2.13) becomes independent of k and can be written as

$$\mathcal{A}_k = \exp\left(\frac{\lambda_\star \sigma(t - t_n)}{\eta\varepsilon}\right) + \frac{\lambda_k}{\lambda_\star}\left(1 - \exp\left(\frac{\lambda_\star \sigma(t - t_n)}{\eta\varepsilon}\right)\right) = 1,$$

so that the inverse $M(t)^{-1}$ defined in (2.12) of $M(t)$ defined in (2.11) becomes:

$$M(t)^{-1} = \exp\left(-\frac{\lambda_\star \sigma(t - t_n)}{\eta\varepsilon}\right)P_0 + \sum_{k=1}^{N'}\mathcal{A}_k^{-1}P_k = \exp\left(-\frac{\lambda_\star \sigma(t - t_n)}{\eta\varepsilon}\right)P_0 + (I - P_0).$$

Besides, in the BGK case, one can observe that $D = P_0 + \lambda_\star I$ or $D - \lambda_\star I = P_0$. It turns out that for the BGK case, the second reconstruction (2.7) and the third reconstruction (2.23) gives the same expression for $S(t)$ defined by (2.10) and (2.25) respectively. Indeed, the reconstruction gives (2.7)

$$
\begin{aligned}
S_j(t) &= \exp\left(\frac{\lambda_\star \sigma(t - t_n)}{\eta\varepsilon}\right)\left(F_{i,j}^n \mathbb{1}_{v_j>0} + F_{i+1,j}^n \mathbb{1}_{v_j<0}\right) \\
&\quad + \left\langle v_j e_j, \int_0^{\frac{\sigma t}{\eta\varepsilon}} \exp\left(\lambda_\star u\right)P_0\left[-\frac{\varepsilon u}{\sigma}\frac{F_{i+1}^n - F_i^n}{\Delta x}\right]du\right\rangle \\
(4.3)\quad &= \exp\left(\frac{\lambda_\star \sigma(t - t_n)}{\eta\varepsilon}\right)\left(F_{i,j}^n \mathbb{1}_{v_j>0} + F_{i+1,j}^n \mathbb{1}_{v_j<0}\right) + \frac{\mathcal{C}(t)v_j}{\lambda_\star}\frac{\varepsilon}{\sigma}\frac{\rho_{i+1}^n - \rho_i^n}{\Delta x},
\end{aligned}
$$

which is the expression that we would have obtained using the third reconstruction (2.23) (let recall the definition (2.15) of $\mathcal{C}(t)$). We can observe that the second term in (4.3) is collinear to V, ie orthogonal to $\mathrm{Ker}\, D$ or $P_0 V = 0$. Solving the system $M(t)F_{i+1/2}(t) = S(t)$ thus leads to:

$$
\begin{aligned}
F_{i+1/2}(t) &= M(t)^{-1}S(t) \\
&= P_0\left\{F_{i,j}^n\mathbb{1}_{v_j>0} + F_{i+1,j}^n\mathbb{1}_{v_j<0}\right\}_{j=1,\ldots,2N} + \mathbf{0} \\
&\quad + (I - P_0)\exp\left(\frac{\lambda_\star\sigma(t-t_n)}{\eta\varepsilon}\right)\left\{F_{i,j}^n\mathbb{1}_{v_j>0} + F_{i+1,j}^n\mathbb{1}_{v_j<0}\right\}_{j=1,\ldots,2N} + \frac{\varepsilon}{\sigma}\frac{\rho_{i+1}^n - \rho_i^n}{\Delta x}\frac{\mathcal{C}(t)V}{\lambda_\star} \\
&= \left(1 - \exp\left(\frac{\lambda_\star\sigma(t-t_n)}{\eta\varepsilon}\right)\right)\left(\rho_i^{+,n} + \rho_{i+1}^{-,n}\right)\mathbf{1} \\
&\quad + \exp\left(\frac{\lambda_\star\sigma(t-t_n)}{\eta\varepsilon}\right)\left\{F_{i,j}^n\mathbb{1}_{v_j>0} + F_{i+1,j}^n\mathbb{1}_{v_j<0}\right\}_{j=1,\ldots,2N} + \frac{\varepsilon}{\sigma}\frac{\rho_{i+1}^n - \rho_i^n}{\Delta x}\frac{\mathcal{C}(t)V}{\lambda_\star}.
\end{aligned}
$$

We observe that this expression and the one obtained in [16] match except for the last term in which half-slopes reconstructions have been used for ρ in [16]. This is due to the fact that in [16], the BGK model requires a reconstruction for F and for ρ which is not the case in this work.

4.3. Some remarks for the Fokker-Planck operator. To apply our scheme to Fokker-Planck operator, we need to give a discretization of $\mathcal{D}$. Moreover, we discuss how the interface value $F_{i+1/2}(t)$ behave in this case.

4.3.1. Discretization of the collision operator. First, let recall that for the continuous Fokker-Planck operator defined by (1.3), the eigenvalues are $\lambda_k = -k(k+1), k \in \mathbb{N}$. The zero eigenvalue is labelled by $k = 0$ and the eigenvalue λ_1 is associated with the eigenfunction v (which can easily recovered from (1.3)). Let us remark that we will choose $\lambda_\star = \lambda_1$ since this is the eigenvalue which enables to get the correct diffusion limit.

Second, let introduce a velocity discretization D of $\mathcal{D}$ given by (1.3). With the definitions introduced above, we use (we recall that we have chosen $c = 1$)

$$(4.4) \qquad (\mathcal{D}f)(v_j) \approx (DF)_j = \frac{1}{\Delta v^2}\Big((1 - v_{j+1/2}^2)(F_{j+1} - F_j) - (1 - v_{j-1/2}^2)(F_j - F_{j-1})\Big),$$

where we defined $v_{j+1/2} = (v_j + v_{j+1})/2$. Using this approximation, one can check that the first two eigenvalues $\lambda_0 = 0$ and $\lambda_1 = -2$ are exactly recovered.

PROPOSITION 4.1. *The discretization* (4.4) *satisfies* $D\mathbf{1} = 0$ *and* $DV = -2V$. *Besides D is symmetric, its non diagonal coefficients are positive and $I + \delta D$ is bistochastic for $\delta > 0$ small enough.*

Proof. The fact that $\lambda_0 = 0$ is true is a direct consequence of the conservative form of the discretization (4.4) so that $D\mathbf{1} = 0$.

Regarding the relation, let replace F_j by v_j in (4.4):

$$\begin{aligned}
(DV)_j &= \frac{1}{\Delta v^2}\Big((1 - v_{j+1/2}^2)(v_{j+1} - v_j) - (1 - v_{j-1/2}^2)(v_j - v_{j-1})\Big)\\
&= \frac{1}{\Delta v^2}\Big((1 - v_{j+1/2}^2)\Delta v - (1 - v_{j-1/2}^2)\Delta v\Big)\\
&= \frac{1}{\Delta v}(-v_{j+1/2}^2 + v_{j-1/2}^2) = -2v_j,
\end{aligned}$$

so that $DV = -2V$ and V is an eigenvector associated to the eigenvalue $\lambda_\star = -2$.

By construction, we clearly have that D is symmetric, its non diagonal coefficients are positive and $I + \delta D$ is a bistochastic matrix for $\delta > 0$ small enough. $\qquad\square$

4.3.2. Interface value. In the Fokker-Planck case, the eigenstructure of D is more involved than in the BGK case so that the third reconstruction (2.23) is required to avoid the knowledge of the projectors P_k. However, it can be noticed that for the second reconstruction (2.7), the term ⓑ defined in (2.18) vanishes. Indeed, thanks to Prop. 4.1, one has $(D - \lambda_\star I)V = 0$, even if $F_{i,j}^n$ is employed for the gradient.

4.4. Some remarks for the scattering operator. For scattering operator defined by (1.2), the resulting matrix D must be symmetric negative with only $\mathbf{1}$ in the kernel and positive coefficients apart on the diagonal. Spectrum and projectors are not always exactly known, thus $\lambda_\star$ has to be numerically determined from (1.10) and in this case reconstruction (2.24) is fully justified. Up to our knowledge, there is no physically relevant non isotropic 1D collision operator. Therefore, to test our approach, we artificially define the following scattering matrix $D = \frac{1}{10}D_1$ with D_1 the periodic Laplacian matrix given by :

$$(D_1)_{i,j} = \frac{1}{\Delta v^2}\begin{cases} -2, & \text{if } i = j,\\ 1, & \text{if } i = j - 1[N_v] \text{ or } i = j + 1[N_v] \end{cases} \quad (N_v = 2N)$$

which owns the required properties (negativity, kernel, positive coefficients). The factor $1/10$ is chosen so that the $\lambda_\star$ value in this case is between the one of BGK model and the one FP model. The eigenvalues of D_1 are known:

$$\lambda_k = -\frac{2}{\Delta v^2}\left(1 - \cos\left(\frac{2\pi k}{N_v}\right)\right), \quad k = 0, 1, \ldots, N_v - 1.$$

The numerical value for $\lambda_\star(D)$ is obtained by the code for $N_v = 100$ is roughly -1.49835.

REMARK 4.2. *The vector V is not an eigenvector since eigenvectors are discrete Fourier modes. Using Fourier analysis one gets that when Δv tends to zero:*

- *D_1 tends to the Laplacian so that U obeys to $U''(v) = v$ ($v \in [-1, 1]$) with periodic boundary conditions $U(1) = U(-1)$, U and satisfies $\frac{1}{2} \int_{-1}^{1} U(v) dv = 0$ (to be in the orthogonal of the constants). The solution U is $U(v) = \frac{1}{6} v^3 - \frac{1}{6} v$.*
- *One can compute the corresponding pseudo-eigenvalue $\lambda_\star$ by considering the limit as $\Delta v \to 0$ in (1.10). Using $\int_{-1}^{1} U(v) v dv = -1/45$ and $\int_{-1}^{1} v^2 dv = 1/3$, one has*

$$\lambda_\star(D_1) := \frac{\langle V, V \rangle}{\langle U, V \rangle} \xrightarrow[\Delta v \to 0]{} \left(\frac{1}{2} \int_{-1}^{1} v^2 dv \right) \bigg/ \left(\frac{1}{2} \int_{-1}^{1} U(v) v dv \right) = -15.$$

Since $D = \frac{1}{10} D_1$, $\lambda_\star(D)$ converges to $\lambda_\star^\infty(D) = -1.5$ in the continuous limit.

5. Numerical results.

We now provide some numerical results to illustrate the properties of the new UGKS to solve (1.1) for generalized collision kernels, namely the scattering operator (1.2) and the Fokker-Planck operator (1.3). In all cases presented below, the equation (1.1) is equipped with periodic conditions in space (more details on other boundary conditions can be found in [16] and the new scheme only needs slight adaptations on the boundaries). The initial particle distribution function and density are chosen far from equilibrium:

$$f_0(x, v) = \exp\left(-(x - 0.5)^2 - 10(1 - v)^2 \right), \quad 0 \le x \le 1, \ -1 \le v \le 1.$$

$$\rho_0(x) = \frac{1}{2} \int_{-1}^{1} f_0(x, v) \, dv = C \exp\left(-(x - 0.5)^2 \right),$$

with $\quad C = \frac{1}{2} \int_{-1}^{1} \exp\left(-10(1 - v)^2 \right) dv \approx 0.14 \qquad$ (see figure (5.1)).
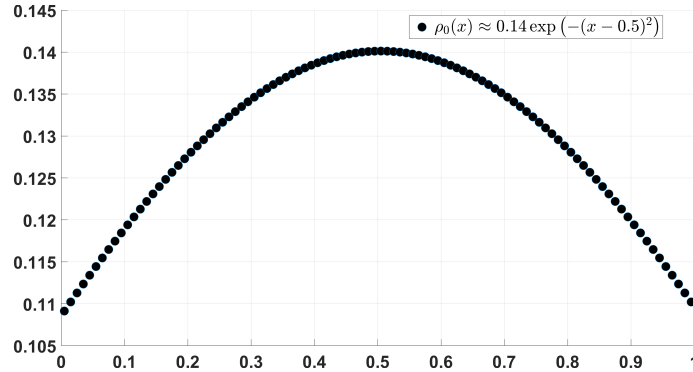

FIG. 5.1. *Initial density $x \mapsto \rho_0(x)$*

The exact solution of the transport problem without any collision is

(5.1) $$f(t, x, v) = f_0\big((x - vt) \bmod 1, v\big) = \exp\left(-\big((x - vt) \bmod 1 - 0.5\big)^2 - 10(1 - v)^2 \right),$$

while the exact solution for density in the diffusive limit with diffusion coefficient $\kappa = \dfrac{1}{3\sigma |\lambda_\star|}$ is:

(5.2) $$\rho(t, x) = C \int_0^1 \left(\sum_{j=-\infty}^{\infty} \frac{1}{\sqrt{4\pi\kappa t}} \exp\left(-\frac{(x - y + j)^2}{4\kappa t} \right) \right) \exp\left(-(y - 0.5)^2 \right) dy.$$

18

.

In the sequel, we focus on the comparison of different collision operators and their corresponding diffusion coefficients κ: the scattering operator (1.2) for which $\kappa = \frac{2}{9\sigma}$ (since $\lambda_\star = -3/2$), the Fokker-Planck operator (1.3) for which $\kappa = \frac{1}{6\sigma}$ (since $\lambda_\star = -2$) and BGK operator for which $\kappa = \frac{1}{3\sigma}$ (since $\lambda_\star = -1$). Regarding the stability, it is empirically found that the new UGKS has the same stability condition as the standard one studied in [16, 18]. The time step is thus chosen so that the natural condition $\Delta t = C_1 \Delta x^2 + C_2 \eta \Delta x$ is empirically satisfied. We refer to the work of Vigier ([18]) for more discussions on the subject.

In the following tests, the mesh in space uses $N_x = 100$ cells whereas $N_v = 100$ cells in velocity ($N = 50$) are considered. The quantity σ will always be set to one. Hence, from the empirical CFL condition, the time step is set to $\Delta t = 10^{-5}$.

The next subsections investigate the three different regimes (transport, intermediate and diffusive) for the three models at different times. In the legends BGK means BGK model, FP means Fokker-Planck model and SC refers to the scattering model.

5.1. Transport regime: $\eta = 1$ and $\varepsilon = 100$. First, we consider the so-called transport regime in (1.1) with $\eta = 1$ and $\varepsilon = 100$. In this regime, the collision part is weakened due to the large value of ε and we capture the transport of the density: initially the mean velocity of the distribution is uniform and strictly positive so that a translation of the bump to the right is expected. In Figure 5.2, we can observe that due to the first order scheme in transport, diffusion in space and smoothing of the bump occur. In this regime, there is almost no collisions so that consequently the three kernels of collisions give the same results at intermediate time $t_i = 0.05$ and final time $t_f = 0.1$.
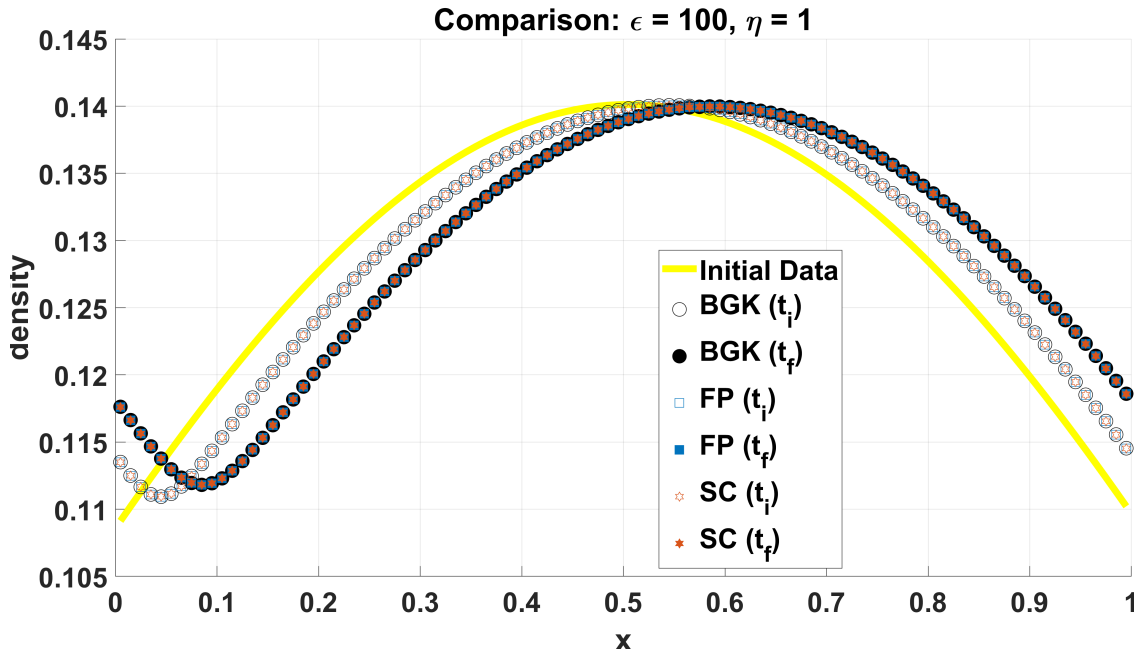

FIG. 5.2. *Transport regime ($\varepsilon = 100, \eta = 1$): density $x \mapsto \rho(t, x)$ for $t_i = 0.05, t_f = 0.1$ computed by UGKS for BGK, Fokker-Planck and scattering collision kernels.*

We can also compare with exact solution of the transport solution given by (5.1) and denoted by Transp at intermediate time and final time: we can see in Figure 5.3 that the three models have exactly the same behavior and that the error comes from the first order upwind scheme.
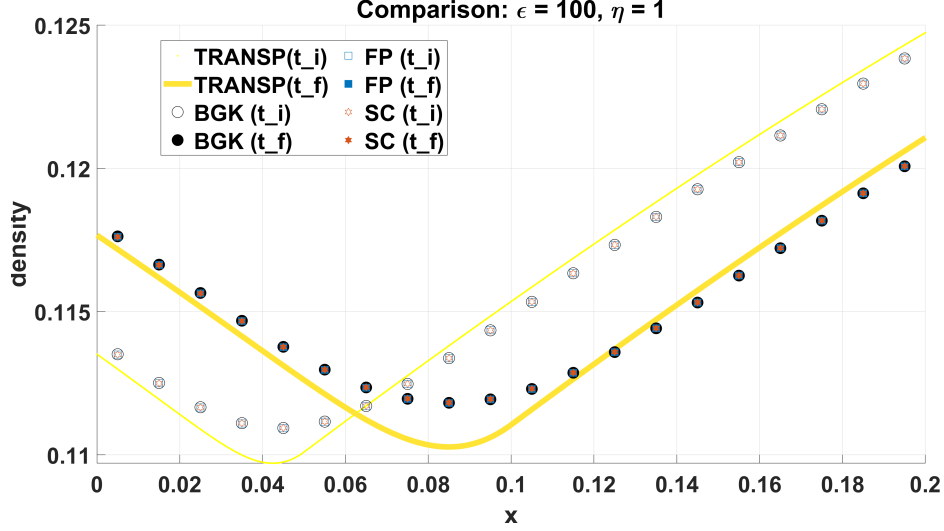
19

FIG. 5.3. *Transport regime ($\varepsilon = 100, \eta = 1$): density $x \mapsto \rho(t,x)$ for $t_i = 0.05, t_f = 0.1$ computed by UGKS for BGK, Fokker-Planck and scattering collision kernels compared with the exact solution.*

5.2. Intermediate regime: $\eta = 0.1$ **and** $\varepsilon = 0.1$. Now, we investigate the intermediate regime for which $\eta = \varepsilon = 0.1$ in (1.1). In this regime both transport and collisions are acting: transport is initially acting (the bump has stopped moving at time $t_i = 0.05$) and afterwards the bump is only damped till $t_f = 0.1$. Consequently, due to the transport part which is the same for all the models, the bump goes to the right but due to the different diffusion coefficient, the damping is different according to the collision operator: the diffusion coefficient of the Fokker-Planck model is half the one of BGK model and three quarters of the scattering model. Thus, one can observe on Figure 5.4 that the BGK solution (black bullets) is more flattened than the one obtained by the scattering operator (orange stars) which is itself more flattened than the Fokker-Planck operator (blue squares) at time $t_f = 0.1$.
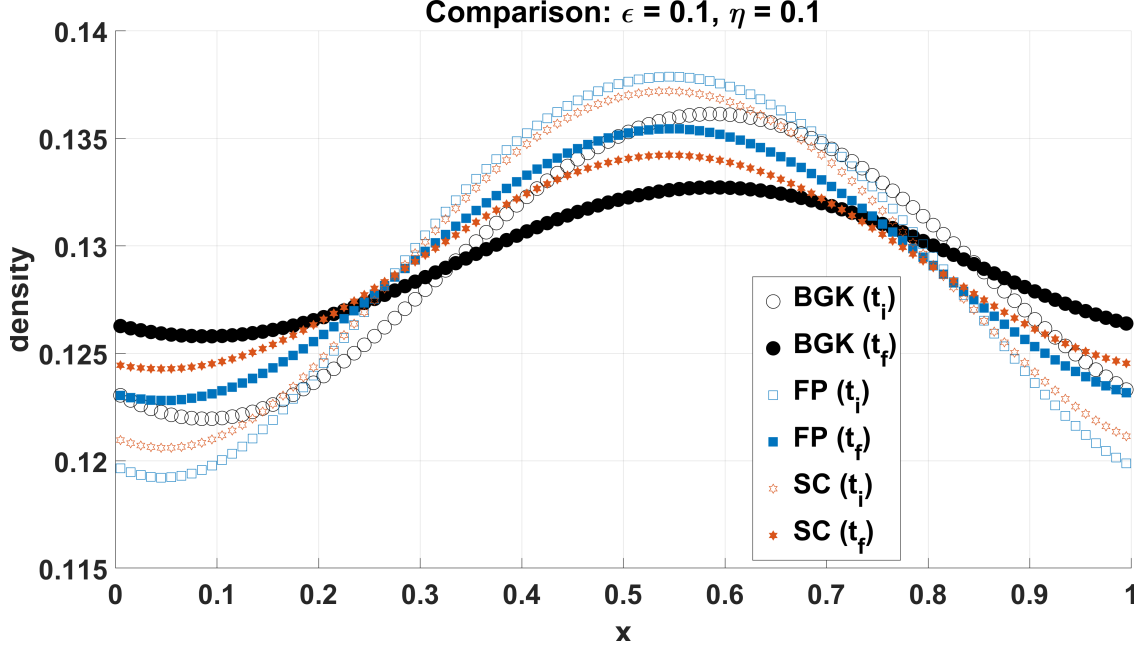
FIG. 5.4. *Intermediate regime ($\varepsilon = \eta = 0.1$): density $x \mapsto \rho(t, x)$ for $t_i = 0.05, t_f = 0.1$ computed by UGKS for BGK, Fokker-Planck and scattering collision kernels.*

5.3. Diffusive regime $\eta = 10^{-4}$ and $\varepsilon = 10^{-4}$. We finally consider the diffusion regime $\eta = \varepsilon = 10^{-4}$. In this regime, the effect of the transport is negligible and the solution is supposed to be closed to the one of the diffusion model (1.5) for which the diffusion coefficient is different according to the collision operator. As a consequence, since the diffusion coefficient in the Fokker-Planck model is half the one of the BGK model, the Fokker-Planck model at time t_f should be very close to the BGK model at time $t_i = t_f/2$. This is observed in Figure 5.5 where the Fokker-Planck density (blue squares) is superimposed with the BGK one (black circles). Moreover, in Figure 5.5, the densities obtained by the three collision operators at t_i and t_f are gradually damped from Fokker-Planck at t_i to BGK at t_f.
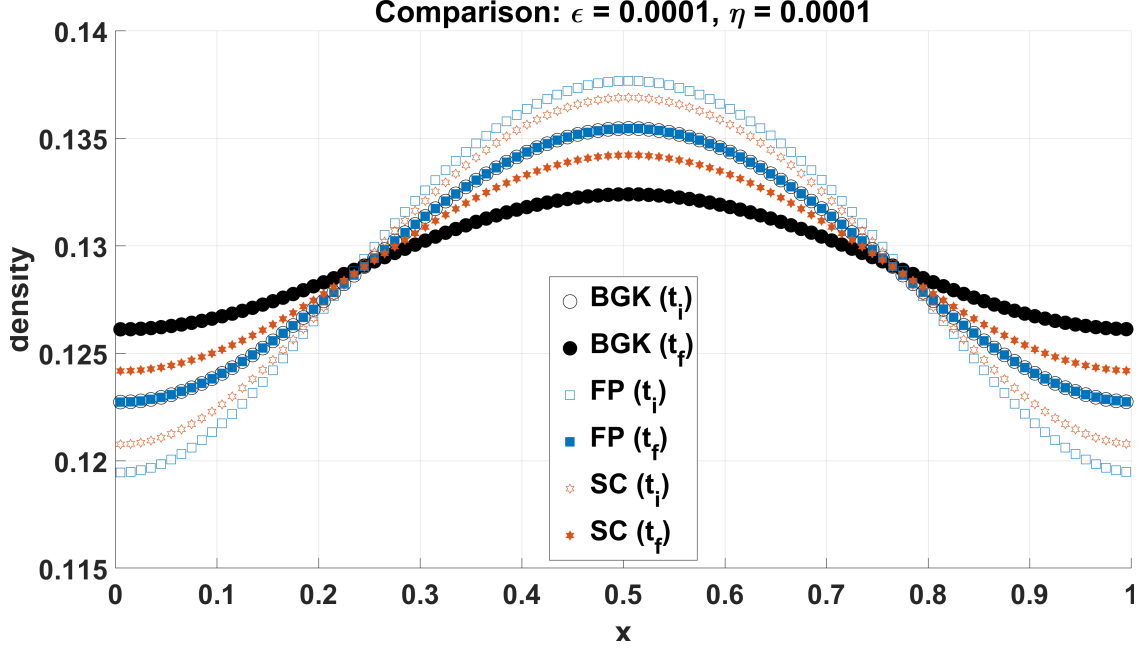
FIG. 5.5. *Diffusive regime ($\varepsilon = \eta = 10^{-4}$): density $x \mapsto \rho(t,x)$ for $t = 0.05$ and $t = 0.1$ computed by UGKS for BGK, Fokker-Planck and scattering collision kernels.*

Similarly, if we set $t_f = 0.075$, the scattering model (orange stars) at time t_f gives very similar result as the BGK model (black circles) at time $t_i = 0.05$ as observed in Figure 5.6. Indeed, the diffusion coefficient for the scattering operator is $\lambda_{\star,SC} = -1.5$ whereas $\lambda_{\star,BGK} = -1$ for the BGK, so that $\frac{t_f}{\lambda_{\star,SC}} = \frac{t_i}{\lambda_{\star,BGK}}$.
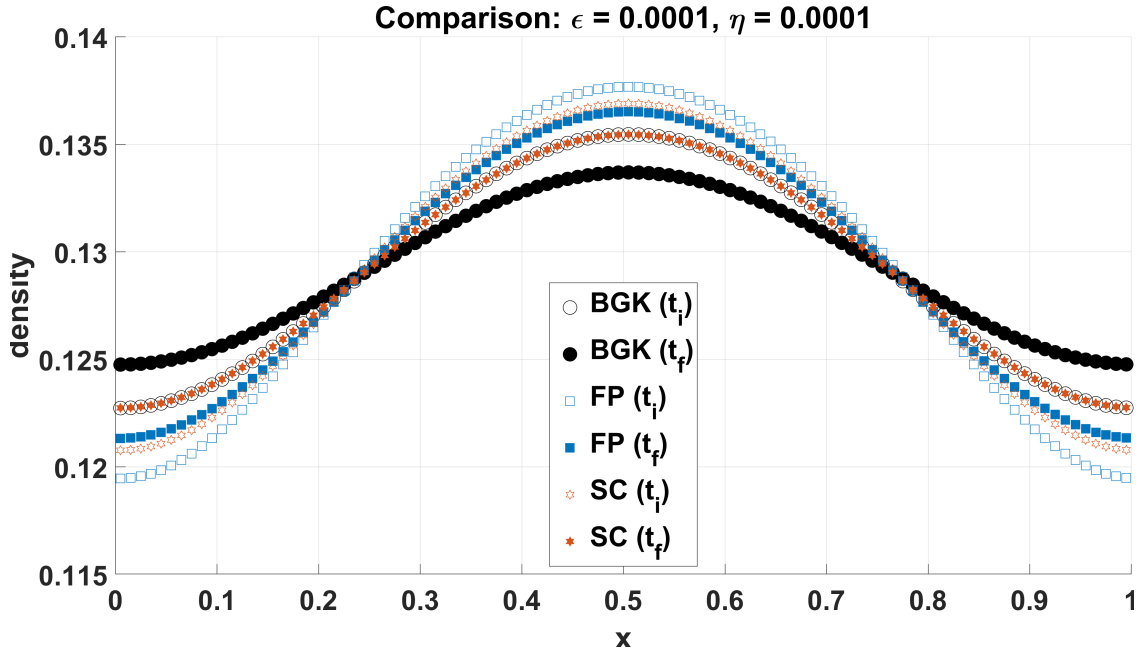

FIG. 5.6. *Diffusive regime ($\varepsilon = \eta = 10^{-4}$): density $x \mapsto \rho(t,x)$ for $t = 0.05$ and $t = 0.075$ computed by UGKS for BGK, Fokker-Planck and scattering collision kernels.*

22

5.3.1. Comparison with exact solution. We can also compare the exact solutions of the Fokker-Planck and scattering operators with their respective analytical solutions (5.2), denoted by $DIFF_{FP}$ and $DIFF_{SC}$, at the final time $t = 0.1$. In Figure 5.7, one can observe that the results are quite satisfactory after 10^4 time steps, given the small number of velocity points and the use of a first-order time scheme.
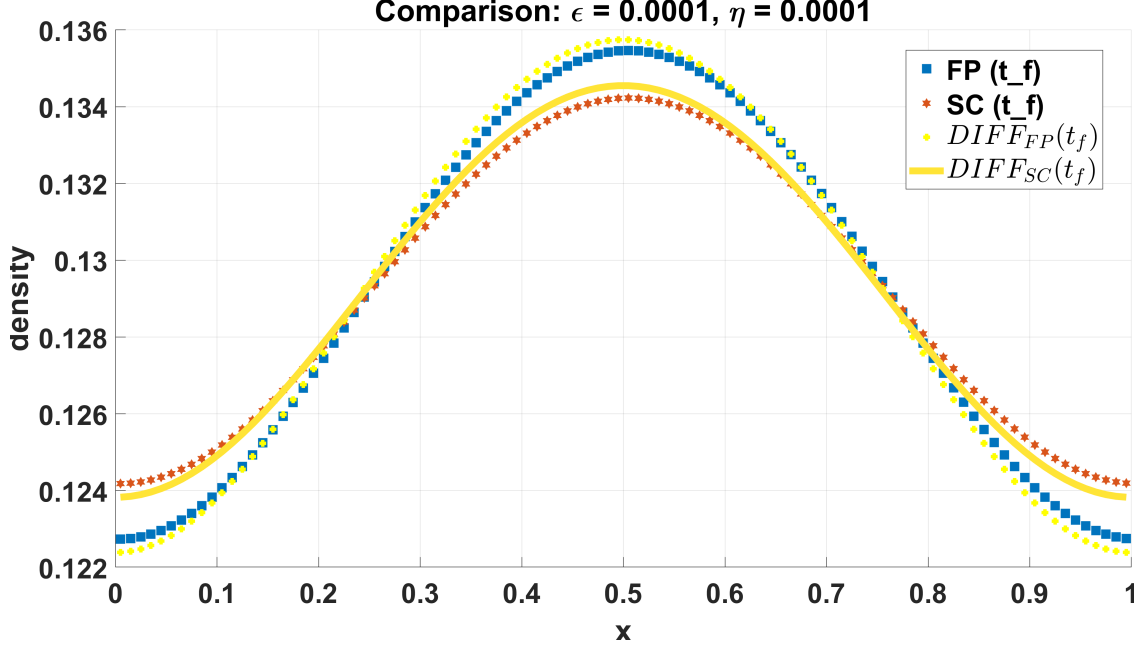

FIG. 5.7. *Diffusive regime* ($\varepsilon = \eta = 10^{-4}$)*: density* $x \mapsto \rho(t, x)$ *for* $t = 0.1$ *computed by UGKS for Fokker-Planck and scattering collision kernels compared with their exact solutions.*

Conclusion. In this work, we have developed and analysed a generalized Unified Gas Kinetic Scheme (UGKS) designed to handle linear kinetic equations in the diffusive limit, with a particular focus on extending the original framework to accommodate a broad class of collision operators, such as the BGK, Fokker–Planck, and scattering models. Building on the asymptotic preserving (AP) structure of the classical UGKS, our formulation is based on a penalized Duhamel representation and a suitable spatial reconstruction that preserve the essential asymptotic behaviors without relying on costly spectral decomposition.

From the theoretical standpoint, we formally established the preservation of the diffusion limit as $\varepsilon = \eta \to 0$. In particular, we showed that the scheme yields the appropriate diffusion equation for each collision operator. The practical effectiveness of the method was demonstrated through several numerical experiments covering a range of physical regimes.

The implicit treatment of the stiff collision term leads to local linear systems that are well-posed, and can be solved using standard solvers. The proposed interface value computation, essential for flux evaluations, preserves both accuracy and efficiency.

These results collectively demonstrate that the generalized UGKS scheme provides a versatile numerical framework for simulating general kinetic models across multiple scales. As a future work, we may explore the extension of the scheme to nonlinear problems and rigorous stability analysis.

REFERENCES

[1] M. BENNOUNE, M. LEMOU, AND L. MIEUSSENS, *Uniformly stable numerical schemes for the Boltzmann equation preserving the compressible Navier–Stokes asymptotics*, Journal of Computational Physics, 227 (2008), pp. 3781–3803.

[2] C. Buet, S. Cordier, B. Lucquin-Desreux, and S. Mancini, *Diffusion limit of the Lorentz model: asymptotic preserving schemes*, ESAIM: Mathematical Modelling and Numerical Analysis, 36 (2002), pp. 631–655.

[3] J.-A. Carrillo, T. Goudon, and P. Lafitte, *Simulation of fluid and particles flows: Asymptotic preserving schemes for bubbling and flowing regimes*, Journal of Computational Physics, 227 (2008), pp. 7929–7951.

[4] J. A. Carrillo, T. Goudon, P. Lafitte, and F. Vecil, *Numerical schemes of diffusion asymptotics and moment closures for kinetic equations*, Journal of Scientific Computing, 36 (2008), pp. 113–149.

[5] F. Filbet and S. Jin, *An asymptotic preserving scheme for the ES-BGK model of the Boltzmann equation*, Journal of Scientific Computing, 46 (2011), pp. 204–224, https://doi.org/10.1007/s10915-010-9394-x.

[6] L. Gosse, *Transient radiative transfer in the grey case: Well-balanced and asymptotic-preserving schemes built on Case's elementary solutions*, Journal of Quantitative Spectroscopy and Radiative Transfer, 112 (2011), pp. 1995–2012.

[7] S. Jin and C. D. Levermore, *Fully-discrete numerical transfer in diffusive regimes*, Transport theory and statistical physics, 22 (1993), pp. 739–791.

[8] S. Jin and D. Levermore, *The discrete-ordinate method in diffusive regimes*, Transport theory and statistical physics, 20 (1991), pp. 413–439.

[9] S. Jin, L. Pareschi, and G. Toscani, *Uniformly accurate diffusive relaxation schemes for multiscale transport equations*, SIAM Journal on Numerical Analysis, 38 (2000), pp. 913–936.

[10] A. Klar, *An asymptotic-induced scheme for nonstationary transport equations in the diffusive limit*, SIAM Journal on Numerical Analysis, 35 (1998), pp. 1073–1094.

[11] A. Klar and C. Schmeiser, *Numerical passage from radiative heat transfer to nonlinear diffusion models*, Mathematical Models and Methods in Applied Sciences, 11 (2001), pp. 749–767.

[12] P. Lafitte and G. Samaey, *Asymptotic-preserving projective integration schemes for kinetic equations in the diffusion limit*, SIAM Journal on Scientific Computing, 34 (2012), pp. A579–A602.

[13] M. Lemou and L. Mieussens, *A new asymptotic preserving scheme based on micro-macro formulation for linear kinetic equations in the diffusion limit*, SIAM Journal on Scientific Computing, 31 (2008), pp. 334–368.

[14] C. Liu and K. Xu, *A unified gas kinetic scheme for continuum and rarefied flows V: multiscale and multi-component plasma transport*, Communications in Computational Physics, 22 (2017), pp. 1175–1223.

[15] C. Liu, K. Xu, Q. Sun, and Q. Cai, *A Unified Gas-Kinetic Scheme for continuum and rarefied flow IV: Full Boltzmann and model equations*, Journal of Computational Physics, 314 (2016), pp. 305–340, https://doi.org/10.1016/j.jcp.2016.03.008.

[16] L. Mieussens, *On the Asymptotic Preserving property of the Unified Gas Kinetic Scheme for the diffusion limit of linear kinetic models*, Journal of Computational Physics, 253 (2013), pp. 138–156.

[17] W. Sun, S. Jiang, K. Xu, and S. Li, *An asymptotic preserving unified gas kinetic scheme for frequency-dependent radiative transfer equations*, Journal of Computational Physics, 302 (2015), pp. 222–238.

[18] T. Vigier, *Méthodes numériques préservant l'asymptotique pour des modèles aux moments de la physique des plasmas*, theses, Université de Bordeaux, Dec. 2024, https://theses.hal.science/tel-04901697.

[19] K. Xu and J.-C. Huang, *A unified gas-kinetic scheme for continuum and rarefied flows*, Journal of Computational Physics, 229 (2010), pp. 7747–7764.

[20] Y. Zhu and K. Xu, *The first decade of unified gas kinetic scheme*, arXiv preprint arXiv:2102.01261, (2021).