*draft, for submission to SINUM, October 2, 2009*

**Analysis of an asymptotic preserving scheme for linear kinetic equations in the diffusion limit**

Jian-Guo Liu[1], Luc Mieussens[2]

**Abstract.** We present a mathematical analysis of the asymptotic preserving scheme proposed in [M. Lemou and L. Mieussens, *SIAM J. Sci. Comput.*, 31, pp. 334–368, 2008] for linear transport equations in kinetic and diffusive regimes. We prove that the scheme is uniformly stable and accurate with respect to the mean free path of the particles. This property is satisfied under an explicitly given CFL condition. This condition tends to a parabolic CFL condition for small mean free paths, and is close to a convection CFL condition for large mean free paths. Our analysis is based on very simple energy estimates.

**Key words.** transport equations, diffusion limit, asymptotic preserving schemes, stiff terms, stability analysis

**AMS subject classifications.** 65M06, 35B25, 82C80, 41A60

# 1 Introduction

Particle systems are often described at the microscopic level by kinetic models (neutron transport, radiative transfer, electrons in semi-conductors, or rarefied gas dynamics). Simulating such systems by using kinetic models can be computationally very expensive, but modern super computers now enable realistic simulations. When the mean free path of the particles is very small as compared to the (macroscopic) size of the computational domain, the kinetic model can be very well approximated by a much simpler macroscopic model (diffusion equation, Rosseland approximation, Euler and Navier-Stokes equations), that can be numerically solved much faster.

However, there are many cases where the ratio mean free path/macroscopic size (the so-called “Knudsen number” in rarefied gas dynamics, denoted by $\varepsilon$ in this paper) is not constant: depending on the geometry of the boundaries, or on the boundary conditions, this ratio may vary in time, and in space. In such multiscale situations, usual kinetic solvers are often useless: for stability and accuracy reasons, they must resolve the microscopic scales, which is computationally to expensive in “fluid” zones (where $\varepsilon$ is small). By contrast, macroscopic solvers are faster but may be inaccurate in “kinetic” zones (where $\varepsilon$ is large).

[1]Department of Physics and Department of Mathematics, Duke University, Durham, NC 27708, USA (`Jian-Guo.Liu@duke.edu`)

[2]Institut de Mathématiques de Bordeaux (UMR 5251), Université de Bordeaux, 351, cours de la Libération, 33405 Talence cedex, France (`Luc.Mieussens@math.u-bordeaux1.fr`)

This is why many people have been working for more than 20 years on a kind of multiscale kinetic schemes: the asymptotic-preserving (AP) schemes. Such schemes are uniformly stable with respect to $\varepsilon$ (thus their computational complexity does not depend on $\varepsilon$), and are consistent with the macroscopic model when $\varepsilon$ goes to 0 (the limit of the scheme is a scheme for the macroscopic model).

Up to our knowledge, AP schemes have first been studied (for steady problems) in neutron transport by Larsen, Morel and Miller [22], Larsen and Morel [21], and then by Jin and Levermore [10, 11]. For unstationary problems, the difficulty is the time stiffness due to the collision operator. To avoid the use of expensive fully implicit schemes, two classes of semi-implicit time discretizations have been proposed by Klar [16] and Jin, Pareschi and Toscani [15] (see preliminary works in [14, 9] and extensions in [13, 12, 25, 17, 18]. Similar ideas have also been used by Buet et al. in [4].

In [23], Lemou and Mieussens have proposed a new AP scheme based on the micro-macro decomposition of the distribution function into microscopic and macroscopic components (similar schemes have also been proposed by Klar and Schmeiser [19], and more recently by Carrillo et al. [5, 6]). A coupled system of equations is obtained for these two components without any linearity assumption. The decomposition only uses basic properties of the collision operator that are common to most of kinetic equations (namely conservation and equilibrium properties). Then this system is solved with a suitable time semi-implicit discretization and space finite differences on staggered grids. While almost all the schemes mentioned before are based on very similar ideas, the approach proposed in [23] has been shown to be very general, since it applies to kinetic equations for both diffusion and hydrodynamic regimes (for the diffusion regime, see [23] for linear transport equations and [3] for the non-linear Kac equation, for the hydrodynamic regime, see [2] for the Boltzmann equation). We also mention the work of Degond, Liu and Mieussens [7] who proposed a similar approach (micro-macro decomposition) to design macroscopic diffusion models with kinetic upscalings: this approach also leads to AP schemes, at least for a semi-discrete time discretization (no AP space discretization was studied in this paper).

While many different schemes have been proposed in the past few years, it appears that the rigorous proof of their AP property looks rather difficult and is seldom investigated. Up to our knowledge, there are only two papers on this subject. Klar and Unterreiter [20] have proved that a scheme similar to that of [16] and [15] for the linear transport equation is uniformly stable. However, their proof is based on a von Neumann analysis, and hence is restricted to a one dimensional equation with constant coefficients in a periodic domain. Gosse and Toscani have proposed in [8] an AP scheme based on a rather different idea (discretization of the collision term as a non-conservative product and use of well-balanced Godunov schemes): they have been able to prove a uniform stability property, still in the linear case, but also a strong positivity property. However, their scheme is based on techniques (like approximation of the steady solution) that are difficult to generalize to other equations.

In this paper, we propose a very simple stability proof for the AP scheme of [23] and we exhibit an explicit CFL condition. This condition is uniform with respect to $\varepsilon$, and gives a standard parabolic CFL condition $\Delta t = O(\Delta x^2)$ when $\varepsilon$ goes to 0. While a stability property was already proved in [23] for this scheme, this was only for a simple two-velocity model (telegraph equation), by using von Neumann analysis. Here, our proof is based on energy estimates that are more general than von Neumann analysis, and hence is valid for one-dimensional linear equations with non-constant coefficients, continuous velocity variable, in the whole space. By the same technique, we are also able to prove uniform error estimates.

Our paper is organized as follows. In section 2, we introduce a general linear equation, we present its discretization by the AP scheme, and we give the main features of this scheme. Then, in section 3, we give our main stability result and its proof. The error estimates are given in section 4.

# 2 An AP scheme for the linear transport equation

Linear transport equation is a model for the evolution of particles in some medium (neutron transport, linear radiative transfer, ...). Generally, this model reads, in scaled variables

$$\varepsilon\partial_t\phi + \Omega\cdot\nabla_r\phi = \frac{\sigma}{\varepsilon}L\phi - \varepsilon\sigma_A\phi + \varepsilon S,$$

where $\phi(t,r,\Omega)$ is the number density of particles in the position-direction phase space that depends on time $t$, position $r=(x,y,z)\in\mathbb{R}^3$, and angular direction of propagation of particles $\Omega\in S^2$. Moreover, $\sigma$ is the total cross section, $\sigma_A$ is the absorption cross section, and $S$ is an internal source of particles, which is independent of $\Omega$. The linear operator $L$ models the scattering of the particles by the medium and acts only on the angular dependence of $\phi$. This simple model does not allow for particles of possibly different energy (or frequency); it is called "one-group" or "monoenergetic" equation. The parameter $\varepsilon$ is a scale factor that measures the ratio between a typical microscopic length (the mean free path of a particle, for instance) to a typical macroscopic length (the size of the computational domain, for instance). See [22] for details.

In this paper, we consider this one-group equation in the slab geometry: we assume that $\phi$ depends only on the slab axis variable $x\in\mathbb{R}$. Then it can be shown that the average of $\phi$ with respect to the $(y,z)$ cosine directions of $\Omega$, denoted by $f(t,x,v)$, satisfies the one-dimensional equation

$$\varepsilon\partial_t f + v\partial_x f = \frac{\sigma}{\varepsilon}Lf - \varepsilon\sigma_A f + \varepsilon S, \tag{1}$$

where $v\in[-1,1]$ is the $x$ cosine direction of $\Omega$. At $t=0$, we have the initial data $f(0,x,v)=f^0(x,v)$. We assume that the cross sections satisfy the inequalities $0<\sigma_m\le\sigma(x)\le\sigma_M$ and $0\le\sigma_A(x)\le\sigma_{AM}$ for every $x$. We do not consider boundary conditions in this paper, and hence $x\in\mathbb{R}$.

The linear operator $L$ is given by

$$Lf(v) = \int_{-1}^{1} s(v,v')(f(v')-f(v))\,dv', \tag{2}$$

where the kernel $s$ is such that $0 < s_m \leq s(v,v') \leq s_M$ for every $v,v'$ in $[-1,1]$. We also assume that $s$ satisfies $\int_{-1}^{1} s(v,v')\,dv' = 1$, and that it is symmetric: $s(v,v') = s(v,v')$. For the sequel, it is useful to define the operator $[.]$ such that $[\phi] = \frac{1}{2}\int_{-1}^{1}\phi(v)\,dv$ is the average of every velocity dependent function $\phi$. With these assumptions, it is standard [1] to state the following properties of $L$:

**Proposition 2.1.**

- *$[L\phi] = 0$ for every $\phi$ in $L^2([-1,1])$*
- *the null space of $L$ is $\mathcal{N}(L) = \{\phi = [\phi]\}$ (constant functions)*
- *the rank of $L$ is $\mathcal{R}(L) = \mathcal{N}^{\perp}(L) = \{\phi \text{ s.t } [\phi] = 0\}$*
- *$L$ is non-positive self-adjoint in $L^2([-1,1])$ and we have*

$$[\phi L\phi] \leq -2s_m\left[\phi^2\right] \tag{3}$$

  *for every $\phi \in \mathcal{N}^{\perp}(L)$*
- *$L$ admits a pseudo-inverse from $\mathcal{N}^{\perp}(L)$ onto $\mathcal{N}^{\perp}(L)$, denoted by $L^{-1}$*
- *the orthogonal projection from $L^2([-1,1])$ onto $\mathcal{N}^{\perp}(L)$ is $[.]$*

When $\varepsilon$ becomes small (the "diffusion" regime), it is well known that the solution $f$ of (1) tends to its own average density $\rho = [f]$, which is a solution of the asymptotic diffusion limit

$$\partial_t\rho - \partial_x\kappa\partial_x\rho = -\sigma_A\rho + S, \tag{4}$$

where the diffusion coefficient is $\kappa(x) = -\frac{[vL^{-1}v]}{\sigma(x)}$. An asymptotic preserving scheme for the linear kinetic equation (1) is a numerical scheme that discretizes (1) in such a way that it leads to a correct discretization of the diffusion limit (4) when $\varepsilon$ is small.

Now, we summarize the results obtained in [23]. By using the micro-macro decomposition $f = \rho + \varepsilon g$, where $\rho = [f]$ and $g$ is such that $[g] = 0$, we derived the micro-macro model for (1) that reads

$$\partial_t\rho + \partial_x[vg] = -\sigma_A\rho + S, \tag{5a}$$

$$\partial_t g + \frac{1}{\varepsilon}(I - [.])(v\partial_x g) = -\frac{\sigma}{\varepsilon^2}Lg - \frac{1}{\varepsilon^2}v\partial_x\rho, \tag{5b}$$

with initial data $\rho^0 = [f^0]$ and $\varepsilon g^0 = f^0 - \rho^0$. This system if formally equivalent to (1).

Then, we proposed the following numerical scheme for this system. We choose a time step $\Delta t$ and times $t_n = n\Delta t$, and two staggered grids of step $\Delta x$ and nodes $x_i = i\Delta x$ and $x_{i+\frac{1}{2}} = (i+\frac{1}{2})\Delta x$. We use the approximated values $\rho_i^n \approx \rho(t_n, x_i)$ and $g_{i+\frac{1}{2}}^n(v) \approx g(t_n, x_{i+\frac{1}{2}}, v)$, and the scheme reads

$$\frac{\rho_i^{n+1}-\rho_i^n}{\Delta t}+\left[v\frac{g_{i+\frac12}^{n+1}-g_{i-\frac12}^{n+1}}{\Delta x}\right]=-\sigma_{A,i}\rho_i^{n+1}+S_i, \tag{6a}$$

$$\frac{g_{i+\frac12}^{n+1}-g_{i+\frac12}^{n}}{\Delta t}+\frac{1}{\varepsilon\Delta x}(I-[.])\left(v^+(g_{i+\frac12}^n-g_{i-\frac12}^n)+v^-(g_{i+\frac32}^n-g_{i+\frac12}^n)\right) \tag{6b}$$
$$=-\frac{\sigma_{i+\frac12}}{\varepsilon^2}Lg_{i+\frac12}^{n+1}-\frac{1}{\varepsilon^2}v\frac{\rho_{i+1}^n-\rho_i^n}{\Delta x},$$

where $v^{\pm}=\frac{v\pm|v|}{2}$. Note that as in the continuous case, this scheme preserves the zero average of $g$:

**Proposition 2.2.** *If the initial data $g^0$ satisfies $\left[g^0_{i+\frac12}\right]=0$ for every $i$, then for every $n$ and $i$, we have*

$$\left[g^n_{i+\frac12}\right]=0. \tag{7}$$

*Proof.* Apply the average operator $[.]$ to (6b): since we have $[I-[.]]=0$ (obvious), $[L]=0$ (proposition 2.1), and $[v]=0$ (obvious), this relation yields $\left[g^{n+1}_{i+\frac12}\right]-\left[g^n_{i+\frac12}\right]=0$, which gives the result. □

In scheme (6), the upwind discretization of $(I-[.])(v\partial_x g)$ is to insure stability in the kinetic regime, while the centered approximations of $\partial_x[vg]$ and $v\partial_x\rho$ are to capture the diffusion limit. Indeed, it is clear that when $\varepsilon$ goes to 0, we have from (6b) $g^{n+1}_{i+\frac12}=-\frac{1}{\sigma_{i+\frac12}}L^{-1}\left(v\frac{\rho^n_{i+1}-\rho^n_i}{\Delta x}\right)+O(\varepsilon)$. Consequently, the flux of $g^{n+1}_{i+\frac12}$ is

$$\left[vg^{n+1}_{i+\frac12}\right]=-\frac{1}{\sigma_{i+\frac12}}\left[vL^{-1}v\right]\frac{\rho^n_{i+1}-\rho^n_i}{\Delta x}+O(\varepsilon),$$

where we have used the facts that $\rho_i^n$ does not depend on $v$ and that $L$—and hence $L^{-1}$ too—only applies to functions of $v$. Then, using this relation in (6a), we get

$$\frac{\rho_i^{n+1}-\rho_i^n}{\Delta t}-\frac{1}{\Delta x}\left(\kappa_{i+\frac12}\frac{\rho^n_{i+1}-\rho^n_i}{\Delta x}-\kappa_{i-\frac12}\frac{\rho^n_i-\rho^n_{i-1}}{\Delta x}\right)=-\sigma_{A,i}\rho_i^n+S_i, \tag{8}$$

with $\kappa_{i+\frac12}=-\frac{\left[vL^{-1}v\right]}{\sigma_{i+\frac12}}$, which is the usual 3-points stencil explicit scheme for the diffusion equation (4).

Of course, this property is true only if $\Delta t$ can be chosen independently of $\varepsilon$, or in other words, if the scheme is uniformly stable with respect to $\varepsilon$. In [23], we have proved that this scheme is indeed uniformly stable under some CFL condition, in the simpler case of the telegraph equation (in which we have only two discrete velocities $v=\pm1$, and $s\equiv1$, $\sigma\equiv1$, $\sigma_A=S\equiv0$). In the following section, we extend this result to the general equation (1) for scheme (6).

# 3 Uniform stability

We give our main result of stability for scheme (6). Without lost of generality, we assume $S = 0$ (source free case).

**Theorem 3.1.** *If $\Delta t$ satisfies the following CFL condition*

$$\Delta t \leq \frac{1}{3}\left(\tilde{\sigma}\Delta x^2 + 2\varepsilon\Delta x\right), \tag{9}$$

*with $\tilde{\sigma} = 2s_m\sigma_m$, then the sequences $\rho^n$ and $g^n$ defined by scheme (6) satisfy the energy estimate*

$$\sum_{i\in\mathbb{Z}} (\rho_i^n)^2 \Delta x + \varepsilon^2 \sum_{i\in\mathbb{Z}} \left[\left(g^n_{i+\frac{1}{2}}\right)^2\right] \Delta x \leq \sum_{i\in\mathbb{Z}} \left(\rho_i^0\right)^2 \Delta x + \varepsilon^2 \sum_{i\in\mathbb{Z}} \left[\left(g^0_{i+\frac{1}{2}}\right)^2\right] \Delta x$$

*for every $n$, and hence the scheme (6) is stable.*

Note that CFL condition (9) can be viewed as an average of a diffusive CFL condition $\Delta t \leq \tilde{\sigma}\Delta x^2$ (needed for the diffusion scheme (8)) and of a convection CFL $\Delta t \leq \varepsilon\Delta x$. It shows that the scheme is stable uniformly in $\varepsilon$, that is to say a diffusive CFL condition $\Delta t \leq C\Delta x^2$ is sufficient for stability for small $\varepsilon$, while a convection CFL is sufficient for $\varepsilon = O(1)$.

Moreover, while our CFL condition (9) is valid for every $\varepsilon$, we point out that it is not optimal for each $\varepsilon$. This might be the price to be paid for getting a uniform condition. For instance, in the simpler case of constant cross sections ($\sigma \equiv \sigma_m$) and kernel ($s = s_m \equiv \frac{1}{2}$), the diffusion coefficient of the diffusion equation (8) is $\kappa = \frac{1}{3\sigma}$, and the optimal CFL condition for scheme (8) is $\Delta t \leq \frac{\Delta x^2}{2\kappa}$. However, CFL condition (9) for our AP scheme reads (for $\varepsilon = 0$) $\Delta t \leq \frac{2}{9}\frac{\Delta x^2}{2\kappa}$, which is 0.2 as small as the optimal CFL.

**Remark 3.1.** If we have a source term $S \neq 0$, then the same CFL condition naturally gives a linear growth of the energy. Since this is standard and does not lead to any additional difficulty, we do not consider this case here.

**Remark 3.2.** In the case of a time explicit discretization of the absorption term (that is to say when $-\sigma_{A,i}\rho_i^{n+1}$ is replaced by $-\sigma_{A,i}\rho_i^n$ in (6), we have the same result with a CFL condition which is a bit more complicated. Namely, the scheme is stable if

$$\Delta t \leq \min\left(\frac{2}{1+\sigma_{A,M}}, \frac{3}{3+\sigma_{A,M}}\Delta t_S\right), \tag{10}$$

where $\Delta t_S$ is the maximum time step allowed by CFL condition (9) for the scheme with implicit (or zero) absorption term.

This theorem is proved in the following sections: section 3.1 contains compact notations and useful lemma, and section 3.2 contains the derivation of our energy estimate.

## 3.1 Notations and useful lemma

We give some useful notations for norms and inner products that are used in our analysis. For every grid function $\mu = (\mu_i)_{i\in\mathbb{Z}}$ we define:

$$\|\mu\|^2 = \sum_{i\in\mathbb{Z}} \mu_i^2 \Delta x. \tag{11}$$

For every velocity dependent grid function $v \in [-1,1] \mapsto \phi(v) = (\phi_{i+\frac{1}{2}}(v))_{i\in\mathbb{Z}}$, we define:

$$|||\phi||| = \sum_{i\in\mathbb{Z}} \left[\phi_{i+\frac{1}{2}}^2\right] \Delta x. \tag{12}$$

If $\phi$ and $\psi$ are two velocity dependent grid functions, we define their inner product:

$$\langle \phi\,,\,\psi\rangle = \sum_{i\in\mathbb{Z}} \left[\phi_{i+\frac{1}{2}}\psi_{i+\frac{1}{2}}\right] \Delta x. \tag{13}$$

Now we give some notations for the finite difference operators that are used in scheme (6). For every grid function $\phi = (\phi_{i+\frac{1}{2}})_{i\in\mathbb{Z}}$, we define the following one-sided operators:

$$D^-\phi_{i+\frac{1}{2}} = \frac{\phi_{i+\frac{1}{2}} - \phi_{i-\frac{1}{2}}}{\Delta x} \quad \text{and} \quad D^+\phi_{i+\frac{1}{2}} = \frac{\phi_{i+\frac{3}{2}} - \phi_{i+\frac{1}{2}}}{\Delta x} \tag{14}$$

We also define the following centered operators:

$$D^c\phi_{i+\frac{1}{2}} = \frac{\phi_{i+\frac{3}{2}} - \phi_{i-\frac{1}{2}}}{2\Delta x} \quad \text{and} \quad D^0\phi_i = \frac{\phi_{i+\frac{1}{2}} - \phi_{i-\frac{1}{2}}}{\Delta x}(= D^-\phi_{i+\frac{1}{2}}). \tag{15}$$

Finally, for every grid function $\mu = (\mu_i)_{i\in\mathbb{Z}}$, we define the following centered operator:

$$\delta^0\mu_{i+\frac{1}{2}} = \frac{\mu_{i+1} - \mu_i}{\Delta x}. \tag{16}$$

**Lemma 3.1** (Centered form of the upwind operator)**.** *For every grid function $\phi = (\phi_{i+\frac{1}{2}})_{i\in\mathbb{Z}}$, we have:*

$$\left(v^+D^- + v^-D^+\right)\phi_{i+\frac{1}{2}} = vD^c\phi_{i+\frac{1}{2}} - \frac{\Delta x}{2}|v|D^-D^+\phi_{i+\frac{1}{2}}.$$

*Proof.* This result is easily obtained by using the relations $v^\pm = \frac{v\pm|v|}{2}$ and the identity $D^- + D^+ = 2D^c$. □

**Lemma 3.2.** *For every grid function $\phi = (\phi_{i+\frac{1}{2}})_{i\in\mathbb{Z}}$, we have:*

$$\sum_{i\in\mathbb{Z}} \left(D^+\phi_{i+\frac{1}{2}}\right)^2 \Delta x \le \frac{4}{\Delta x^2}\sum_i \phi_{i+\frac{1}{2}}^2 \Delta x.$$

*Proof.* Expand the left-hand side, use the inequality $(a+b)^2 \leq 2(a^2+b^2)$, and then use the change of index $i+\frac{3}{2} \to i+\frac{1}{2}$. $\square$

**Lemma 3.3** (Estimate for the adjoint upwind operator)**.** *For every positive real number $\alpha$ and for every velocity dependent grid functions $\phi$ and $\psi$, we have:*

$$\left|\left\langle \left(v^+D^+ + v^-D^-\right)\psi\,,\,\phi\right\rangle\right| \leq \alpha|||\phi|||^2 + \frac{1}{4\alpha}|||\,|v|D^+\psi|||^2.$$

*Proof.* From Young's inequality, we obtain for any positive real number $\alpha$:

$$\left|\left\langle \left(v^+D^+ + v^-D^-\right)\psi\,,\,\phi\right\rangle\right| \leq \alpha|||\phi|||^2 + \frac{1}{4\alpha}|||\left(v^+D^+ + v^-D^-\right)\psi|||^2. \tag{17}$$

Now, the second term of the right-hand side of this inequality can be written:

$$\frac{1}{4\alpha}|||\left(v^+D^+ + v^-D^-\right)\psi|||^2 = \frac{1}{4\alpha}\sum_{i\in\mathbb{Z}}\frac{1}{2}\left(\int_{-1}^{0} v^2(D^-\psi_{i+\frac{1}{2}})^2\,dv + \int_0^1 v^2(D^+\psi_{i+\frac{1}{2}})^2\,dv\right)\Delta x.$$

Then, with a simple changing of index, $D^-$ can be replaced by $D^+$ in the first integral, which gives:

$$\begin{aligned}\frac{1}{4\alpha}|||\left(v^+D^+ + v^-D^-\right)\psi|||^2 &= \frac{1}{4\alpha}\sum_{i\in\mathbb{Z}}\frac{1}{2}\int_{-1}^{1} v^2(D^+\psi_{i+\frac{1}{2}})^2\,dv\Delta x\\ &= \frac{1}{4\alpha}|||\,|v|D^+\psi|||^2.\end{aligned}$$

Finally, using this inequality in (17) gives the result. $\square$

**Lemma 3.4** (Discrete integration by parts)**.** *For every grid functions $\phi = (\phi_{i+\frac{1}{2}})_{i\in\mathbb{Z}}$, $\psi = (\psi_{i+\frac{1}{2}})_{i\in\mathbb{Z}}$, and $\mu = (\mu_i)_{i\in\mathbb{Z}}$, we have:*

$$\begin{aligned}&\sum_{i\in\mathbb{Z}}\mu_i D^0\phi_i\Delta x = -\sum_{i\in\mathbb{Z}}\left(\delta^0\mu_{i+\frac{1}{2}}\right)\phi_{i+\frac{1}{2}}\Delta x,\\ &\sum_{i\in\mathbb{Z}}\psi_{i+\frac{1}{2}}D^-\phi_{i+\frac{1}{2}}\Delta x = -\sum_{i\in\mathbb{Z}}\left(D^+\psi_{i+\frac{1}{2}}\right)\phi_{i+\frac{1}{2}}\Delta x,\\ &\sum_{i\in\mathbb{Z}}\phi_{i+\frac{1}{2}}D^c\phi_{i+\frac{1}{2}}\Delta x = 0.\end{aligned}$$

*Proof.* These results are simply obtained by using obvious changing of indexes and the definitions of the finite difference operators given in (14–16). $\square$

**Lemma 3.5.** *If $g \in L^2([-1,1])$ then*

$$[vg]^2 \leq \frac{1}{2}\left[|v|g^2\right]$$

*Proof.* We just note that

$$\begin{aligned}
[vg]^2 &= \frac{1}{4}\left(\int_{-1}^{1} vg\,dv\right)^2 = \frac{1}{4}\left(\int_{-1}^{1} \operatorname{sign}(v)\sqrt{|v|}\sqrt{|v|}g\,dv\right)^2 \\
&\leq \frac{1}{4}\int_{-1}^{1}\left(\operatorname{sign}(v)\sqrt{|v|}\right)^2 dv \int_{-1}^{1}\left(\sqrt{|v|}g\right)^2 dv \qquad \text{by Cauchy-Schwarz inequality} \\
&= \frac{1}{2}\left[|v|g^2\right].
\end{aligned}$$

□

## 3.2 Energy estimates

Since the time discretization of absorption term $-\sigma_A\rho$ is implicit, it plays no role in the energy estimate. Consequently, to simplify the proof, we consider the case where there is no absorption (see remark 3.3 at the end of this section). We proceed in five short steps.
*Step 1.*
Here we derive a first energy relation. With the finite difference operators defined in (14)–(16), the scheme can be written in the following compact form:

$$\frac{\rho_i^{n+1}-\rho_i^n}{\Delta t} + D^0\left[vg_i^{n+1}\right] = 0 \tag{18a}$$

$$\frac{g_{i+\frac{1}{2}}^{n+1}-g_{i+\frac{1}{2}}^n}{\Delta t} + \frac{1}{\varepsilon}(I-[.])\left(v^+D^- + v^-D^+\right)g_{i+\frac{1}{2}}^n = \frac{\sigma_{i+\frac{1}{2}}}{\varepsilon^2}Lg_{i+\frac{1}{2}}^{n+1} - \frac{1}{\varepsilon^2}v\,\delta^0\rho_{i+\frac{1}{2}}^n. \tag{18b}$$

We define the energy of system (5) as $\int_{\mathbb{R}}\rho^2\,dx + \varepsilon^2\int\left[g^2\right]dx$. It is clear that the scheme can be proved to be stable if the energy at time $n+1$ can be controlled by the energy at time $n$. Consequently, we first multiply (18a) by $\rho_i^{n+1}$, then we take the sum over $i\in\mathbb{Z}$, and finally, we use the standard equality $a(a-b) = \frac{1}{2}(a^2-b^2+|a-b|^2)$ to get:

$$\frac{1}{2\Delta t}\left(\|\rho^{n+1}\|^2 - \|\rho^n\|^2 + \|\rho^{n+1}-\rho^n\|^2\right) + \sum_{i\in\mathbb{Z}}\rho_i^{n+1}D^0\left[vg_i^{n+1}\right]\Delta x = 0. \tag{19a}$$

Second, we multiply (18b) by $g_{i+\frac{1}{2}}^{n+1}$, we take the velocity average, we sum over $i\in\mathbb{Z}$, and we get:

$$\begin{aligned}
&\frac{1}{2\Delta t}\left(|||g^{n+1}|||^2 - |||g^n|||^2 + |||g^{n+1}-g^n|||^2\right) + \frac{1}{\varepsilon}\left\langle g^{n+1}\,,\,(I-[.])\left(v^+D^-+v^-D^+\right)g^n\right\rangle \\
&= \frac{1}{\varepsilon^2}\left\langle g^{n+1}\,,\,\sigma Lg^{n+1}\right\rangle - \frac{1}{\varepsilon^2}\sum_{i\in\mathbb{Z}}\left[vg_{i+\frac{1}{2}}^{n+1}\right]\delta^0\rho_{i+\frac{1}{2}}^n\Delta x.
\end{aligned} \tag{19b}$$

Now we use relation (7): since $\left[g_{i+\frac{1}{2}}^{n+1}\right] = 0$ for every $i$, a simple expansion of the inner product of the left-hand side of (19b) shows that it can be reduced to:

$$
\begin{aligned}
&\left\langle g^{n+1}\,,\,(I-[.])\left(v^+D^-+v^-D^+\right)g^n\right\rangle\\
&=\left\langle g^{n+1}\,,\,\left(v^+D^-+v^-D^+\right)g^n\right\rangle-\sum_{i\in\mathbb{Z}}\left[g^{n+1}_{i+\frac{1}{2}}\right]\left[\left(v^+D^-+v^-D^+\right)g^n_{i+\frac{1}{2}}\right]\Delta x\\
&=\left\langle g^{n+1}\,,\,\left(v^+D^-+v^-D^+\right)g^n\right\rangle.
\end{aligned}
\tag{20}
$$

Moreover, we can use relation (3) and the assumptions of the cross section and the kernel to estimate the inner product of the right-hand side of (19b) as follows:

$$
\begin{aligned}
\left\langle g^{n+1}\,,\,\sigma L g^{n+1}\right\rangle&=\sum_{i\in\mathbb{Z}}\sigma_{i+\frac{1}{2}}\left[g^{n+1}_{i+\frac{1}{2}}Lg^{n+1}_{i+\frac{1}{2}}\right]\Delta x\\
&\leq-\underbrace{2s_m\sigma_m}_{\tilde{\sigma}}|||g^{n+1}|||^2.
\end{aligned}
\tag{21}
$$

Consequently, we add up (19a) and $\varepsilon^2$ times (19b), then we use (20), (21), and the discrete integration by parts of lemma 3.4 to get our preliminary energy estimate:

$$
\begin{aligned}
&\frac{1}{2\Delta t}\left(\|\rho^{n+1}\|^2-\|\rho^n\|^2+\|\rho^{n+1}-\rho^n\|^2\right)+\sum_{i\in\mathbb{Z}}\rho^{n+1}_i D^0\left[vg^{n+1}_i\right]\Delta x\\
&+\frac{\varepsilon^2}{2\Delta t}\left(|||g^{n+1}|||^2-|||g^n|||^2+|||g^{n+1}-g^n|||^2\right)+\varepsilon\left\langle g^{n+1}\,,\,\left(v^+D^-+v^-D^+\right)g^n\right\rangle\\
&\leq-\tilde{\sigma}|||g^{n+1}|||^2+\sum_{i\in\mathbb{Z}}\left[vD^0g^{n+1}_i\right]\rho^n_i\Delta x.
\end{aligned}
\tag{22}
$$

*Step 2.*
In this step, we show how the $\rho^{n+1}-\rho^n$ term can be eliminated in (22). First, it is useful to write $\rho^n_i$ in the right-hand side of (22) as $(\rho^n_i-\rho^{n+1}_i)+\rho^{n+1}_i$: indeed, the terms $\sum_{i\in\mathbb{Z}}\rho^{n+1}_i D^0\left[vg^{n+1}_i\right]$ and $\sum_{i\in\mathbb{Z}}\left[vD^0g^{n+1}_i\right]\rho^{n+1}_i$ in the left and right-hand sides cancel out and we obtain:

$$
\begin{aligned}
&\frac{1}{2\Delta t}\left(\|\rho^{n+1}\|^2-\|\rho^n\|^2+\|\rho^{n+1}-\rho^n\|^2\right)\\
&+\frac{\varepsilon^2}{2\Delta t}\left(|||g^{n+1}|||^2-|||g^n|||^2+|||g^{n+1}-g^n|||^2\right)+\varepsilon\left\langle g^{n+1}\,,\,\left(v^+D^-+v^-D^+\right)g^n\right\rangle\\
&=-\tilde{\sigma}|||g^{n+1}|||^2+\sum_{i\in\mathbb{Z}}\left[vD^0g^{n+1}_i\right](\rho^n_i-\rho^{n+1}_i)\Delta x.
\end{aligned}
\tag{23}
$$

Now, we use the following Young inequality:

$$
\sum_{i\in\mathbb{Z}}\left[vD^0g^{n+1}_i\right](\rho^n_i-\rho^{n+1}_i)\Delta x\leq\alpha\|\rho^{n+1}-\rho^n\|^2+\frac{1}{4\alpha}\sum_{i\in\mathbb{Z}}\left[vD^0g^{n+1}_i\right]^2\Delta x.
\tag{24}
$$

Then the $\rho^{n+1} - \rho^n$ terms cancel out in (23) if $\alpha = \frac{1}{2\Delta t}$ and we get

$$\begin{aligned} &\frac{1}{2\Delta t}\left(\|\rho^{n+1}\|^2 - \|\rho^n\|^2\right) + \frac{\varepsilon^2}{2\Delta t}\left(|||g^{n+1}|||^2 - |||g^n|||^2 + |||g^{n+1} - g^n|||^2\right) \\ &+ \varepsilon\left\langle g^{n+1}, \left(v^+D^- + v^-D^+\right)g^n\right\rangle \le -\tilde{\sigma}|||g^{n+1}|||^2 + \frac{\Delta t}{2}\sum_{i\in\mathbb{Z}}\left[vD^0 g^{n+1}_{i+\frac{1}{2}}\right]^2\Delta x. \end{aligned} \tag{25}$$

*Step 3.*
Here, we work on the inner product of (25) to show that the $g^{n+1} - g^n$ terms can also be eliminated. First, we insert $g^{n+1}$ in this inner product to get:

$$\begin{aligned} &\left\langle g^{n+1}, \left(v^+D^- + v^-D^+\right)g^n\right\rangle \\ &= \left\langle g^{n+1}, \left(v^+D^- + v^-D^+\right)g^{n+1}\right\rangle + \left\langle g^{n+1}, \left(v^+D^- + v^-D^+\right)\left(g^n - g^{n+1}\right)\right\rangle \\ &= A + B, \end{aligned} \tag{26}$$

and we rewrite terms $A$ and $B$ as follows. For $A$, we use the centered form of the upwind operator (lemma 3.1) and the discrete integration by parts of lemma 3.4 to get:

$$\begin{aligned} A &= \left\langle g^{n+1}, vD^c g^{n+1}\right\rangle - \frac{\Delta x}{2}\left\langle g^{n+1}, |v|D^-D^+g^{n+1}\right\rangle \\ &= \frac{\Delta x}{2}\left\langle D^+g^{n+1}, |v|D^+g^{n+1}\right\rangle \\ &= \frac{\Delta x}{2}\sum_{i\in\mathbb{Z}}\left[|v|\left(D^+g^{n+1}_{i+\frac{1}{2}}\right)^2\right]\Delta x. \end{aligned} \tag{27}$$

For $B$, we also use the discrete integration by parts of lemma 3.4 to get:

$$B = -\left\langle\left(v^+D^+ + v^-D^-\right)g^{n+1}, g^n - g^{n+1}\right\rangle. \tag{28}$$

Then we apply the inequality of lemma 3.3 to $B$ to get

$$|B| \le \alpha|||g^{n+1} - g^n|||^2 + \frac{1}{4\alpha}||||v|D^+g^{n+1}|||^2. \tag{29}$$

Therefore, using (25), (26), (27) and (29), we see that the $g^{n+1} - g^n$ terms cancel out in (25) if $\alpha = \frac{\varepsilon}{2\Delta t}$, and we get

$$\begin{aligned} &\frac{1}{2\Delta t}\left(\|\rho^{n+1}\|^2 - \|\rho^n\|^2\right) + \frac{\varepsilon^2}{2\Delta t}\left(|||g^{n+1}|||^2 - |||g^n|||^2\right) \\ &+ \varepsilon\frac{\Delta x}{2}\sum_{i\in\mathbb{Z}}\left[|v|\left(D^+g^{n+1}_{i+\frac{1}{2}}\right)^2\right]\Delta x - \frac{\Delta t}{2}||||v|D^+g^{n+1}|||^2 \\ &\le -\tilde{\sigma}|||g^{n+1}|||^2 + \frac{\Delta t}{2}\sum_{i\in\mathbb{Z}}\left[vD^0 g^{n+1}_{i+\frac{1}{2}}\right]^2\Delta x. \end{aligned} \tag{30}$$

*Step 4.*
Now, we show how all the $D^+g^{n+1}$ and the $D^0g^{n+1}$ terms can be controlled by $|||g^{n+1}|||$. First, note that the term $\frac{\Delta t}{2}|||\,|v|D^+g^{n+1}|||^2$ of the left-hand side of (30) can be estimated as follows:

$$\begin{aligned}\frac{\Delta t}{2}|||\,|v|D^+g^{n+1}|||^2 &= \frac{\Delta t}{2}\sum_{i\in\mathbb{Z}}\left[|v|^2\left(D^+g^{n+1}_{i+\frac{1}{2}}\right)^2\right]\Delta x\\ &\leq \frac{\Delta t}{2}\sum_{i\in\mathbb{Z}}\left[|v|\left(D^+g^{n+1}_{i+\frac{1}{2}}\right)^2\right]\Delta x,\end{aligned} \tag{31}$$

since $|v|\leq 1$. Moreover, using lemma 3.5 and a change of indices shows that the last term of the right-hand side of (30) satisfies

$$\frac{\Delta t}{2}\sum_{i\in\mathbb{Z}}\left[vD^0g^{n+1}_{i+\frac{1}{2}}\right]^2\Delta x \leq \frac{\Delta t}{4}\sum_{i\in\mathbb{Z}}\left[|v|\left(D^+g^{n+1}_{i+\frac{1}{2}}\right)^2\right]\Delta x. \tag{32}$$

Finally, we use these two estimates in (30) to obtain:

$$\begin{aligned}&\frac{1}{2\Delta t}\left(\|\rho^{n+1}\|^2-\|\rho^n\|^2\right)+\frac{\varepsilon^2}{2\Delta t}\left(|||g^{n+1}|||^2-|||g^n|||^2\right)\\ &\leq -\tilde{\sigma}|||g^{n+1}|||^2+\left(\frac{3\Delta t}{4}-\varepsilon\frac{\Delta x}{2}\right)\sum_{i\in\mathbb{Z}}\left[|v|\left(D^+g^{n+1}_{i+\frac{1}{2}}\right)^2\right]\Delta x.\end{aligned} \tag{33}$$

Now, taking the positive part of the factor $(\frac{3\Delta t}{4}-\varepsilon\frac{\Delta x}{2})$ of the right-hand side of (33), we have the estimate

$$\begin{aligned}\left(\frac{3\Delta t}{4}-\varepsilon\frac{\Delta x}{2}\right)\sum_{i\in\mathbb{Z}}\left[|v|\left(D^+g^{n+1}_{i+\frac{1}{2}}\right)^2\right]\Delta x &\leq \left(\frac{3\Delta t}{4}-\varepsilon\frac{\Delta x}{2}\right)^+\sum_{i\in\mathbb{Z}}\left[\left(D^+g^{n+1}_{i+\frac{1}{2}}\right)^2\right]\Delta x\\ &\leq \left(\frac{3\Delta t}{4}-\varepsilon\frac{\Delta x}{2}\right)^+\frac{4}{\Delta x^2}|||g^{n+1}|||^2,\end{aligned} \tag{34}$$

where we have used $|v|\leq 1$ and the estimate of lemma 3.2.

*Step 5.*
Finally, estimates (33) and (34) show that

$$\frac{1}{2\Delta t}\left(\|\rho^{n+1}\|^2-\|\rho^n\|^2\right)+\frac{\varepsilon^2}{2\Delta t}\left(|||g^{n+1}|||^2-|||g^n|||^2\right)\leq\left(\left(\frac{3\Delta t}{4}-\varepsilon\frac{\Delta x}{2}\right)^+\frac{4}{\Delta x^2}-\tilde{\sigma}\right)|||g^{n+1}|||^2.$$

This means that we have the final energy estimate

$$\|\rho^{n+1}\|^2+\varepsilon^2|||g^{n+1}|||^2\leq\|\rho^n\|^2+\varepsilon^2|||g^n|||^2$$

if $\Delta t$ is such that

$$\left(\frac{3\Delta t}{4} - \varepsilon\frac{\Delta x}{2}\right)^+ \frac{4}{\Delta x^2} \leq \tilde{\sigma}.$$

Since $\tilde{\sigma} \geq 0$, an equivalent condition is $(\frac{3\Delta t}{4} - \varepsilon\frac{\Delta x}{2})\frac{4}{\Delta x^2} \leq \sigma$, which gives the sufficient condition

$$\Delta t \leq \frac{\Delta x^2 \tilde{\sigma}}{3} + \frac{2}{3}\varepsilon\Delta x,$$

which proves the theorem.

**Remark 3.3.** As explained at the beginning of this section, when the absorption term is non zero and is discretized implicitly, its contribution $-\sum_{i\in\mathbb{Z}} \sigma_{A,i}(\rho_i^{n+1})^2\Delta x$ to the energy estimate is non-positive and plays no role in the previous analysis. However, if we use instead an explicit discretization, then our analysis has to be modified. The difference now is that there is the additional term $-\sum_{i\in\mathbb{Z}} \sigma_{A,i}\rho_i^{n+1}\rho_i^n\Delta x$ in the right-hand side of (22). The idea of step 2 can be applied to this term (replace $\rho^n$ by $(\rho^n - \rho^{n+1}) + \rho^{n+1}$ and use a Young inequality) so that the $\rho^{n+1} - \rho^n$ terms cancel out, but now with $\alpha = \frac{1}{2\Delta t(1+\sigma_{A,M})}$. The other steps are the same, except that some coefficients are different. In order to shorten the paper, these details are left to the reader.

# 4 Error estimates

In this section, we simplify the presentation by taking constant total cross section $\sigma$ and kernel $s \equiv \frac{1}{2}$, no absorption ($\sigma_A \equiv 0$), and no source term ($S \equiv 0$). These assumptions are not restrictive at all, and our analysis could be directly applied in the general case.

Let $T > 0$ be some finite time. For this study, we assume that the exact solution $(\rho, g)$ of (5) has the following regularity:

$$\begin{aligned} &\rho \in C^2([0,T], H^1(\mathbb{R})) \cap C^0([0,T], H^3(\mathbb{R})) \\ &g \in C^2([0,T], L^2([-1,1], H^1(\mathbb{R}))) \cap C^0([0,T], L^2([-1,1], H^3(\mathbb{R}))). \end{aligned} \tag{35}$$

Note that this assumption implies that $g$ is uniformly bounded with respect to $\varepsilon$, which is quite strong. In particular, this excludes the case of initial or transition layers. Indeed, if, for instance, $f$ is not isotropic at $t = 0$, then $\rho(t = 0) \neq f(t = 0)$, and hence $g(t = 0) = \frac{1}{\varepsilon}(f - \rho)|_{t=0} = O(\varepsilon^{-1})$ cannot be uniformly bounded with respect to $\varepsilon$.

With this assumption, we can obtain the following result.

**Theorem 4.1.** *If $\Delta t$ satisfies the following condition*

$$\Delta t \leq \frac{\Delta x^2\sigma}{6} + \frac{2}{3}\varepsilon\Delta x, \tag{36}$$

*then for every time $T > 0$, there exists a constant $C$ independent of $\Delta t$, $\Delta x$, and $\varepsilon$, such that the numerical solution obtained by scheme (6) satisfies the following error estimate*

$$\max_{n, n\Delta t \le T} \left( \sum_{i\in\mathbb{Z}} |\rho(t_n, x_i) - \rho_i^n| \Delta x + \varepsilon \sum_{i\in\mathbb{Z}} \left[ |g(t_n, x_{i+\frac{1}{2}}, v) - g^n_{i+\frac{1}{2}}(v)| \right] \Delta x \right)$$
$$\le C \left( (1+\varepsilon^2)\Delta t + \Delta x^2 + \varepsilon \Delta x \right).$$

This theorem is proved in the following sections: in section 4.1, we first derive the truncation error of the scheme, then in section 4.2, we apply the same analysis as for the stability result to prove the theorem.

## 4.1 Truncation error

Let $a_i^n$ and $\frac{1}{\varepsilon^2} b_i^n$ be the truncation errors of scheme (6), that is to say the reminders obtained by inserting the exact solution of (5) in relations (6):

$$\frac{\rho(t_{n+1}, x_i) - \rho(t_n, x_i)}{\Delta t} + \left[ v \frac{g(t_{n+1}, x_{i+\frac{1}{2}}) - g(t_{n+1}, x_{i-\frac{1}{2}})}{\Delta x} \right] = a_i^n \tag{37a}$$

$$\frac{g(t_{n+1}, x_{i+\frac{1}{2}}, .) - g(t_n, x_i, .)}{\Delta t} \tag{37b}$$

$$+ \frac{1}{\varepsilon \Delta x} (I - [.]) \left( v^+ (g(t_n, x_{i+\frac{1}{2}}, v) - g(t_n, x_{i-\frac{1}{2}}, v) + v^- (g(t_n, x_{i+\frac{3}{2}}, v) - g(t_n, x_{i+\frac{1}{2}}, v)) \right) \tag{37c}$$

$$= -\frac{\sigma}{\varepsilon^2} g(t_{n+1}, x_{i+\frac{1}{2}}, v) - \frac{1}{\varepsilon^2} v \frac{\rho(t_n, x_{i+1}) - \rho(t_n, x_i)}{\Delta x} + \frac{1}{\varepsilon^2} b_i^n.$$

We can prove the following estimate of these truncation errors:

**Lemma 4.1.** *There exists a constant $\tilde{C}$ independent of $\Delta t$, $\Delta x$ and $\varepsilon$, such that*

$$\|a^n\| + |||b^n||| \le \tilde{C} \left( (1+\varepsilon^2)\Delta t + \Delta x^2 + \varepsilon \Delta x \right)$$

*for every $n$*

This lemma is proved by using standard simple techniques (Taylor-Lagrange formula of different orders). But to simplify the paper, the proof—which is a bit long—is given in appendix A.

Now, let $\tilde{\rho}_i^n$ and $\tilde{g}^n_{i+\frac{1}{2}}$ be the convergence errors, that is to say the sequences defined by

$$\tilde{\rho}_i^n = \rho(t_n, x_i) - \rho_i^n \quad \text{and} \quad \tilde{g}^n_{i+\frac{1}{2}}(v) = g(t_n, x_{i+\frac{1}{2}}, v) - g^n_{i+\frac{1}{2}}(v).$$

Then these sequences satisfy the "perturbed" scheme, written in the following compact form:

$$\frac{\tilde\rho_i^{n+1}-\tilde\rho_i^n}{\Delta t}+D^0\left[v\tilde g_i^{n+1}\right]=a_i^n \tag{38a}$$

$$\frac{\tilde g_{i+\frac12}^{n+1}-\tilde g_{i+\frac12}^n}{\Delta t}+\frac{1}{\varepsilon}(I-[.])\left(v^+D^-+v^-D^+\right)\tilde g_{i+\frac12}^n=-\frac{\sigma}{\varepsilon^2}\tilde g_{i+\frac12}^{n+1}-\frac{1}{\varepsilon^2}v\,\delta^0\tilde\rho_{i+\frac12}^n+\frac{1}{\varepsilon^2}b_i^n, \tag{38b}$$

with the homogeneous initial data $\tilde\rho_i^0=0$ and $\tilde g_{i+\frac12}^0=0$ for every $i$.

## 4.2 Analysis of the convergence error

In this section, we apply the same analysis as for the stability result to prove that $\tilde\rho^n$ and $\tilde g^n$ can be controlled by the truncation errors.

*Step 1.*

We multiply (38a) by $\tilde\rho_i^{n+1}$ and take the sum over $i$ to get

$$\frac{1}{2\Delta t}\left(\|\tilde\rho^{n+1}\|^2-\|\tilde\rho^n\|^2+\|\tilde\rho^{n+1}-\tilde\rho^n\|^2\right)+\sum_{i\in\mathbb{Z}}\tilde\rho_i^{n+1}D^0\left[v\tilde g_i^{n+1}\right]\Delta x=\sum_{i\in\mathbb{Z}}\tilde\rho_i^{n+1}a_i^n\Delta x. \tag{39a}$$

Second, we multiply (38b) by $\tilde g_{i+\frac12}^{n+1}$, we take the velocity average, we sum over $i$, which yields

$$\begin{aligned}&\frac{1}{2\Delta t}\left(|||\tilde g^{n+1}|||^2-|||\tilde g^n|||^2+|||\tilde g^{n+1}-\tilde g^n|||^2\right)+\frac{1}{\varepsilon}\left\langle\tilde g^{n+1}\,,\,(I-[.])\left(v^+D^-+v^-D^+\right)\tilde g^n\right\rangle\\&=-\frac{\sigma}{\varepsilon^2}|||\tilde g^{n+1}|||^2-\frac{1}{\varepsilon^2}\sum_{i\in\mathbb{Z}}\left[v\tilde g_{i+\frac12}^{n+1}\right]\delta^0\tilde\rho_{i+\frac12}^n\Delta x+\frac{1}{\varepsilon^2}\left\langle\tilde g^{n+1}\,,\,b^n\right\rangle.\end{aligned} \tag{39b}$$

Finally, we add up (39a) and $\varepsilon^2$ times (39b), we use (20) and lemma 3.4 to get

$$\begin{aligned}&\frac{1}{2\Delta t}\left(\|\tilde\rho^{n+1}\|^2-\|\tilde\rho^n\|^2+\|\tilde\rho^{n+1}-\tilde\rho^n\|^2\right)+\sum_{i\in\mathbb{Z}}\tilde\rho_i^{n+1}D^0\left[v\tilde g_i^{n+1}\right]\Delta x\\&+\frac{\varepsilon^2}{2\Delta t}\left(|||\tilde g^{n+1}|||^2-|||\tilde g^n|||^2+|||\tilde g^{n+1}-\tilde g^n|||^2\right)+\varepsilon\left\langle\tilde g^{n+1}\,,\,\left(v^+D^-+v^-D^+\right)\tilde g^n\right\rangle\\&=\sum_{i\in\mathbb{Z}}\tilde\rho_i^n a_i^n\Delta x-\sigma|||\tilde g^{n+1}|||^2+\sum_{i\in\mathbb{Z}}\left[vD^0\tilde g_{i+\frac12}^{n+1}\right]\tilde\rho_{i+\frac12}^n\Delta x+\left\langle\tilde g^{n+1}\,,\,b^n\right\rangle.\end{aligned} \tag{40}$$

*Step 2.*

Here, we can copy the steps 2 to 4 of the stability analysis (see section 3) for relation (40). Therefore, skipping the details, we just give the resulting energy estimate:

$$\begin{aligned}&\frac{1}{2\Delta t}\left(\|\tilde\rho^{n+1}\|^2-\|\tilde\rho^n\|^2\right)+\frac{\varepsilon^2}{2\Delta t}\left(|||\tilde g^{n+1}|||^2-|||\tilde g^n|||^2\right)\\&\le-\sigma|||\tilde g^{n+1}|||^2+\left(\frac{3\Delta t}{4}-\varepsilon\frac{\Delta x}{2}\right)\sum_{i\in\mathbb{Z}}\left[|v|\left(D^+\tilde g_{i+\frac12}^{n+1}\right)^2\right]\Delta x\\&\quad+\sum_{i\in\mathbb{Z}}\tilde\rho_i^n a_i^n\Delta x+\left\langle\tilde g^{n+1}\,,\,b^n\right\rangle.\end{aligned} \tag{41}$$

*Step 3.*
We estimate the scalar products of the right-hand side of (41) by using two different Young inequalities:

$$\begin{aligned}&\sum_{i\in\mathbb{Z}} \tilde{\rho}_i^n a_i^n \Delta x \le \frac{1}{2}\left(\|\tilde{\rho}^{n+1}\|^2 + \|a^n\|^2\right)\\ &\left\langle \tilde{g}^{n+1}, b^n \right\rangle \le \frac{\sigma}{2}|||\tilde{g}^{n+1}|||^2 + \frac{1}{2\sigma}\|b^n\|^2.\end{aligned} \tag{42}$$

Moreover, by taking the positive part of the factor $(\frac{3\Delta t}{4} - \varepsilon\frac{\Delta x}{2})$ in (41), we have the estimate

$$\left(\frac{3\Delta t}{4} - \varepsilon\frac{\Delta x}{2}\right)\sum_{i\in\mathbb{Z}}\left[|v|\left(D^+\tilde{g}^{n+1}_{i+\frac{1}{2}}\right)^2\right]\Delta x \le \left(\frac{3\Delta t}{4} - \varepsilon\frac{\Delta x}{2}\right)^+ \frac{4}{\Delta x^2}|||\tilde{g}^{n+1}|||^2. \tag{43}$$

Now, we can use the inequalities (42) and (43) to obtain the energy estimate

$$\begin{aligned}&\|\tilde{\rho}^{n+1}\|^2 + \varepsilon^2|||\tilde{g}^{n+1}|||^2\\ &\quad\le (1+\Delta t)\|\tilde{\rho}^n\|^2 + \varepsilon^2|||\tilde{g}^n|||^2 + \left(-\frac{\sigma}{2} + \frac{4}{\Delta x^2}\left(\frac{3\Delta t}{4} - \varepsilon\frac{\Delta x}{2}\right)^+\right)2\Delta t|||\tilde{g}^{n+1}|||^2\\ &\qquad + \Delta t\|a^n\|^2 + \frac{\Delta t}{\sigma}|||b^n|||^2.\end{aligned} \tag{44}$$

Note that the factor of $|||g^{n+1}|||^2$ in the right-hand side of (44) is non-positive if $\frac{4}{\Delta x^2}\left(\frac{3\Delta t}{4} - \varepsilon\frac{\Delta x}{2}\right)^+ \le \frac{\sigma}{2}$. Since $\sigma > 0$, an equivalent condition is $\frac{4}{\Delta x^2}\left(\frac{3\Delta t}{4} - \varepsilon\frac{\Delta x}{2}\right) \le \frac{\sigma}{2}$, which gives the sufficient condition

$$\Delta t \le \frac{\Delta x^2\sigma}{6} + \frac{2}{3}\varepsilon\Delta x,$$

which is a bit more restrictive than the CFL condition (9). In that case, the energy estimate (44) can be simplified in

$$\begin{aligned}&\|\tilde{\rho}^{n+1}\|^2 + \varepsilon^2|||\tilde{g}^{n+1}|||^2\\ &\quad\le (1+\Delta t)\left(\|\tilde{\rho}^n\|^2 + \varepsilon^2|||\tilde{g}^n|||^2\right) + \Delta t C_\sigma\left(\|a^n\|^2 + |||b^n|||^2\right),\end{aligned}$$

where $C_\sigma = 1 + \frac{1}{\sigma}$.

Now, by using a simple recursion, we obtain

$$\begin{aligned}&\|\tilde{\rho}^n\|^2 + \varepsilon^2|||\tilde{g}^n|||^2\\ &\quad\le (1+\Delta t)^n\left(\|\tilde{\rho}^0\|^2 + \varepsilon^2|||\tilde{g}^0|||^2\right)\\ &\qquad + \Delta t C_\sigma\sum_{k=1}^{n}\left(\|a^{n-k}\|^2 + |||b^{n-k}|||^2\right)(1+\Delta t)^{k-1}.\end{aligned}$$

If $n$ is such that $n\Delta t \leq T$, we can use the classical inequality $(1+\Delta t)^n \leq e^{n\Delta t} \leq e^T$. Moreover, we can use lemma 4.1 and the fact that $\tilde{\rho}^0 = \tilde{g}^0 = 0$ to get

$$\begin{aligned}
&\|\tilde{\rho}^n\|^2 + \varepsilon^2 |||\tilde{g}^n|||^2 \\
&\quad \leq e^T C_\sigma T \tilde{C} \left((1+\varepsilon^2)\Delta t + \Delta x^2 + \varepsilon \Delta x\right)^2 ,
\end{aligned}$$

and hence

$$\begin{aligned}
&\|\tilde{\rho}^n\| + \varepsilon |||\tilde{g}^n||| \\
&\quad \leq C \left((1+\varepsilon^2)\Delta t + \Delta x^2 + \varepsilon \Delta x\right) ,
\end{aligned}$$

where $C = e^{\frac{T}{2}} \sqrt{C_\sigma T \tilde{C}}$ is independent of $\Delta t$, $\Delta x$, and $\varepsilon$. This concludes the proof of the theorem.

# 5 Conclusion

In this paper, we have proposed a very simple stability proof for the recent AP scheme of [23]. An explicit CFL condition that can be used in a computational code has been found: it insures that the scheme is stable and accurate, independently of $\varepsilon$. This condition gives a classical parabolic CFL condition $\Delta t = O(\Delta x^2)$ when $\varepsilon$ goes to 0. Our proof uses very basic and simple arguments (energy estimates and Young inequalities), and is valid for one-dimensional linear equations with non-constant coefficients, continuous velocity variable, in the whole space. Our technique applies to general linear collision operators like operators of neutron transport or linear radiative transfer. By the same technique, we have also proved uniform error estimates. We mention that, in a work in preparation [24], we are able to apply our method to a simple non-linear problem coming from radiative transfer.

In the future, the analysis of this scheme for initial boundary-value problems will be investigated. It would also be important to extend the scheme and its analysis to 2 or 3 dimensional problems.

# References


[1] C. Bardos, R. Santos, and R. Sentis. Diffusion approximation and computation of the critical size. *Trans. Amer. Math. Soc.*, 284(2):617–649, 1984.

[2] M. Bennoune, M. Lemou, and L. Mieussens. Uniformly stable numerical schemes for the Boltzmann equation preserving the compressible Navier-Stokes asymptotics. *J. Comput. Phys.*, 227(8):3781–3803, 2008.

[3] M. Bennoune, M. Lemou, and L. Mieussens. An asymptotic preserving scheme for the Kac model of the boltzmann equation in the diffusion limit. To appear in Contin. Mech. Thermodyn., 2009.

[4] C. Buet, S. Cordier, B. Lucquin-Desreux, and S. Mancini. Diffusion limit of the Lorentz model: asymptotic preserving schemes. *M2AN Math. Model. Numer. Anal.*, 36(4):631–655, 2002.

[5] J. A. Carrillo, T. Goudon, and P. Lafitte. Simulation of fluid and particles flows: asymptotic preserving schemes for bubbling and flowing regimes. *J. Comput. Phys.*, 227(16):7929–7951, 2008.

[6] J. A. Carrillo., T. Goudon, P. Lafitte, and F. Vecil. Numerical schemes of diffusion asymptotics and moment closures for kinetic equations. *J. Sci. Comput.*, 36(1):113–149, 2008.

[7] P. Degond, J.-G. Liu, and L. Mieussens. Macroscopic fluid models with localized kinetic upscaling effects. *Multiscale Modeling & Simulation*, 5(3):940–979, 2006.

[8] L. Gosse and G. Toscani. Asymptotic-preserving & well-balanced schemes for radiative transfer and the Rosseland approximation. *Numer. Math.*, 98(2):223–250, 2004.

[9] S. Jin. Efficient asymptotic-preserving (AP) schemes for some multiscale kinetic equations. *SIAM J. Sci. Comput.*, 21(2):441–454, 1999.

[10] S. Jin and C. D. Levermore. The discrete-ordinate method in diffusive regimes. *Transport Theory Statist. Phys.*, 20(5-6):413–439, 1991.

[11] S. Jin and C. D. Levermore. Fully discrete numerical transfer in diffusive regimes. *Transport Theory Statist. Phys.*, 22(6):739–791, 1993.

[12] S. Jin and L. Pareschi. Discretization of the multiscale semiconductor Boltzmann equation by diffusive relaxation schemes. *J. Comput. Phys.*, 161(1):312–330, 2000.

[13] S. Jin and L. Pareschi. Asymptotic-preserving (AP) schemes for multiscale kinetic equations: a unified approach. In *Hyperbolic problems: theory, numerics, applications, Vol. I, II (Magdeburg, 2000)*, volume 141 of *Internat. Ser. Numer. Math., 140*, pages 573–582. Birkhäuser, Basel, 2001.

[14] S. Jin, L. Pareschi, and G. Toscani. Diffusive relaxation schemes for multiscale discrete-velocity kinetic equations. *SIAM J. Numer. Anal.*, 35(6):2405–2439, 1998.

[15] S. Jin, L. Pareschi, and G. Toscani. Uniformly accurate diffusive relaxation schemes for multiscale transport equations. *SIAM J. Numer. Anal.*, 38(3):913–936, 2000.

[16] A. Klar. An asymptotic-induced scheme for nonstationary transport equations in the diffusive limit. *SIAM J. Numer. Anal.*, 35(3):1073–1094, 1998.

[17] A. Klar. An asymptotic preserving numerical scheme for kinetic equations in the low Mach number limit. *SIAM J. Numer. Anal.*, 36(5):1507–1527, 1999.

[18] A. Klar. A numerical method for kinetic semiconductor equations in the drift-diffusion limit. *SIAM J. Sci. Comput.*, 20(5):1696–1712, 1999.

[19] A. Klar and C. Schmeiser. Numerical passage from radiative heat transfer to nonlinear diffusion models. *Math. Models Methods Appl. Sci.*, 11(5):749–767, 2001.

[20] A. Klar and A. Unterreiter. Uniform stability of a finite difference scheme for transport equations in diffusive regimes. *SIAM J. Numer. Anal.*, 40(3):891–913, 2002.

[21] A. W. Larsen and J. E. Morel. Asymptotic solutions of numerical transport problems in optically thick, diffusive regimes. II. *J. Comput. Phys.*, 83(1):212–236, 1989.

[22] A. W. Larsen, J. E. Morel, and W. F. Miller Jr. Asymptotic solutions of numerical transport problems in optically thick, diffusive regimes. *J. Comput. Phys.*, 69(2):283–324, 1987.

[23] M. Lemou and L. Mieussens. A new asymptotic preserving scheme based on micro-macro formulation for linear kinetic equations in the diffusion limit. *SIAM J. Sci. Comput.*, 31(1):334–368, 2008.

[24] J.-G. Liu and L. Mieussens. Analysis of an asymptotic preserving scheme for the nonlinear radiative transfer equation. in preparation.

[25] G. Naldi and L. Pareschi. Numerical schemes for kinetic equations in diffusive regimes. *Appl. Math. Lett.*, 11(2):29–35, 1998.

# A Estimate of the truncation errors: proof of lemma 4.1

First, we remind (with no proof) that standard Taylor expansions give the following estimates for some time finite differences.

**Lemma A.1.** *(i) If $\psi \in C^1([0,T])$, then*

$$|\psi(t_{n+1}) - \psi(t_n)| \leq \Delta t \max_{[0,T]} |\psi'|.$$

*(ii) If $\psi \in C^2([0,T])$, then*

$$|\frac{\psi(t_{n+1}) - \psi(t_n)}{\Delta t} - \psi'(t_n)| \leq \Delta t \max_{[0,T]} |\psi''|.$$

Then we also have the following estimates for some space finite differences.

**Lemma A.2.** *(i) If $\phi \in H^1(\mathbb{R})$ and $\Delta x \leq 1$, then*

$$\sum_{i\in\mathbb{Z}} \phi(x_i)^2 \Delta x \leq 2\|\phi\|^2_{H^1}.$$

*(ii) If $\phi \in H^2(\mathbb{R})$, then*

$$\sum_{i\in\mathbb{Z}} |\frac{\phi(x_{i+1}) - \phi(x_i)}{\Delta x} - \phi'(x_i)|^2 \Delta x \leq \frac{\Delta x^2}{3} \|\phi''\|_{L^2}^2.$$

*(iii) If $\phi \in H^3(\mathbb{R})$, then*

$$\sum_{i\in\mathbb{Z}} |\frac{\phi(x_{i+1}) - \phi(x_i)}{\Delta x} - \phi'(x_{i+\frac{1}{2}})|^2 \Delta x \leq \frac{\Delta x^4}{320} \|\phi'''\|_{L^2}^2.$$

*Proof.* For (i), we write the difference of the continuous and discrete $L^2$ norms of $\phi$ as

$$\begin{aligned}
&\int_{\mathbb{R}} \phi(x)^2\,dx - \sum_{i\in\mathbb{Z}} \phi(x_i)^2 \Delta x \\
&= \sum_{i\in\mathbb{Z}} \int_{x_i}^{x_{i+1}} (\phi(x)^2 - \phi(x_i)^2)\,dx = \sum_{i\in\mathbb{Z}} \int_{x_i}^{x_{i+1}} \left( \int_{x_i}^{x} \frac{d}{dy}(\phi(y)^2)\,dy \right) dx \\
&= \sum_{i\in\mathbb{Z}} \int_{x_i}^{x_{i+1}} \left( \int_{x_i}^{x} 2\phi'(y)\phi(y)\,dy \right) dx.
\end{aligned}$$

Then, using a simple Young inequality, we get:

$$\begin{aligned}
\sum_{i\in\mathbb{Z}} \phi(x_i)^2 \Delta x &\leq \int_{\mathbb{R}} \phi(x)^2\,dx + \sum_{i\in\mathbb{Z}} \int_{x_i}^{x_{i+1}} \left( \int_{x_i}^{x} \left(\phi'(y)^2 + \phi(y)^2\right)\,dy \right) dx \\
&\leq \int_{\mathbb{R}} \phi(x)^2\,dx + \sum_{i\in\mathbb{Z}} \int_{x_i}^{x_{i+1}} \left( \int_{x_i}^{x_{i+1}} \left(\phi'(y)^2 + \phi(y)^2\right)\,dy \right) dx \\
&= \|\phi\|_{L^2}^2 + \Delta x \left( \|\phi'\|_{L^2}^2 + \|\phi\|_{L^2}^2 \right) \\
&\leq (1 + \Delta x)\|\phi\|_{H^1}^2.
\end{aligned}$$

For (ii), we use the Taylor-Lagrange formula up to first order to get

$$\frac{\phi(x_{i+1}) - \phi(x_i)}{\Delta x} - \phi'(x_i) = -\frac{1}{\Delta x} \int_{x_i}^{x_{i+1}} (x - x_{i+1})\phi''(x)\,dx.$$

Then a simple Cauchy-Schwarz inequality gives the following:

$$\left| \frac{\phi(x_{i+1}) - \phi(x_i)}{\Delta x} - \phi'(x_i) \right|^2 \leq \frac{\Delta x}{3} \int_{x_i}^{x_{i+1}} \phi''(x)^2\,dx.$$

Finally, we get the result by multiplying by $\Delta x$ and taking the sum over $i \in \mathbb{Z}$.

For (iii), we use the Taylor-Lagrange formula up to second order to get the following two relations:

$$\phi(x_{i+1}) - \phi(x_{i+\frac{1}{2}}) = \frac{1}{2}\phi'(x_{i+\frac{1}{2}}) + \frac{\Delta x^2}{8}\phi''(x_{i+\frac{1}{2}}) + \int_{x_{i+\frac{1}{2}}}^{x_{i+1}} \frac{(x-x_{i+1})^2}{2\Delta x}\phi'''(x)\,dx,$$

$$\phi(x_i) - \phi(x_{i+\frac{1}{2}}) = -\frac{1}{2}\phi'(x_{i+\frac{1}{2}}) + \frac{\Delta x^2}{8}\phi''(x_{i+\frac{1}{2}}) - \int_{x_i}^{x_{i+\frac{1}{2}}} \frac{(x-x_i)^2}{2\Delta x}\phi'''(x)\,dx.$$

Then, we take the difference of these relations to obtain:

$$\frac{\phi(x_{i+1}) - \phi(x_i)}{\Delta x} - \phi'(x_{i+\frac{1}{2}}) = \int_{x_{i+\frac{1}{2}}}^{x_{i+1}} \frac{(x-x_{i+1})^2}{2\Delta x}\phi'''(x)\,dx + \int_{x_i}^{x_{i+\frac{1}{2}}} \frac{(x-x_i)^2}{2\Delta x}\phi'''(x)\,dx.$$

Now, using a Young then a Cauchy-Schwarz inequalities gives

$$\left|\frac{\phi(x_{i+1}) - \phi(x_i)}{\Delta x} - \phi'(x_{i+\frac{1}{2}})\right|^2 \leq \frac{\Delta x^3}{320}\int_{x_i}^{x_{i+1}} \phi'''(x)^2\,dx.$$

Again, the final result is obtained by multiplying by $\Delta x$ and taking the sum over $i \in \mathbb{Z}$. $\square$

Now we study the truncation error $a_i^n$ which is defined by (37a). Since $(\rho, g)$ is the exact solution, we have:

$$a_i^n = \left(\frac{\rho(t_{n+1}, x_i) - \rho(t_n, x_i)}{\Delta t} - \partial_t\rho(t_n, x_i)\right) + \left[v\left(\frac{g(t_{n+1}, x_{i+\frac{1}{2}}) - g(t_{n+1}, x_{i-\frac{1}{2}})}{\Delta x} - \partial_x g(t_{n+1}, x_i)\right)\right]$$

Then we estimate the norm of $a^n$ as follows: we use a Young inequality, lemma A.1, (iii) of lemma A.2, and lemma 3.5, and we obtain

$$\sum_{i\in\mathbb{Z}} |a_i^n|^2 \Delta x \leq 2\left(\sum_{i\in\mathbb{Z}} \Delta t^2 \max_{[0,T]} |\partial_{tt}\rho(t, x_i)|^2\,\Delta x + \frac{1}{2}\left[\frac{\Delta x^4}{320}\,\|\partial_{xxx} g(t_{n+1}, ., v)\|_{L^2}^2\right]\right).$$

By using (i) of lemma A.2, the first term of the right-hand side of the previous estimate can be controlled by a norm of $\rho$ (if $\Delta x \leq 1$) and we get

$$\sum_{i\in\mathbb{Z}} |a_i^n|^2 \Delta x \leq 2\left(2\Delta t^2 \|\rho\|_X^2 + \frac{1}{4}\frac{\Delta x^4}{320}\|g\|_{\mathcal{Y}}^2\right),$$

where $X = C^2([0,T], H^1(\mathbb{R}))$ and $\mathcal{Y} = C^0([0,T], L^2([-1,1], H^3(\mathbb{R})))$. Consequently, it is clear that $\|a^n\| \leq \tilde{C}_1(\Delta t + \Delta x^2)$, where $\tilde{C}_1$ depends only on the norms of $\rho$ and $g$.

Now, we study the truncation error $b_i^n$ which is defined by (37c). Again, since $(\rho, g)$ is the exact solution, we have:

$$\begin{aligned}
b_i^n =& \varepsilon^2 \left( \frac{g(t_{n+1}, x_{i+\frac{1}{2}}, v) - g(t_n, x_i, v)}{\Delta t} - \partial_t g(t_n, x_i, v) \right) \\
&+ \varepsilon(I - [.]) \left( v^+ \left( \frac{g(t_n, x_{i+\frac{1}{2}}, v) - g(t_n, x_{i-\frac{1}{2}}, v)}{Dx} - \partial_x g(t_n, x_{i-\frac{1}{2}}, v) \right) \right. \\
&\qquad \left. + v^- \left( \frac{g(t_n, x_{i+\frac{3}{2}}, v) - g(t_n, x_{i+\frac{1}{2}}, v)}{Dx} - \partial_x g(t_n, x_{i+\frac{1}{2}}, v) \right) \right) \\
&+ \sigma \left( g(t_{n+1}, x_{i+\frac{1}{2}}, v) - g(t_n, x_{i+\frac{1}{2}}, v) \right) \\
&+ v \left( \frac{\rho(t_n, x_{i+1}) - \rho(t_n, x_i)}{\Delta x} - \partial_x \rho(t_n, x_{i+\frac{1}{2}}) \right)
\end{aligned}$$

The sequel is similar, though a but longer, to what we did for $a_i^n$. By using again lemmas A.1 and A.2, and some standard inequalities (Young and Jensen), the reader can easily find that the following estimate is satisfied:

$$\sum_{i \in \mathbb{Z}} \left[ |b_i^n|^2 \right] \Delta x \leq 4 \left( \varepsilon^4 2 \Delta t^2 \|g\|_{\mathcal{X}}^2 + \varepsilon^2 2 \frac{4}{3} \Delta x^2 \|g\|_{\mathcal{Y}}^2 + \sigma^2 2 \Delta t^2 \|g\|_{\mathcal{X}}^2 + \frac{\Delta x^4}{5 \times 2^6} \|\rho\|_Y^2 \right),$$

where $\mathcal{X} = C^2([0, T], L^2([-1, 1], H^1(\mathbb{R})))$ and $Y = C^0([0, T], H^3(\mathbb{R}))$. Therefore, we have $|||b^n||| \leq \tilde{C}_2 \left( (1 + \varepsilon^2) \Delta t + \varepsilon \Delta x + \Delta x^2 \right)$, where $C_2$ depends only on the norms of $\rho$ and $g$. Finally, the estimates of $\|a^n\|$ and $|||b^n|||$ allow us to complete the proof.