Commun. Comput. Phys.
doi: 10.4208/cicp.290615.020316a

Vol. **20**, No. 3, pp. 551-582
September 2016

Extended Hydrodynamic Models and Multigrid Solver of a Silicon Diode Simulation

Zhicheng Hu[1], Ruo Li[2] and Zhonghua Qiao[1,*]

[1] *Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Hong Kong.*
[2] *HEDPS & CAPT, LMAM & School of Mathematical Sciences, Peking University, Beijing, China.*

Received 29 June 2015; Accepted (in revised version) 2 March 2016

Abstract. Extended hydrodynamic models for carrier transport are derived from the semiconductor Boltzmann equation with relaxation time approximation of the scattering term, by using the globally hyperbolic moment method and the moment-dependent relaxation time. Incorporating the microscopic relaxation time and the applied voltage bias, a formula is proposed to determine the relaxation time for each moment equation, which sets different relaxation rates for different moments such that higher moments damp faster. The resulting models would give more satisfactory results of macroscopic quantities of interest with a high-order convergence to those of the underlying Boltzmann equation as the involved moments increase, in comparison to the corresponding moment models using a single relaxation time. In order to simulate the steady states efficiently, a multigrid solver is developed for the derived moment models. Numerical simulations of an n^+-n-n^+ silicon diode are carried out to demonstrate the validation of the presented moment models, and the robustness and efficiency of the designed multigrid solver.

AMS subject classifications: 35L75, 35Q20, 65N55, 82D37

Key words: Boltzmann transport equation, extended hydrodynamic model, moment-dependent relaxation time, multigrid, semiconductor device simulation.

1 Introduction

Numerical simulations of carrier transport in semiconductors are of great interest in the design of modern devices. As the characteristic length of device shrinks into submicron scale regime, the famous drift-diffusion (DD) model [38] and its augmented version, such

*Corresponding author. *Email addresses:* huzhicheng1986@gmail.com (Z. Hu), rli@math.pku.edu.cn (R. Li), zqiao@polyu.edu.hk (Z. Qiao)

 ©2016 Global-Science Press

as the high-field model [19], become more and more inadequate to describe the important transport phenomena [22]. In such situations, one has to turn back to consider the fundamental Boltzmann transport equation (BTE), which gives a mesoscopic description of the charged particles in a semiclassical approximation. It follows, however, a great growth of the computational cost for numerical simulations based on the BTE in comparison to the DD model. This limits the widespread use of the BTE in a practical computer-aided design (CAD) of semiconductors, and nowadays to develop some appropriate macroscopic models, that describe the transport phenomena well with a significant reduction of the computational cost, is still receiving considerable attention. Following this direction, many models have been raised, such as energy-transport equations, hydrodynamic equations and high-order hydrodynamic equations, see [22] and [28] for a detailed description. One of powerful and systematic approaches to derive the macroscopic models might be employing Grad's moment method [21] with an appropriate closure ansatz for the resulting moment system. As pointed out in [22], the closure relations are crucial for the successful usage of such method. A well-posed entropy maximum closure has been proposed in [35]. Yet this closure is only of theoretical interest, for it depends on the analytical expression of the distribution function, which is unavailable in general for the system involving many moments. By imposing the derived quasi-linear moment system to satisfy the essential hyperbolicity, a new closure without any additionally empirical parameters has been presented in the context of the Boltzmann equation [7, 8] and the Wigner equation [9]. The systematic derivation makes it possible to employ the hyperbolic moment system with moments up to arbitrary order as one of promising models for the practical semiconductor device simulation. Moreover, the hyperbolic moment system is expected to converge to the underlying BTE with a high-order rate as the involved moments increase, for the Grad moment expansion can be viewed as a certain Hermite spectral discretization of the distribution function.

As a preliminary application of the hyperbolic moment method therein, the simulation of a simple n^+-n-n^+ diode, where the carrier transport can be depicted by the one dimensional BTE with the relaxation time approximation scattering term, has been carried out in [26]. It is shown that, for the diode with a channel of 400nm at the applied voltage bias lower than 0.5V, the 5th-order hyperbolic moment system is able to give the macroscopic quantities of physical interest, including current-voltage (I-V) characteristic curve, electron density, mean velocity, potential and electric field, that agree well with the reference obtained by the discrete velocity method (DVM) of the underlying BTE. However, oscillations, that may introduce numerical instability of the method, is also observed in those solutions, especially the temperature in the case the voltage bias larger than 0.5V is applied. This makes the expected convergence of the system can not be evidently verified, and restricts the application of the system to a more realistic device.

With careful and comprehensive observations, it is noticed that the scattering part of the entire moment system, derived in [26], employs the same microscopic relaxation time of the BTE scattering term. This is typically different from the usual approach of deriving hydrodynamic and extended hydrodynamic models, where a set of one single

relaxation time is adopted for each macroscopic quantity, see e.g., [1,2,4]. In order to give accurate models, these macroscopic relaxation times have to be carefully calibrated from the Monte Carlo results. As a result, the computational cost for the matching procedure of all macroscopic relaxation times will become very expensive, if the moment system of a very high order is considered.

On the other hand, according to the spectral viscosity method [36] and the filtered spherical harmonics method [34], it is possible to prevent the oscillations and strengthen the stability of the method by gradually decreasing macroscopic relaxation times such that higher moments damp faster. Motivated by this, we propose in the present paper an empirical formula, that incorporates the microscopic relaxation time and the applied voltage bias with an additional single parameter, to give the relaxation time for each moment equation. The required parameter is actually insensitive, and could be easily determined via fitting solutions of the underlying BTE. With such a moment-dependent relaxation time, it can be seen that oscillations in the solution of the modified hyperbolic moment system disappear, and the expected high-order convergence of the system can be well validated.

Apparently, the numerical scheme proposed in [26] is still suitable for the modified moment models. For simulations of steady states, however, the state-of-the-art numerical techniques for efficiency improvement have not been sufficiently explored, since the explicit time-stepping scheme therein takes a long time integration to achieve the steady state. As one of most powerful approaches, multigrid methods [6,24] have been successfully applied for various problems, see e.g., [5,25,29,33,37]. In this paper, we are also concerned with the development of multigrid solver for the derived models. In [25], a nonlinear multigrid (NMG) iteration, which has a unified framework for the hyperbolic moment system of arbitrary order, has been developed and demonstrated a significant improvement in efficiency for the simulation of microflows. Unfortunately, this NMG iteration can not be directly extended to the model under our current consideration, for a coupled Poisson equation must be solved to provide the self-consistent potential and electric field. Due to the importance of maintaining self-consistency throughout the iterative procedure, we develop an NMG iteration for the coupled problem, where the Gummel method using the Richardson iteration with a symmetric Gauss-Seidel (SGS) acceleration for the moment system part is utilized as the smoother. In order to efficiently prepare a good initial guess, which is also important for the multigrid approach, the interpolation of the solution improved by several NMG iterations on the coarser grid is employed for it. Recursively applying this strategy then gives a full multigrid (FMG) solver.

The rest part of this paper is organized as follows. In Section 2, the model equations, including the one dimensional semiconductor Boltzmann equation, the self-consistent Poisson equation, and the hyperbolic moment system of arbitrary order with moment-dependent relaxation time, are introduced, followed with a unified discretization. Section 3 describes the details of the Gummel iteration on the single grid level. Using the Gummel iteration as smoother, in Section 4, the nonlinear multigrid iteration and the re-

sulting full multigrid solver are presented. Numerical simulations of a classical diode are then carried out in Section 5 to demonstrate the validation, robustness and efficiency of the proposed solver. Some concluding remarks are given in the last section.

2 Model equations

2.1 Semiconductor Boltzmann equation

The carrier transport, in simplified semiconductor devices such as diode, where the important features of the charge transport are given in the direction parallel to the force field, can be well modeled by the one dimensional Boltzmann transport equation [15]. In steady state, it reads

$$v\frac{\partial f}{\partial x}+\frac{q_e}{m^*}E(x)\frac{\partial f}{\partial v}=Q(f), \tag{2.1}$$

where $f(x,v)$ is the probability density function (distribution function) of an electron at position $x\in[0,L]$ and velocity $v\in\mathbb{R}$, in which L is the length of the device. The constants q_e and m^* are, respectively, the charge unit and the effective mass of electron. The electric field E is defined by

$$E(x)=-\frac{\Phi_x}{q_e}, \tag{2.2}$$

where Φ is the self-consistent potential determined by the coupled Poisson equation

$$-\epsilon_0\Phi_{xx}=q_e^2(\rho-C(x)), \tag{2.3}$$

with the boundary conditions

$$\Phi(0)=0, \qquad \Phi(L)=-q_eV_{bias}, \tag{2.4}$$

where ϵ_0 is the permittivity of the device material, V_{bias} is the applied voltage bias, $C(x)$ is the doping density, and ρ is the electron number density given by

$$\rho(x)=\int_{\mathbb{R}}f(x,v)\,\mathrm{d}v. \tag{2.5}$$

Some other macroscopic quantities of electrons, such as mean velocity u, temperature θ, are related to f by

$$\rho(x)u(x)=\int_{\mathbb{R}}vf(x,v)\,\mathrm{d}v, \qquad \frac{1}{2}\rho(x)\theta(x)=\frac{1}{2}\int_{\mathbb{R}}|v-u|^2f(x,v)\,\mathrm{d}v. \tag{2.6}$$

The right-hand side of (2.1) is the scattering term representing all kinds of scattering mechanisms, which are of essential importance when modeling devices [31,39]. Taking

into account only scattering with background impurities by the low density approximation, the scattering term can be replaced by a linear relaxation time approximation [15] as

$$Q(f)=\frac{1}{\tau}(f^{\mathrm{eq}}-f), \tag{2.7}$$

in which the relaxation time τ is computed by

$$\tau=\frac{m^*\mu}{q_e}, \tag{2.8}$$

and the local equilibrium distribution f^{eq} is a Maxwellian given by

$$f^{\mathrm{eq}}(x,v)=\frac{\rho(x)}{(2\pi\theta_L)^{1/2}}\exp\left(-\frac{|v|^2}{2\theta_L}\right), \tag{2.9}$$

where μ is the mobility and θ_L is the lattice temperature with $\theta_L=\frac{k_B}{m^*}T_L$. Here, k_B is the Boltzmann constant, and T_L is the lattice temperature in Kelvin. Some choices of the mobility for different characterizations can be found in [18], and in the present study we would like to employ the electric-field-dependent formula

$$\mu(E)=\frac{2\mu_0}{1+\sqrt{1+4(\mu_0|E|/v_d)^2}}, \tag{2.10}$$

where μ_0 and v_d stand for the low-field mobility and the saturation speed respectively.

For practical applications, the BTE (2.1) has to be equipped with proper boundary conditions at the source contact $x=0$ and the drain contact $x=L$. In order to preserve the electric neutrality numerically, we adopt the boundary conditions used in [16, 20] in our simulations. Precisely, we have

$$f\left(0^-,v\right)=c_l f\left(0^+,v\right), \qquad f\left(L^+,v\right)=c_r f\left(L^-,v\right), \tag{2.11}$$

where the coefficients c_l and c_r are determined by

$$c_l=\frac{C(0)}{\displaystyle\int_{\mathbb{R}} f\left(0^+,v\right)\mathrm{d}v}, \qquad c_r=\frac{C(L)}{\displaystyle\int_{\mathbb{R}} f\left(L^-,v\right)\mathrm{d}v}. \tag{2.12}$$

In other words, the distribution function at the boundary is rescaled to satisfy the electric neutrality.

2.2 Extended hydrodynamic models

Following Grad's moment method [21], we expand the distribution function $f(x,v)$ into a series as

$$f(x,v)=\sum_{\alpha=0}^{\infty} f_\alpha(x)\mathcal{H}_\alpha^{[u(x),\theta(x)]}(v), \tag{2.13}$$

where $f_\alpha(x)$ are the coefficients, and $\mathcal{H}_\alpha^{[\cdot,\cdot]}(\cdot)$ are the basis functions defined by

$$\mathcal{H}_\alpha^{[u,\theta]}(v)=\frac{1}{\sqrt{2\pi}}\theta^{-\frac{\alpha+1}{2}}He_\alpha(\xi)\exp\left(-\xi^2/2\right),\quad \forall \alpha\in\mathbb{N},\ v\in\mathbb{R}, \tag{2.14}$$

in which $\xi=(v-u)/\sqrt{\theta}$, and $He_n(\cdot)$ is the Hermite polynomial of degree n, i.e.,

$$He_n(x)=(-1)^n\exp\left(x^2/2\right)\frac{\mathrm{d}^n}{\mathrm{d}x^n}\exp\left(-x^2/2\right).$$

Note the basis functions $\mathcal{H}_\alpha^{[u,\theta]}(v)$ are dependent on f via the local mean velocity $u(x)$ and temperature $\theta(x)$. With such an expansion, the first three coefficients have the following simple relations

$$f_0=\rho,\qquad f_1=f_2=0. \tag{2.15}$$

It is useful to introduce $\mathcal{F}^{[u,\theta]}$ and $\mathcal{F}_M^{[u,\theta]}$ to denote, respectively, the linear spaces spanned by $\mathcal{H}_\alpha^{[u,\theta]}(v)$ for all $\alpha\in\mathbb{N}$ and $\alpha\le M$. Accordingly, $\mathcal{F}_M^{[u,\theta]}$ forms a finite dimensional subspace of $L^2(\mathbb{R},\exp((v-u)^2/(2\theta))\mathrm{d}v)$, and $\mathcal{F}_M^{[u,\theta]}\subset\mathcal{F}_{M+1}^{[u,\theta]}\subset\cdots\subset\mathcal{F}^{[u,\theta]}$. If the equilibrium distribution $f^{\mathrm{eq}}\in\mathcal{F}^{[u,\theta]}$, then we have the following expansion

$$f^{\mathrm{eq}}(x,v)=\sum_{\alpha=0}^{\infty} f_\alpha^{\mathrm{eq}}(x)\mathcal{H}_\alpha^{[u(x),\theta(x)]}(v), \tag{2.16}$$

where the coefficients f_α^{eq} can be calculated recursively by

$$f_0^{\mathrm{eq}}=\rho,\qquad f_1^{\mathrm{eq}}=-\rho u,\qquad f_\alpha^{\mathrm{eq}}=\frac{\theta_L-\theta}{\alpha}f_{\alpha-2}^{\mathrm{eq}}-\frac{u}{\alpha}f_{\alpha-1}^{\mathrm{eq}},\qquad \alpha\ge 2. \tag{2.17}$$

The detailed derivation of the above formula is presented in the Appendix.

Plugging (2.13) and (2.16) into the BTE (2.1), and matching the coefficients of the same basis function, we will get an infinite system of equations for the mean velocity u, the temperature θ and the coefficients f_α. To make it numerically solvable, we choose a positive integer $M\ge 2$ and select the finite system consisting of the first $M+1$ equations. This is equivalent to approximating the distribution function f in the space $\mathcal{F}_M^{[u,\theta]}$, i.e.,

$$f(x,v)\approx\sum_{\alpha=0}^{M} f_\alpha(x)\mathcal{H}_\alpha^{[u(x),\theta(x)]}(v). \tag{2.18}$$

Unfortunately, such a finite system is not closed, since the coefficient f_{M+1} appears in the last equation. An appropriate closure should be employed to replace it by an expression involving u, θ and the first $M+1$ coefficients. Due to the essential role of hyperbolicity in the local well-posedness of the system, we adopt the globally hyperbolic closure proposed in [7,8] in this paper. As a result, the final system of the first $M+1$ equations becomes

$$\begin{aligned}&\theta\frac{\partial f_{\alpha-1}}{\partial x}+u\frac{\partial f_\alpha}{\partial x}+(1-\delta_{\alpha,M})(\alpha+1)\frac{\partial f_{\alpha+1}}{\partial x}+\frac{\partial u}{\partial x}\Big(\theta f_{\alpha-2}+uf_{\alpha-1}+(1-\delta_{\alpha,M})(\alpha+1)f_\alpha\Big)\\&\quad+\frac{1}{2}\frac{\partial\theta}{\partial x}\Big(\theta f_{\alpha-3}+uf_{\alpha-2}+(1-\delta_{\alpha,M})(\alpha+1)f_{\alpha-1}\Big)\\&=\frac{q_e}{m^*}Ef_{\alpha-1}+\frac{f_\alpha^{\rm eq}-f_\alpha}{\tau},\qquad 0\le\alpha\le M,\end{aligned}\tag{2.19}$$

where δ_{ij} is the Kronecker delta symbol, and f_α is taken as zero if $\alpha=1,2$ or $\alpha<0$. Let $\boldsymbol{w}=(\rho,u,\theta,f_3,\cdots,f_M)^T$, then the above system can be reformulated into a quasi-linear form as

$$\mathbf{M}(\boldsymbol{w})\frac{\partial\boldsymbol{w}}{\partial x}=\mathbf{G}(\boldsymbol{w})+\boldsymbol{Q}(\boldsymbol{w}),\tag{2.20}$$

where the left-hand side corresponds to the left-hand side of (2.19), $\mathbf{G}(\boldsymbol{w})$ corresponds to the first term of the right-hand side of (2.19), and $\boldsymbol{Q}(\boldsymbol{w})$ corresponds to the second term of the right-hand side of (2.19), respectively. As proved in [7], the matrix $\mathbf{M}$ is diagonalizable and its eigenvalues are $u+c_j\sqrt{\theta}$, where c_j is the jth root of the Hermite polynomial $He_{M+1}(x)$. That is, the system (2.20), or equivalently the system (2.19), is globally hyperbolic, and hence is referred to as the hyperbolic moment system of order M. From the macroscopic point of view, the hyperbolic moment system (2.19) can be viewed as an extended hydrodynamic model for semiconductors. In particular, it gives the classical hydrodynamic model when $M=2$. Alternatively, in terms of numerical methods, the system (2.19) is actually discretized from the underlying BTE (2.1) by the Hermite spectral method in velocity space. Thereby, it is expected that the system (2.19), or in other words, the truncated expansion (2.18), has a very fast convergence to the exact solution of the BTE, as the order M goes to infinity when the solution is smooth.

However, it has been observed in Fig. 3 (see also in [26]) that numerical solutions of the system (2.19) might contain oscillations in the local region, where the temperature has a very large gradient. Consequently, a very high-order moment system, e.g., $M=30$ in Fig. 3, is required to get a satisfactory approximation of the exact solution of the BTE. On the other hand, these oscillations may imply some instability of the moment method when M is large. These strongly restrict the application of the hyperbolic moment system.

In our preliminary study, we found that oscillations in the solution disappear, providing that the relaxation time τ in the system (2.19) is modified to a moment-dependent relaxation time τ_α for the αth equation such that higher moments damp faster. This idea comes from the spectral viscosity method [36], which avoids possible instability of the

spectral method. Similar technique has already been used in the field of radiative transport [34]. Based on a number of numerical experiments, we propose the modification of τ_α as follows

$$\tau_\alpha^{-1}=\begin{cases}\tau^{-1}, & \text{if } \alpha=0,\\ \left(1+\beta V_{bias}(\alpha-1)^2\right)\tau^{-1}, & \text{if } 1\le\alpha\le M,\end{cases} \tag{2.21}$$

where β is a user-defined parameter. In our numerical tests, β is insensitive and is often set as 0.1. With the above modification, the moment system (2.19) is modified to

$$\begin{aligned}&\theta\frac{\partial f_{\alpha-1}}{\partial x}+u\frac{\partial f_\alpha}{\partial x}+(1-\delta_{\alpha,M})(\alpha+1)\frac{\partial f_{\alpha+1}}{\partial x}+\frac{\partial u}{\partial x}\Big(\theta f_{\alpha-2}+uf_{\alpha-1}+(1-\delta_{\alpha,M})(\alpha+1)f_\alpha\Big)\\&\quad+\frac{1}{2}\frac{\partial\theta}{\partial x}\Big(\theta f_{\alpha-3}+uf_{\alpha-2}+(1-\delta_{\alpha,M})(\alpha+1)f_{\alpha-1}\Big)\\&=\frac{q_e}{m^*}Ef_{\alpha-1}+\frac{f_\alpha^{\rm eq}-f_\alpha}{\tau_\alpha},\quad 0\le\alpha\le M.\end{aligned} \tag{2.22}$$

It is apparent that the first two equations in (2.22) remain the same as those in (2.19) and other equations are modified slightly. Therefore, most properties of (2.19) such as global hyperbolicity are preserved for the modified moment system (2.22). We point out that (2.21) is just an empirical formula for τ_α and the underlying mechanism of this behavior remains unclear. However, we will show in Section 5 that the modified moment system (2.22) completely prevents oscillations, and could give a more accurate approximation to the BTE than the original hyperbolic moment system (2.19).

2.3 Discretization

In order to develop a unified solver for the system with moments up to arbitrary order, we discretize the modified moment system (2.22) under the framework of the NRxx method which has been studied in [10, 11, 13, 14]. The main idea is considering the discretization formally based on the truncated expansion (2.18), rather than directly based on the moments $\boldsymbol{w}$. More specifically, suppose the spatial domain $[0,L]$ is divided by a mesh

$$0=x_0<x_1<\cdots<x_{N-1}<x_N=L,$$

and let $\Delta x_i=x_{i+1}-x_i$. Then the finite volume discretization of the BTE (2.1) over the ith cell $[x_i,x_{i+1}]$ can be written in a general framework as

$$\frac{F(f_i(v),f_{i+1}(v))-F(f_{i-1}(v),f_i(v))}{\Delta x_i}=G(f_i(v))+Q(f_i(v)), \tag{2.23}$$

where $f_i(v)\in\mathcal{F}_M^{[u_i,\theta_i]}$ is the approximation of the distribution function on the ith cell, i.e.,

$$f_i(v)=\sum_{\alpha=0}^{M}f_{i,\alpha}\mathcal{H}_\alpha^{[u_i,\theta_i]}(v),$$

and u_i, θ_i are corresponding mean velocity and temperature, respectively, such that the relation (2.15) holds for the coefficients $f_{i,\alpha}$.

The left-hand side of (2.23) is discretized from the convection term, and the numerical flux $F(f_i,f_{i+1})$, defined at the right boundary of the ith cell, is constructed specially to reflect the hyperbolic closure of the moment system. In our numerical experiments, the numerical flux presented in [12] is used. To be specific, $F(f_i,f_{i+1})=\hat{F}_{i+1/2}+\tilde{F}^{-}_{i+1/2}$ and $F(f_{i-1},f_i)=\hat{F}_{i-1/2}+\tilde{F}^{+}_{i-1/2}$ in the ith cell's discretization. Here, $\hat{F}_{i+1/2}$ is the HLL numerical flux defined by

$$\hat{F}_{i+1/2}=\begin{cases} vf_i(v), & 0\le\lambda^L, \\ \dfrac{\lambda^R vf_i(v)-\lambda^L vf_{i+1}(v)+\lambda^L\lambda^R(f_{i+1}(v)-f_i(v))}{\lambda^R-\lambda^L}, & \lambda^L<0<\lambda^R, \\ vf_{i+1}(v), & 0\ge\lambda^R, \end{cases} \tag{2.24}$$

and $\tilde{F}^{\pm}_{i+1/2}$ is the numerical flux for the hyperbolic closure defined by

$$\tilde{F}^{-}_{i+1/2}=\begin{cases} 0, & 0\le\lambda^L, \\ -\dfrac{\lambda^L}{\lambda^R-\lambda^L}g_{i+1/2}(v), & \lambda^L<0<\lambda^R, \\ g_{i+1/2}(v), & 0\ge\lambda^R, \end{cases} \tag{2.25}$$

and

$$\tilde{F}^{+}_{i+1/2}=\begin{cases} -g_{i+1/2}(v), & 0\le\lambda^L, \\ -\dfrac{\lambda^R}{\lambda^R-\lambda^L}g_{i+1/2}(v), & \lambda^L<0<\lambda^R, \\ 0, & 0\ge\lambda^R, \end{cases} \tag{2.26}$$

where λ^L (λ^R) is the minimal (maximal) one among the eigenvalues of $\mathbf{M}(\boldsymbol{w}_i)$ and $\mathbf{M}(\boldsymbol{w}_{i+1})$, and $g_{i+1/2}(v)$ is a specific function determined from hyperbolic closure and the DLM theory [32]. It is emphasized that in the ith cell's discretization (2.23), both numerical fluxes, $F(f_{i-1},f_i)$ and $F(f_i,f_{i+1})$, are computed and approximated in the space $\mathcal{F}_M^{[u_i,\theta_i]}$, i.e.,

$$\begin{aligned} F(f_{i-1},f_i)&=\sum_{\alpha=0}^{M}F_\alpha(f_{i-1},f_i)\mathcal{H}_\alpha^{[u_i,\theta_i]}(v), \\ F(f_i,f_{i+1})&=\sum_{\alpha=0}^{M}F_\alpha(f_i,f_{i+1})\mathcal{H}_\alpha^{[u_i,\theta_i]}(v), \end{aligned} \tag{2.27}$$

where $F_\alpha(f_{i-1},f_i)$ and $F_\alpha(f_i,f_{i+1})$ are the corresponding expansion coefficients. Since $f_{i\pm1}(v)\in\mathcal{F}_M^{[u_{i\pm1},\theta_{i\pm1}]}$, the computation of the above formulae requires a transformation

between $\mathcal{F}_M^{[u_{i\pm1},\theta_{i\pm1}]}$ and $\mathcal{F}_M^{[u_i,\theta_i]}$. A fast transformation between two spaces, $\mathcal{F}_M^{[u,\theta]}$ and $\mathcal{F}_M^{[\tilde{u},\tilde{\theta}]}$, has already been provided in [10]. Indeed, such a transformation is the core of the NRxx method. And in this paper, we might call this transformation, when it is necessary, without explicitly pointing out.

Similar to the numerical flux, the terms $G(f_i)$ and $Q(f_i)$ in (2.23), which are respectively corresponding to the discretization of the electric force term and the scattering term, are also approximated in $\mathcal{F}_M^{[u_i,\theta_i]}$, i.e.,

$$G(f_i(v))=\sum_{\alpha=0}^{M} G_{i,\alpha}\mathcal{H}_\alpha^{[u_i,\theta_i]}(v), \tag{2.28}$$

$$Q(f_i(v))=\sum_{\alpha=0}^{M} Q_{i,\alpha}\mathcal{H}_\alpha^{[u_i,\theta_i]}(v). \tag{2.29}$$

From the modified moment system (2.22), we can easily deduce that $G_{i,\alpha}=q_eE_if_{i,\alpha-1}/m^*$ and $Q_{i,\alpha}=(f_{i,\alpha}^{\rm eq}-f_{i,\alpha})/\tau_{i,\alpha}$. Then substituting (2.27)-(2.29) into (2.23), and matching the coefficients of the same basis function $\mathcal{H}_\alpha^{[u_i,\theta_i]}(v)$, we get a system which is a discretization of the modified moment system (2.22) on the ith cell.

It remains to provide the ith cell's electric field E_i in the above discretization. To this end, we discretize the self-consistent Poisson equation (2.3) by the standard cell-centered finite difference method, which gives rise to the following system

$$-\frac{2}{\Delta x_i}\left(\frac{\Phi_{i+1}-\Phi_i}{\Delta x_i+\Delta x_{i+1}}-\frac{\Phi_i-\Phi_{i-1}}{\Delta x_{i-1}+\Delta x_i}\right)=\lambda(\rho_i-C_i),\quad i=0,1,\cdots,N-1, \tag{2.30}$$

with $\lambda=q_e^2/\epsilon_0$. Then the electric field on the ith cell is calculated by the central difference formula of (2.2) as

$$E_i=-\frac{\Phi_{i+1/2}-\Phi_{i-1/2}}{q_e\Delta x_i}=-\frac{1}{q_e}\left(\frac{\Phi_{i+1}-\Phi_i}{\Delta x_i+\Delta x_{i+1}}+\frac{\Phi_i-\Phi_{i-1}}{\Delta x_{i-1}+\Delta x_i}\right). \tag{2.31}$$

3 The Gummel iteration

In this section, we will present the algorithm to solve the discretized moment system (2.23) in conjunction with the self-consistent Poisson system (2.30) on a single grid.

For simplicity, let us define the local residual of the moment system on the ith cell as

$$R_i(v;f_{i-1},f_i,f_{i+1};E_i)=\frac{F(f_i(v),f_{i+1}(v))-F(f_{i-1}(v),f_i(v))}{\Delta x_i}-G(f_i(v))-Q(f_i(v)). \tag{3.1}$$

Then the discretization (2.23) can be rewritten into

$$R_i(v;f_{i-1},f_i,f_{i+1};E_i)=r_i(v),\quad i=0,1,\cdots,N-1, \tag{3.2}$$

with $r_i(v)\equiv 0$. Here we introduce a function $r_i(v)\in\mathcal{F}_M^{[u_i,\theta_i]}$, which is independent of all $f_j(v)$ and E_j, $j=0,1,\cdots,N-1$, to make the problem slightly more general. In addition, the discretized Poisson system (2.30) is reformulated into matrix form as

$$\mathbf{A}\boldsymbol{\Phi}-\lambda\boldsymbol{\rho}=\boldsymbol{s}, \tag{3.3}$$

where $\mathbf{A}$ stands for the tridiagonal coefficient matrix, $\boldsymbol{\Phi}$ and $\boldsymbol{\rho}$ are column vectors with their ith components representing Φ_i and ρ_i respectively. As for the right-hand side $\boldsymbol{s}$, its ith component is $s_i=-\lambda C_i$, yet it is introduced for the same reason as $r_i(v)$.

It is clear that (3.2) is a nonlinear system coupled with (3.3) via $\boldsymbol{\rho}$ and $\boldsymbol{E}$, which is the column vector of the discrete electric field E_i. To solve this problem, there are mainly two types of iterative method. One updates the distribution function and potential simultaneously by solving the coupled system (3.2) and (3.3) as a whole, e.g., the full Newton method [3]. The other one updates two parts of the solution alternately by solving (3.2) and (3.3) independently and alternately, with one system's solution as input for the other. In semiconductor device simulations, the latter approach is known as the Gummel iteration method [23,39], and is more universal since it is much simpler. In this paper, we also focus on the Gummel type method for the coupled system (3.2) and (3.3). Generally, to update a given distribution function f^{old} with $f_i^{\text{old}}(v)\in\mathcal{F}_M^{[u_i^{\text{old}},\theta_i^{\text{old}}]}$, and its corresponding potential $\boldsymbol{\Phi}^{\text{old}}$, as well as electric field $\boldsymbol{E}^{\text{old}}$, a single Gummel iteration consists of three steps as presented in the following algorithm.

Algorithm 1 A single Gummel iteration

1. Get the updated distribution function f^{new} with $f_i^{\text{new}}(v)\in\mathcal{F}_M^{[u_i^{\text{new}},\theta_i^{\text{new}}]}$ from f^{old} and the moment system (3.2) with electric field $\boldsymbol{E}^{\text{old}}$ by a certain method.
2. Extracting $\boldsymbol{\rho}^{\text{new}}$ from f^{new} and substituting it into (3.3), solve the resulting linear system for the updated potential $\boldsymbol{\Phi}^{\text{new}}$.
3. Compute the updated electric field $\boldsymbol{E}^{\text{new}}$ by substituting $\boldsymbol{\Phi}^{\text{new}}$ into (2.31).

Repeating the above algorithm until convergence, we obtain a single grid solver of the Gummel method, which is collected as calling Algorithm 1 in Algorithm 2.

The advantage of the Gummel method is that the existing solvers for each of the two systems, (3.2) and (3.3), could be directly applied for the first two steps of Algorithm 1. Mathematically in this algorithm, step 2 is much easier than step 1, especially in one dimensional case that the tridiagonal linear system (3.3) could be solved by an efficient direct method such as the famous Thomas algorithm. Therefore, the rest of this section is devoted to the method for step 1 of Algorithm 1.

We would first introduce a simple relaxation method, namely, Richardson iteration, which reads

$$f_i^{n+1}(v)=f_i^n(v)+\omega\left(r_i(v)-R_i(v;f_{i-1}^n,f_i^n,f_{i+1}^n;E_i^n)\right),\quad i=0,1,\cdots,N-1, \tag{3.4}$$

Algorithm 2 Outer loop of iterative methods

1. Given an initial distribution function f^0 with $f_i^0(v)\in\mathcal{F}_M^{[u_i^0,\theta_i^0]}$, solve (3.3) for the initial potential $\boldsymbol{\Phi}^0$ and calculate electric field $\boldsymbol{E}^0$ by (2.31). Let $n:=0$.
2. Perform a certain iteration, such as Algorithm 1, to get the updated solution f^{n+1} and $\boldsymbol{\Phi}^{n+1}$, as well as $\boldsymbol{E}^{n+1}$.
3. Calculate the global residual $\tilde{R}$ with $\tilde{R}_i(v)=r_i(v)-R_i(v;f_{i-1}^{n+1},f_i^{n+1},f_{i+1}^{n+1};E_i^{n+1})$, and its L^2 norm, which is defined by

$$\left\|\tilde{R}\right\|=\sqrt{\frac{1}{L}\left(\sum_{i=0}^{N-1}\left\|\tilde{R}_i\right\|^2\Delta x_i\right)}, \tag{3.5}$$

   where the local residual norm is computed using the L^2 norm of the linear space $\mathcal{F}_M^{[u_i^{n+1},\theta_i^{n+1}]}$, that is,

$$\left\|\tilde{R}_i\right\|=\sqrt{\int\left(\tilde{R}_i(v)\right)^2\exp\left(\frac{\left|v-u_i^{n+1}\right|^2}{2\theta_i^{n+1}}\right)\mathrm{d}v}. \tag{3.6}$$

   By using the orthogonality of basis functions, it follows that

$$\left\|\tilde{R}_i\right\|=\sqrt{\sum_{\alpha=0}^{M}C_\alpha\left|\tilde{R}_{i,\alpha}\right|^2}, \tag{3.7}$$

   where $C_\alpha=(2\pi)^{-1/2}\left(\theta_i^{n+1}\right)^{-\alpha-1/2}\alpha!$, and $\tilde{R}_{i,\alpha}$ is the expansion coefficients of $\tilde{R}_i(v)$ in $\mathcal{F}_M^{[u_i^{n+1},\theta_i^{n+1}]}$.
4. If the global residual norm $\left\|\tilde{R}\right\|$ is smaller than a given tolerance *Tol*, that indicates convergence, then stop; otherwise, let $n:=n+1$, and return to step 2.

where ω is an appropriate parameter such that the sequence f^n converges. In our implementation, it numerically consists of two steps as follows:

1. Compute an intermediate distribution function $f_i^*(v)$ in $\mathcal{F}_M^{[u_i^n,\theta_i^n]}$, i.e.,

$$f_i^*(v)=f_i^n(v)+\omega\left(r_i(v)-R_i(v;f_{i-1}^n,f_i^n,f_{i+1}^n;E_i^n)\right).$$

2. Compute the new macroscopic velocity u_i^{n+1} and temperature θ_i^{n+1} from $f_i^*(v)$, then project $f_i^*(v)$ into $\mathcal{F}_M^{[u_i^{n+1},\theta_i^{n+1}]}$ to obtain $f_i^{n+1}(v)$.

If the parameter ω is selected according to the CFL condition

$$\omega \max_i \frac{\lambda_{\max,i}}{\Delta x_i} < 1, \tag{3.8}$$

where $\lambda_{\max,i}$ is the largest value among the absolute values of all eigenvalues of $\mathbf{M}(\boldsymbol{w}_i^n)$, then the Richardson iteration (3.4) corresponds to an explicit time-integration scheme. Consequently, the resulting Gummel iteration becomes a transient method which is inherently more robust [3].

However, as a steady-state computation, we may still give up time accuracy in the interest of speed of convergence, while keep robustness. To this end, the parameter ω in (3.4) is modified slightly to a local parameter ω_i according to the local CFL condition

$$\omega_i \frac{\lambda_{\max,i}}{\Delta x_i} < 1. \tag{3.9}$$

In order to enhance the robustness of the method, the choice of ω_i should also take into account the positivity preservation of the local density and temperature, which can be found in [25]. Such a local parameter might accelerate the iteration significantly, especially when the non-uniform grid is used [27]. On the other hand, the local parameter is more appropriate when the Richardson iteration (3.4) is revised to an SGS version for further improvement of efficiency. Concretely, the SGS-type Richardson iteration is formulated into two loops as

1. Loop i increasingly from 0 to $N-1$, and update the ith distribution function by

$$f_i^{n+\frac{1}{2}}(v) = f_i^n(v) + \omega_i \left(r_i(v) - R_i(v; f_{i-1}^{n+\frac{1}{2}}, f_i^n, f_{i+1}^n; E_i^n) \right). \tag{3.10}$$

2. Loop i decreasingly from $N-1$ to 0, and update the ith distribution function by

$$f_i^{n+1}(v) = f_i^{n+\frac{1}{2}}(v) + \omega_i \left(r_i(v) - R_i(v; f_{i-1}^{n+\frac{1}{2}}, f_i^{n+\frac{1}{2}}, f_{i+1}^{n+1}; E_i^n) \right). \tag{3.11}$$

Remark 3.1. The Gummel method using the SGS-type Richardson iteration could converge in general several times as fast as the one using the original Richardson iteration. Yet the method behaves similar to the one using the original Richardson iteration and most of other single grid methods including the time-integration scheme. To be precise, the total iterations increases linearly as the grid number increases.

Remark 3.2. From the implementation point of view, both the SGS-Newton iteration and the NMG iteration proposed in [25] could be used instead of the SGS-type Richardson iteration. However, we observe that the resulting Gummel method usually generates a more wildly oscillating sequence of residual, and even does not converge unless the initial guess is good enough. The reason should be that the variation of $\boldsymbol{\rho}$, produced in

these iterations, is too large such that the decoupled Poisson system (3.3) fails to provide an appropriate potential, that is compatible with the computation in the next step. In other words, such a Gummel iteration could not maintain the self-consistency well as the iteration goes on.

4 Multigrid solver

It is well known that a well-designed multigrid algorithm for linear problem could produce convergence rates independent of grid size, indicating an essentially efficiency improvement on a certain sufficiently fine grid. So in this section, we turn to follow the multigrid method to develop a solver as efficient as possible.

As can be seen, what we need to solve is a nonlinear system. Since the system might be very large when the high-order moment system is considered, a global linearization would be too difficult to be implemented efficiently. Owing to this, we adopt the nonlinear multigrid approach [24], known also as the full approximation storage (FAS) multigrid method [5], instead of the global-linearization-based multigrid method. A great benefit is that the single grid methods mentioned in previous section can be directly employed as the smoothing operator. Besides, the framework of the NMG algorithm is quite standard, and can be found in a variety of textbooks, e.g., [6, 24]. Only two remaining key points, the coarse grid correction and the operators (restriction and prolongation) between the fine and coarse grids, need to be further specified for the target problem, that is, the coupled system (3.2) and (3.3). In the following subsections, we first consider the two-grid case to illustrate these ingredients. Then we give a complete multigrid algorithm by recursively applying the two-grid iteration. Finally, we will discuss the importance of the initial guess for an iterative method, followed with a full multigrid solver.

4.1 Coarse grid correction

For convenience, we introduce subscripts h and H to denote operators and variables related to the fine and coarse grids, respectively. In view of the importance of the self-consistency during the iterative procedure, we must consider the coarse grid correction for both the distribution function and the potential simultaneously.

Let us first rewrite the fine grid problem resulting from (3.2) and (3.3) into a global form as

$$\begin{cases} R_h(f_h;\boldsymbol{E}_h)=r_h, & (4.1)\\ \mathbf{A}_h\boldsymbol{\Phi}_h-\lambda\boldsymbol{\rho}_h=\boldsymbol{s}_h. & (4.2)\end{cases}$$

After performing several smoothing steps of the Gummel iteration, suppose we have the approximate solution $\bar{f}_h$ and $\bar{\boldsymbol{\Phi}}_h$ as well as $\bar{\boldsymbol{E}}_h$ for the above problem. Then according to

the NMG framework, the corresponding coarse grid problem is given as

$$\begin{cases} R_H(f_H;\boldsymbol{E}_H)=r_H \triangleq R_H(\tilde{I}_h^H \bar{f}_h;\bar{\boldsymbol{E}}_H)+I_h^H\left(r_h-R_h(\bar{f}_h;\bar{\boldsymbol{E}}_h)\right), & (4.3)\\ \mathbf{A}_H\boldsymbol{\Phi}_H-\lambda\boldsymbol{\rho}_H=\boldsymbol{s}_H \triangleq \mathbf{A}_H\tilde{\tilde{I}}_h^H\bar{\boldsymbol{\Phi}}_h-\lambda\bar{\boldsymbol{\rho}}_H+\hat{I}_h^H(\boldsymbol{s}_h-\mathbf{A}_h\bar{\boldsymbol{\Phi}}_h+\lambda\bar{\boldsymbol{\rho}}_h), & (4.4)\end{cases}$$

where $\square_h^H$ are the restriction operators from the fine grid to the coarse grid, $\bar{\boldsymbol{E}}_H$ is calculated from $\tilde{\tilde{I}}_h^H\bar{\boldsymbol{\Phi}}_h$ by the formula (2.31), and $\bar{\boldsymbol{\rho}}_h$ and $\bar{\boldsymbol{\rho}}_H$ are respectively the first moments of $\bar{f}_h$ and $\tilde{I}_h^H\bar{f}_h$. As for the coarse grid operators R_H and $\mathbf{A}_H$, they are analogous to the fine grid counterparts R_h and $\mathbf{A}_h$, in other words, both R_H and $\mathbf{A}_H$ are constructed respectively, by re-discretizing the moment system (2.22) and the Poisson equation (2.3). It follows that the coarse grid problem (4.3) and (4.4) can be solved using the same method, the Gummel iteration, as the fine grid problem.

Remark 4.1. The last term of the right-hand side of (4.4) actually vanishes, since at each Gummel iteration as stated in Algorithm 1, the tridiagonal linear system (4.2) is solved for the potential by the direct method after the distribution function has been updated.

Now suppose we have the solution f_H, $\boldsymbol{\Phi}_H$ and $\boldsymbol{E}_H$ for the coarse grid problem (4.3) and (4.4), then the fine grid distribution function $\bar{f}_h$ is corrected as

$$\tilde{f}_h=\bar{f}_h+I_H^h\left(f_H-\tilde{I}_h^H\bar{f}_h\right), \tag{4.5}$$

where I_H^h is the prolongation operator of the distribution function from the coarse grid to the fine grid. Certainly, the fine grid potential $\bar{\boldsymbol{\Phi}}_h$ can be corrected similarly as the above formula, yet this might introduce a little inconsistency between the updated fine grid potential and distribution function $\tilde{f}_h$. So we alternatively solve (4.2) for the updated potential $\tilde{\boldsymbol{\Phi}}_h$, providing the first moments $\tilde{\boldsymbol{\rho}}_h$ of $\tilde{f}_h$, that is,

$$\tilde{\boldsymbol{\Phi}}_h=\mathbf{A}_h^{-1}(\boldsymbol{s}_h+\lambda\tilde{\boldsymbol{\rho}}_h). \tag{4.6}$$

Then the updated fine grid electric field $\tilde{\boldsymbol{E}}_h$ is computed by substituting $\tilde{\boldsymbol{\Phi}}_h$ into (2.31).

4.2 Restriction and prolongation

In our implementation, the coarse grid is generated in a standard way by merging fine grid cells. In particular, we have the coarse grid point $x_{H,i}$ coincides with the fine grid point $x_{h,2i}$. With this property, the restriction on the ith coarse grid cell $[x_{H,i},x_{H,i+1}]$, for a cell-centered function, can be constructed locally dependent on the function's values on the corresponding fine grid cells, i.e., $[x_{h,2i},x_{h,2i+1}]$ and $[x_{h,2i},x_{h,2i+2}]$.

Although in the general NMG framework, the restriction operator $\tilde{I}_h^H$ may be different from I_h^H, we employ the restriction operator proposed in [25] for both $\tilde{I}_h^H$ and I_h^H. For any fine grid function g_h with $g_{h,i}\in\mathcal{F}_M^{[u_{h,i},\theta_{h,i}]}$, this operator produces its restriction with the

moments conserved on each coarse grid cell. As the calculation is a little tedious, we refer to [25] for the details.

The restriction operator $\tilde{\tilde{I}}_h^H$, which transfers the fine grid potential $\bar{\boldsymbol{\Phi}}_h$ onto the coarse grid, is much simpler. To be specific, we define the restriction of $\bar{\boldsymbol{\Phi}}_h$ by taking its weighted average on each coarse grid cell, that is,

$$\left(\tilde{\tilde{I}}_h^H \bar{\boldsymbol{\Phi}}_h\right)_{H,i} = \frac{\bar{\Phi}_{h,2i}\Delta x_{h,2i} + \bar{\Phi}_{h,2i+1}\Delta x_{h,2i+1}}{\Delta x_{H,i}}. \tag{4.7}$$

Apart from the restriction operators, it remains to give the prolongation operator I_H^h, for which the simple identity operator is employed, as in [25]. Then the correction formula (4.5) is rewritten as

$$\tilde{f}_h = \delta \bar{f}_h + f_H, \tag{4.8}$$

where $\delta \bar{f}_h = \bar{f}_h - \tilde{I}_h^H \bar{f}_h$.

4.3 Nonlinear multigrid algorithm

It is straightforward to show that the coarse grid problem (4.3) and (4.4) itself can also be solved by the two-grid method using a much coarser grid correction. In other words, recursively applying the two-grid strategy will give rise to the complete NMG algorithm.

For a given grid, a sequence of grids will be generated by merging grid cells level by level. Let the total grid levels is $K+1$, and introduce subscripts h_k, $k=0,1,\cdots,K$, to denote operators and variables related to the kth level grid, where h_0 and h_K correspond, respectively, to the coarsest and the finest grid. Then the $(k+1)$th level NMG iteration, denoted by $[f_{h_k}^{n+1}, \boldsymbol{\Phi}_{h_k}^{n+1}, \boldsymbol{E}_{h_k}^{n+1}] = \mathrm{NMG}_k(f_{h_k}^n, \boldsymbol{\Phi}_{h_k}^n, \boldsymbol{E}_{h_k}^n; r_{h_k}, \boldsymbol{s}_{h_k})$, is given in Algorithm 3.

An NMG solver is then obtained by performing this NMG iteration instead of the Gummel iteration for the second step of Algorithm 2.

In contrast to the linear multigrid approach, a direct method for the exact solution of the coarsest grid problem is unavailable in the NMG method. Since the coarsest grid problem is a nonlinear system with the coarsest grid operators R_{h_0} and $\mathbf{A}_{h_0}$ analogous to the corresponding operators on other grid levels, the Gummel iteration is again applied to it for the coarsest grid solution. Ideally, the Gummel iteration is performed until convergence such that the coarsest grid problem is accurately solved. However, it would lead to too much Gummel iterations to make the whole NMG solver inefficient, if the coarsest grid number N_{h_0} is not small enough. As a matter of fact, this would usually occur, since the decoupled approach makes the Gummel iteration usually generate a wildly oscillating sequence of residual and even do not converge on a very coarse grid. Therefore, we only perform ν_3 steps of the Gummel iteration for the coarsest grid problem, where ν_3 is a positive integer a little larger than the smoothing steps $\nu_1+\nu_2$. In such a case, we have observed that the W-cycle NMG approach is much faster than the V-cycle NMG approach for all of our numerical tests, so the W-cycle NMG solver is adopted in this paper.

Algorithm 3 A $(k+1)$th level NMG iteration

1. If $k=0$, call the coarsest grid solver, which will be given later, to have a solution as $f_{h_0}^{n+1}$, $\mathbf{\Phi}_{h_0}^{n+1}$ and $\boldsymbol{E}_{h_0}^{n+1}$; otherwise, go to the next step.
2. Pre-smoothing: perform ν_1 steps of the Gummel iteration (Algorithm 1) to obtain a new approximation $\bar{f}_{h_k}$, $\bar{\mathbf{\Phi}}_{h_k}$ and $\bar{\boldsymbol{E}}_{h_k}$.
3. Coarse grid correction:
   (a) Compute the fine grid residual as $\bar{R}_{h_k}=r_{h_k}-R_{h_k}(\bar{f}_{h_k};\bar{\boldsymbol{E}}_{h_k})$.
   (b) Compute the coarse grid approximation $\bar{f}_{h_{k-1}}$, $\bar{\mathbf{\Phi}}_{h_{k-1}}$ by the restriction operators as
   $$\bar{f}_{h_{k-1}}=\bar{I}_{h_k}^{h_{k-1}}\bar{f}_{h_k},\quad \bar{\mathbf{\Phi}}_{h_{k-1}}=\tilde{I}_{h_k}^{h_{k-1}}\bar{\mathbf{\Phi}}_{h_k},$$
   and $\bar{\boldsymbol{E}}_{h_{k-1}}$ by substituting $\bar{\mathbf{\Phi}}_{h_{k-1}}$ into the formula (2.31).
   (c) Compute the right-hand side of the coarse grid problem (4.3) and (4.4) as
   $$r_{h_{k-1}}=I_{h_k}^{h_{k-1}}\bar{R}_{h_k}+R_{h_{k-1}}(\bar{f}_{h_{k-1}};\bar{\boldsymbol{E}}_{h_{k-1}}),\quad \boldsymbol{s}_{h_{k-1}}=\mathbf{A}_{h_{k-1}}\bar{\mathbf{\Phi}}_{h_{k-1}}-\lambda\bar{\boldsymbol{\rho}}_{h_{k-1}}.$$
   (d) Recursively call the NMG algorithm (repeat γ times with $\gamma=1$ for a so-called V-cycle, $\gamma=2$ for a W-cycle, and so on) as
   $$[\tilde{f}_{h_{k-1}},\tilde{\mathbf{\Phi}}_{h_{k-1}},\tilde{\boldsymbol{E}}_{h_{k-1}}]=\mathrm{NMG}_{k-1}^{\gamma}(\bar{f}_{h_{k-1}},\bar{\mathbf{\Phi}}_{h_{k-1}},\bar{\boldsymbol{E}}_{h_{k-1}};r_{h_{k-1}},\boldsymbol{s}_{h_{k-1}}).$$
   (e) Compute the difference $\delta\bar{f}_{h_k}$ with
   $$\delta\bar{f}_{h_k,2i}=\bar{f}_{h_k,2i}-\bar{f}_{h_{k-1},i},\quad \delta\bar{f}_{h_k,2i+1}=\bar{f}_{h_k,2i+1}-\bar{f}_{h_{k-1},i},\qquad i=0,1,\cdots,N_{h_{k-1}}-1,$$
   and update the fine grid distribution by
   $$\tilde{f}_{h_k}=\delta\bar{f}_{h_k}+\tilde{f}_{h_{k-1}}.$$
   (f) Update the fine grid potential by
   $$\tilde{\mathbf{\Phi}}_{h_k}=\mathbf{A}_{h_k}^{-1}(\boldsymbol{s}_{h_k}+\lambda\tilde{\boldsymbol{\rho}}_{h_k}),$$
   and compute the new fine grid electric field $\tilde{\boldsymbol{E}}_{h_k}$ by the formula (2.31).
4. Post-smoothing: perform ν_2 steps of the Gummel iteration to obtain the new approximation $f_{h_k}^{n+1}$, $\mathbf{\Phi}_{h_k}^{n+1}$ and $\boldsymbol{E}_{h_k}^{n+1}$.

A remaining technical issue is that how much levels for the NMG solver is suitable in a practical application. As mentioned above, the coarsest grid number N_{h_0} should be large enough to guarantee convergence of the solver. On the other hand, we have also observed that when N_{h_0} is very large, the solver is usually one time faster as total levels of the NMG solver increases one for most of our simulation tests, due to the great improvement of the convergence rate along with the increasement of the total levels. Based on these, the choice of the total levels should take into account both efficiency and robustness of the solver, and it will be finally given in the concrete examples.

4.4 Full multigrid solver

For an iterative method, a good initial guess can not only enhance the robustness of the method, but also accelerate the convergence, especially for the NMG solver. The issue is that how to get a good initial guess efficiently.

The main approach starts from a general guess on the coarsest grid, which is successively improved by several NMG iterations and interpolated on to the finer grid as the initial solution of the finer grid problem. This results in the so-called FMG solver, which is given in the following algorithm.

Algorithm 4 The FMG solver

1. Given an initial guess of $f_{h_0}^0$, $\boldsymbol{\Phi}_{h_0}^0$ and $\boldsymbol{E}_{h_0}^0$ on the coarsest grid. let $k:=0$.
2. If $k=0$, go to the next step; otherwise, prepare the initial solution on the kth level grid. To be specific, first compute the distribution function $f_{h_k}^0$ from the solution on the $(k-1)$th level grid by interpolation, for which the prolongation operator is employed again, i.e.,

$$f_{h_k}^0 = I_{h_{k-1}}^{h_k} f_{h_{k-1}}^{l_{k-1}}.$$

   Then compute the initial potential $\boldsymbol{\Phi}_{h_k}^0$ by

$$\boldsymbol{\Phi}_{h_k}^0 = \mathbf{A}_{h_k}^{-1}(\boldsymbol{s}_{h_k} + \lambda \boldsymbol{\rho}_{h_k}^0),$$

   and the corresponding electric field $\boldsymbol{E}_{h_k}^0$ by the formula (2.31).
3. Perform l_k steps of the $(k+1)$th level NMG iteration to obtain a new solution on the kth level grid, i.e.,

$$[f_{h_k}^{l_k}, \boldsymbol{\Phi}_{h_k}^{l_k}, \boldsymbol{E}_{h_k}^{l_k}] = \mathrm{NMG}_k^{l_k}(f_{h_k}^0, \boldsymbol{\Phi}_{h_k}^0, \boldsymbol{E}_{h_k}^0; r_{h_k}, \boldsymbol{s}_{h_k}),$$

   where $r_{h_k} \equiv 0$, and $\boldsymbol{s}_{h_k}$ is the vector with $s_{h_k,i} = -\lambda C_{h_k,i}$, $i = 0,1,\cdots,N_{h_k}-1$.
4. If $k=K$, stop; otherwise, let $k:=k+1$, return to step 2.

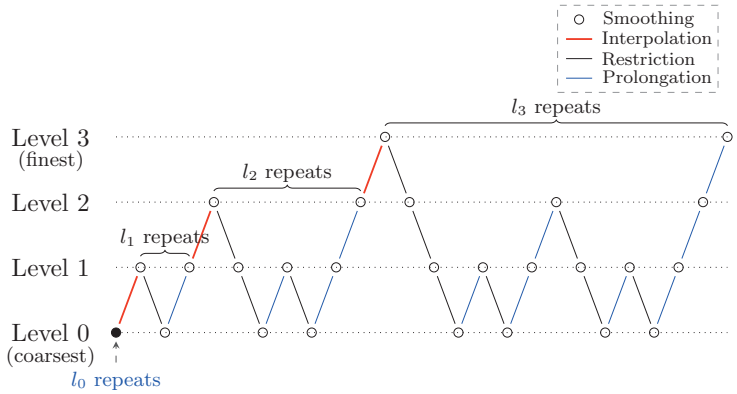


Figure 1: Diagram of the W-cycle FMG solver.

A 4th level W-cycle FMG solver is illustrated in Fig. 1. At each grid level, the iterative steps l_k is determined adaptively such that the convergence is achieved for the $(k+1)$th level NMG iteration. In other words, the FMG solver gives rise to the exact solution of the coupled system (3.2) and (3.3) on all grid levels. Of course, the choice of l_k can be relaxed to improve the FMG solver, if only the solution on the finest grid is expected. However, due to the high convergence rate of the solver, it is difficult to determine an optimal choice of l_k, and this usually improve the solver a little. Note that the single grid solver of the Gummel iteration (1st level NMG iteration) is actually applied for $f_{h_0}^{l_0}$, $\boldsymbol{\Phi}_{h_0}^{l_0}$ and $\boldsymbol{E}_{h_0}^{l_0}$, which indicates the coarsest grid should be chosen coarse enough such that the Gummel iteration converges quickly.

Besides the above FMG solver, sometimes we have other approaches to get a good initial guess. When investigating the convergence of the solution of the coupled system (3.2) and (3.3) with respect to the order M, the interpolation of the solution of the coupled system with a lower order is a good choice of the initial guess. For the computation of the I-V curves, the solution for the lower voltage bias is also commonly used as the initial value for the problem with a higher voltage bias.

5 Numerical examples

As test problem we consider a classic n^+-n-n^+ silicon diode, see Fig. 2 (left), of a total length $L=600$nm with a channel of 400nm located in the middle of the device, for which has been frequently investigated in the literature [15,17,18,29,30] and references therein. The doping profile is given by

$$C(x)=\begin{cases}5\times10^{17}\,\mathrm{cm}^{-3}, & n^+\text{-region},\\ 2\times10^{15}\,\mathrm{cm}^{-3}, & n\text{-region},\end{cases} \tag{5.1}$$

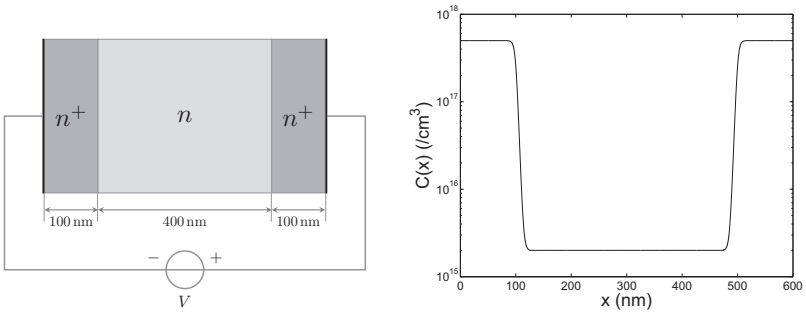


Figure 2: Schematic representation (left) and doping profile (right) of an n^+-n-n^+ silicon diode.

with a smooth transition of width 50nm at the junctions between n^+ and n regions, see Fig. 2 (right). The values of other parameters for the silicon device are taken by $m^* = 0.26\times 0.9109$ (10^{-30}kg), $q_e = 0.1602$ (10^{-18}C), $k_B = 0.138046\times 10^{-4}$ (10^{-18}J/K), $\epsilon_0 = 11.7\times 8.85418$ $(10^{-18}\text{F}/\mu\text{m})$, $T_L = 300\text{K}$, $\mu_0 = 1799\text{cm}^2/(\text{Vs})$, and $v_d = 1.5117\times 10^7\text{cm/s}$.

The simulation is performed under various voltage bias from 0 to 3V. The tolerance *Tol*, indicating the achievement of steady states, is set to 10^{-14}. For the NMG iteration described in Algorithm 3, the smoothing steps are set to $\nu_1 = \nu_2 = 2$, and the coarsest grid problem is solved by only $\nu_3 = 5$ steps of the Gummel iteration in each calling. In most cases, the computation starts from a local Maxwellian with

$$\rho^0(x) = C(x), \qquad u^0(x) = 0, \qquad \theta^0(x) = \theta_L, \tag{5.2}$$

on a coarsest grid with the total 60 or 120 cells, then the FMG solver described in Algorithm 4 is applied. It is almost no difference in efficiency for the FMG solver with either 60 or 120 cells on the coarsest grid. For the grid with more less cells, the single grid solver of the Gummel iteration would always not converge. It has also been observed that the Gummel iteration for the system with a higher order, e.g., $M = 10$, might not even converge on the grid with 120 cells, as the applied voltage bias V_{bias} increases larger than 2V. In such cases, the coarsest grid with 240 cells should be adopted for the NMG iteration and the FMG solver. However, the single grid solver on this grid requires a bit much Gummel iterations to obtain the convergence, which slightly reduces the efficiency of the FMG solver. Thereby, when the applied voltage bias V_{bias} is larger than 2V, the FMG solver with the coarsest grid of 120 cells is first applied to the system with a lower order such as $M = 3,4$. Then the solution is interpolated as the initial guess for the higher order system, and the NMG solver with the coarsest grid of 240 cells is performed.

In the following subsections, we first investigate the convergence of the solution with respect to the system's order M to demonstrate the validation of the hyperbolic moment system with the moment-dependent relaxation time, then study the convergence behavior of the residual to illustrate the main features of the presented NMG iteration (Algorithm 3) and the corresponding FMG solver (Algorithm 4).

5.1 Convergence of solution

For comparison, the solution, obtained by performing a long time integration of the coupled Boltzmann-Poisson system using an explicit time-stepping scheme with the DVM and an upwind finite-difference spatial discretization, is adopted as reference. The velocity space $\mathbb{R}$ is truncated to a finite interval $[-8,8]\times 4.4492\times 10^7\,\mathrm{cm/s}$, and is discretized by 1600 uniform cells. It has been checked that the selected discrete velocity space is accurate enough to recover the solution of the problem. As the calculation of the reference is really too cumbersome, a relatively sparse grid is used for the spatial discretization, that is, the spatial grid with 1200 uniform cells is used for the computation of the I-V curves, while the grid with 8000 uniform cells is used for the comparison of other macroscopic variables, such as electron density, mean velocity, and temperature.

In [26], numerical convergence for the coupled problem of the moment system (2.19) and the Poisson equation (2.3), with respect to both the spatial grid size and the order of the moment system, has been investigated. Due to stronger numerical diffusion of the higher-order moment system and the instability introduced by the oscillation of the temperature, the convergence with respect to the system's order has been verified indeed not very well. In order to eliminate the effects caused by the spatial discretization, we investigate this convergence again on a quite fine spatial grid, i.e., a grid with 30720 uniform cells.

Fig. 3 presents the solution of electron density and temperature for the coupled problem of the moment system (2.19) and the Poisson equation (2.3) at the applied voltage bias $V_{bias}=0.5\mathrm{V}$. The I-V curves for this coupled problem is shown in Fig. 4 (left). Although the convergence with respect to the system's order M can be observed in most of the region, we observe that in the region around the second junction, the temperature has a very large gradient, and may contain oscillations, which can not be eliminated by refining the spatial grid, and lead to incorrect resolution of other macroscopic variables such as electron density. As a result, a system with relatively high order such as $M=30$ is required to give a satisfactory solution. However, the appearance of oscillations leads to instability of the method and even makes the solver breakdown, for instance, when $M=15$ at $V_{bias}=0.5\mathrm{V}$, and $M=10$ at $V_{bias}\geq 0.6\mathrm{V}$. This indicates that the application of the hyperbolic moment system (2.19) is extremely limited.

Now we turn to investigate the convergence behavior for the coupled problem of the modified moment system (2.22) and the Poisson equation (2.3). As the numerical convergence with respect to the spatial grid size could easily be verified, similar to those in [26], we omit any discussion on the spatial convergence, and only focus on the convergence with respect to the system's order in this paper. Similarly, a spatial grid with 30720 uniform cells is employed. The resulting I-V curves, as shown in Fig. 4 (right), is absolutely much better than those for the coupled problem of the original moment system (2.19) and the Poisson equation (2.3).

For the solution of macroscopic variables of the coupled problem of the modified moment system (2.22) and the Poisson equation (2.3), Fig. 5 presents the electron density,

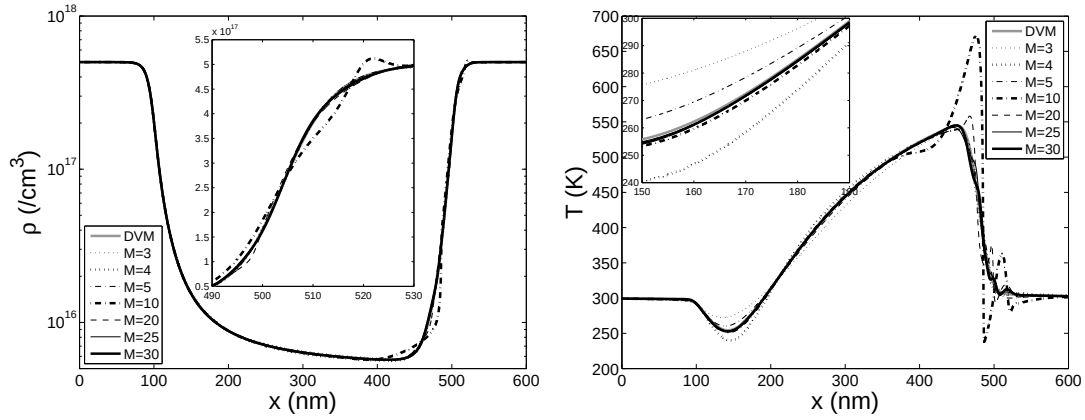


Figure 3: Electron density (left) and temperature (right) for the coupled problem of the moment system (2.19) and the Poisson equation (2.3), at the applied voltage bias $V_{bias}=0.5$V.

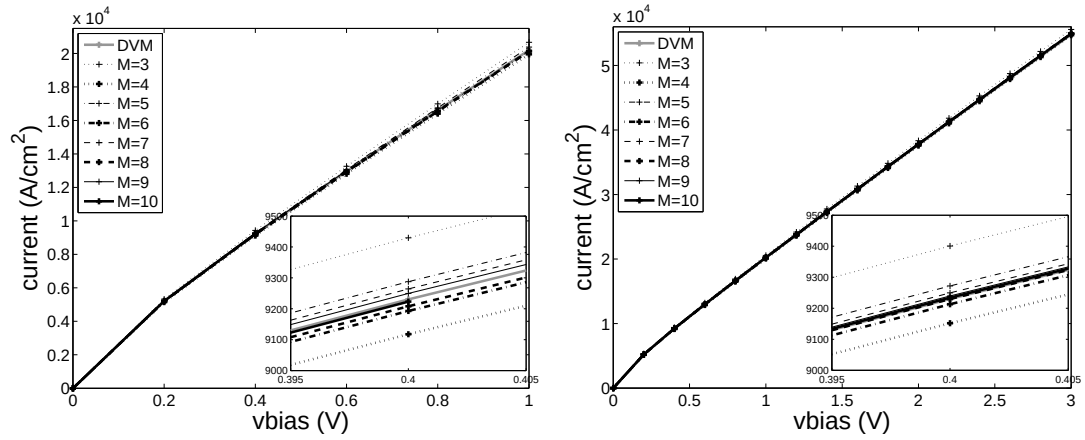


Figure 4: I-V curves. *Left*: the coupled problem of the moment system (2.19) and the Poisson equation (2.3); *right*: the coupled problem of the modified moment system (2.22) and the Poisson equation (2.3).

mean velocity, potential, and electric field at the applied voltage bias $V_{bias}=0.5$V. The corresponding temperature is plotted in Fig. 6. At a larger applied voltage bias $V_{bias}=2$V, these macroscopic variables are shown in Figs. 7 and 8. Note at the latter voltage bias, the solver for the original moment system with the order $M\geq 6$ is failure to give any results, due to the wild oscillations of the temperature. As can be seen, with the moment-dependent relaxation time (2.21), the oscillations disappear for the modified moment system of all order, and the solution is greatly improved in comparison to the solution of the original coupled problem. Moreover, these numerical results show a high order convergence with respect to the order M of the system, as expected for the spectral method. It is also found that the modified moment system (2.22) of order $M=10$ is enough to give these macroscopic variables of interest for various applied voltage biases in a comparable accuracy, compared with the solution of the DVM.

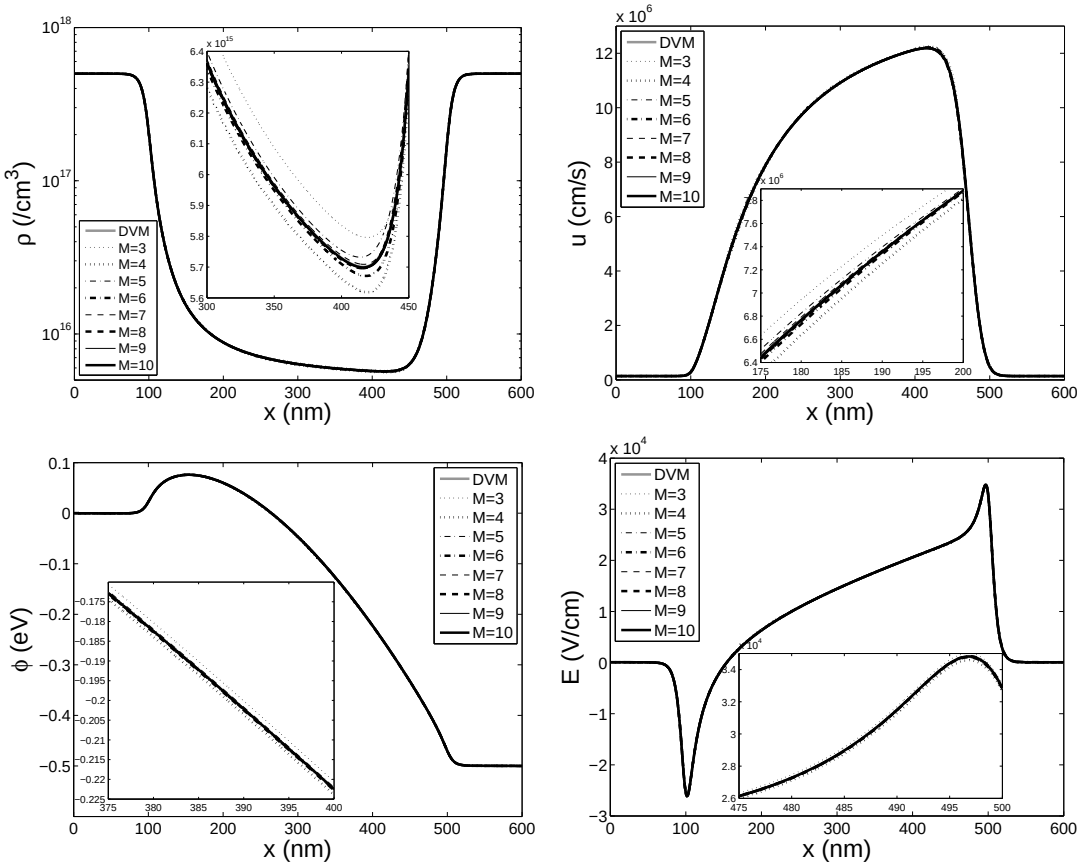


Figure 5: Electron density (top left), mean velocity (top right), potential (bottom left), and electric field (bottom right) for the coupled problem of the modified moment system (2.22) and the Poisson equation (2.3), at the applied voltage bias $V_{bias}=0.5$V.

At last, it is pointed out that the parameter β in the moment-dependent relaxation formula (2.21) is insensitive. For other choices of β, such as $\beta=0.2$ and 0.5, the temperature of the modified moment system (2.22) and the Poisson equation (2.3) at the applied voltage bias $V_{bias}=0.5$V is plotted in Fig. 9 (left and middle), which shows subtle difference from Fig. 6 where $\beta=0.1$. While with other formula for the moment-dependent relaxation time τ_α, the temperature might still have oscillations. For example, if τ_α is assumed as follows

$$\tau_\alpha^{-1}=\begin{cases}\tau^{-1}, & \text{if } \alpha=0,\\ (1+\beta V_{bias}(\alpha-1))\tau^{-1}, & \text{if } 1\le\alpha\le M,\end{cases} \tag{5.3}$$

the corresponding temperature with $\beta=0.1$ is presented in Fig. 9 (right), which shows that the oscillation is still observed, although it is suppressed much in comparison to Fig. 3.

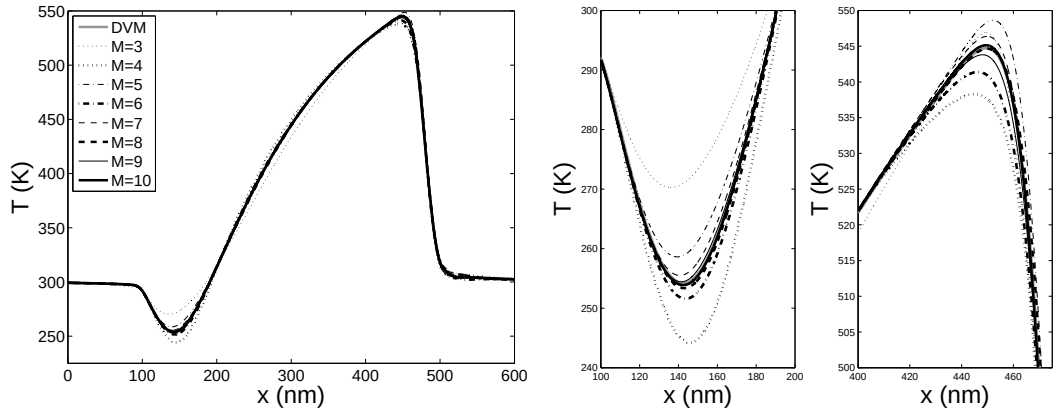


Figure 6: Temperature (left) and its zoom (right) for the coupled problem of the modified moment system (2.22) and the Poisson equation (2.3), at the applied voltage bias $V_{bias}=0.5\text{V}$.

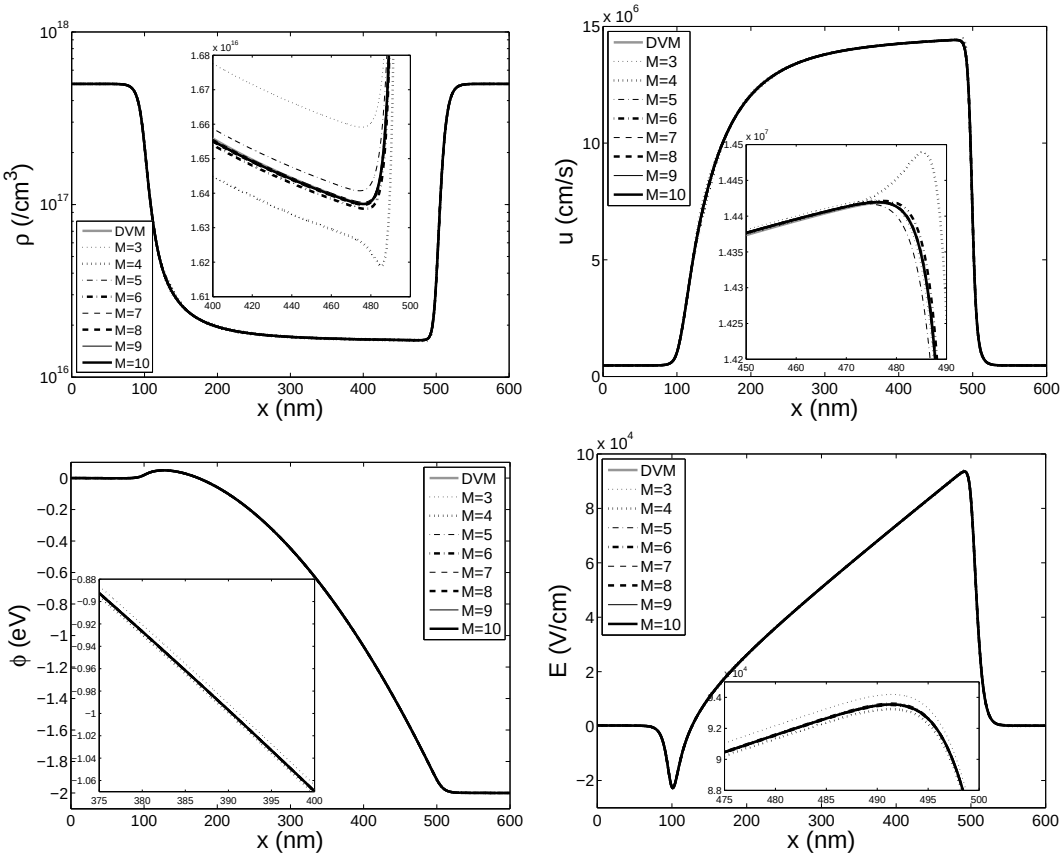


Figure 7: Electron density (top left), mean velocity (top right), potential (bottom left), and electric field (bottom right) for the coupled problem of the modified moment system (2.22) and the Poisson equation (2.3), at the applied voltage bias $V_{bias}=2\text{V}$.

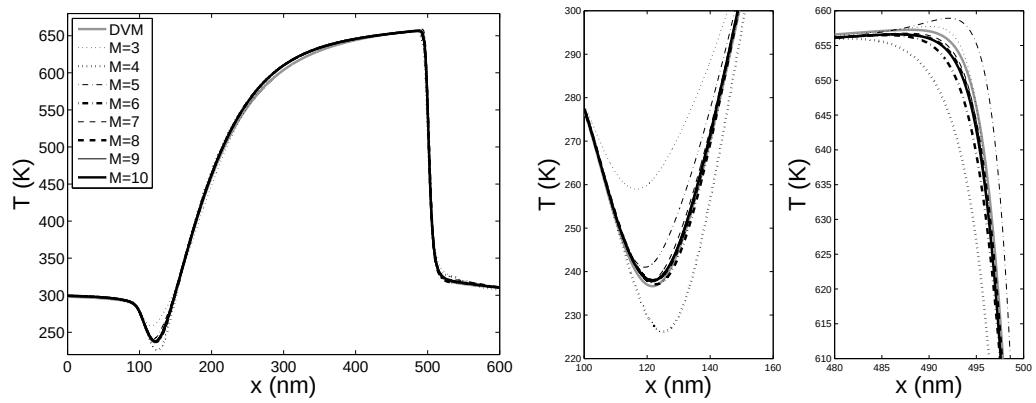


Figure 8: Temperature (left) and its zoom (right) for the coupled problem of the modified moment system (2.22) and the Poisson equation (2.3), at the applied voltage bias $V_{bias}=2$V.

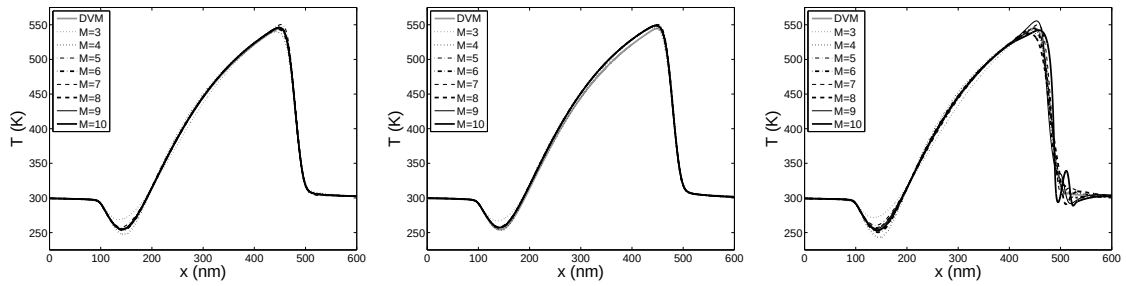
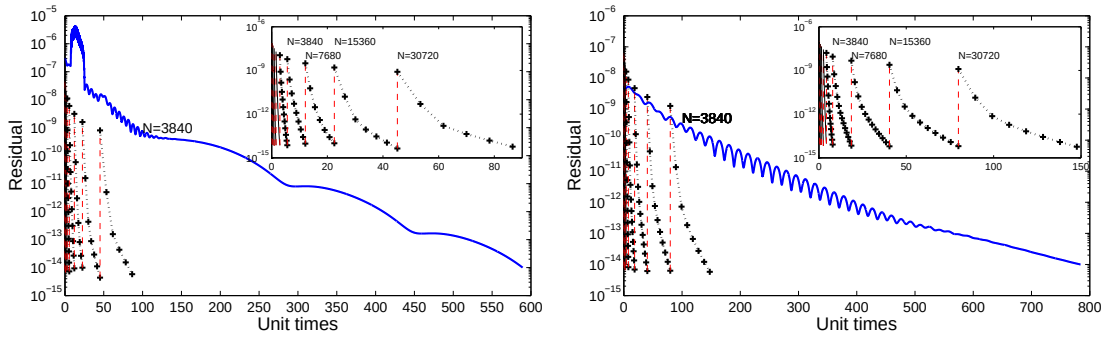


Figure 9: Temperature of the modified moment system (2.22) and the Poisson equation (2.3) at the applied voltage bias $V_{bias}=0.5$V. *Left*: The modification (2.21) with $\beta=0.2$; *Middle*: The modification (2.21) with $\beta=0.5$; *Right*: The modification (5.3) with $\beta=0.1$.

5.2 Numerical efficiency

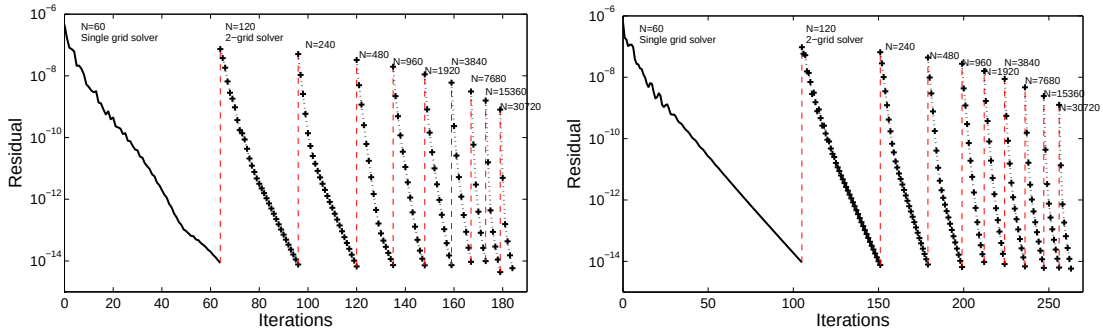
In view of that the modified moment system has only $M+1$ moments while the degree of freedom in velocity space by the DVM is much larger, we have that the computational cost is considerably reduced by the modified moment system even a time-integration scheme is applied for the steady-state solution, in comparison to the DVM. Thereby, we only consider the improvement in efficiency by the NMG iteration and the FMG solver in comparison to the single grid solver. As the FMG solver exhibits similar features for all of the tests, we only show some of results in this paper.

Fig. 10(a) presents the convergence histories in terms of unit times, which are generated by both the FMG solver and the single grid solver for the coupled problem of the modified moment system (2.22) with $M=5$ and the Poisson equation (2.3) at the applied voltage bias $V_{bias}=0.5$V. Both computations begin with the same initial Maxwellian (5.2). As can be seen, the sequence of the residual oscillates wildly at the beginning iterations for the single grid solver. This oscillation introduces the instability of the solver, and might make the solver breakdown for the problem with a higher order, e.g., $M=10$. In



(a) $M=5$ (b) $M=10$

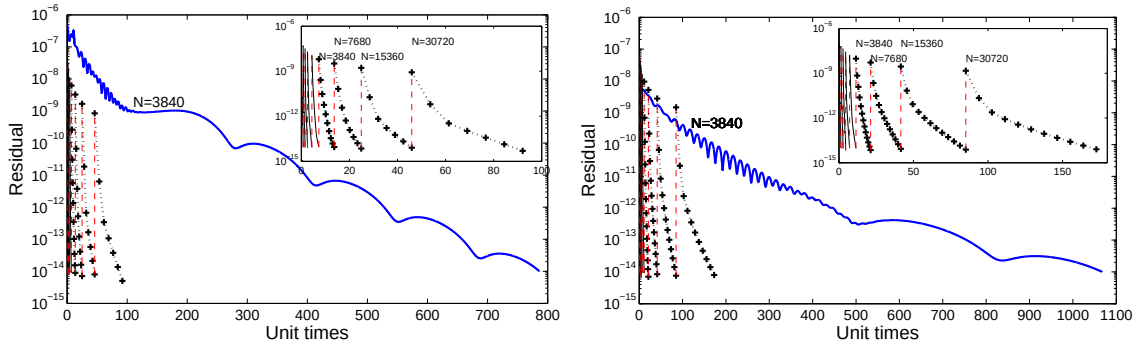
Figure 10: Convergence histories in terms of unit times for the coupled problem of the modified moment system (2.22) and the Poisson equation (2.3), at the applied voltage bias $V_{bias}=0.5$V. Black curve or marker '+': FMG solver; blue curve: single grid solver. All computations start from the same initial Maxwellian (5.2), except the single grid solver for $M=10$, which starts from the interpolation of the steady-state solution on the coarser grid $N=1920$.



(a) $M=5$ (b) $M=10$

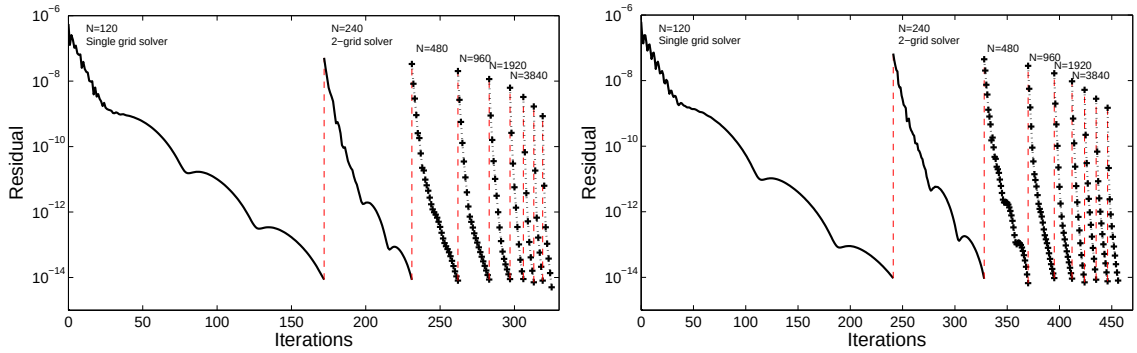
Figure 11: Convergence histories in terms of iterations for the coupled problem of the modified moment system (2.22) and the Poisson equation (2.3), at the applied voltage bias $V_{bias}=0.5$V.

this case, we change to employ the approximation interpolated from the steady-state solution on the coarser grid as the initial guess, as shown in Fig. 10(b). It turns out that using the coarse grid solution for the initial guess can enhance the robustness of the single grid solver, yet it is also observed that a good initial guess does not help much for the single grid solver in improving efficiency, since the total iterations is only reduced a little. However, as shown in Fig. 11, we observe that the total NMG iterations gradually decreases as the grid level increases for the FMG solver, which indicates that for the NMG iteration and the resulting FMG solver, a good initial guess can not only enhance the robustness of the solver, but also improve the efficiency of the solver greatly. Recalling that the total iterations for the single grid solver with the same initial guess increases linearly, in the general case, as the grid number increases, we conclude that the FMG solver is substantially more efficient than the single grid solver on the finer grid. For instance, the



(a) $M=5$ (b) $M=10$

Figure 12: Convergence histories in terms of unit times for the coupled problem of the modified moment system (2.22) and the Poisson equation (2.3), at the applied voltage bias $V_{bias}=1$V. Black curve or marker '+': FMG solver; blue curve: single grid solver. All computations start from the same initial Maxwellian (5.2), except the single grid solver for $M=10$, which starts from the interpolation of the steady-state solution on the coarser grid $N=1920$.



(a) $M=5$ (b) $M=10$

Figure 13: Convergence histories in terms of iterations for the coupled problem of the modified moment system (2.22) and the Poisson equation (2.3), at the applied voltage bias $V_{bias}=1$V.

FMG solver for the steady-state solution on the uniform grid with $N=3840$ is dozens of times faster than the corresponding single grid solver, as seen in Fig. 10.

At the applied voltage bias $V_{bias}=1$V, similar convergence histories can be found in Figs. 12 and 13, implying that the FMG solver drastically improves the efficiency in comparison to the single grid solver.

6 Concluding remarks

We modify the hyperbolic moment models derived from the underlying semiconductor Boltzmann equation by a moment-dependent relaxation time. An empirical formula,

involving the microscopic relaxation time and the applied voltage bias with one single parameter, has been proposed to determine the relaxation time for each moment equation. Numerical experiments show that the modified moment models completely prevent the possible oscillations in the solution and give a high-order convergence to the underlying BTE. Furthermore, an NMG iteration, using the Gummel method as its smoother, has been developed for these moment models. Numerical simulations have also demonstrated the robustness and the drastic improvement in efficiency of the NMG iteration and the resulting FMG solver. Although only steady-state solution is considered throughout the work, the NMG iteration can certainly be used to improve the efficiency for time evolution simulations, where an implicit time-integration scheme is used.

Besides the above performance, the method explored in the present paper does also have at least the following characteristics.

- The modified moment models inherit most good properties of the original hyperbolic moment models, including the systematic derivation, global hyperbolicity, high-order convergence, etc.
- Numerical framework is unified for the model of arbitrary order and the smoothing operator. As a result, one can get improved accuracy when necessary by using a sufficient high-order model. Acceleration of some other single grid iterations can also be obtained by using it as the smoother in the NMG iteration.
- The proposed method can be extended without much effort to the case when quantum effects are important, thanks to the similarity between the BTE and its quantum counterpart, the Wigner equation, as see in [9].

In summary, this is a preliminary attempt on the application of hyperbolic moment method and corresponding multigrid solver to semiconductor device simulations. For simplicity, only simulations of an n^+-n-n^+ silicon diode with a channel of 400nm, modeled basically by the BTE with the relaxation time approximation of scattering term, have been carried out, to investigate the performance of the proposed models and solver. It should be noted that, nowadays, many commercial devices such as MOSFETs are fabricated on the scale of tens of nanometers. For such devices, the relaxation time approximation might fail to properly represent the actual scattering mechanisms, and a more fundamental scattering mechanism, such as the electron-phonon scattering, must be considered. Nevertheless, we believe that the proposed method is a promising way and gives us some guidelines to construct accurate and robust macroscopic models with efficient solver for nowadays's nanoscale devices. In our future work, we will consider simulations of these more realistic devices. The extensions of the proposed solver to the multi-dimensional case and to the case considering quantum effects are also under consideration. Benefit from the unified numerical framework, it is confident that the proposed method can be extended to more realistic devices and the multi-dimensional case without much effort.

Acknowledgments

We would like to express our deep thanks to the anonymous reviewers whose valuable comments and suggestions help us improve this article greatly. The research of Z. Hu is supported by the Hong Kong Research Council ECS grant No. 509213. The research of R. Li is supported in part by the National Basic Research Program of China (2011CB309704) and the National Science Foundation of China (11325102, 91330205). The research of Z. Qiao is partially supported by the Hong Kong Research Council ECS grant No. 509213.

Appendix: Calculation of $f_\alpha^{\mathbf{eq}}$

The derivation of (2.17) will be given here. For simplicity, the spatial variable x is omitted, since the derivation is independent of the spatial position. Suppose $f^{\mathrm{eq}} \in \mathcal{F}^{[\boldsymbol{u},\theta]}$. It follows that

$$f^{\mathrm{eq}}(\boldsymbol{v}) = \frac{\rho}{(2\pi\theta_L)^{D/2}} \exp\left(-\frac{|\boldsymbol{v}|^2}{2\theta_L}\right) = \sum_{\alpha\in\mathbb{N}^D} f_\alpha^{\mathrm{eq}} \mathcal{H}_\alpha^{[\boldsymbol{u},\theta]}(\boldsymbol{v}), \tag{A.1}$$

where $D=1,2,3$ is the dimension of the velocity space. Define

$$\mathcal{P}_\alpha^{[\boldsymbol{u},\theta]}(\boldsymbol{v}) = \mathcal{H}_\alpha^{[\boldsymbol{u},\theta]}(\boldsymbol{v}) \exp\left(\frac{|\boldsymbol{v}-\boldsymbol{u}|^2}{2\theta}\right). \tag{A.2}$$

Then multiplying both side of (A.1) by $\mathcal{P}_\beta^{[\boldsymbol{u},\theta]}(\boldsymbol{v})$ and using the orthogonality of Hermite polynomials, we have

$$\begin{aligned} f_\alpha^{\mathrm{eq}} &= \frac{1}{C_{\theta,\alpha}} \int_{\mathbb{R}^D} \mathcal{P}_\alpha^{[\boldsymbol{u},\theta]}(\boldsymbol{v}) f^{\mathrm{eq}}(\boldsymbol{v}) \,\mathrm{d}\boldsymbol{v} \\ &= \frac{1}{C_{\theta,\alpha}} \frac{\rho}{(2\pi)^{D/2}} \int_{\mathbb{R}^D} \mathcal{P}_\alpha^{[\boldsymbol{u},\theta]}\left(\sqrt{\theta_L}\boldsymbol{\xi}_L\right) \exp\left(-|\boldsymbol{\xi}_L|^2/2\right) \mathrm{d}\boldsymbol{\xi}_L, \end{aligned} \tag{A.3}$$

where $\boldsymbol{\xi}_L = \boldsymbol{v}/\sqrt{\theta_L}$, and

$$C_{\theta,\alpha} = \int_{\mathbb{R}^D} \mathcal{H}_\alpha^{[\boldsymbol{u},\theta]}(\boldsymbol{v}) \mathcal{P}_\alpha^{[\boldsymbol{u},\theta]}(\boldsymbol{v}) \,\mathrm{d}\boldsymbol{v} = \frac{\alpha_1!\cdots\alpha_D!}{(2\pi)^{D/2}\theta^{|\alpha|+D/2}}.$$

Note the following recursion relation

$$\mathcal{P}_\alpha^{[\boldsymbol{u},\theta]}(\boldsymbol{v}) = \theta^{-1}(v_i-u_i)\mathcal{P}_{\alpha-e_i}^{[\boldsymbol{u},\theta]}(\boldsymbol{v}) - \theta^{-1}(\alpha_i-1)\mathcal{P}_{\alpha-2e_i}^{[\boldsymbol{u},\theta]}(\boldsymbol{v}) \tag{A.4}$$

holds for any α, if there exists an i, $1\leq i\leq D$, such that $\alpha_i\geq 2$. Substituting (A.4) into (A.3), we get

$$\begin{aligned} f_\alpha^{\rm eq} &= \frac{1}{C_{\theta,\alpha}}\frac{\rho}{(2\pi)^{D/2}}\int_{\mathbb{R}^D}\Big(\theta^{-1}\Big(\sqrt{\theta_L}\xi_{L,i}-u_i\Big)\mathcal{P}_{\alpha-e_i}^{[\boldsymbol{u},\theta]}\Big(\sqrt{\theta_L}\boldsymbol{\xi}_L\Big) \\ &\quad -\theta^{-1}(\alpha_i-1)\mathcal{P}_{\alpha-2e_i}^{[\boldsymbol{u},\theta]}\Big(\sqrt{\theta_L}\boldsymbol{\xi}_L\Big)\Big)\exp\left(-|\boldsymbol{\xi}_L|^2/2\right)\mathrm{d}\boldsymbol{\xi}_L \\ &= \frac{1}{C_{\theta,\alpha}}\frac{\rho}{(2\pi)^{D/2}}\int_{\mathbb{R}^D}\frac{\sqrt{\theta_L}}{\theta}\xi_{L,i}\mathcal{P}_{\alpha-e_i}^{[\boldsymbol{u},\theta]}\Big(\sqrt{\theta_L}\boldsymbol{\xi}_L\Big)\exp\left(-|\boldsymbol{\xi}_L|^2/2\right)\mathrm{d}\boldsymbol{\xi}_L \\ &\quad -\frac{u_i}{\alpha_i}f_{\alpha-e_i}^{\rm eq}-\frac{\theta}{\alpha_i}f_{\alpha-2e_i}^{\rm eq}. \end{aligned} \tag{A.5}$$

Applying integration by parts to the integral yields

$$\begin{aligned} f_\alpha^{\rm eq} &= \frac{1}{C_{\theta,\alpha}}\frac{\rho}{(2\pi)^{D/2}}\int_{\mathbb{R}^D}\frac{\sqrt{\theta_L}}{\theta}\frac{\partial}{\partial\xi_{L,i}}\mathcal{P}_{\alpha-e_i}^{[\boldsymbol{u},\theta]}\Big(\sqrt{\theta_L}\boldsymbol{\xi}_L\Big)\exp\left(-|\boldsymbol{\xi}_L|^2/2\right)\mathrm{d}\boldsymbol{\xi}_L \\ &\quad -\frac{u_i}{\alpha_i}f_{\alpha-e_i}^{\rm eq}-\frac{\theta}{\alpha_i}f_{\alpha-2e_i}^{\rm eq} \\ &= \frac{1}{C_{\theta,\alpha}}\frac{\rho}{(2\pi)^{D/2}}\int_{\mathbb{R}^D}\frac{\theta_L}{\theta}\frac{\partial}{\partial v_i}\mathcal{P}_{\alpha-e_i}^{[\boldsymbol{u},\theta]}\Big(\sqrt{\theta_L}\boldsymbol{\xi}_L\Big)\exp\left(-|\boldsymbol{\xi}_L|^2/2\right)\mathrm{d}\boldsymbol{\xi}_L \\ &\quad -\frac{u_i}{\alpha_i}f_{\alpha-e_i}^{\rm eq}-\frac{\theta}{\alpha_i}f_{\alpha-2e_i}^{\rm eq}. \end{aligned} \tag{A.6}$$

The differential relation of Hermite polynomials shows that

$$\frac{\partial}{\partial v_i}\mathcal{P}_{\alpha-e_i}^{[\boldsymbol{u},\theta]}(\boldsymbol{v})=(\alpha_i-1)\theta^{-1}\mathcal{P}_{\alpha-2e_i}^{[\boldsymbol{u},\theta]}(\boldsymbol{v}). \tag{A.7}$$

Inserting (A.7) into (A.6), we obtain

$$f_\alpha^{\rm eq}=\frac{\theta_L-\theta}{\alpha_i}f_{\alpha-2e_i}^{\rm eq}-\frac{u_i}{\alpha_i}f_{\alpha-e_i}^{\rm eq}. \tag{A.8}$$

It remains to calculate $f_\alpha^{\rm eq}$ when $\alpha=0,e_i,e_i+e_j,e_1+e_2+e_3$ with $i\neq j$, $i,j=1,\cdots,D$. A routine computation gives rise to

$$\begin{aligned} &f_0^{\rm eq}=\rho, \\ &f_{e_i}^{\rm eq}=-\rho u_i, \\ &f_{e_i+e_j}^{\rm eq}=\rho u_iu_j,\quad i\neq j, \\ &f_{e_1+e_2+e_3}^{\rm eq}=-\rho u_1u_2u_3. \end{aligned} \tag{A.9}$$

For the special case $D=1$, the formula (A.8) and (A.9) immediately give the recursion relation (2.17).

References

[1] A. M. Anile and V. Romano. Hydrodynamical modeling of charge carrier transport in semiconductors. *Meccanica*, 35(3):249–296, 2000.

[2] A. M. Anile, V. Romano, and G. Russo. Extended hydrodynamical model of carrier transport in semiconductors. *SIAM J. Appl. Math.*, 61(1):74–101, 2000.

[3] B. A. Biegel and J. D. Plummer. Comparison of self-consistency iteration options for the Wigner function method of quantum device simulation. *Phys. Rev. B*, 54(11):8070–8082, 1996.

[4] K. Bløtekjær. Transport equations for electrons in two-valley semiconductors. *IEEE Trans. Electron Devices*, 17(1):38–47, Jan 1970.

[5] A. Brandt. Multi-level adaptive solutions to boundary-value problems. *Math. Comp.*, 31(138):333–390, 1977.

[6] A. Brandt and O. E. Livne. *Multigrid Techniques: 1984 Guide with Applications to Fluid Dynamics*. Classics in Applied Mathematics. SIAM, revised edition, 2011.

[7] Z. Cai, Y. Fan, and R. Li. Globally hyperbolic regularization of Grad's moment system in one dimensional space. *Comm. Math. Sci.*, 11(2):547–571, 2013.

[8] Z. Cai, Y. Fan, and R. Li. Globally hyperbolic regularization of Grad's moment system. *Comm. Pure Appl. Math.*, 67(3):464–518, 2014.

[9] Z. Cai, Y. Fan, R. Li, T. Lu, and Y. Wang. Quantum hydrodynamics models by moment closure of Wigner equation. *J. Math. Phys.*, 53:103503, 2012.

[10] Z. Cai and R. Li. Numerical regularized moment method of arbitrary order for Boltzmann-BGK equation. *SIAM J. Sci. Comput.*, 32(5):2875–2907, 2010.

[11] Z. Cai, R. Li, and Z. Qiao. NRxx simulation of microflows with Shakhov model. *SIAM J. Sci. Comput.*, 34(1):A339–A369, 2012.

[12] Z. Cai, R. Li, and Z. Qiao. Globally hyperbolic regularized moment method with applications to microflow simulation. *Computers and Fluids*, 81:95–109, 2013.

[13] Z. Cai, R. Li, and Y. Wang. An efficient NRxx method for Boltzmann-BGK equation. *J. Sci. Comput.*, 50(1):103–119, 2012.

[14] Z. Cai, R. Li, and Y. Wang. Numerical regularized moment method for high Mach number flow. *Commun. Comput. Phys.*, 11(5):1415–1438, 2012.

[15] J. A. Carrillo, I. M. Gamba, and C.-W. Shu. Computational macroscopic approximations to the one-dimensional relaxation-time kinetic system for semiconductors. *Physica D*, 146:289–306, 2000.

[16] J.A. Carrillo, I.M. Gamba, A. Majorana, and C.-W. Shu. 2D semiconductor device simulations by WENO-Boltzmann schemes: Efficiency, boundary conditions and comparison to Monte Carlo methods. *J. Comput. Phys.*, 214(1):55–80, 2006.

[17] J.A. Carrillo, I.M. Gamba, O. Muscato, and C.-W. Shu. Comparison of Monte Carlo and deterministic simulations of a silicon diode. In *IMA Volumes in Mathematics and Its Applications*, pages 75–84. Springer-Verlag, 2004.

[18] C. Cercignani, I. M. Gamba, J. W. Jerome, and C.-W. Shu. Device benchmark comparisons via kinetic, hydrodynamic, and high-hield models. *Comput. Methods Appl. Mech. Engrg.*, 181:381–392, 2000.

[19] C. Cercignani, I. M. Gamba, and C. D. Levermore. High field approximations to a Boltzmann-Poisson system and boundary conditions in a semiconductor. *Appl. Math. Lett.*, 10(4):111–117, 1997.

[20] S. Chen, W. E, Y. Liu, and C.-W. Shu. A discontinuous Galerkin implementation of a domain decomposition method for kinetic-hydrodynamic coupling multiscale problems in gas

dynamics and device simulations. *J. Comput. Phys.*, 225(2):1314 – 1330, 2007.
[21] H. Grad. On the kinetic theory of rarefied gases. *Comm. Pure Appl. Math.*, 2(4):331–407, 1949.
[22] T. Grasser, T.-W. Tang, H. Kosina, and S. Selberherr. A review of hydrodynamic and energy-transport models for semiconductor device simulation. *Proceedings of The IEEE*, 91(2):251–274, 2003.
[23] H. K. Gummel. A self-consistent iterative scheme for one-dimensional steady state transistor calculations. *IEEE Trans. Electron Devices*, 11(10):455–465, 1964.
[24] W. Hackbusch. *Multi-Grid Methods and Applications*. Springer-Verlag, Berlin, 1985. second printing 2003.
[25] Z. Hu and R. Li. A nonlinear multigrid steady-state solver for 1D microflow. *Computers and Fluids*, 103:193–203, 2014.
[26] Z. Hu, R. Li, T. Lu, Y. Wang, and W. Yao. Simulation of an n^+-n-n^+ diode by using globally-hyperbolically-closed high-order moment models. *J. Sci. Comput.*, 59(3):761–774, 2014.
[27] T. Imamura, K. Suzuki, T. Nakamura, and M. Yoshida. Acceleration of steady-state lattice Boltzmann simulations on non-uniform mesh using local time step method. *J. Comput. Phys.*, 202(2):645663, Jan 2005.
[28] A. Jüngel. *Transport Equations for Semiconductors*. Number 773 in Lecture Notes in Physics. Springer-Verlag, Berlin, 2009.
[29] R. Kannan. An implicit LU-SGS spectral volume method for the moment models in device simulations: Formulation in 1D and application to a p-multigrid algorithm. *Int. J. Numer. Meth. Biomed. Engng.*, 27:1362–1375, 2011.
[30] Y. X. Liu and C.-W. Shu. Error analysis of the semi-discrete local discontinuous Galerkin method for semiconductor device simulation models. *Sci. China Math.*, 53(12):32553278, 2010.
[31] T. Lu, G. Du, X. Liu, and P. Zhang. A finite volume method for the multi subband Boltzmann equation with realistic 2D scattering in DG MOSFETs. *Commun. Comput. Phys.*, 10:305–338, 2011.
[32] G. Dal Maso, P. G. LeFloch, and F. Murat. Definition and weak stability of nonconservative products. *J. Math. Pures Appl.*, 74(6):483–548, 1995.
[33] D. J. Mavriplis. Multigrid solution of the steady-state lattice Boltzmann equation. *Computers and Fluids*, 35:793–804, 2006.
[34] R. G. McClarren and C. D. Hauck. Robust and accurate filtered spherical harmonics expansions for radiative transfer. *J. Comput. Phys.*, 229(16):5597–5614, 2010.
[35] H. Struchtrup. Extended moment method for electrons in semiconductors. *Physica A: Statistical Mechanics and its Applications*, 275(12):229–255, 2000.
[36] E. Tadmor. Convergence of spectral methods for nonlinear conservation laws. *SIAM J. Numer. Anal.*, 26(1):30–44, 1989.
[37] J. Tölke, M. Krafczyk, and E. Rank. A multigrid-solver for the discrete Boltzmann equation. *J. Stat. Phys.*, 107(1/2):573–591, 2002.
[38] W. Van Roosbroeck. Theory of the flow of electrons and holes in germanium and other semiconductors. *Bell System Technical Journal*, 29(4):560–607, 1950.
[39] D. Vasileska, S. M. Goodnick, and G. Klimeck. *Computational Electronics: Semiclassical and Quantum Device Modeling and Simulation*. CRC Press, 2010.