Modelling Simul. Mater. Sci. Eng. **8** (2000) 133–149. Printed in the UK PII: S0965-0393(00)10901-5

Relaxation-time approximations of quasi-hydrodynamic type in semiconductor device modelling

R V N Melnik† and Hao He‡

† Mathematical Modelling of Industrial Processes, CSIRO Mathematical and Information Sciences—Sydney, Macquarie University Campus, North Ryde, NSW 2113, Australia
‡ Department of Theoretical Physics, School of Physics, University of Sydney, NSW 2006, Australia

E-mail: roderick.melnik@cmis.csiro.au

Received 26 September 1999, accepted for publication 19 December 1999

Abstract. We analyse mathematical models for the description of carrier transport in semiconductors as a hierarchy of models constructed on the basis of the relaxation-time concept. In this hierarchy we focus on a reasonable compromise between drift-diffusion, hydrodynamic, and kinetic models. This compromise is provided by non-local quasi-hydrodynamic mathematical models describing non-equilibrium physical processes in semiconductor devices. Details of the normalization procedure for the quasi-hydrodynamic system will be given along with a transformation of the energy-balance equations to provide computationally convenient forms.

1. Introduction: mathematical models for electron–hole plasma

The origin of the mathematical modelling of transport phenomena, including transport phenomena in semiconductors, is the Liouville equation for the evolution of the position-velocity probability density $f(\boldsymbol{x}, \boldsymbol{v}, t)$:

$$\partial_t f + \boldsymbol{v} \cdot \nabla_{\boldsymbol{x}} f + \frac{1}{m^*} \boldsymbol{\mathcal{F}} \cdot \nabla_{\boldsymbol{v}} f = 0 \qquad t > 0, \quad \boldsymbol{x} \in \mathbb{R}^{3M}_{\boldsymbol{x}}, \quad \boldsymbol{v} \in \mathbb{R}^{3M}_{\boldsymbol{v}} \tag{1.1}$$

where the position, $\boldsymbol{x} \in \mathbb{R}^3_{\boldsymbol{x}}$, and velocity vectors, $\boldsymbol{v} \in \mathbb{R}^3_{\boldsymbol{v}}$, of a charge carrier (say, the electron) are functions of time t, m^* is the effective carrier mass, M is the number of carriers in the system and $\boldsymbol{\mathcal{F}}$ is the driving force. Model (1.1) has to be supplemented by the conditions for the probability density

$$f(\boldsymbol{x}, \boldsymbol{v}, t)|_{t=0} \geqslant 0 \qquad \iint_{\mu} f(\boldsymbol{x}, \boldsymbol{v}, t)|_{t=0} \, \mathrm{d}\boldsymbol{x} \, \mathrm{d}\boldsymbol{v} = 1 \tag{1.2}$$

with the integration in (1.2) over the whole $6M$-dimensional $(\boldsymbol{x}, \boldsymbol{v})$ space, denoted by μ. The definition of the driving force in the context of semiconductor device theory typically has the following form [11]

$$\boldsymbol{\mathcal{F}} = -q\boldsymbol{E} \qquad \text{or} \qquad \boldsymbol{\mathcal{F}} = -q(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}) \tag{1.3}$$

where q is the elementary *positive* charge, $\boldsymbol{E}$ is the electric field vector and $\boldsymbol{B}$ is the magnetic induction vector.

In its essence, the model (1.1)–(1.3) is a reformulation of Newton's second law in terms of probability density. Using the Hamiltonian-canonical-system arguments this model can also be formulated in the $(\boldsymbol{x}, \boldsymbol{p})$-space with $\boldsymbol{p} \in \mathbb{R}^{3M}_{\boldsymbol{p}}$ as the momentum vector. We also note that using the concept of primitive cells of the reciprocal lattice, known as Brillouin zones, we can

0965-0393/00/020133+17$30.00 © 2000 IOP Publishing Ltd

incorporate quantum effects into the model. Such effects become important if the period of the crystal lattice is of the order of 10^{-8} cm. Then the ions in the crystal lattice induce a lattice potential which may significantly influence the motion of charged particles. Formally, the classical model (1.1)–(1.3) follows from quantum models in the limit of the reduced Planck constant $\hbar \to 0$. We assume that in model (1.1)–(1.3) quantum effects are taken indirectly into account by the effective mass of electrons.

If we exclude certain strong singularities of $\mathcal{F}$ at finite $\boldsymbol{x}$, $\boldsymbol{v}$ and t ('collisions of ensembles'), then the mathematical analysis of model (1.1)–(1.3) in reflexive Banach spaces (such as L^2) is known in the literature (see references in [11]). However, in the general case, Liouville-type models are hardly effective in the engineering practice of semiconductor device modelling due to the following reasons: (i) $6M$-dimensional μ-space is unrealistic for modelling many of today's devices; (ii) adequate models for the driving force as a combination of short-range and long-range interactions are not readily available [11, 15]. As an alternative, it is a common practice to use the Bogolubov–Born–Green–Kirkwood–Yvon (BBGKY) hierarchy for the description of transport phenomena. First, in the BBGKY hierarchy we consider a limiting case of collisionless systems that mathematically corresponds to the situation when $M \to \infty$. In other words, those points that represent physical particles are smeared out, forming a 'continuous' phase fluid in the space μ. Systems of this type are thought of as a large ensemble of weakly interacting particles and only long-range forces (like Coulomb forces) are considered. The mathematical model for such situations can be thought of as a single-particle Liouville equation supplemented by an effective field equation that represents the averaged effect of many-body physics. A typical model of this type is the Vlasov equation that can be written in terms of the probability of existence $[F(\boldsymbol{x}, \boldsymbol{v}, t)]$ of a particle at the state $(\boldsymbol{x}, \boldsymbol{v})$ at time t as follows

$$\partial_t F + \boldsymbol{v} \cdot \nabla_{\boldsymbol{x}} F + \frac{1}{m^*} \mathcal{F}_{\text{eff}} \cdot \nabla_{\boldsymbol{v}} F = 0 \qquad t > 0, \quad \boldsymbol{x} \in \mathbb{R}^3_{\boldsymbol{x}}, \quad \boldsymbol{v} \in \mathbb{R}^3_{\boldsymbol{v}} \tag{1.4}$$

where $\mathcal{F}_{\text{eff}}$ is defined analogously to (1.3). For example,

$$\mathcal{F}_{\text{eff}} = -q \boldsymbol{E}_{\text{eff}} \tag{1.5}$$

with the effective electric strain

$$\boldsymbol{E}_{\text{eff}}(\boldsymbol{x}, t) = \boldsymbol{E}_{\text{ext}}(\boldsymbol{x}, t) + \int_{\mathbb{R}^3_{\bar{\boldsymbol{x}}}} n(\boldsymbol{x}, t) \boldsymbol{E}_{\text{int}}(\boldsymbol{x}, \bar{\boldsymbol{x}}) \, \mathrm{d}\bar{\boldsymbol{x}} \tag{1.6}$$

where $n(\boldsymbol{x}, t)$ is the number of charged particles per unit volume in an infinitesimal neighbourhood of $\boldsymbol{x}$ at time t, and $\boldsymbol{E}_{\text{int}}$ and $\boldsymbol{E}_{\text{ext}}$ are the internal and external electric strains, respectively.

In the general case, equations (1.4)–(1.6) are supplemented by the Maxwell system. If we consider only the Coulomb force, then

$$\boldsymbol{E}_{\text{int}} = -\frac{q}{4\pi\epsilon} \frac{\boldsymbol{x} - \boldsymbol{y}}{\|\boldsymbol{x} - \boldsymbol{y}\|^3} \qquad \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^3_{\boldsymbol{x}}, \quad \boldsymbol{x} \neq \boldsymbol{y}. \tag{1.7}$$

Under the assumption that $\boldsymbol{E}_{\text{eff}} = -\nabla_{\boldsymbol{x}} \varphi$ (φ is the electric field potential), the effective field equation for an electron system is reducible to the Poisson equation

$$-\epsilon\epsilon_0 \Delta\varphi = q(N - n) \tag{1.8}$$

where $N(\boldsymbol{x}, t)$ is the overall concentration of doping (accounting for background ions), and ϵ and ϵ_0 are the dielectric permittivities of the considered material and vacuum, respectively.

We note that in the semiconductor context, function $F(\boldsymbol{x}, \boldsymbol{v}, t)$ from (1.4) may be interpreted as the number of charged particles per unit volume in an infinitesimal neighbourhood of $(\boldsymbol{x}, \boldsymbol{v})$ at time t. Equation (1.4) is the nonlinear equation with nonlinearity

defined by the effective field equation such as (1.8). The problem of dimensionality is overcome for model (1.4)–(1.8), but the difficulty with this model lies in the definition of the driving force. Indeed, the integration of equation (1.4) leads to an idealized macroscopic conservation law

$$q\partial_t n - \operatorname{div} \boldsymbol{J} = 0 \tag{1.9}$$

where $\boldsymbol{J} = -q\int_{\mathbb{R}^3_v} \boldsymbol{v} F \, \mathrm{d}\boldsymbol{v}$ is the current density. Equation (1.9) can be satisfied only in a collisionless environment. In the general case, model (1.4)–(1.8) is not appropriate in the large-time scale modelling.

More realistic in the semiconductor device context is the consideration of systems with collisions. Indeed, for sufficiently large time scales, the motion of carriers decisively depends on scattering (i.e. on the short-range forces, which in classical situations leads to particle collisions). Scattering effects can only be included *statistically* in the collision operator $Q(F)$ of the Boltzmann equation, written with respect to the distribution function (the number density) for which we use the same notation $F(\boldsymbol{x}, \boldsymbol{v}, t)$:

$$\partial_t F + \boldsymbol{v} \cdot \nabla_{\boldsymbol{x}} F + \frac{1}{m^*}\boldsymbol{\mathcal{F}}_{\text{eff}} \cdot \nabla_{\boldsymbol{v}} F = Q(F). \tag{1.10}$$

Nonlinearity in equation (1.10) is defined by the effective field model and by the model for collisions. Originally introduced for dilute gases, equation (1.10) is based on the observation that the rate of change of F, caused by the effective force $\boldsymbol{\mathcal{F}}_{\text{eff}}$, vanishes along the characteristics defined by Newton equations of motion

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \boldsymbol{v} \qquad \frac{\mathrm{d}(m^*\boldsymbol{v})}{\mathrm{d}t} = \boldsymbol{\mathcal{F}}_{\text{eff}}. \tag{1.11}$$

As a result, we think about the collision effect as the *instantaneous* scattering from one state to another through a very fast change $\|\triangle \boldsymbol{v}\|$ in the velocity (momentum) vector and a very slow change $\|\triangle \boldsymbol{x}\|$ in the position vector. Formally, the situation when

$$\|\triangle \boldsymbol{v}\| \to \infty \qquad \text{and} \qquad \|\triangle \boldsymbol{x}\| \to 0 \tag{1.12}$$

is not excluded. Strictly speaking, we have to show that the Pauli principle, asserting that *two electrons cannot occupy the same state* $(\boldsymbol{x}, \boldsymbol{v})$ *at the same time*, is not violated. In the general case it might be a difficult task which is intrinsically connected with the stability issues of the mathematical model. For example, if the considered system is a Coulomb system, then one has to show that there are at most two electrons per volume $(2\pi)^3$ in the phase space.

Fortunately, in applications the situation (1.12) is prohibited by allowing for *non-zero relaxation time*. Then the Pauli principle can be satisfied if the collision operator is chosen in the following form [11]:

$$Q(F) = \int_B \{s(\boldsymbol{x}, \boldsymbol{v}', \boldsymbol{v})F(\boldsymbol{x}, \boldsymbol{v}', t)[1 - F(\boldsymbol{x}, \boldsymbol{v}, t)] \\ - s(\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{v}')F(\boldsymbol{x}, \boldsymbol{v}, t)[1 - F(\boldsymbol{x}, \boldsymbol{v}', t)]\} \, \mathrm{d}\boldsymbol{v}' \tag{1.13}$$

where s is the scattering (or transition) rate for a particle in the position $\boldsymbol{x}$ from velocity $\boldsymbol{v}'$ to velocity $\boldsymbol{v}$ at time t. The velocity of a particle in (1.13) is a function of the wave vector $\boldsymbol{k} = \boldsymbol{p}\hbar$ ($\boldsymbol{p}$ is the momentum vector) which corresponds to a specific energy band of the Brillouin zone B (see [25]), and hence the rate of transition can be considered as a function of $\boldsymbol{x}$ and $\boldsymbol{k}$. Then the relaxation time (average time between consecutive collisions at $(\boldsymbol{x}, \boldsymbol{k})$) can be introduced as follows

$$\tau(\boldsymbol{x}, \boldsymbol{k}) = 1/\lambda(\boldsymbol{x}, \boldsymbol{k}) \qquad \boldsymbol{k} \in B \text{ where } \lambda = \int_B s(\boldsymbol{x}, \boldsymbol{k}, \boldsymbol{k}') \, \mathrm{d}\boldsymbol{k}' \tag{1.14}$$

where integral (1.14) is taken over the Brillouin zone of the lattice.

The well-posedness of model (1.10) with the effective field equation defined by (1.8), is typically discussed in the literature under quite excessive smoothness requirements on function s. However, in reality transition rates are highly non-regular functions. The type of nonlinearity of the model, as well as the regularity of the transition rates, depend on the mechanism of scattering that can be formalized mathematically only using physical parametrization. Today's technology develops devices with active regions of characteristic dimensions below 1 μm that operate in electric fields $\sim 10^6$ V cm^{-1}, leading to serious mathematical modelling difficulties. Since quantum effects may not be negligible in these situations, in principle we have to describe arbitrary mixed quantum states that may not be represented by a single wave function. In addition, the energy-wave vector function may have several minima (energy-valleys) and we usually have to approximate the energy-band structure. As a result, common assumptions on the parabolic band structure (formally obtainable by scaling procedures) and on the equality of effective masses in the different directions cannot be rigorously justified in the semiconductor context.

Due to the sensitivity of model (1.10) to the definition of $Q(F)$, which is the subject of approximation, the success in modelling semiconductor devices essentially depends on the consistency of function $\tau(\boldsymbol{x}, \boldsymbol{k})$ to a specific practical situation. The applicability range of mathematical models in semiconductor device theory is eventually determined by certain functional relationships between function $\tau(\boldsymbol{x}, \boldsymbol{k})$ and other characteristics of semiconductor plasma [15].

Following this idea, we organize this paper as follows.

- In section 2 we analyse mathematical models for the description of carrier transport in semiconductors as a hierarchy of models constructed on the basis of the relaxation-time concept. Stability issues for mathematical models with collision operators are also considered in this section.
- In section 3 we consider the quasi-hydrodynamic model and compare this model with the conventional drift-diffusion and kinetic models.
- In section 4 we address the problem of critical variation of the magnitudes of dependent variables in the quasi-hydrodynamic model and focus on the normalization procedure for this model.
- Section 5 is devoted to problems of effective approximations of the energy-balance equations. In particular, in this section we provide details of the transformation of these equations amenable to computational efficiency.
- Conclusions and future directions are discussed in section 6.

The present paper complements the paper [15] where numerical procedures for the quasi-hydrodynamic model, stability conditions, and computational convergence were discussed with numerical examples. Using a general framework for the construction of the relaxation-time-based hierarchy of mathematical models in semiconductor device theory, in this paper we provide the foundation for modelling non-equilibrium processes in semiconductor plasma with non-local models.

2. Classification of models on the basis of relaxation-time concepts

Physical properties of semiconductor plasma are characterized by a number of fundamental lengths, such as De Broglie's wavelength ($\lambda = h/(m^*\tilde{v})$), the length of momentum relaxation ($\lambda_p = \tilde{v}\tau_p$), or the length of energy relaxation ($\lambda_\omega = \tilde{v}\sqrt{\tau_p\tau_\omega}$). Here $\tilde{v}$ is the characteristic velocity of charge carrier motion, $h = 2\pi\hbar$, and τ_p and τ_ω are the momentum and energy relaxation times, respectively (see details in [15]).

Our interest is limited to the devices with characteristic dimension l for which at least one of the following inequalities holds

$$l \gg \lambda \qquad l \gg \lambda_p \qquad l \gg \lambda_\omega \tag{2.1}$$

which means that we do not take into account quantum effects [15]. In this case the relaxation of mathematical models in semiconductor device theory may be provided by comparing the role of collisions with other scattering mechanisms and we have to define the range of model applicability with respect to the mean time between the collision introduced by (1.14) that characterizes the momentum-and-energy exchange speed. Initially, this consideration leads to two limiting cases, namely kinetic models (KM) and hydrodynamic models (HDM). The former group of models may be efficient in the case when

$$\tau_p \leqslant \tau_\omega \ll \tau \tag{2.2}$$

while the former group is confined to the case

$$\tau \ll \tau_p \ll \tau_\omega \tag{2.3}$$

(see the more detailed discussion in [15]).

As follows from (2.2) and (2.3), kinetic and hydrodynamic models belong to two distinct and generally non-overlapping classes of mathematical models for semiconductor device modelling. Strictly speaking, neither perturbation techniques nor the method of moments can lead to the rigorous derivation of hydrodynamic models from the Boltzmann kinetic equation unless simplified physical assumptions are made. Nevertheless, a reduction of kinetic equations to a low-dimensional system is computationally desirable. This requires considering stability issues for mathematical models with collision operators.

The perturbation technique, used for such a reduction, is usually based on the expansion of the solution with respect to a dimensionless parameter, typically the scaled mean free path defined as $\alpha = \lambda_p/l$. This technique, known as the Hilbert expansion, works well for small electric fields. More precisely, the Boltzmann equation after rescaling has the following form

$$\alpha(\partial_t F + \boldsymbol{v} \cdot \nabla_{\boldsymbol{x}} F) - \boldsymbol{E}_{\text{eff}} \cdot \nabla_{\boldsymbol{v}} F = Q(F). \tag{2.4}$$

The chosen time scaling for the Boltzmann equation defines a relationship between the collision operator $Q(F)$ and the driving force $\mathcal{F}$ [11]. In this case, F_0, the leading term in the expansion $F = \sum_{i=0}^{\infty} \alpha^i F_i$, has to satisfy the following equation

$$-\boldsymbol{E}_{\text{eff}} \cdot \nabla_{\boldsymbol{v}} F_0 = Q(F_0). \tag{2.5}$$

Unfortunately, this relationship may lead to the runaway phenomenon, the occurrence of which depends on the collision frequency, i.e., on the *physical* mechanisms of scattering. This is a clear indication of the need to correct the leading term F_0 in the Hilbert expansion. However, equations (2.4) and (2.5) (or a modification of the latter) are not independent of each other and form a system of coupled equations. This leads to considerable difficulties in finding a correction to F_0. Resolving these difficulties in a semiconductor device context ultimately leads to a hyperbolic-type equation for concentrations, the solution of which may have discontinuities. These discontinuities may be formally eliminated via 'viscosity' arguments similar to the derivation of the Navier–Stokes system from the kinetic equations. Although in many applications such a parabolic smoothing is often acceptable from an engineering point of view [12], its rigorous justification in the semiconductor context requires a certain connection between diffusion coefficients, D_n, D_p, and the drift mobility of carriers μ_n, μ_p. For example, if thermal equilibrium is assumed with the absolute carrier temperature T, then such a connection is often described by the classical Einstein relation

$$D_n/\mu_n = D_p/\mu_p = \varphi_T \tag{2.6}$$

where $\varphi_T = k_b T/q$ is referred to as the thermal voltage (or thermal potential) and k_b is the Boltzmann constant [15]. Using appropriate assumptions, different modifications and generalizations of (2.6) have been proposed in the literature [13]. However, for sufficiently high electron fields the dependency (2.6) as well as its modifications can be violated. Hence, it is quite natural to require that for the description of non-equilibrium and non-local processes in semiconductor plasma, high-field phenomena have to be modelled in a way compatible with experiment [6].

During recent years attempts have been made to improve hydrodynamic models using the method of moments and taking into account moments of higher orders [29]. This allows us to take into account the energy flow. Attempts have also been made to obtain new improved expressions for the current density [6]. In the final analysis, in order to rigorously derive the hydrodynamic model or its modifications from kinetic equations one has to know some *a priori* information on the solution of the Boltzmann equation and to use mathematical assumptions compatible with the physical situation under investigation. Let us consider, for example, the electro-hydrodynamic model for an electron system (see [3, 5, 15, 18] and references therein):

$$\frac{\partial \boldsymbol{z}}{\partial t} = \boldsymbol{\zeta} + \left(\frac{\partial \boldsymbol{z}}{\partial t}\right)_{\mathrm{col}} \tag{2.7}$$

where

$$\boldsymbol{z} = (n, \boldsymbol{v}, W)^T \qquad \boldsymbol{\zeta} = (\mathcal{F}_1, \boldsymbol{\mathcal{F}}_2, \mathcal{F}_3)^T \qquad \mathcal{F}_1 = -\nabla \cdot (n\boldsymbol{v}) \tag{2.8}$$

$$\boldsymbol{\mathcal{F}}_2 = -\boldsymbol{v} \cdot \nabla \boldsymbol{v} - q\boldsymbol{E}_{\mathrm{eff}}/m_n^* - \nabla(nT_n)/(m_n^* n) \tag{2.9}$$

$$\mathcal{F}_3 = -\nabla \cdot (\boldsymbol{v}W) - qn\boldsymbol{v} \cdot \boldsymbol{E}_{\mathrm{eff}} - \nabla \cdot (\boldsymbol{v}nT_n) - \nabla \cdot \boldsymbol{q} \tag{2.10}$$

T_n is the electron temperature given in energetic units, n is the electron concentration, $\boldsymbol{v}$ is their averaged velocity, W is the energy density (typically modelled by $W = 3nT_n/2 + mn\|\boldsymbol{v}\|^2$), $\boldsymbol{q}$ is the heat flow (typically modelled by the Fourier law $\boldsymbol{q} = -k\nabla T_n$) and m_n^* is the effective electron mass.

The system (2.7) couples electrical, mechanical and thermal fields. The first term in the rhs of (2.7) approximates the thermo-electromagnetic field effect, whereas the second term is of a 'mechanical' nature and approximates collisions caused by lattice vibrations, impurities, crystal imperfections etc. Equations (2.7) have to be supplemented by the field equation, for example the Poisson equation (1.8). Although equations (2.7) are similar to the Euler equations, in contrast to the latter they include source terms modelling relaxation processes and electric field effects. In the general case, the type of the differential equations (2.7) changes with respect to the functional dependency between 'collision terms', $(\partial_t n)_{\mathrm{col}}$, $(\partial_t \boldsymbol{v})_{\mathrm{col}}$ and $(\partial_t T_n)_{\mathrm{col}}$. These terms cannot be evaluated explicitly. In applications, the collision terms are typically approximated by relaxation-time approximations. In such cases hierarchy of the mathematical models with respect to the dependencies between τ, τ_ω and τ_p is the most natural.

The roots of the main difficulties arising in semiconductor device modelling lie in the adequate definition of the approximate relationship between the collision operator and the driving force. In its essence, such an *approximate* relationship determines the range of the mathematical model applicability in the solution of practical problems. For example, the reduction of kinetic equations to hydrodynamic-type models such as (2.7) may be reasonably justified only for systems with strong collision. Assuming that this is the case, we can model semiconductor plasma using the analogy with an ensemble of interacting particles. A similar situation takes place in fluid dynamics when we model non-viscous, non-compressible fluid (vortices). In the reduction procedure we ignore part of the information about the system. This results in the difficulties well known in multiparticle quantum theory (MQT) [7]. When studying a large dynamic system it is important to have information on its ground state

energy, i.e., information on the eigenfunction that corresponds to the lowest eigenvalue. In the reduction to hydrodynamic models we modify the original problem considering 'particle cloud' motion instead of the interaction between particles. This may lead to a significant deviation between the results obtained with hydrodynamic models and experiments [21]. If we consider a large Coulomb system, the shape of the cloud is typically obtained assuming that the system of particles in its ground state behaves like a Thomas–Fermi gas, i.e., such as a classical gas supplemented by the Pauli principle. However, the Thomas–Fermi theory only gives the leading asymptotic terms of the ground state system energy, which represents the quasi-classical energy, whereas the second term represents the quantum spectrum of Coulomb singularities. In this case, the well-posedness of the mathematical formulation of the problem can be established if one can show the positiveness of the system Hamiltonian. In turn, this can be established *only* under certain relationships between the fine structure constant, $\alpha_{\rm f} = q^2/\hbar c$, or the scaled mean free path between two scattering events ([11], p 86), α, the number of nuclei, K, their charges, Z, and the number of electrons, N, in the system. From a practical point of view, if the perturbation technique is used, one has to obtain a lower bound for the 'perturbation parameter' α, which is a function of Z, N and K. In the general case, without such a bound one cannot guarantee that the system is stable. Since in practice the value of N and K are typically finite, it is the connection between α ($\alpha_{\rm f}$) and Z that holds the key to the stability problem solution in applications (see, for example, [8, 9]).

The model (2.7) can be simplified to the drift-diffusion model (DDM). The name of the latter model came from the type of dependence of the *current densities* on carrier densities and electric field. For the DDM the current densities are *the sums* of drift terms (with the mobilities μ_n and μ_p) and diffusion terms (with the diffusion coefficient D_n, D_p) which are connected by the Einstein-type relation [see (2.6)]. As is the case for hydrodynamic-type models, mobility coefficients in the DDM cannot be evaluated explicitly. In addition, the scaling of the problem (and as a result the definition of a perturbation parameter α in the Hilbert expansion) decisively depends on the choice of the reference time and the reference field strength (i.e. on the ratio between the thermal voltage, φ_T, and the reference length, l_0). In modern applications of semiconductors the reference field strength may be very large. In such cases the analysis of the problem based on the Hilbert expansion cannot adequately account for high-field effects. Strictly speaking, the DDM can only be applied when $\alpha \to 0$. However, through the technological advances connected with the miniaturization and the use of materials other than silicon, the mean path becomes *larger* compared to the device size. To meet these new challenges we need to develop the next generation of mathematical models and, consequently, efficient numerical methods for their solutions.

3. Quasi-hydrodynamic models as a reasonable compromise between drift-diffusion, hydrodynamic, and kinetic models

3.1. Assumptions and the range of applicability

There are many reasons in favour of the development of models other than hydrodynamic, kinetic and drift-diffusion types (see details in [15]). One of them is that a wide area of applications is confined to the situation, which may not overlap with (2.2) or (2.3), when

$$\tau_p \leqslant \tau \ll \tau_\omega. \tag{3.1}$$

In such cases the most reasonable candidate in the relaxation-time hierarchy of mathematical models is the quasi-hydrodynamic model. Typically, for this kind of model we assume

- 'small anisotropy', i.e., we require

$$m_n^* \|\boldsymbol{v}\|^2/2 \ll 3T_n/2 \tag{3.2}$$

- quadratic law of dispersion and the parabolicity of the energy bands, i.e., the effective masses of carriers are scalar constants (m_n^* and m_p^* for the electron and the hole, respectively).

Both of these assumptions may be questionable. However, the first assumption allows us to represent an approximation to the average energies in the form

$$\langle \varepsilon_n \rangle = 3T_n/2 \qquad \langle \varepsilon_p \rangle = 3T_p/2 \tag{3.3}$$

and implies that all quantities that depend on $\langle \varepsilon_n \rangle$ (or $\langle \varepsilon_p \rangle$) become functions of temperature (for example, $\mu_n = \mu_n(T_n)$, $D_n = D_n(T_n)$ etc). The second assumption helps us to deal with multi-extrema situations in energy-wave vector-functions. In the general case, much complexity in this area of mathematical modelling stems from the multi-temperature character of semiconductor plasma. It is a well known fact that the energy spectrum of electrons in semiconductors (including unipolar field-effect transistors of submicrometer size or Si-MOS structures) has a multi-valley character. This leads to the difference in dynamic properties of carriers that belong to different valleys. Such a difference may be considerable in high electromagnetic fields. This phenomenon can be explained physically with respect to a specific material. For example, for GaAs semiconductors a *difference* in effective masses and locations of energy minima of the central Γ-valley and lateral L- and X-valleys may play a crucial role. In this case, assumptions on the equality of effective masses in different directions can hardly be justified ([11], p 69). Another difficulty lies in the differences of dynamic properties of carriers that belong to different valleys. They may be connected with differences in the orientation of these valleys with respect to the electromagnetic field (as is the case for Si-based devices). Whenever this is the case, we have several options. For example, we may use the kinetic equation for the description of carriers in different valleys of the conductance band. Alternatively, we may use a multi-temperature quasi-hydrodynamic model. In the latter case we assume that carrier collisions guarantee 'almost Maxwellization' of carriers that belong to different valleys and that EHP is non-degenerate, which means that the energy density of carriers (for example, electrons) in the ith valley is defined by $(\bar{\varepsilon}_n)_i = 3n_i(T_n)_i/2$ (as before the temperature is given in energy units). Since in this case collisions lead to the relaxation to a inter-valley balance (which is defined by the common effective temperature of carriers of each sign) it is reasonable to consider one-component (for each type of carriers) models. This agreement will be adopted for our further consideration.

3.2. Limiting cases and the comparison with other models

It is common practice to use an analogy with conservation laws for the physical interpretation of hydrodynamic equations such as (2.21). Mathematically such laws may be satisfied only approximately and the first three moments of the Boltzmann equation may not be sufficient for the adequate description of semiconductor physics. Different attempts to extend the hydrodynamic model by including energy flux conservation have recently been reported in the literature [29].

In this paper we follow a different direction. It is clear that for modelling of a number of non-equilibrium processes we have to use the energy-balance law, which can be derived as a third moment of the kinetic equation. It is reasonable to supplement the fundamental semiconductor system by this equation without the momentum conservation equation. Of course, in this case in the continuity equation we have to introduce thermal dependencies

of coefficients on the energy variable (3.3). The first model of this type was introduced by Stratton [24] for taking into account 'hot carrier' effects and for more accurate estimation of their contribution to the current. This model, often referred to as the energy-balance model (EBM) or the hot-carrier transport equations (HCTE), has been developed by many authors. Attempts have also been made to use different types of simplified hydrodynamic models (SHDM) (see [1, 10, 27, 28] and references therein). If relaxation times are obtained by fitting specified velocity-field and temperature-field characteristics, the distinction between EBM and SHDM loses its significance [1]. Both types of model, EBM and SHDM, may be referred to as quasi-hydrodynamic models (QHDM). For models of *quasi-hydrodynamic* type *mobility becomes a function of carrier energy rather than local electric field strength as in the DDM*. Mathematically this leads to a strong coupling between continuity and energy-balance equations. As a result of this coupling serious mathematical and computational difficulties arise in the analysis of such models. These difficulties have their origin in physical models, which require the description of essentially non-equilibrium behaviour of electron–hole plasma with parameters (such as concentrations, velocity, energy) that are non-locally connected with the electric field strength.

In practice we often observe a good agreement between the results obtained from HDM and QHD models [23]. This can be explained by the fact that in most practical situations the condition $\tau_p \ll \tau_\omega$ will be satisfied and this condition is the same for both models. Hence, although internal physical processes predicted by the classical DDMs and EBMs, HDMs, SHDMs are different in principle, such a difference may not manifest itself on the output characteristics of devices. As a result, there are many artificial approaches for the modification of DDM. In the majority of cases assumptions made under such approaches can hardly be justified (for example, the assumption on the equality between carrier temperature and lattice temperature or neglecting thermoflux of carriers). The area of application of DDM is restricted at least by elements with dimensions exceeding lengths of relaxation for momentum and energy. This restriction does not allow us to account for non-equilibrium and non-locality of EHP. If effective temperatures of the semiconductor structure are considered as local functions of the electromagnetic field (it may often be an acceptable approximation if (2.1) holds), the quasi-hydrodynamic model turns into DDM for which the carrier velocity is a local function of the electric field. For example, for the electron system we may have

$$\boldsymbol{v}_n = \mu_n(\boldsymbol{E}_{\text{eff}})\boldsymbol{E}_{\text{eff}} \tag{3.4}$$

where $\mu_n = q\tau_p/m_n^*$ is the low-field mobility and τ_p, as above, the moment relaxation time. We recall that the mobility coefficient μ_n is determined by scattering mechanisms (i.e. carrier collisions with phonons and lattice vibrations, impurities and crystal imperfections), the approximation of which is often tedious ([22], p 75 and appendix 23). The linear dependence of the velocity on the electric field, expressed by (3.4), is a consequence of Newton's second law of motion (1.11). The rhs of (1.11), $\mathcal{F}_{\text{eff}}$, has to approximate a combined effect of thermomechanical and electromagnetic forces. For example, in the simplest case of a one-electron system evolving in a low-electric field, we may use the following approximation

$$\mathcal{F}_{\text{eff}} \approx \mathcal{F}_0 = -q\boldsymbol{E}_{\text{eff}} - \frac{m_n^*\boldsymbol{v}}{\tau_p}. \tag{3.5}$$

Since this approximation is not appropriate in high-electric fields, attempts have been made to improve it. One such attempt is expressed by the second equation of system (2.7), where two extra terms were introduced to the model (3.5)

$$\mathcal{F}_2 = \mathcal{F}_0 - \boldsymbol{v}\cdot\nabla\boldsymbol{v} - \frac{1}{nm_n^*}\nabla(nT_n). \tag{3.6}$$

However, it is well known that in describing the dynamics of one-electron systems in high electric fields, the concept of energy relaxation time, τ_ω, becomes inevitable (see [22], p 73). From a mathematical point of view we arrive at the evolutionary equation for the energy density

$$\frac{\partial W}{\partial t} = \mathcal{F}_3 + \left(\frac{\partial W}{\partial t}\right)_{\text{col}} \tag{3.7}$$

where the simplest approximations of the rhs of this equation can be written in the form

$$\mathcal{F}_3 \approx \mathcal{F}_3^0 = -qn\boldsymbol{v} \cdot \boldsymbol{E}_{\text{eff}} \qquad \left(\frac{\partial W}{\partial t}\right)_{\text{col}} \approx -\frac{W - 3nT_n/2}{\tau_\omega}. \tag{3.8}$$

An improved approximation of $\mathcal{F}_3$ is provided by the third equation in (2.7) (see [3, 22])

$$\mathcal{F}_3 \approx \mathcal{F}_3^0 - \nabla \cdot (\boldsymbol{v}W) - \nabla \cdot (\boldsymbol{v}nT_n) - \nabla \cdot \boldsymbol{q}. \tag{3.9}$$

The idea of replacing the equation for momentum by the energy-balance equation is central to the development of quasi-hydrodynamic-type models. For the construction of such models we use the equations of conservation of mass and energy, but not momentum.

Quasi-hydrodynamic models have at least two limiting cases (for two-type carrier models):

- effective masses of different type carriers are of the same order (then it is reasonable to talk about common temperature of charge carriers, which may be different from the lattice temperature);
- effective masses of different type carriers are sharply different, but times between collision for the same type of carriers are fairly small (then it is reasonable to talk about temperatures for carriers of different type, which may not coincide between themselves as well as with the lattice temperature).

As mentioned above, the similarity between hydrodynamic and quasi-hydrodynamic models manifests itself in the fact that the condition $\tau_p \ll \tau_\omega$ is the same for both models. In other words, both models can be applied in the case of *weakly non-elastic* scattering. This happens if momentum scattering takes place on impurity atoms (under arbitrary energy scattering mechanism) or if energy scattering, as well as (quasi-)momentum scattering, is induced by interaction between charge carriers and acoustic/piezoelectric oscillation of the lattice. The scattering on optical phonons will be weakly non-elastic if the average energy of the charge carrier essentially exceeds the energy of optical phonons. In non-degenerate EHP the relaxation time is inversely proportional to the carrier concentration (within a slowly changing factor). Therefore, both hydrodynamic and quasi-hydrodynamic models may be applied only under high enough concentrations that exceed certain critical values, n_ω and n_p. Above these values scattering of energy and (quasi-)momentum (respectively) at the cost of between-electron (or between-hole) collisions becomes dominant. In order to evaluate these values in practice we scale the ratios τ_ω/τ and τ_p/τ to 1, leading to the approximate relation $n_\omega/n_p \approx \tau_p/\tau_\omega$.

Quasi-hydrodynamic models may be effectively applied even if not applicable in the rigorous sense of the word, i.e., when concentration of charge is not large enough compared to n_ω. This approach provides a reasonable approximation if we are interested only in physical quantities obtainable by averaging relatively smooth energy functions (the error of approximation is connected only with the identification of the average energy with (3.3)). In this case we chose the inertial system in such a way that the system of charge carriers is in equilibrium (see (3.3)). In any other inertial system the average energy would contain a term connected with the kinetic energy of motion of the charge carrier system as a whole (drift energy). In this inertial system the rhs of the energy-balance equation takes a form similar to that from the heat transfer theory, but with a relaxation factor. For example, for the electron

system we have $\langle \partial_t \mathcal{E}_n \rangle = (T_n - T_l)/\tau_\omega$, where T_l denotes the temperature of the lattice. EHP and the crystal lattice play the role of subsystems that exchange energy. Whenever we expect $T_n > T_l$ the 'hot carriers' terminology acquires the intuitive sense.

3.3. Mathematical model and physical parametrization

Following [2, 4, 10, 14, 15], in the space–time region $\bar{G}^R = \{(x, t) : 0 \leqslant x \leqslant L, 0 \leqslant t \leqslant T\}$ we consider the following quasi-hydrodynamic model for semiconductor device modelling

$$\begin{cases} \partial_{xx}\varphi = q(n - p - N)/\epsilon\epsilon_0 \\ \partial_t n - \partial_x J_n/q = F \\ \partial_t p + \partial_x J_p/q = F \\ \partial_t \bar{\mathcal{E}}_n + \partial_x Q_n = -J_n \partial_x \varphi + P_n \\ \partial_t \bar{\mathcal{E}}_p + \partial_x Q_p = -J_p \partial_x \varphi + P_p \end{cases} \tag{3.10}$$

where expressions for densities of carrier currents, J_n, J_p, and flux energies, Q_n and Q_p, have the following form

$$J_n = -qn\mu_n \partial_x \varphi + q\partial_x(D_n n) \qquad J_p = -qp\mu_p \partial_x \varphi - \partial_x(D_p p) \tag{3.11}$$

$$Q_n = \beta_n T_n n \mu_n \partial_x \varphi - \beta_n \partial_x [T_n D_n n]/q \qquad Q_p = -\beta_p T_p p \mu_p \partial_x \varphi - \beta_p \partial_x [T_p D_p p]/q \tag{3.12}$$

$\bar{\mathcal{E}}_n = 3nT_n/2$, $\bar{\mathcal{E}}_p = 3pT_p/2$ are the average densities of the electron and hole systems, respectively, and F is an approximation to the contribution of the generation–recombination (and, possibly, ionization) processes.

All our notation in this paper coincide with those used in [15]. For the convenience of the reader, we recall that β_n and β_p in (3.12) are the Peltier coefficients, β_n and β_p in (3.12), defined by

$$\beta_n = 2.5 + \xi_n \qquad \beta_p = 2.5 + \xi_p \tag{3.13}$$

where $\xi_n = \mathrm{d}\ln \mu_n(T_n)/\mathrm{d}\ln T_n$, $\xi_p = \mathrm{d}\ln \mu_p(T_p)/\mathrm{d}\ln T_p$ [2]. The rhs of the energy-balance equations have the forms

$$P_n = n(T_l - T_n)/\tau_\omega^n \qquad P_p = p(T_l - T_p)/\tau_\omega^p \tag{3.14}$$

where temperature is considered in energy units, and τ_ω^n, τ_ω^p are the average energy relaxation times for electrons and holes, respectively. Their approximations are discussed in detail in [15].

Average mobilities of carriers are set to

$$\mu_n = \mu_n^0 (T_n/T_l)^q \qquad \mu_p = \mu_p^0 (T_p/T_l)^q \tag{3.15}$$

$$\tau_\omega^n = \tau_{\omega,0}^n (T_l/T_n)^s \qquad \tau_\omega^p = \tau_{\omega,0}^p (T_l/T_p)^s \tag{3.16}$$

where q and s are determined by the dominant relaxation mechanisms of the momentum and energy [4, 16].

As for the dependencies of the diffusion coefficients on carrier temperatures, we admit that the numerical procedures developed in [15] can easily be generalized to the general type of dependence

$$D_n(T_n)/\mu_n(T_n) = \tilde{f}_1(T_n) \qquad D_p(T_p)/\mu_p(T_p) = \tilde{f}_2(T_p). \tag{3.17}$$

Initial conditions for the model are

$$n(x, 0) = n_0(x) \qquad p(x, 0) = p_0(x) \qquad T_n(x, 0) = T_p(x, 0) = T_l \qquad 0 \leqslant x \leqslant L. \tag{3.18}$$

We assume that the functions $n_0(x)$ and $p_0(x)$ in the initial conditions are defined as equilibrium values of densities for electrons and holes, that is

$$p_0(x)n_0(x) = n_{ie}^2 \qquad n_0(x) - p_0(x) - N = 0 \tag{3.19}$$

where n_{ie} is the effective intrinsic concentration of carriers (see details of its approximation in [15]).

Boundary conditions depend on the type of modelling structure, and for definiteness we assume:

- equality of carrier temperature and lattice temperature

$$T_n(0,t) = T_p(L,t) = T_l \tag{3.20}$$

- conditions of quasi-neutrality and infinite velocity of recombination (thermodynamic equilibrium):

$$p - n + N = 0 \qquad pn = n_{ie}^2 \qquad x \in \partial G^R = \{0, L\} \tag{3.21}$$

  from where it is easy to get

$$n = \frac{N}{2} + \sqrt{\left(\frac{N}{2}\right)^2 + n_{ie}^2} \qquad p = -\frac{N}{2} + \sqrt{\left(\frac{N}{2}\right)^2 + n_{ie}^2} \qquad x \in \partial G^R = \{0, L\}. \tag{3.22}$$

For the potential, boundary conditions are standard [11]

$$\varphi(0,t) = 0 \qquad \varphi(L,t) = U + \varphi_{\text{cont}} \tag{3.23}$$

where U is the applied voltage and φ_{cont} is the contact potential difference determined by the formula $\varphi_{\text{cont}} = \varphi_T \ln(n(t, L)/n_{ie})$, (obtained as a consequence of $n = n_{ie}\exp((\varphi - \varphi_n)/\varphi_T)$ by setting $\varphi_n = U$). In other words, we assume that the bias is applied at the right contact, while the left contact is grounded. In this case we require the conjugating conditions

$$\varphi(0,0) = 0 \qquad \varphi(L,0) = U + \varphi_{\text{cont}} \tag{3.24}$$

to be satisfied.

As in [15], the recombination model takes into account recombination on defects induced by dopants (the Shockley–Read–Hall recombination) and between-zone Auger recombination:

$$F(n,p) = \frac{pn - n_{ie}^2}{\tau_n(p + n_{ie}) + \tau_p(n + n_{ie})} + (pn - n_{ie}^2)(c_n n + c_p p) \tag{3.25}$$

where the carrier lifetimes and coefficients of Auger recombination are set as follows

$$\tau_n = 1.7 \times 10^{-5}\ \text{s} \qquad \tau_p = 3.95 \times 10^{-4}\ \text{s}$$
$$c_n = 2.9 \times 10^{-31}\ \text{cm}^6\ \text{s}^{-1} \qquad c_p = 1.2 \times 10^{-31}\ \text{cm}^6\ \text{s}^{-1}.$$

For the problem where impact ionization plays a significant role we have to add the velocity of ionization term

$$G_p - G_n = \alpha_p J_p - \alpha_n J_n \tag{3.26}$$

where

$$J_n = -qnv_n \qquad J_p = qpv_p \tag{3.27}$$

and α_n, α_p are field-dependent carrier ionization rates defined as the number of electron–hole pairs generated by an electron/hole per unit distance travelled [25].

4. Normalization procedure and challenges in the computational treatment of non-local models

In this section we present details of the normalization procedure used in [15]. Let us introduce the following dimensionless variables

$$x' = x/L \qquad t' = t/t_* \qquad \varphi' = \varphi/\varphi_* \qquad n' = n/n_* \qquad p' = p/n_*$$
$$T'_c = T_c/T_* \qquad N' = N/n_* \qquad J'_c = J_c/J_* \qquad F' = F/F_* \qquad \mu'_n = \mu_c/\mu_*$$
$$Q'_c = Q_c/Q_* \qquad \tau'_c = \tau_c/t_* \qquad \alpha'_c = \alpha/\alpha_* \qquad c'_c = c_c/c_* \tag{4.1}$$

where quantities with the subindex '*' are critical values to be defined below, and variables with the subindex 'c' (carriers) are equally applied to electrons or holes. We perform non-dimensionalization of problem (3.10)–(3.12) in four steps.

Step 1. From the Poisson equation written in the dimensionless variables

$$\frac{\epsilon\epsilon_0}{L^2}\varphi_* \frac{\partial^2\varphi'}{\partial x'^2} = qn_*(n' - p' - N') \tag{4.2}$$

we obtain

$$\frac{\epsilon\epsilon_0}{L^2}\varphi_* = qn_*. \tag{4.3}$$

Step 2. The continuity equations (as an example we consider the continuity equation for electrons)

$$\frac{n_*}{t_*}\frac{\partial n'}{\partial t'} - \frac{1}{q}\frac{J_*}{L}\frac{\partial J'_n}{\partial x'} = F_* F' \tag{4.4}$$

leads to the relationships for two other normalized factors

$$\frac{n_* qL}{t_* J_*} = 1 \qquad \frac{F_* qL}{J_*} = 1. \tag{4.5}$$

The formula for the current density

$$J_* J'_n = -\frac{qn_*\mu_*\varphi_*}{L} n'\mu'_m \frac{\partial\varphi'}{\partial x'} + \frac{T_*\mu_* n_*}{L}\frac{\partial}{\partial x'}(T'_n\mu'_n n') \tag{4.6}$$

and relationships (4.3) and (4.5) allow us to obtain

$$n_* = \frac{\epsilon\epsilon_0 T_*}{L^2 q^2} \qquad J_* = \frac{\epsilon\epsilon_0 T_*}{Lqt_*} \qquad t_* = \frac{qL^2}{\mu_* T_*} \qquad \varphi_* = T_*/q. \tag{4.7}$$

Step 3. Then, from the energy-balance equation we get

$$\frac{3}{2}\frac{n_* T_*}{t_*}\frac{n'T'}{\partial t'} + \frac{Q_*}{L}\frac{\partial Q'_n}{\partial x'} = -\frac{J_*\varphi_*}{L} J'_n \frac{\partial\varphi'}{\partial x'} + \frac{n_* T_*}{t_*} n' \frac{1 - T'_n}{\tau'_\omega} \tag{4.8}$$

that leads to only one relationship

$$Q_* = (n_* T_* L)/t_*. \tag{4.9}$$

(the other is satisfied automatically due to (4.7)).

The current density formula

$$Q_* Q'_n = \frac{T_* n_* \mu_* \varphi_*}{L}\beta_n T'_n n' \mu'_n \frac{\partial\varphi'}{\partial x'} - \frac{\beta_n}{q}\frac{T_* D_* n_*}{L}\frac{\partial}{\partial x'}[T'_n D'_n n'] \tag{4.10}$$

requires $D_* = \mu_* T_*$. The critical values of μ and T_c have been set to one (in mobility coefficient units) and T_l (in energy units), respectively.

Step 4. Finally, we consider the recombination–generation–ionization term in dimensionless variables

$$F_*F' = n_*^2[p'n' - (n'_{ie})^2]\left\{\frac{1}{t_*n_*[\tau'_n(p'+n'_{ie})+\tau'_p(n'+n'_{ie})} + c_*n_*(c'_n n' + c_p p')\right\} + \frac{1}{q}J_*\alpha_*[\alpha'_p J'_p - \alpha'_n J'_n]. \tag{4.11}$$

This leads to

$$\frac{n_*}{t_*}\frac{qL}{J_*} = 1 \qquad \frac{c_*n_*^3 qL}{J_*} = 1 \qquad \frac{J_*}{q}\alpha_*\frac{qL}{J_*} = 1 \tag{4.12}$$

from which we easily obtain normalized factors for the Auger recombination coefficients and for the ionization rates

$$c_* = 1/(n_*^2 t_*) \qquad \alpha_* = 1/L. \tag{4.13}$$

The above deliberation leads to the following normalized factors that we use in this paper

$$\begin{aligned} &T_* = T_l k_b/q \qquad n_* = (\epsilon\epsilon_0 T_*)/(L^2q^2) \qquad \varphi_* = qn_*L^2/(\epsilon\epsilon_0) = T_*/q \\ &D_* = \mu_* T_*, \quad t_* = qL^2/(\mu_* T_*) \qquad c_* = 1/(n_*^2 t_*) \qquad \alpha_* = 1/L \\ &J_* = \epsilon\epsilon_0 T_*/(Lqt_*) \qquad Q_* = n_* T_* L/t_* \qquad \mu_* = 1. \end{aligned} \tag{4.14}$$

The constants and normalized factors used in our computational experiments are summarized in [15, see appendix].

Using (4.14), after simple transformations we get the following normalized system

$$\begin{cases} \partial_{xx}\varphi = n - p - N \\ \partial_t n - \partial_x J_n = F \\ 3/2\partial_t(nT_n) + \partial_x Q_n = -J_n\partial_x\varphi + P_n \\ \partial_t p + \partial_x J_p = F \\ 3/2\partial_t(pT_p) + \partial_x Q_p = -J_p\partial_x\varphi + P_p \end{cases} \tag{4.15}$$

where

$$J_n = -n\mu_n\partial_x\varphi + \partial_x(T_n\mu_n n) \qquad J_p = -p\mu_p\partial_x\varphi - \partial_x(T_p\mu_p p) \tag{4.16}$$

$$Q_n = \beta_n T_n n\mu_n\partial_x\varphi - \beta_n\partial_x[T_n D_n n] \qquad Q_p = -\beta_p T_p p\mu_p\partial_x\varphi - \beta_p\partial_x[T_p D_p p]. \tag{4.17}$$

All scaled quantities in (4.15)–(4.17) are denoted by the same symbols as their unscaled counterparts. The system (4.15)–(4.17) is considered in the normalized space–time region $\bar{G} = \{(x,t) : 0 \leqslant x \leqslant 1, 0 \leqslant t \leqslant T/t_*\}$ and is supplemented by the normalized initial and boundary conditions:

$$n(x,0) = n_0(x)/n_* \qquad p(x,0) = p_0(x)/n_* \qquad T_n(x,0) = T_p(x,0) = 1 \tag{4.18}$$

$$p - n + N = 0 \qquad pn = n_{ie}^2 \qquad T_n = T_p = 1 \qquad x \in \partial G = \{0, 1\} \tag{4.19}$$

$$\varphi(0,t) = 0 \qquad \varphi(1,t) = (U + \varphi_{\text{cont}})/\varphi_*. \tag{4.20}$$

It is assumed that the condition $J_n(x,0) = J_p(x,0) = 0$ and the normalized conjugating conditions $\varphi(0,0) = 0$, $\varphi(1,0) = (U + \varphi_{\text{cont}})/\varphi_*$ are satisfied.

Since the primary numerical difficulty in the solution of system (4.15)–(4.20) lies with the energy-balance equations, the next section is devoted to the transformation of these equations to forms amenable to computational efficiency.

5. Transformation of the energy-balance equations to forms amenable to computational efficiency

As we pointed out in [15], in contrast to the continuity equation, the energy-balance equation *cannot be readily reduced* to a 'divergent' or 'conservation' form [2, 19, 20]. In the semiconductor-device modelling context, the main problem with the energy-balance equation lies with the presence of the product between the current density and the electric field strength ($J_n \times E$ or $J_p \times E$), that has a 'non-divergent' structure. Hence, one cannot immediately apply the general theory developed for the construction of monotone difference schemes [17, 19, 20]. However, since the product between the current density and the electric field strength provides the key to the non-local coupling between the effective carrier temperature and the electric field, the problem of its efficient approximation has to be dealt with. In this section we show how the energy-balance equations can be reduced to an 'almost-divergent' form which can then be used for their effective discretization. In our companion paper [15] we provided an example of such a discretization.

Following [2, 10, 26], let us transform the energy fluxes to forms where all derivatives of ε_n and ε_p are 'covered' by the symbol of divergence. We start from the energy-balance equation for the system of electrons

$$\partial_t \bar{\varepsilon}_n = -\partial_x Q_n - J_n \partial_x \varphi + P_n. \tag{5.1}$$

The rhs of (5.1) is transformed as follows

$$\begin{aligned}
-\frac{\partial Q_n}{\partial x} - J_n \frac{\partial \varphi}{\partial x} + P_n &= \frac{\partial}{\partial x}\left[\beta_n \frac{\partial (T_n D_n n)}{\partial x} - \beta_n n T_n \mu_n \frac{\partial \varphi}{\partial x}\right] - \frac{\partial}{\partial x}[\mu_n n T_n]\frac{\partial \varphi}{\partial x} + n\mu_n \left(\frac{\partial \varphi}{\partial x}\right)^2 \\
+P_n &= \beta_n \frac{\partial}{\partial x}\left[\frac{\partial (T_n D_n n)}{\partial x}\right] - \frac{\partial}{\partial x}[\mu_n n T_n]\frac{\partial \varphi}{\partial x} - \mu_n n T_n \frac{\partial^2 \varphi}{\partial x^2} + n T_n \mu_n \frac{\partial^2 \varphi}{\partial x^2} \\
&\quad -\beta_n \frac{\partial}{\partial x}[\mu_n n T_n]\frac{\partial \varphi}{\partial x} - \beta_n \mu_n n T_n \frac{\partial^2 \varphi}{\partial x^2} + n\mu_n \left(\frac{\partial \varphi}{\partial x}\right)^2 + P_n \\
&= \frac{\partial}{\partial x}\left[\beta_n \frac{\partial (D_n n T_n)}{\partial x} - (1+\beta_n)\mu_n n T_n \frac{\partial \varphi}{\partial x}\right] \\
&\quad + n T_n \left[\mu_n \frac{\partial^2 \varphi}{\partial x^2} + \frac{\mu_n}{T_n}\left(\frac{\partial \varphi}{\partial x}\right)^2 + \frac{1 - T_n}{\tau_{\omega^n} T_n}\right] = \frac{\partial Q_n^*}{\partial x} + S_n(T_n, \varphi)\mathcal{E}_n
\end{aligned} \tag{5.2}$$

where

$$\varepsilon_n = n T_n \qquad S_n = \mu_n(T_n)\partial_{xx}\varphi + \mu_n(T_n)(\partial_x \varphi)^2 + (1 - T_n)/(\tau_\omega^n(T_n)T_n) \tag{5.3}$$

$$Q_n^* = \beta_n \partial_x (D_n(T_n)\varepsilon_n) - (1+\beta_n)\mu_n(T_n)\varepsilon_n \partial_x \varphi. \tag{5.4}$$

Therefore, equation (5.1) can be written in the form

$$3\partial_t \varepsilon_n / 2 = \partial_x Q_n^* + S_n(T_n, \varphi)\varepsilon_n. \tag{5.5}$$

In a similar manner we transform the energy-balance equation for the system of holes

$$\partial_t \bar{\varepsilon}_p = -\partial_x Q_p - J_p \partial_x \varphi + P_p. \tag{5.6}$$

Since

$$\begin{aligned}
-\frac{\partial}{\partial x}&\left[-\beta_p T_p p \mu_p \frac{\partial \varphi}{\partial x} - \beta_p \frac{\partial}{\partial x}[T_p D_p p]\right] - \left[-p\mu_p \frac{\partial \varphi}{\partial x} - \frac{\partial}{\partial x}(D_p p)\right]\frac{\partial \varphi}{\partial x} + P_p \\
&= \beta_p \frac{\partial}{\partial x}\left[\frac{\partial (T_p D_p p)}{\partial x}\right] + \frac{\partial}{\partial x}(\mu_p T_p p)\frac{\partial \varphi}{\partial x} - \mu_p p T_p \frac{\partial^2 \varphi}{\partial x^2} + p T_p \mu_p \frac{\partial^2 \varphi}{\partial x^2} \\
&\quad +\beta_p \frac{\partial}{\partial x}[\mu_p T_p p]\frac{\partial \varphi}{\partial x} + \beta_p \mu_p p T_p \frac{\partial^2 \varphi}{\partial x^2} + p\mu_p \left(\frac{\partial \varphi}{\partial x}\right)^2 + P_p
\end{aligned}$$

$$= \frac{\partial}{\partial x}\left\{\beta_p \frac{\partial(D_p T_p p)}{\partial x} + \mu_p p T_p \frac{\partial \varphi}{\partial x} + \beta_p T_p p \mu_p \frac{\partial \varphi}{\partial x}\right\} + \mu_p p T_p \frac{\partial^2 \varphi}{\partial x^2}$$
$$- p\mu_p \left(\frac{\partial \varphi}{\partial x}\right)^2 + P_p = \frac{\partial}{\partial x}\left\{\beta_p \frac{\partial(D_p T_p p)}{\partial x} + (1+\beta_p)\mu_p p T_p T_p \frac{\partial \varphi}{\partial x}\right\}$$
$$- \mu_p p T_p \frac{\partial^2 \varphi}{\partial x^2} + p\mu_p \left(\frac{\partial \varphi}{\partial x}\right)^2 + P_p = \frac{\partial Q_p^*}{\partial x} + S_p(T_p, \varphi)\varepsilon_p \tag{5.7}$$

we rewrite equation (5.6) in the following form

$$3\partial_t \varepsilon_p / 2 = \partial_x Q_p^* + S_p(T_p, \varphi)\varepsilon_n \tag{5.8}$$

where

$$\varepsilon_p = pT_p \qquad S_p = -\mu_p(T_p)\partial_{xx}\varphi + \mu_p(T_p)(\partial_x \varphi)^2 + (1 - T_p)/(\tau_\omega^p(T_p)T_p) \tag{5.9}$$

$$Q_p^* = \beta_p \partial_x (D_p(T_p)\varepsilon_p) + (1+\beta_p)\mu_p(T_p)\varepsilon_p \partial_x \varphi. \tag{5.10}$$

The representations (5.5) and (5.8) allow us to construct efficient numerical discretizations for non-local models applied to semiconductor device simulation. In particular, the construction of monotone exponential difference schemes [19, 20] based on these representations, as well as results of numerical simulations for typical semiconductor devices, have already been discussed in the literature (see [15] and references therein).

6. Conclusions and future directions

In this paper we considered a hierarchy of semiconductor device models using the relaxation time concept. In order to describe non-local, non-equilibrium processes in electron–hole semiconductor plasma, we focused on the quasi-hydrodynamic model which provides a reasonable compromise between kinetic, hydrodynamic and drift-diffusion models. The normalization procedure, and issues of the approximation of fluxes for this model were discussed in detail. The energy-balance equations for the electron and hole systems were reduced to convenient forms for computational implementation. These forms lead to effective discretizations of non-local models. A special case of such discretizations based on exponential difference schemes, the problems of the computational stability of the algorithmic realizations of the proposed schemes, and the results of numerical simulations can be found in our companion paper [15]. However, the application of the representations discussed in this paper is not restricted to the construction of exponential difference schemes. Such representations provide a useful framework in constructing other effective discretization procedures in modelling semiconductor devices with non-local models.

Acknowledgments

The authors were supported by grant USQ-PTRP 17989 and by Australian Research Council small grant 17906. We thank Dr David Smith for his assistance at the final stage of preparation of this paper and an anonymous referee for helpful comments.

References

[1] Apanovich Y *et al* 1994 Steady-state and transient analysis of submicron devices using energy balance and simplified hydrodynamic models *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems* **13** 702–11

[2] Birukova L J *et al* 1989 Simulation algorithms for computing processes in electron plasma of submicron semiconductor devices *Math. Modelling* **1** 11–22

[3] Blotekjaer K 1970 Transport equations for electrons in two-valley semiconductors *IEEE Trans. Electron Devices* **17** 38–47
[4] Cook R K 1983 Numerical simulation of hot-carrier transport in silicon bipolar transistors *IEEE Trans. Electron Devices* **30** 1103–10
[5] Gardner C L, Jerome J W and Rose D J 1989 Numerical methods for the hydrodynamic device model: subsonic flow *IEEE Trans. Computer-Aided Design* **8** 501–7
[6] Hansch W and Miura-Mateush 1985 A new current relation for hot electron transport *NEMACODE IV: Proc. 4th Int. Conf. on the Numerical Analysis of Semiconductor Devices and Integrated Circuits (Trinity College, Dublin, Ireland)* ed J J H Miller (Boole Press) pp 311–14
[7] Ivrii V Ja and Sigal I M 1993 Asymptotics of the ground state energies of large Coulomb systems *Ann. Math.* **138** 243–335
[8] Lieb E H and Yau H-T 1988 The stability and instability of relativistic matter *Commun. Math. Phys.* **118** 177–213
[9] Lieb E H, Siedentop H and Solovej J P 1997 Stability and instability of relativistic electrons in classical electromagnetic fields *J. Stat. Phys.* **89** 37–57
[10] Lyumkis E D *et al* 1992 Transient semiconductor device simulation including energy balance equation *COMPEL—Int. J. Computation Math. Electric. Electron. Eng.* **11** 311–25
[11] Markowich P A, Ringhofer C A and Schmeiser C 1991 *Semiconductor Equations* (Berlin: Springer)
[12] Markowich P A 1995 Diffusion approximation of nonlinear electron phonon collision mechanisms *Math. Modelling Numerical Anal. (M^2 AN)* **29** 857–69
[13] Marshak A H and van Vliet K M 1978 Electrical current in solids with position-dependent band structure *Solid-State Electron.* **21** 417–27
[14] Melnik R V N 1995 Semi-implicit finite-difference schemes with flow correction for quasi-hydrodynamic models of semiconductor devices *Eng. Simul.* (London: Gordon and Breach) **12** 856–65
[15] Melnik R V N and He H 2000 Modelling nonlocal processes in semiconductor devices with exponential difference schemes *J. Eng. Math.* at press
[16] Nikolaeva V A, Ryzhii V I and Cheverushkin B N 1988 A numerical method for the simulation of two-dimensional semiconductor structures in the quasi-hydrodynamic approximation *Sov. Phys. Dokl* **33** 110–12
[17] Oran E S and Boris J P 1981 *Numerical Simulation of Reactive Flow* (Amsterdam: Elsevier)
[18] Rudan M, Odeh F and White J 1987 Numerical solution of the hydrodynamic model for a one-dimensional semiconductor device *COMPEL—Int. J. Computation Math. Electric. Electron. Eng.* **6** 151–70
[19] Samarskii A A 1984 *Theorie der Differenzenverfahren* (Series: Mathematik und ihre Anwendengen in Physik und Technik) (Aufl. Leipzig: Academische Verlagesellschaft)
[20] Samarskii A A and Nikolaev E S 1989 *Numerical Methods for Grid Equations* (Basle, Boston: Birkhauser)
[21] Scrobohaci P G and Tang T-W 1994 Modeling of the hot electron subpopulation and its application to impact ionization in submicron silicon devices *IEEE Trans. Electron Devices* **41** 1197–212
[22] Shur M 1990 *Physics of Semiconductor Devices* (Englewood Cliffs, NJ: Prentice Hall)
[23] Snowden C M and Loret D 1987 Two-dimensional hot-electron models for short-gate-length GaAs MESFET's *IEEE Trans. Electron Devices* **34** 212–23
[24] Stratton R 1962 Diffusion of hot and cold electrons in semiconductor barriers *Phys. Rev.* B **126** 2002–14
[25] Sze S M 1981 *Physics of Semiconductor Devices* (New York: Wiley)
[26] Tang T-W 1984 Extension of the Scharfetter–Gummel algorithm to the energy balance equation *IEEE Trans. Electron. Devices* **31** 1912–14
[27] Tang T-W, Ou X X and Navon D X 1985 Prediction of the velocity overshoot by a nonlocal hot-carrier transport model *NEMACODE IV: Proc. 4th Int. Conf. on the Numerical Analysis of Semiconductor Devices and Integrated Circuits (Trinity College, Dublin, Ireland)* ed J J H Miller (Boole Press) pp 519–24
[28] Tang T-W and Ieong M-K 1995 Discretization of flux densities in device simulations using optimum artificial diffusivity *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems* **14** 1309–15
[29] Zhang Y and Nokali M El 1993 A hydrodynamic transport model and its applications in semiconductor device simulation *Solid-State Electron.* **36** 1689–96