A CRITICAL REVIEW OF THE FUNDAMENTAL SEMICONDUCTOR EQUATIONS

By

G. Baccarani
F. Odeh
A. Gnudi
and
D. Ventura

IMA Preprint Series # 1019
August 1992

A Critical Review of the Fundamental Semiconductor Equations

G. Baccarani, F. Odeh* A.Gnudi, D. Ventura

Dipartimento di Elettronica, Informatica e Sistemistica
Università di Bologna, viale Risorgimento 2
40136 Bologna, Italy

**IBM Research Division, T. J. Watson Research Center*
Yorktown Heights, NY 10598, USA

Abstract

Physical models for numerical device simulation are reviewed in an engineering perspective, and their derivation from more general physical principles is examined. More specifically, the limitations of the basic semiconductor equations within the drift-diffusion model is discussed and its validity range is assessed with emphasis on silicon devices. An approximate solution to the Boltzmann Transport Equation (BTE) is obtained by using the Hilbert expansion method near equilibrium conditions; however, in contrast to previous treatments, however, the drift and diffusion term are linked together in order to overcome the usual limitation of small external fields. The solution is carried out up to the third term of the expansion, thus leading to a correction of the usual drift-diffusion current density. This correction yields a nonlinear velocity-field relationship and accounts for velocity overshoot.

1. Introduction

The drift-diffusion model of current transport in semiconductors is to date well established [1]. Numerical solvers based on the above model have successfully been used to predict new device effects and to help the designing of advanced processes and devices [2-3]. Yet, its mathematical justification under general non-homogeneous conditions is still unclear, since its derivation from the Boltzmann Transport Equation (BTE) usually requires the assumption of a small electric field and small carrier gradients [4]. Even in the simplest electron device, i.e. the p-n junction, such a condition is almost never fulfilled: fields as large as 10^5 V/cm, well exceeding $k_B T/q\lambda$ (where λ is the carrier mean-free path) are typically found even under equilibrium conditions. It is therefore important to assess more carefully the validity of the model from a mathematical standpoint.

The main aim of this paper is to identify the limits of the drift-diffusion model in the context of device simulation. We use the Hilbert expansion method [4] to derive the model but, in contrast to previous approaches, we link the drift and diffusion terms together in order to properly scale their combination near equlibrium conditions. In doing that, the validity of the model is assessed and found to hold so long as the gradient of the quasi-Fermi potential is small compared with $(k_B T/q\lambda)$. In the second part of the paper, we consider the equations governing the higher-order terms, and solve them up to third order. Hence, we find the correction to the current density.

The paper is organized as follows: section 2 defines the mathematical problem and discusses the limitations of the BTE in the context of electron transport in semiconductors. The scaling of the BTE is performed in section 3. Section 4 is devoted to the evaluation of the first-order solution of the BTE, and to a discussion of its limitations. The second- and third-order solutions to the BTE are derived in section 5, where the correction to the current density is identified. Finally, conclusions are drawn in section 6.

2. Semiconductor Equations

The starting point of a first-principles transport theory is the quantum mechanical Liouville-Von Neumann equation for the density matrix. Under a number of simplifying assumptions, the latter reduces to the well-known Boltzmann transport equation (BTE) reported below

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \mathbf{u} \cdot \mathrm{grad} f \mp \frac{q}{\hbar}\mathbf{E} \cdot \mathrm{grad} f = \mathcal{S} f \tag{2.1}$$

where $f(\mathbf{r}, \mathbf{k}, t)\, d^3r\, d^3k$ represents the number of electrons/holes in the elementary volume $d^3r\, d^3k$ of the phase space at the time t, and $\mathbf{u}$ is the group velocity. In (2.1) the total derivative df/dt is evaluated along the classical trajectory $\mathbf{r} = \mathbf{r}(t)$, $\mathbf{k} = \mathbf{k}(t)$, i.e. in the absence of collisions. The $\mp$ sign refers to the electron and hole cases, respectively. The collision term on the RHS of (2.1) represents the rate of change of f due to collisions; its expression is the following

$$\mathcal{S}f = \int [S(\mathbf{r},\mathbf{k}',\mathbf{k})f(\mathbf{r},\mathbf{k}',t) - S(\mathbf{r},\mathbf{k},\mathbf{k}')f(\mathbf{r},\mathbf{k},t)]\, d^3k' , \tag{2.2}$$

provided a non degenerate electron gas is assumed. In (2.2) $S(\mathbf{r},\mathbf{k},\mathbf{k}')\, d^3k'$ represents the scattering probability per unit time from the state $\mathbf{k}$ at $\mathbf{r}$ to any other state $\mathbf{k}'$ contained in d^3k'. In what folllows we shall assume that, due to the symmetry properties of the S function, the operator $\mathcal{S}$ is even, i.e.

$$\mathcal{S}[f(-\mathbf{k})] = (\mathcal{S}f)(-\mathbf{k}) \tag{2.3}$$

Actually, the above assumption is only correct for isotropic scattering mechanisms, such as acoustic- and optical-phonon interactions. If impurity scattering plays a role, we shall assume an equivalent isotropic impurity scattering which provides the same momentum relaxation time as the real anisotropic scattering mechanism.

In (2.1), the electric field $\mathbf{E} = -\operatorname{grad}\varphi$ is governed by Poisson's equation

$$\operatorname{div}(\epsilon_s \operatorname{grad}\varphi) = q\,(p - n + N_D - N_A) \tag{2.4}$$

where φ is the electric potential, ϵ_s the semiconductor dielectric constant, n and p the electron and hole concentrations, N_D and N_A the donor and acceptor impurity densities, respectively. It should be mentioned that, within a device, E may be large and fast varying both in space and time, depending upon the nature of the device and its boundary conditions.

The problem of device simulation is thus defined by a couple of BTEs (2.1) for both electrons and holes and by the coupling equation (2.4). For the sake of simplicity, we shall consider only the BTE for electrons and neglect generation-recombination effects, which play a negligible role for most micron-size devices. From the definition of the distribution function f, the electron concentration and the current density are given by

$$n(\mathbf{r},t) = \int f(\mathbf{r},\mathbf{k},t)\, d^3k \tag{2.5,a}$$

$$\mathbf{J}_n(\mathbf{r},t) = -q \int \mathbf{u}\, f(\mathbf{r},\mathbf{k},t)\, d^3k \tag{2.5,b}$$

Eq. (2.1) basically represents a continuity equation in the phase space which relies upon the single-particle approximation. The basic assumptions underlying Eqs. (2.1) and (2.2) can be summarized as follows (see e.g. [5]):

i) The single-particle approximation holds, i.e. the interaction among carriers is assumed to be weak.
ii) The band theory and the effective-mass theorem apply to the semiconductor.
iii) The electric field is slowly varying over a scale length comparable with the physical dimensions of the wave packet describing the motion of the particle; therefore a classical description of electron dynamics is possible via the Ehrenfest theorem.
iv) Only point collisions are considered in time and in space. Therefore, no energy is gained by the particle from the electric field during collisions.
v) Electron-electron interaction is neglected in the scattering integral (2.2).
vi) The scattering probability is independent of the external field.
vii) The electron/hole gas is not degenerate.
viii) The effect of the magnetic field is neglected.

In essence, Eq. (2.1) relies upon a semi-classical picture of carrier transport in semiconductors. Electron motion results from a sequence of drifts in the electric field, followed by scattering events. The drifting time, the type of scattering process, and the final state are random quantities which can be expressed in terms of transition rates due to the various processes; however, the free flight is deterministic and depends on the spatial distribution of the electric field.

The BTE is an integro-differential equation which, unfortunately, allows for a closed-form solution only in extremely special cases. The most successful technique to attack the problem is based on a Monte Carlo approach [6-7], whereby hot-electron effects could be successfully interpreted in most semiconductor materials. Other techniques based on a variational approach [8] or iterative procedures [9-11] could help in solving certain specific problems, but turned out to be of a scarse general applicability.

If equation (2.1) is integrated over the first Brillouin zone, we find a continuity equation in the physical space, namely

$$\frac{\partial n}{\partial t} - \frac{1}{q}\,\mathrm{div}\mathbf{J}_n = 0 \tag{2.6}$$

where eqs. (2.5) have been accounted for. In order that eq. (2.6) can be self-consistently solved with Poisson's equation (2.4), an expression for the current density $\mathbf{J}_n$ must be found against φ, n and their derivatives. In principle, this can be achieved by looking for approximate solutions to eq. (2.1).

3. Scaling of the BTE

The approximate, or asymptotic, analysis of a differential equation is usually accomplished by first introducing a scaling process, the aim of which is that of assessing the relative importance of the various terms and of identifying the operating conditions under which some terms can be considered of a higher order with respect to others. In doing so, a set of normalizing constants, possibly related to the physical processes being described by the equation, must be identified. Here a problem arises, for eqs. (2.1) and (2.4) do not uniquely identify a characteristic time or length: the latters are rather determined by the specific nature of the device under consideration and its boundary conditions. The scaling which follows pertains to the regime of "near equilibrium" in a sense which is given later.

We assume as a characteristic length the device active length L and as a characteristic time the transit time of the carriers travelling at the velocity $v_d = \mu(k_B T/qL)$; hence, the transit time turns out to be $\tau_t = L/v_d$. Next, we normalize the group velocity $\mathbf{u}$ by the thermal velocity $v_{th} = (k_B T/m^*)^{1/2}$ and the crystal momentum $\mathbf{k}$ by $k_{th} = (m^*/\hbar)v_{th}$. Finally, the scattering rates are normalized by τ_{po}^{-1}, where

$$\tau_{po} = \frac{\langle \tau_{co}(\mathbf{u}), u^2 f_0 \rangle}{\langle 1, u^2 f_0 \rangle} \tag{3.1}$$

and $\tau_{co}(\mathbf{u})$ is the inverse scattering rate for an undoped semiconductor material. In (3.1) brackets denote an averaging, i.e. a scalar product of L_2 functions over the velocity space. Also, f_0 is the equilibrium distribution function. We now define the following normalized physical quantities:

$$t^* = t/\tau_t \;\; ; \quad \mathbf{r}^* = \mathbf{r}/L \;\; ; \quad \mathbf{u}^* = \mathbf{u}/v_{th} \;\; ; \quad \mathbf{k}^* = \mathbf{k}/k_{th} \;\; ; \quad \mathcal{S}^* = \tau_{po}\mathcal{S}$$

and, by substituting the normalized variables into (2.1), we find

$$\epsilon^2 \frac{\partial f^*}{\partial t^*} + \epsilon (\mathbf{u}^* \cdot \mathrm{grad}_{\mathbf{r}^*} f^* - \mathbf{E}^* \cdot \mathrm{grad}_{\mathbf{k}^*} f^*) = \mathcal{S}^* f^* \tag{3.2}$$

where f^* is the new distribution function defined in the $(\mathbf{r}^*, \mathbf{k}^*)$ space, $\mathbf{E}^*$ is the normalized field

$$\mathbf{E}^* = \frac{qL\mathbf{E}}{k_B T} \tag{3.3}$$

and

$$\epsilon = \sqrt{\frac{\tau_{po}}{\tau_t}} = \frac{\lambda_o}{L} \tag{3.4}$$

where $\lambda_o = v_{th}\tau_{po}$ is the carrier mean-free path for an undoped semiconductor. We shall attempt an approximate solution of eq. (3.2) under the assumptions that: *i)* $\epsilon \ll 1$ and, *ii)* the device is weakly perturbed with respect to equilibrium. This means that the applied voltage is not too large, and we defer for now the estimate of an acceptable value for the voltage itself. We want to stress, however, that we do not impose any constraint on the local values of the force acting upon the carriers, apart from those which ensure the applicability of the Ehrenfest theorem; nor do we assume a slowly-varying impurity concentration in (2.4) and within the scattering operator $\mathcal{S}^*$. From a mathematical standpoint, this implies that the two terms in parenthesis on the l.h.s. of eq. (3.2) may be individually large. These terms, however, cancel in equilibrium, and it is reasonable to assume that under the assumption of a small perturbation their algebraic sum remains of the same order as that of f^*.

4. First-order solution to the BTE

In this paragraph we attempt an approximate solution to the BTE and, in order to simplify the analysis, we assume spherical energy bands. For the sake of convenience, we omit henceforth the superscript * to denote normalized variables. It is now worth noting that $\lambda_o \simeq 10\,\mathrm{nm}$ in silicon and, assuming $L \simeq 1\,\mu\mathrm{m}$, then $\epsilon \simeq 10^{-2}$. The form of eq. (3.2) then suggests an expansion of the f function in a power series of the form [4]

$$f = \sum_{j=0}^{\infty} \epsilon^j f^{(j)} \tag{4.1}$$

and, by substituting (4.1) into (3.2) and equating coefficients of the same power, we find the following set of equations

$$\mathcal{S} f^{(0)} = 0 \tag{4.2, a}$$

$$\mathcal{S} f^{(1)} = \mathbf{u} \cdot \mathrm{grad} f^{(0)} - \mathbf{E} \cdot \mathrm{grad} f^{(0)} \tag{4.2, b}$$

$$\mathcal{S} f^{(2)} = \frac{\partial f^{(0)}}{\partial t} + \mathbf{u} \cdot \mathrm{grad} f^{(1)} - \mathbf{E} \cdot \mathrm{grad} f^{(1)} \tag{4.2, c}$$

....

Substituting the power series expansion (4.1) into eq. (2.5,a) gives

$$n(\mathbf{r}, t) = \sum_{j=0}^{\infty} n^{(j)}(\mathbf{r}, t) \tag{4.3, a}$$

where

$$n^{(j)}(\mathbf{r}, t) \equiv \epsilon^j \int f^{(j)}(\mathbf{r}, \mathrm{k}, t)\, d^3k \tag{4.3, b}$$

The solution of eq. (4.2,a) is

$$f^{(0)}(\mathbf{r}, \mathrm{k}, t) = \exp(-\varepsilon(\mathrm{k}))\, g(\mathbf{r}, t) \tag{4.4}$$

where the normalized energy ε is related to the crystal momentum by the band function

$$\gamma(\varepsilon) = \varepsilon\,(1 + \alpha\varepsilon) = \frac{1}{2}\,\mathbf{k}\cdot\mathbf{k} \tag{4.5}$$

and $g = g(\mathbf{r}, t)$ is an arbitrary function of position and time. The arbitrariness of the g function comes from assumption *iv)* above that only point collisions may occur in time and in space.

From (4.3,b), the zero-order solution $f^{(0)}$ may be expressed as

$$f^{(0)}(\mathbf{r},\mathbf{k},t) = N_c^{-1}\exp(-\varepsilon(\mathbf{k}))\,n^{(0)}(\mathbf{r},t) \tag{4.6}$$

where

$$N_c = (2\pi)^{3/2}\left(1 + \frac{15}{4}\,\alpha\right) \tag{4.7}$$

The term $(15/4)\alpha$ in eq. (4.7) represents the correction due to the non-parabolicity of the band. Because of the small value of α, the above correction is typically of the order of 5%. Under the condition that the r.h.s. of eqs. (4.2) belong to the range of the operator $\mathcal{S}$, $f^{(j)}$ may be formally represented as

$$f^{(j)} = \mathcal{S}^{-1}\left(\frac{\partial f^{(j-2)}}{\partial t} + \mathbf{u}\cdot\mathrm{grad}\,f^{(j-1)} - \mathbf{E}\cdot\mathrm{grad}\,f^{(j-1)}\right) \tag{4.8}$$

with $j = 1, 2, 3, \ldots\ldots$, where we use the notation $f^{(-1)} \equiv 0$, and where the functions $f^{(j)}$ are uniquely determined by the normalizing conditions

$$\left\langle 1, f^{(j)}\right\rangle = 0 \quad \text{for}\, j \geq 1 \tag{4.9}$$

Using (4.9) one obtains

$$n(\mathbf{r},t) = n^{(0)}(\mathbf{r},t) \tag{4.10}$$

We note that, since the r.h.s. of (4.2) are alternatively even and odd functions in $\mathbf{k}$, the parity properties of the $\mathcal{S}$ operator ensure us that $f^{(j)}$s are alternatively even and odd as well.

If we consider now eq. (4.2,b), then using (4.6) and (4.10) leads to the equation

$$\mathcal{S}f^{(1)} = N_c^{-1}\exp(-\varepsilon(\mathbf{k}))\,\mathbf{u}\cdot(\mathrm{grad}\,n + n\mathbf{E}) \tag{4.11}$$

whose solution is

$$f^{(1)}(\mathbf{r},\mathbf{k},t) = -\,N_c^{-1}\tau_c(\mathbf{r},\mathbf{k})\,\exp(-\varepsilon(\mathbf{k}))\,\mathbf{u}\cdot(\mathrm{grad}\,n + n\mathbf{E}) \tag{4.12}$$

where $\tau_c(\mathbf{r},\mathbf{k})$ is the inverse scattering rate normalized to τ_{po}. The $\mathbf{r}$ dependence of τ_c is due to the position-dependent impurity concentration. From the definition of the current density (2.5,b) we find

$$\mathbf{J}_n^{(1)}(\mathbf{r},t) = -\int \epsilon\,\mathbf{u}\,f^{(1)}(\mathbf{r},\mathbf{k},t)\,d^3k = \epsilon\,D_n(\mathbf{r})\,(\mathrm{grad}\,n + n\mathbf{E}) \tag{4.13}$$

where the normalized diffusivity is

$$D_n(\mathbf{r}) = \tau_p(\mathbf{r})\,(1 - 5\,\alpha) \tag{4.14,a}$$

and $\tau_p(\mathbf{r})$ is a suitable weighted average of the collision time given by

$$\tau_p(\mathbf{r}) = \frac{\langle \tau_c(\mathbf{r},\mathbf{k})\,u^2\rangle}{\langle u^2\rangle} \tag{4.14,b}$$

and may be identified with the momentum relaxation time. The correction factor 5α accounts for the non-parabolicity of the band. The latter, which is usually neglected in elementary theory, turns out to be of the order of 6%. Using (4.13) we can express $f^{(1)}$ as a function of $\mathbf{J}_n^{(1)}$, namely

$$f^{(1)}(\mathbf{r},\mathbf{k},t) = -\,N_c^{-1}\,\tau_c(\mathbf{r},\mathbf{k})\,\exp(-\varepsilon(\mathbf{k}))\,\mathbf{u}\cdot\frac{1}{\epsilon D_n(\mathbf{r})}\,\mathbf{J}_n^{(1)} \tag{4.15}$$

In order to conclude that $f^{(0)} + \epsilon\,f^{(1)}$ represents a meaningful 1^{st} order solution to eq. (3.2), we must compare the relative magnitudes of $f^{(0)}$ and $\epsilon\,f^{(1)}$. Such a comparison is made problematic by the nature of the two functions at hand, which turn out to be even and odd, respectively. By observing that the mean quadratic value of each component of the group velocity is 1, it is reasonable to demand that

$$\left\langle \left| \epsilon\,\frac{f^{(1)}\mathbf{u}}{f^{(0)}} \right| \right\rangle \ll 1 \tag{4.16}$$

which, from (4.12), reads

$$\left| \epsilon\,\tau_p(\mathbf{r})\,\frac{1}{n}\,(\mathrm{grad}n + n\mathbf{E}) \right| \ll 1 \tag{4.17}$$

In order to better clarify the meaning of (4.17), we define a normalized quasi-Fermi potential φ_n in the standard fashion

$$n = n_i\,\exp(\varphi - \varphi_n) \tag{4.18}$$

Inequality (4.17) then becomes

$$|\epsilon\,\tau_p(\mathbf{r})\,(\mathrm{grad}\varphi_n)| \ll 1 \tag{4.19}$$

which dictates that the change in the normalized quasi-Fermi potential φ_n over a carrier mean-free path $\lambda = \tau_p(\mathbf{r})\lambda_o$ is $o(1)$. Equivalently, (4.19) implies that the carrier velocity is much smaller than thermal velocity, i.e., $v \ll v_{th}$. (4.19), on the other hand, does not imply that the electric potential variation over a carrier mean-free path is $o(1)$. Such a condition is often violated within real devices even in equilibrium. If φ_n is a smooth function of position, as is the case in forward-biased p-n junctions, the limiting condition on the applied voltage V_a turns out to be

$$V_a \ll \frac{1}{\epsilon} \tag{4.20}$$

which is not especially restrictive. On the other hand, eq. (4.19) is likely to be violated when one considers diffusion near an absorbing boundary, such as carrier transport within the base of a bipolar transistor. In this case, the distribution function tends to become fairly skewed and highly distorted with respect to equilibrium, which is reflected by the increase of the average carrier velocity. Under such circumstances, eq. (4.13) is expected to break down, and higher-order corrections to the distribution, which we derive below, will become more important.

5. Second-order solution to the BTE

Having determined $f^{(0)}$ and $f^{(1)}$, we can now address eq. (4.2,c). The solvability condition for (4.2,c), namely

$$\left\langle 1, \mathcal{S} f^{(2)} \right\rangle = 0 \tag{5.1}$$

implies that

$$\frac{\partial n}{\partial t} - \frac{1}{\epsilon} \operatorname{div} \mathbf{J}_n^{(1)} = 0 \tag{5.2}$$

which is equivalent to (2.5) so long as $\mathbf{J}_n^{(1)} \simeq \mathbf{J}_n$.

For the solution of eq. (4.2,c), we assume a one-dimensional geometry. From eqs. (4.6), (4.10) and (4.15) we find:

$$\mathcal{S} f^{(2)} = N_c^{-1} \exp(-\varepsilon) \left\{ \left[\frac{\partial n}{\partial t} - \frac{1}{\epsilon} \frac{\tau_c(x,k)}{D_n(x)} u_x^2 \frac{\partial J_{nx}^{(1)}}{\partial x} \right] + \right.$$

$$\left. \frac{1}{\epsilon} \frac{\tau_c(x,k)}{D_n(x)} \left[\gamma'^{-1} - \left(1 - \frac{1}{\tau_c} \frac{\mathrm{d}\tau_c}{\mathrm{d}\varepsilon} + \gamma'^{-1} \gamma'' \right) u_x^2 \right] E_x J_{nx}^{(1)} \right\} \tag{5.3}$$

Eq. (5.3) has been obtained under the assumption that the ratio τ_c/τ_p, and therefore D_n/τ_p, is actually independent of x. Such a condition would be exact if τ_c could be expressed as the product of a function of position times a function of energy. Under such circumstances, the averaging process (4.14,b) leading to τ_p would not alter the x-dependence of τ_c. Strictly speaking, such a condition is not fulfilled when impurity scattering plays an important role; on the other hand, it is reasonable to assume a weak x-dependence of the above ratio, so that eq. (5.3) is expected to be near-exact.

The r.h.s. of eq. (5.3) is an even function of k and, since $\mathcal{S}$ is an even operator, $f^{(2)}$ must be an even function of k. Thus, we may expand $f^{(2)}$ in spherical harmonics as follows

$$f^{(2)}(x, \mathrm{k}, t) = f_0^{(2)}(x, k, t) + f_2^{(2)}(x, k, t) \frac{1}{2} \left(3 \cos^2 \theta - 1 \right) \tag{5.4}$$

where some obvious symmetry on $f^{(2)}$ is implied. Defining

$$A \equiv N_c^{-1} \exp(-\varepsilon) \left(\frac{\partial n}{\partial t} + \frac{1}{\epsilon} \frac{\tau_c(x,k)}{D_n(x)} \gamma'^{-1} E_x J_{nx}^{(1)} \right) \tag{5.5,a}$$

$$B \equiv N_c^{-1} \exp(-\varepsilon) \frac{1}{\epsilon} \frac{\tau_c(x,k)}{D_n(x)}$$

$$\left[\frac{\partial J_{nx}^{(1)}}{\partial x} + \left(1 - \frac{1}{\tau_c} \frac{\mathrm{d}\tau_c}{\mathrm{d}\varepsilon} + \gamma'^{-1} \gamma'' \right) E_x J_{nx}^{(1)} \right] u^2 \tag{5.5,b}$$

where u is the velocity magnitude, we write eq. (5.3) in the form

$$\mathcal{S}f^{(2)} = A - B\cos^2\theta = \left(A - \frac{1}{3}B\right) - \frac{2}{3}B\left(\frac{1}{2}(3\cos^2\theta - 1)\right) \tag{5.6}$$

and balance the coefficients of the spherical harmonics of order 0 and 2, to obtain

$$\left(\mathcal{S}f^{(2)}\right)_0 = \left(A - \frac{1}{3}B\right) \tag{5.7,a}$$

$$\left(\mathcal{S}f^{(2)}\right)_2 = \frac{2}{3}B \tag{5.7,b}$$

Using the energy-conserving properties of the scattering mechanisms, the l.h.s. of (5.7) may now be explicitly evaluated. For (5.7,a), only inelastic scattering (i.e., optical phonon) does actually contribute and we find

$$c_{op}\left\{N_{op}^{+}\,g(\varepsilon+\varepsilon_{op})\,f_0^{(2)}(\varepsilon+\varepsilon_{op}) - \left[N_{op}^{+}\,g(\varepsilon-\varepsilon_{op}) + N_{op}g(\varepsilon+\varepsilon_{op})\right]f_0^{(2)}(\varepsilon) +\right.$$

$$\left. N_{op}\,g(\varepsilon-\varepsilon_{op})\,f_0^{(2)}(\varepsilon-\varepsilon_{op})\right\} = \left(A - \frac{1}{3}B\right) \tag{5.8}$$

where we have defined $f_0^{(2)}(\varepsilon) \equiv g(\varepsilon) \equiv 0$ for $\varepsilon < 0$. In (5.8) ε_{op} is the normalized optical-phonon energy; $N_{op} = [\exp(\varepsilon_{op})-1]^{-1}$ is the phonon occupation number; $N_{op}^{+} = N_{op} + 1$; c_{op} is the optical-phonon scattering probability per unit time, and $g(\varepsilon)$ is the density-of-states function given by

$$g(\varepsilon) = \frac{2}{\sqrt{\pi}}\,\gamma^{1/2}\,\gamma' \tag{5.9}$$

Since eq. (5.8) is a difference equation with variable coefficients, an exact solution would generally require a numerical approach. In what follows, we shall assume that $g(\varepsilon+\varepsilon_{op})$ and $g(\varepsilon-\varepsilon_{op})$ may be approximated by $g(\varepsilon)$ at the l.h.s. of (5.8). Such an approximation is expected to be reasonable for large values of the energy ε, due to the weak ε dependence of the g function. Furthermore, it does not alter the solution of the associated homogeneous equation, which turns out to be proportional to $f^{(0)}$. Under the above simplifying assumption, eq. (5.8) becomes

$$f_0^{(2)}(\varepsilon+\varepsilon_{op}) - (1+\delta)f_0^{(2)}(\varepsilon) + \delta f_0^{(2)}(\varepsilon-\varepsilon_{op}) =$$

$$\left[c_{op}\,N_{op}^{+}\,g(\varepsilon)\right]^{-1}\left(A - \frac{1}{3}B\right) \equiv F(\varepsilon) \tag{5.10}$$

In (5.10) δ has been defined as follows

$$\delta = (N_{op}/N_{op}^{+}) = \exp(-\varepsilon_{op}) \tag{5.11}$$

Using the expressions for A, B given by (5.5), the r.h.s. of (5.10) turns out to be

$$F(\varepsilon) = \left(\frac{2}{\sqrt{\pi}} c_{op} N_{op}^{+} N_c \gamma^{1/2} \gamma'\right)^{-1} \left\{ \left[\frac{\partial n}{\partial t} - \frac{2\varepsilon}{3\epsilon \gamma'^2} \frac{\tau_c(x,\varepsilon)}{D_n(x)} \frac{\partial J_{nx}^{(1)}}{\partial x}\right] + \right.$$

$$\left. \frac{1}{\epsilon} \frac{\tau_c(x,\varepsilon)}{D_n(x)} \left[\gamma'^{-1} - \frac{2}{3}\left(1 - \frac{1}{\tau_c}\frac{d\tau_c}{d\varepsilon} + \gamma'^{-1}\gamma''\right) \varepsilon \gamma'^{-2}\right] E_x J_{nx}^{(1)} \right\} \exp(-\varepsilon) \quad (5.12)$$

In principle, eq. (5.10) can now be solved by means of the z-transform technique. However, to get a transparent form of the solution $f_0^{(2)}$, we make some simplification of the rather complicated function $F(\varepsilon)$. First we assume parabolic bands and obtain

$$F(\varepsilon) = \left(\frac{2}{\sqrt{\pi}} c_{op} N_{op}^{+} N_c\right)^{-1} \left\{ \left[\frac{\partial n}{\partial t} - \frac{2\varepsilon}{3\epsilon} \frac{\tau_c(x,\varepsilon)}{\tau_p(x)} \frac{\partial J_{nx}^{(1)}}{\partial x}\right] + \right.$$

$$\left. \frac{1}{\epsilon} \frac{\tau_c(x,\varepsilon)}{\tau_p(x)} \left[1 - \frac{2}{3}\left(1 - \frac{1}{\tau_c}\frac{d\tau_c}{d\varepsilon}\right) \varepsilon\right] E_x J_{nx}^{(1)} \right\} \varepsilon^{-1/2} \exp(-\varepsilon) \quad (5.13)$$

Next, we assume steady-state conditions, so that the first term in brakets equals zero. Finally, some energy dependence for $\tau_c(\varepsilon)$ must be considered. In analogy with previous assumptions, we take

$$\tau_c(x,\varepsilon) = \tau_o(x)\, \varepsilon^{-\beta} \quad (5.14, a)$$

with

$$\beta \simeq \frac{1}{2} \quad (5.14, b)$$

The resulting expression for $F(\varepsilon)$ is

$$F(\varepsilon) = \frac{\pi/4}{\epsilon\, c_{op} N_{op}^{+} N_c} \left(\varepsilon^{-1} - 1\right) \exp(-\varepsilon)\, E_x J_{nx}^{(1)} \quad (5.15)$$

The singularity in the above expression is due to (5.14) and the simplifying assumption that $g(\varepsilon + \varepsilon_{op}) \simeq g(\varepsilon)$, which was used to get a difference equation with constant coefficients from (5.8). Therefore, we cannot expect the r.h.s. of (5.15) to be correct at low energy. On the other hand, we are looking for a solution of (5.10) only for $\varepsilon \geq \varepsilon_{op}$ and, for large energies, ε^{-1} can be neglected with respect to 1, so that, if we drop the term ε^{-1} we still expect our solution to be asymptotically correct. Hence, the simplified final form of the difference equation (5.10) becomes

$$f_0^{(2)}(\varepsilon + \varepsilon_{op}) - (1+\delta) f_0^{(2)}(\varepsilon) + \delta f_0^{(2)}(\varepsilon - \varepsilon_{op}) =$$

$$- \frac{\pi/4}{\epsilon\, c_{op}\, N_{op}^{+}\, N_c} \exp(-\varepsilon)\, E_x J_{nx}^{(1)} \tag{5.16}$$

As shown in Appendix A, the solution of (5.16) which also satisfies the normalizing condition (4.9), is

$$f_0^{(2)} = \frac{\pi/4}{\epsilon\, N_c\, c_{op}\, \varepsilon_{op}} \left(\varepsilon - \frac{3}{2}\right) \exp(-\varepsilon)\, E_x J_{nx}^{(1)} \tag{5.17}$$

Now $f_2^{(2)}$ may be trivially obtained from (5.7,b), and turns out to be

$$f_2^{(2)} = \frac{2}{3}\, \tau_c(x, \varepsilon)\, B =$$

$$\frac{\sqrt{\pi}}{\epsilon N_c} \left(\varepsilon^{1/2} + \frac{1}{2}\varepsilon^{-1/2}\right) \exp(-\varepsilon)\, E_x J_{nx}^{(1)} \tag{5.18}$$

Then, from (5.4), the expression of the $f^{(2)}$ function becomes

$$f^{(2)}(x, k, \theta) = \frac{\pi/4}{\epsilon\, N_c\, c_{op}\, \varepsilon_{op}} \left(\varepsilon - \frac{3}{2}\right) \exp(-\varepsilon)\, E_x J_{nx}^{(1)} +$$

$$\frac{\sqrt{\pi}}{\epsilon N_c} \left(\varepsilon^{1/2} + \frac{1}{2}\varepsilon^{-1/2}\right) \exp(-\varepsilon)\, E_x J_{nx}^{(1)} \left(\frac{1}{2}\left(3\cos^2\theta - 1\right)\right) \tag{5.19}$$

Eq. (5.19) holds in the limit of $\varepsilon \gg \varepsilon_{op}$, and brakes down for $\varepsilon \simeq \varepsilon_{op}$. Nevertheless, the form of eq. (5.19) allows us to conclude that the average energy in excess of 3/2 turns out to be proportional to the local product $E_x J_{nx}^{(1)}$, as may be seen by evaluating its statistical average

$$n\,\langle\varepsilon\rangle = \frac{1}{N_c} \int \varepsilon \exp(-\varepsilon) \mathrm{d}^3 k + \frac{(\pi/4)\,\epsilon\, E_x J_{nx}^{(1)}}{N_c\, c_{op}\, \varepsilon_{op}} \int \left(\varepsilon^2 - \frac{3}{2}\varepsilon\right) \exp(-\varepsilon)\, \mathrm{d}^3 k$$

and therefore

$$n\left(\langle\varepsilon\rangle - 3/2\right) = \frac{3\pi}{8} \frac{\epsilon\, E_x J_{nx}^{(1)}}{c_{op}\varepsilon_{op}} = \tau_w\, \epsilon\, E_x J_{nx}^{(1)} \tag{5.20}$$

where the energy relaxation time τ_w has been defined as

$$\tau_w = \frac{3\pi}{8} \left(c_{op}\varepsilon_{op}\right)^{-1} \tag{5.21}$$

Apart from the numerical coefficient $(3\pi/8)$, eq. (5.21) lends itself to an easy interpretation: since $N_{op}^{+} = N_{op} + 1$, the following identity holds:

$$c_{op}\varepsilon_{op} = N_{op}^{+} c_{op}\varepsilon_{op} - N_{op} c_{op}\varepsilon_{op} \tag{5.22}$$

showing that $(c_{op}\varepsilon_{op})$ may be interpreted as the difference between the emission and absorption scattering probabilities per unit time, multiplied by the optical-phonon energy. Therefore, $n\ (\varepsilon)\ (c_{op}\varepsilon_{op})$ represents the rate at which energy relaxes to equilibrium. From (5.21) it appears that, to second order in the distribution function, the energy relaxation time is a constant, independent of the carrier energy. Monte Carlo simulations of current transport in semiconductors show that τ_w is indeed a very weak function of energy. We also notice that eq. (5.20) is a simple form of energy balance under the assumption of a negligible divergence of the energy flow $\mathbf{S}_n$. Since $\operatorname{div}\mathbf{J}_n = 0$, the above assumption is justified for small values of the temperature gradient, which is again consistent with the assumption of a small perturbation with respect to equilibrium.
We now consider the equation for $f^{(3)}$, which is latter is

$$\mathcal{S}f^{(3)} = u_x\,\frac{\partial f^{(2)}}{\partial x} - E_x\,\frac{\partial f^{(2)}}{\partial k_x} \tag{5.23}$$

Since $f^{(3)}$ is an odd function in $\mathbf{k}$, the l.h.s. is simply given by

$$\mathcal{S}f^{(3)} = -\frac{f^{(3)}}{\tau_c(x,k)} \tag{5.24}$$

and the solution of eq. (5.23) is

$$f^{(3)} = -\tau_c(x,k)\left(u_x\,\frac{\partial f^{(2)}}{\partial x} - E_x\,\frac{\partial f^{(2)}}{\partial k_x}\right) \tag{5.25}$$

From (5.25), the second-order correction to the current density can now be evaluated

$$J^{(3)}_{nx} = \int \epsilon^3 f^{(3)} u_x \mathrm{d}^3k \tag{5.26}$$

and the corrected expression of the current density reads

$$J_{nx} \simeq J^{(1)}_{nx}\left\{1 + \epsilon^2\left[\eta_n\frac{\partial E_x}{\partial x} - \zeta_n E_x^2\right]\right\} \tag{5.27}$$

where

$$\eta_n = \frac{\epsilon}{E_x J^{(1)}_{nx}} \int -\tau_c(x,k) f^{(2)}(x,k,\theta) u_x^2 \mathrm{d}^3k \tag{5.28, a}$$

$$\zeta_n = \frac{\epsilon}{E_x J^{(1)}_{nx}} \int -\tau_c(x,k) \frac{\partial f^{(2)}}{\partial k_x} u_x \mathrm{d}^3k \tag{5.28, b}$$

From (5.27) we notice that the correction in brakets is second order in ϵ, and that it is composed of two terms: the first one, proportional to $\partial E_x/\partial x$, was first suggested by Thornber [12] based on heuristic considerations, and it was shown to account for non-static transport effects such as velocity overshoot. The second one, proportional to E_{nx}^2, is a second-order correction to the low-field mobility.

It is worth mentioning that, although eq. (5.19) is only approximate, the linear dependence of $f^{(2)}$ upon the product $E_x J_{nx}^{(1)}$ is exact, because the averaging process implied by eq. (5.8) will inherit the functional dependence of the r.h.s. on $E_x J_{nx}^{(1)}$. A final consideration is that the present procedure can in principle be extended to the multi-dimensional case, leading to a tensor relationship which generalizes eq. (5.27).

6. Summary and conclusions

In this work we have reexamined the derivation of the drift-diffusion model from the BTE. Such a derivation is obtained by means of the Hilbert expansion near equilibrium conditions. Compared with previous approaches, we do not require that the electric field and the carrier gradients are individually small, since such a condition is never fulfilled in real devices. Rather, we assume that the algebraic sum of the drift and diffusion terms within the BTE is $O(\epsilon)$ or, alternatively, that the carrier average velocity is small compared with the thermal velocity. We then solve the set of resulting equations up to the third order in ϵ, and determine under general conditions a correction to the usual expression of current density.

Such a correction is composed of two terms: one of them is proportional to the square of the electric field, and accounts for mobility degradation near $k_B T/q\lambda$. The second one is proportional to $\partial E_x/\partial x$ and accounts for non-static effects such as velocity over- and undershoot occurring in response to a sharp change of the electric field. Finally, a byproduct of the present treatment is an expression for the energy relaxation time, which turns out to be independent of the carrier average energy.

Appendix A

Consider the sequence of samples $n\varepsilon_{op}$, with $n = 0, 1, 2, ...$, and take the unilateral z-transform of eq. (5.16). We find

$$\hat{f}_0^{(2)}(z)\left[z-(1+\delta)+\delta z^{-1}\right] - z f_0^{(2)}(0) = \hat{F}(z) \qquad (A.1)$$

where

$$\hat{f}_0^{(2)}(z) = \sum_{n=0}^{\infty} f_0^{(2)}(n\varepsilon_q)\, z^{-n}$$

and

$$\hat{F}(z) = \sum_{n=0}^{\infty} F(n\varepsilon_q)\, z^{-n}$$

are the unilateral z-transforms of the unknown function $f_0^{(2)}(\varepsilon)$ and the r.h.s. $F(\varepsilon)$, respectively. From (A.1) we get

$$\hat{f}_0^{(2)}(z) = \frac{z\hat{F}(z)}{z^2 - (1+\delta)z + \delta} + \frac{z^2 f_0^{(2)}(0)}{z^2 - (1+\delta)z + \delta} \qquad (A.2)$$

The roots of the denominator in (A.2) are $z_1 = \delta$, $z_2 = 1$, so that eq. (A.2) takes the form

$$\hat{f}_0^{(2)}(z) = \frac{z\hat{F}(z)}{(z-\delta)(z-1)} + f_0^{(2)}(0)\left\{1 + \frac{(1+\delta)z - \delta}{(z-\delta)(z-1)}\right\} \qquad (A.3)$$

Also, from the expression (5.16) of $F(\varepsilon)$, namely

$$F(\varepsilon) = -F_o \exp(-\varepsilon)$$

we find that

$$\hat{F}(z) = -F_o \frac{z}{z + \exp(-\varepsilon_q)} \qquad (A.4)$$

Standard procedures then yield the anti-transform of (A.3) in the form

$$f_0^{(2)}(\varepsilon) = \frac{F_o}{(1-\delta)^2}\left\{(2-\delta)\frac{\varepsilon}{\varepsilon_{op}}\exp(-\varepsilon) - \frac{\varepsilon - \varepsilon_{op}}{\varepsilon_{op}}\exp(-\varepsilon) - u_h(\varepsilon - \varepsilon_{op})\right\} +$$

$$f_0^{(2)}(0)\left\{\delta_d(0) + \frac{1}{(1-\delta)}u_h(\varepsilon - \varepsilon_{op}) - \frac{\delta}{(1-\delta)}\exp(-\varepsilon)\right\} \qquad (A.5)$$

where $u_h(\varepsilon)$ is the Heaviside unit function, and $\delta_d(\varepsilon)$ is the discrete delta function. By imposing on $f_0^{(2)}(\varepsilon)$ the property of vanishing asymptotically, we finally obtain eq. (5.17).

References

1. W. V. van Roosbroeck: "Theory of Flow of Electrons and Holes in Germanium and Other Semiconductors", *Bell Syst. Tech. J.*, vol. 29, pp. 560-607, 1950.
2. S. Selberherr: "Analysis and Simulation of Semiconductor Devices", Springer Verlag, Vienna, 1984.
3. W. L. Engl: "Process and Device Modeling", North Holland, 1986.
4. P. A. Markowich, C. A. Ringhofer, C. Schmeiser: "Semiconductor Equations", Springer Verlag, Wien, 1990.
5. G. Baccarani: "Physics of Submicron Devices", in *Large Scale Integrated Circuit Technology: State of the Art and Prospects*, Eds.: L. Esaki and G. Soncini, pp. 647-669, Nijhoff, The Hague, 1982.

6. W. Fawcett, A. D. Boardman, S. Swain: "Monte Carlo Determination of Electron Transport Properties in Gallium Arsenide", *J. Phys. Chem. Solids*, vol. 31, pp. 1963-1990, 1970.
7. C. Canali, C. Jacoboni, F. Nava, G. Ottaviani, A. Alberigi Quaranta: "Electron Drift Velocity in Silicon", *Phys. Rev. B*, vol. 12, pp. 2265-2284, 1975.
8. M. Kohler: "Behandlung von Nichtgleichgewichtsvorgängen mit Hilfe eines Extremal Prinzips", *Zeitschrift für Physik*, vol. 124, pp. 772-789, 1948.
9. P. A. Wolff: "Theory of Multiplication in Silicon and Germanium", *Phys. Rev. B*, vol. 95, pp. 1415-1420, 1954.
10. G. A. Baraff: "Distribution Function and Ionization Rates for Hot Electrons in Semiconductors", *Phys. Rev.*, vol. 128, pp. 2507-2517, 1962.
11. D. L. Rode: "Electron Mobility in Direct-Gap Polar Semiconductors", *Phys. Rev. B*, vol. 2, pp. 1012-1024, 1970.
12. K. K. Thornber: "Current Equations for Velocity Overshoot", *IEEE Electron Device Lett.*, vol. EDl-3, pp. 69-71, 1982.