

Hydrodynamic Models of Semiconductor Electron Transport at High Fields

M.G. ANCONA*

Center for Nonlinear Studies Los Alamos National Laboratory Los Alamos, NM

(Received September 1, 1993)

Hydrodynamic or continuum descriptions of electron transport have long been used for modeling and simulating semiconductor devices. In this paper, we use classical field theory ideas to discuss the physical foundations of such descriptions as applied specifically to high-field transport regimes. The classical field theory development of these types of models is of interest because it differs significantly from and may be viewed as complementary to conventional derivations based on the Boltzmann equation. After outlining the general field theoretic principles upon which our development of fluid-based high-field transport descriptions is based, we study several specific models both analytically and using numerical simulation. These models provide an overall framework for understanding and extending various theories which have appeared in the literature. Most importantly, they emphasize the importance of including memory or history effects and viscosity in describing high-field transport. In all of this our aim is a unified and synoptic view unencumbered with microscopic details. Obtaining quantitative agreement with specific experiments and/or microscopic simulations is only of secondary importance. We share the view that continuum approaches can provide succinct and computationally-efficient models needed for current and future semiconductor device analysis and engineering. At the same time, we believe that these models need not be phenomenological but can be given solid physical foundation in macroscopic principles.

Key Words: *High-field transport, velocity overshoot, hydrodynamics, continuum theory, submicron semiconductor devices, rate effects, viscosity.*

1. INTRODUCTION

Much effort has been devoted to developing theories of various kinds for modeling and simulating semiconductor devices. And with the device size reductions required for VLSI integration such work has become increasingly important for the further progress of the entire microelectronics enterprise. Unfortunately, the reduction in device sizes has also brought “new” physical phenomena into play which are not described by the traditional modeling approaches, e.g., “high-field” transport effects such as velocity overshoot or quantum transport effects such as tunneling. As a result, there has been a need for engineering-oriented modeling methods which are capable of describing some or all of this “new” submicron device physics. In attempt-

ing to fill this need, the initial and primary response (about 15 years ago) was that it had become necessary to abandon the macroscopic or continuum¹ basis of the traditional approaches and that a microscopic approach was now required. This was most prominently manifested in the intense development of ensemble Monte Carlo methods for solving the Boltzmann transport equation [1]. Subsequently, however, it became increasingly apparent that the computational demands of the microscopic approaches render them largely inappropriate for device design. Thus, macroscopic descriptions have remained of great value from an engineering point-of-view. The purpose of this paper is to provide an

* Permanent address: Code 6813, Naval Research Laboratory, Washington, DC 20375. E-mail: ancona@estdsun4.nrl.navy.mil.

¹In this paper we use the words “macroscopic”, “continuum”, “hydrodynamic” and “fluid-based” more or less interchangeably. Thus our meaning for “hydrodynamic” is the usual one in other fields but differs from the very specific definition—energy-transport models—often given it in electron device modeling. In our terms, even the diffusion-drift description is “hydrodynamic”.

overview and general development of the macroscopic approaches to semiconductor electron transport theory as applied specifically to high field transport regimes.

Because of their computational advantages, fluid-based models of high-field transport have received considerable attention in the literature [2–7]. Now in all of this literature the view is maintained that *the* correct way of obtaining and generalizing such models is via microscopic derivation (phase-space averages) from the Boltzmann equation [8–10]. While this reductionist idea does have physical appeal and is obviously not without value, it does have a very significant drawback: Microscopic derivations can provide only *sufficient conditions* for the validity of macroscopic equations. That is, a microscopic derivation merely shows that a particular macroscopic description is a good approximation if a specific set of microscopic assumptions—which, because of the difficulties of the derivations, are often rather restrictive and of uncertain merit—is satisfied. It should be clear that the microscopic derivation says little about the *necessary conditions* for the validity of the macroscopic description and thus says little about its actual range of applicability. That the microscopically-derived theories are often applied in a phenomenological way, i.e., clearly beyond their supposed range of applicability, merely shows up the weakness of their method of derivation. This central defect of microscopic derivations has long been recognized in other disciplines (originally by Cauchy and others) and has given rise to an alternative approach for formulating macroscopic descriptions known as *classical field theory* [11]. The most important feature of the classical field theory approach is that it focuses directly on the consequences of broad physical principles of conservation and invariance. In this way, it develops general equations which are largely independent of microscopic assumptions and which must therefore hold over a wide range of circumstances. This approach may be viewed as complementary to the conventional one and, from a device engineering standpoint, we believe it to be more useful. Whatever the case, because the classical field theory approach is not widely appreciated in the electron transport modeling community (see Ref. 12 for some exceptions), it is deserving of attention. Presenting this approach and using it as a framework to organize and understand several fluid-based high field transport theories are the primary goals of this paper.

The foundational assumption of a fluid theory is the *continuum assumption* which says that a macroscopic theory is reasonable only when the discrete-

ness of the real physical system can be locally averaged out (in space and/or time) to produce meaningful *density variables*. Thus there must exist a hierarchy of space/time scales:

microscopic scale

≪ scale on which densities can be defined

≪ scale on which densities vary.

The first inequality is the continuum assumption needed to formulate the macroscopic description while the second defines the regime where this description is useful. Note that no explicitly microscopic requirement, e.g., an assumption of a Maxwellian distribution, enters this hierarchy. The limits set by these inequalities are not sharply defined but it is well-known that they tend to be more flexible than one might reasonably expect. (Moreover, in the case of low-energy electron gases the uncertainty principle tends to smooth out discreteness making continuum treatments (of probability densities) applicable at even smaller scales.) In any event, classical field theory assumes that this vague basis is solid, asserts that such densities may be defined and then develops the consequences of this assertion. It is viewed as the task of experiment and/or of microscopic theory to assess the soundness of this foundation. This basis in a set of primitive quantities (the densities) is both the chief drawback and the prime advantage of a classical field theory. On the one hand, it is a drawback because the primitives are “primitive”: They are assumed to be well-defined but this can never be verified *from within the theory*. On the other hand, it provides a simplest possible framework for developing a *consistent* theory in which no explicitly microscopic assumptions need be made.

Given the existence of the densities, e.g., of charge, momentum or energy, classical field theory develops various general conservation equations *constraining* the values of these densities. We say that they *constrain* the values of the densities because these equations (which must be satisfied), are not, in themselves, sufficient to fully *determine* the densities since there are more unknown densities than there are conservation equations. To close the system and allow problems to be solved, an additional set of equations known as *constitutive equations* or *material response functions* must be supplied. These equations describe “material” behavior, e.g., the equation of state of the electron gas, and typically have some range of applicability, e.g., the range over which the gas can be characterized as

ideal. Beyond this range a new constitutive equation must be devised which better describes the (non-ideal) material behavior. It is important to note that the applicability of macroscopic methods is *not* tied to the validity of a certain set of constitutive equations. That is, the balance laws remain valid no matter what the material response. With respect to this paper, should further work reveal our specific descriptions to be inadequate, it is the constitutive equations we devise which are most likely in need of adjustment and not the entire macroscopic framework.

It should not be surprising that there are significant similarities between the hydrodynamic descriptions we develop in this paper and various fluid-based models appearing in the literature. The main route of development of the latter was set by the work of Shur [2] who showed that by simultaneously solving momentum and energy balance equations with appropriate relaxation terms one could describe velocity overshoot. Much work along these lines followed [3–7] mostly based on the derivation of Bløtekjaer [9] and these types of models have come to be referred to as “energy transport models” or simply as “the hydrodynamic model.” In Sec. 2, we use classical field theory methods to develop a fluid description which is quite similar to these models. In fact, when particular constitutive choices are made, the two descriptions become identical. Because of this our discussion of this class of models is brief. More interestingly, in Sec. 3 we discuss models which are possible when heat conduction in the electron gas is insignificant. These models, while bearing some similarity to the phenomenological approach of Thornber [13], are new to the literature and are therefore deserving of more detailed discussion and simulation examples. Finally, a third class of fluid descriptions, multi-fluid models such as that introduced by Wilson [14], are also of interest but will not be discussed here because of space limitations. The best-known example of a multi-fluid model in an electron transport context is the standard diffusion-drift description when both electrons and holes are considered. Multiple-fluid models for high-field transport have the same underlying rationale: To describe situations in which there are several distinct (although interacting) populations of electrons with markedly different transport properties. Such models are often necessary for describing electron transport in III-V semiconductors because of their easily accessible satellite valleys. Because we do not discuss multi-fluid models in this paper, in our simulation examples in Sec. 3 we focus on high-field electron transport phenomena in silicon

and especially on describing velocity overshoot in this material.²

As a final introductory point, we wish to again emphasize that the primary goal of this work is to exhibit the unity and clarity provided by the classical field theory viewpoint. These attributes stem entirely from the simplicity of the approach, its focus on general principles and its clear separation of these principles from the specifics of a given “material’s” response. Not only does this simplicity aid theoretical understanding but it can also lead to new descriptions of important device phenomena. Such descriptions are generally well-suited to engineering applications of design and optimization because of their reduced computational burden. Overall, we expect that fluid-based descriptions like those developed here will remain of value for a large majority of current and future device technologies. Only when the discreteness of the electron begins to play a central role, e.g., as in Coulomb blockade devices, will macroscopic descriptions become entirely inappropriate.

2. HIGH-FIELD TRANSPORT DESCRIPTION WITH ENERGY BALANCE

We consider describing the flow of the population of conduction band electrons through a semiconductor as the flow of a single fluid through a solid (lattice). *The primitives of the single-fluid theory are therefore the quantities which define this electron fluid and its interaction with itself, with the lattice and with the electrostatic field at every point.* We assume an electron fluid with mass density, mn (where m is the free electron mass and n is the number density), and momentum density, $m\mathbf{nu}$ (where \mathbf{u} is the electron fluid velocity), which interacts *with itself* through a surface force density, \mathbf{F}^n , *with the lattice* through a body force density, \mathbf{F}^l , and *with the electrostatic field* through a body force density, \mathbf{F}^e . In addition, we assume the usual electrostatic field variables \mathbf{E} and \mathbf{D} and that the lattice has a fixed ionized impurity density of qN_D , where q is the charge on an electron. Considering an arbitrary fixed volume of semi-

²We do not discuss impact ionization or hot-carrier emission into a gate oxide although for silicon devices these high-field transport phenomena are generally more important than overshoot effects. Impact ionization can, under most circumstances, be included in a hydrodynamic theory (independently of whether energy balance is solved), however, hot-carrier emission can probably only be treated phenomenologically since only electrons in the extreme tail of the energy distribution are emitted.

conductor V with surface S and outwardly directed unit normal \mathbf{n} , the equation of mass balance for the electron gas can be written as

$$(2.1a) \quad \frac{\partial}{\partial t} \int_V mn \, dV = - \int_S \mathbf{n} \cdot mn\mathbf{u} \, dS.$$

In words this says that the time rate of increase of mass in V is balanced by the net inflow of mass through S . Since m and q are fixed, (2.1a) is also an expression of charge balance. Similarly, the equation for momentum balance is

$$(2.1b) \quad \begin{aligned} & \frac{\partial}{\partial t} \int_V mn\mathbf{u} \, dV \\ &= - \int_S \mathbf{n} \cdot (mn\mathbf{u}\mathbf{u}) \, dS \\ & \quad + \int_S \mathbf{F}^n \, dS + \int_V (\mathbf{F}^e + \mathbf{F}^l) \, dV, \\ &= - \int_S \mathbf{n} \cdot (mn\mathbf{u}\mathbf{u}) \, dS \\ & \quad + \int_S \mathbf{n} \cdot \tau^n \, dS - \int_V qn(\mathbf{E} + \mathbf{E}^n) \, dV, \end{aligned}$$

where in the second equality we have used the usual expression for the electrostatic force ($\mathbf{F}^e = -qn\mathbf{E}$), written the force of interaction of the fluid with itself in terms of a stress tensor, τ^n , ($\mathbf{F}^n = \mathbf{n} \cdot \tau^n$) and expressed \mathbf{F}^l in terms of a new quantity, the lattice force per unit charge, \mathbf{E}^n , ($\mathbf{F}^l = -qn\mathbf{E}^n$). In words (2.1b) says that the time rate of increase in momentum in V equals the net inflow of momentum through S plus the supply of momentum through surface and body forces. In order to satisfy conservation of angular momentum it is readily shown that the stress tensor, τ^n , must be symmetric. Finally, the electrostatic equations are familiar with the charge equation taking the form

$$(2.1c) \quad \int_S \mathbf{n} \cdot \mathbf{D} \, dS = \int_V q(N_D - n) \, dV,$$

and Faraday's law reducing to a statement that \mathbf{E} is conservative so that $\mathbf{E} \equiv -\nabla\varphi$, where φ is the electrostatic potential.

In addition to the above equations we have thermodynamic equations. For these, we define the following additional primitives: The specific internal energy densities of the lattice and the electron gas, ε^l and ε^n , two heat flux densities, \mathbf{q}^l and \mathbf{q}^n , two specific external energy sources, s^l and s^n , an en-

tropy density, η , and an entropy flux, Φ . The energy balance equation in the electron gas is then

$$(2.1d) \quad \begin{aligned} & \frac{\partial}{\partial t} \int_V \left(mn\varepsilon^n + \frac{1}{2} mn\mathbf{u} \cdot \mathbf{u} \right) dV \\ &= - \int_S \mathbf{n} \cdot \mathbf{u} \left(mn\varepsilon^n + \frac{1}{2} mn\mathbf{u} \cdot \mathbf{u} \right) dS \\ & \quad + \int_S \mathbf{n} \cdot (\tau^n \cdot \mathbf{u} - \mathbf{q}^n) \, dS \\ & \quad - \int_V qn\mathbf{u} \cdot \mathbf{E} \, dV + \int_V (mns^n - qn\mathbf{u} \cdot \mathbf{E}_r^n) \, dV, \end{aligned}$$

where we have assumed \mathbf{E}^n can be split into a sum of a recoverable or non-dissipative part (\mathbf{E}_r^n) and a dissipative (\mathbf{E}_d^n) part. Note that the former is subtracted from the S^n term since it does no work. Equation (2.1d) says that the time rate of change of the total energy (internal + kinetic) in V is equal to the rate of energy and heat inflow through S plus the rates of working of the various forces and the contribution from external energy sources. The similar energy balance equation for the lattice is

$$(2.1e) \quad \begin{aligned} & \frac{\partial}{\partial t} \int_V \rho\varepsilon^l \, dV = - \int_S \mathbf{n} \cdot \mathbf{q}^l \, dS \\ & \quad + \int_V \mathbf{E} \cdot \frac{\partial \mathbf{P}}{\partial t} \, dV \\ & \quad + \int_V (\rho s^l + qn\mathbf{u} \cdot \mathbf{E}_r^n) \, dV, \end{aligned}$$

where ρ ($\gg mn$) is the mass density of the lattice and \mathbf{P} ($\equiv \mathbf{D} - 4\pi\mathbf{E}$) is its polarization.

From the foregoing equations we can derive both the differential equations and boundary conditions of the field theory. The former follow when the necessary derivatives exist while the latter arise when the differentiability assumptions fail (at boundaries and interfaces) and are readily derived via "pillbox" arguments [11, 15]. For this paper, we omit further discussion of boundary conditions because we wish to focus solely on bulk transport issues. When the differentiability assumptions hold, applying the divergence theorem leads directly to the following

differential equations

$$(2.2a) \quad n_{,t} + \nabla \cdot (n\mathbf{u}) = 0,$$

$$(2.2b) \quad mn \frac{d\mathbf{u}}{dt} = -\nabla p^n + \nabla \cdot \tau_d^n - qn(\mathbf{E} + \mathbf{E}^n),$$

$$(2.2c) \quad \nabla \cdot \mathbf{D} = q(N_D - n),$$

$$(2.2d) \quad mn \frac{d\varepsilon^n}{dt} = -\nabla \cdot \mathbf{q}^n + \tau_d^n : \mathbf{d} + \frac{p^n}{n} \frac{dn}{dt} + qn\mathbf{E}_d^n \cdot \mathbf{u} + mns^n,$$

$$(2.2e) \quad \rho \frac{\partial \varepsilon^l}{\partial t} = -\nabla \cdot \mathbf{q}^l + qn\mathbf{E}_r^n \cdot \mathbf{u} + \mathbf{E} \cdot \frac{\partial \mathbf{P}}{\partial t} + \rho s^l,$$

where $d/dt \equiv \partial/\partial t + \mathbf{u} \cdot \nabla$, \mathbf{d} is the rate-of-deformation tensor defined by $\mathbf{d} \equiv \frac{1}{2}[\nabla\mathbf{u} + (\nabla\mathbf{u})^T]$ and we assume that τ^n (like \mathbf{E}^n) splits into a recoverable pressure ($-p^n\mathbf{I}$) and a dissipative viscous stress (τ_d^n). Adding (2.2d) and (2.2e) we then have the energy balance equation for the combined system

$$(2.2f) \quad \rho \frac{\partial \varepsilon^l}{\partial t} + mn \frac{d\varepsilon^n}{dt} - \mathbf{E} \cdot \frac{\partial \mathbf{P}}{\partial t} - \frac{p^n}{n} \frac{dn}{dt} - qn\mathbf{E}_r^n \cdot \mathbf{u} = \tau_d^n : \mathbf{d} - \nabla \cdot \mathbf{q} + qn\mathbf{E}_d^n \cdot \mathbf{u} + mns^n + \rho s^l,$$

where $\mathbf{q} = \mathbf{q}^n + \mathbf{q}^l$. Finally, following Ref. 16, we can write the rate of entropy production inequality as

$$(2.2g) \quad \rho \frac{d\eta}{dt} + \nabla \cdot \Phi - \frac{mns^n}{T^n} - \frac{\rho s^l}{T^l} \geq 0,$$

where T^l and T^n are the temperatures of the electron gas and lattice, respectively.

As indicated in the Introduction, the differential equations (2.2) and their associated boundary conditions may be viewed as a set of physical constraints on the possible solution functions. That is, the possible solution functions which satisfy these equations are physically plausible in that they satisfy the fundamental principles of charge, mass, momentum and energy balance and do not violate the second law of thermodynamics. However, as noted previously, these constraint equations are not sufficient to *determine* the solution functions because there

are fewer equations than variables. To complete the system, additional equations called *constitutive equations* or *material response functions* must be supplied. These express the particular material behavior under a given set of circumstances. By counting equations (10) and variables (34), one sees that 24 constitutive relations are required for the system (2.2). The form of the thermodynamic equations (2.2d-g) suggests (non-uniquely) that we should have constitutive equations for

$$(2.3a) \quad \mathbf{P}, p^n, \tau_d^n, \mathbf{E}^n, \varepsilon^n, \varepsilon^l, \mathbf{q}^n, \mathbf{q}^l, \eta, \Phi.$$

In general, these quantities will be determined by the *histories* of the electron fluid motion, of its temperature, of the lattice temperature and of the electrostatic field. Thus, in the general case these constitutive equations are *functionals* of the histories of

$$(2.3b) \quad \mathbf{E}, n, \mathbf{u}, T^n, T^l,$$

and possibly of their space derivatives. The forms of these functionals are restricted by the need not to violate the laws of balance and of thermodynamics. Developing the particular consequences of the latter (which is modified in the general case when splittings into recoverable and dissipative parts are not possible [11]) is the main focus of the subject of continuum thermodynamics [11, 15, 16]. For this paper, we omit all of this further development and note only that if T^l is assumed to be uniform and constant and we select

$$p^n = kT^n n, \quad \tau_d^n = 0,$$

$$\mathbf{E}^n = \mathbf{E}_d^n = \frac{\mathbf{u}}{\mu_{LF}} \frac{T^n}{T^l},$$

$$(2.4) \quad \varepsilon^n = \frac{3T^n}{2m^*}, \quad \mathbf{q}^n = -\frac{3}{2}D_{LF}nk\nabla T^n,$$

$$mns^n = -\rho s^l = -\frac{n}{\tau_w} \left[\frac{m^*}{2} \mathbf{u} \cdot \mathbf{u} + \frac{3k}{2}(T^n - T^l) \right],$$

where μ_{LF} and D_{LF} are the low-field mobility and diffusivity, respectively, and τ_w is an ‘‘energy relaxation time’’ which depends on T^n , then (2.2) becomes identical to a usual form of the energy transport model [9, 17]. (Note that in (2.4) m^* is an effective mass which did not appear in the earlier equations; see Sec. 3 for further discussion). As these equations have been widely studied we make no attempt here at an evaluation. We note only that

these equations, while often yielding reasonable results, are known to produce spurious overshoots under certain circumstances [18]. The likely cause is the usual one—a failure of the constitutive equations (2.4)—and some work has appeared which attempts to eliminate the problem by constitutive modifications [19].

3. HIGH-FIELD TRANSPORT DESCRIPTION WITHOUT ENERGY BALANCE

In the field theory of Sec. 2, the energy balance equations (2.2d) and (2.2e) allow one to solve for the heat conduction in both the electron gas and the lattice. That is, when the heat fluxes, \mathbf{q}^l and \mathbf{q}^n , are significant, the temperatures of the lattice (T^l) and of the electron gas (T^n) are dependent variables to be determined by solving the energy balance equations (2.2d) and (2.2e) as part of the boundary value problem for the electron transport. Now it is important to recognize that *when the heat conduction terms are negligible, (2.2d) and (2.2e) need not be solved explicitly*. The reason is perhaps most simply understood by noting that the above transport equations may be viewed as a hierarchy of moment equations. In this hierarchy it is the heat conduction terms which couple the energy equation to the next higher (third-order) moment. In order to truncate the hierarchy at the level of the energy equation one supplies an explicit constitutive equation for the heat conduction in terms of lower order variables and temperature, e.g., (2.4)₅. Implicit in this truncation is the assumption that the heat conduction can, in fact, be expressed in terms of lower-order variables, something that is true only if the coupling between the third and fourth-order moment equations is negligible. In a similar way, if heat conduction is negligible then one can truncate the system at the momentum level (first-order moment) simply by providing a constitutive equation for the stress tensor, the quantity which couples it to the second-order moment. Under this circumstance the energy balance equation contains no additional physical information and need not be solved. In this section, we examine in detail the case when heat conduction in the electron gas is unimportant so that (2.2d) need not be solved explicitly.

It should be emphasized that “hot” electron transport phenomena are not necessarily accompanied by significant electronic heat conduction. For example, in Fig. 1 we show the heat flux (in silicon) that accompanies a homogeneous electrical current

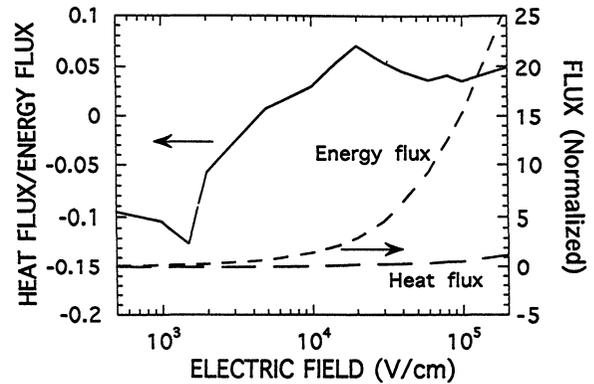


FIGURE 1 Homogeneous Monte Carlo simulation of the heat flux and total energy flux as a function of the applied electric field in silicon. Also shown is the ratio of these two fluxes. The heat flux is seen to be quite small over the entire range of homogeneous fields thus suggesting that a fluid theory without energy balance is often appropriate for silicon.

as computed from the Boltzmann equation using Monte Carlo simulation. The heat flux is seen to be a relatively small (under 10%) proportion of the total energy flux over almost the entire range of electric fields and is actually smaller (and positive [7]) under high-field conditions. (We remark that the proportionality between heat flux and energy flux suggested by Ref. 7 does not hold for our simulation). Thus we conclude that, at least in some high-field situations, heat flux is insignificant. And, as argued above, in such cases there is no need to solve the energy equation in order to describe the transport even though that transport may exhibit “hot”-carrier phenomena such as overshoot. This conclusion is contrary to the assertion of the conventional energy transport theory that in order to describe overshoot phenomena it is necessary to explicitly include differing momentum and energy relaxation times [2].

The question of whether heat conduction in the electron gas is significant or not is obviously situation-dependent and also partially depends on the meaning of “significant.” Under *homogeneous* field conditions, from Fig. 1 it appears that heat conduction is never very important in silicon. However, as emphasized in Ref. 7, conclusions based on homogeneous simulation need not carry over to the inhomogeneous case. For example, in an inhomogeneous situation where the electron distribution is “hot” but the average velocity is small, one might expect heat conduction effects to be larger and an energy balance solution therefore more necessary. Given these questions, perhaps the best basis for the work of this section is that we wish to investigate how well one can do in some important device situations *without*

explicitly solving the energy balance equation. From the point-of-view of device modeling this possibility is especially worthy of study insofar as it can reduce computational requirements.

A. Specification of the Material Response Functions

As in Sec. 2, the differential equations (2.2a–g) and their associated boundary conditions are physical constraints on the possible solution functions but are insufficient to actually determine the solution because there are fewer equations than variables. To complete the system, the “material” response must be quantified by supplying a set of constitutive equations. For the system without energy balance, we have 8 differential equations with 20 variables and hence 12 constitutive equations must be specified. These are equations for \mathbf{P} , \mathbf{E}^n , p^n and τ_d^n in terms of the histories of n , \mathbf{u} , \mathbf{E} and T^l and their space derivatives.

The constitutive equations for \mathbf{P} , \mathbf{E}^n , p^n and τ_d^n are not arbitrary but rather are subject to the restriction that they violate neither the balance laws nor the thermodynamic equations. Concerning the latter, because the heat conduction in the electron gas is negligible, we can regard the system as having a single thermodynamic temperature, $T \equiv T^l = T^n$. Again, this does not mean that we cannot describe “hot” electron behavior. In this case, that the energy balance equation for the combined system (2.2f) is separated into purely recoverable terms on the left side and purely dissipative terms on the right side implies that the second law of thermodynamics may be written immediately as [11, 15]

$$(3.1) \quad \rho \frac{\partial \varepsilon^l}{\partial t} + mn \frac{d\varepsilon^n}{dt} - \frac{p^n}{n} \frac{dn}{dt} - \mathbf{E} \cdot \frac{\partial \mathbf{P}}{\partial t} - qn \mathbf{E}_d^n \cdot \mathbf{u} = \rho T \frac{d\eta}{dt}.$$

Then combining (3.1) with (2.2g) yields the rate of entropy production inequality [11]

$$(3.2) \quad \frac{1}{T} \left[-\frac{\mathbf{q}}{T} \cdot \nabla T + qn \mathbf{E}_d^n \cdot \mathbf{u} + \tau_d^n : \mathbf{d} \right] \geq 0.$$

If, as is reasonable, we assume that ε^n is independent of \mathbf{P} , that ε^l is independent of n and \mathbf{u} , and that $E_r^n = -\chi d\mathbf{u}/dt$ [see discussion in connection with (3.8) below] then the form of (3.1) shows that $\varepsilon^n = \varepsilon^n(n, \eta, u^2)$ and $\varepsilon^l = \varepsilon^l(\mathbf{P}, \eta)$. Alternatively, by defining the thermodynamic state function $\psi \equiv$

$\varepsilon^l - \mathbf{E} \cdot \mathbf{P} - \eta T$, we can convert to a more convenient set of variables with $\varepsilon^n = \varepsilon^n(n, T, u^2)$ and $\psi = \psi(\mathbf{E}, T)$. Then, using the chain rule, (3.1) takes the form

$$(3.3) \quad \left(mn \frac{\partial \varepsilon^n}{\partial n} - \frac{p^n}{n} \right) \frac{dn}{dt} + \left(\eta + mn \frac{\partial \varepsilon^n}{\partial T} + \rho \frac{\partial \psi}{\partial T} \right) \frac{\partial T}{\partial t} + \left(\mathbf{P} + \rho \frac{\partial \psi}{\partial \mathbf{E}} \right) \cdot \frac{\partial \mathbf{E}}{\partial t} + \left(mn \frac{\partial \varepsilon^n}{\partial u^2} + \frac{1}{2} qn \chi \right) \frac{du^2}{dt} = 0.$$

Since all of the material time derivatives in (3.3) are independent and (3.3) holds for arbitrary dn/dt , dT/dt , $\partial \mathbf{E}/\partial t$ and du^2/dt their coefficients must vanish yielding the following *recoverable constitutive equations*

$$(3.4) \quad \mathbf{P} = -\frac{\partial \psi}{\partial \mathbf{E}}, \quad p^n = mn^2 \frac{\partial \varepsilon^n}{\partial n},$$

$$\eta = -\frac{\partial \psi}{\partial T} - mn \frac{\partial \varepsilon^n}{\partial T}, \quad \chi = -\frac{2m}{q} \frac{\partial \varepsilon^n}{\partial u^2}.$$

In addition to these recoverable constitutive equations there are also *dissipative constitutive equations* for \mathbf{E}_d^n , \mathbf{q} and τ_d^n . Thermodynamic information about these equations can be had from (3.2). In particular, (3.2) must hold for all possible values of \mathbf{E}_d^n , \mathbf{q} and τ_d^n and it therefore follows that

$$(3.5) \quad \mathbf{E}_d^n = \frac{\mathbf{u}}{\mu}, \quad \mathbf{q} = -\kappa \nabla T, \quad \tau_d^n = r_v \mathbf{d},$$

where the mobility, μ , thermal conductivity, κ , and viscosity coefficient, r_v , *must all be positive*. In general, these coefficients can depend on the histories of n , \mathbf{u} , \mathbf{E} and T and their space derivatives [11].

The constraints on the constitutive equations represented by (3.4) and (3.5) are the requirements imposed by the laws of thermodynamics on the behavior of any material (including an electron gas) which may be described by our assumed set of primitive quantities. Beyond this, it is presumed that any behavior is possible and it is up to experiment or perhaps microscopic calculation to say exactly which of the infinite number of possibilities is actually realized by the electron gas in a semiconductor.

Now, to introduce some of the issues which arise in further specifying the constitutive equations we first discuss a familiar example, the constitutive equation for the polarization \mathbf{P} in (3.4)₁, or, equivalently, the electric displacement \mathbf{D} . From (3.4)₁ and the definition of \mathbf{D} we have that $\mathbf{D} = \mathbf{D}(\mathbf{E}, T)$. In most cases it is sufficient to assume that this relationship is linear in \mathbf{E} (a linear dielectric), i.e.,

$$(3.6a) \quad \mathbf{D} = \varepsilon_d \mathbf{E},$$

where ε_d is the permittivity, a temperature-dependent *material coefficient* which characterizes the polarizability of the medium. But, of course, there are situations, e.g., at high frequencies, when (3.6a) is not a good model of the material response. In this case one might consider a “rate theory” such as

$$(3.6b) \quad \mathbf{D} = \varepsilon_d^1 \mathbf{E} + \varepsilon_d^2 \frac{\partial \mathbf{E}}{\partial t},$$

or, more generally, a linear functional (convolution integral)

$$(3.6c) \quad \mathbf{D} = \int_{-\infty}^t \varepsilon_d^r(t - \tau) \mathbf{E}(\tau) d\tau,$$

where ε_d^1 and ε_d^2 are material constants and $\varepsilon_d^r(t)$ is a time-dependent material response function. (We note that in the case of (3.6c), the thermodynamic equations are somewhat modified because the breakup into recoverable and dissipative parts becomes impossible [11]). Each of these possible constitutive equations is distinguished by the extent to which \mathbf{D} depends on the *history* of \mathbf{E} : (3.6a) has no history dependence, (3.6b) contains the most recent history (in a Taylor sense) and (3.6c) assumes dependence on the full history. For this paper, (3.6a) is obviously adequate, however, the example is instructive. First, a prime characteristic of the high field transport situations studied in this paper is history dependence (in other constitutive equations) and in describing this we will be led to consider constitutive equations analogous to (3.6b) and (3.6c). Secondly, there is the general point of classical field theory [11] that the constitutive theory is independent of the general balance equations: (2.2c) (or its electromagnetic version, Ampere’s law) holds no matter whether (3.6a), (3.6b) or (3.6c) is used. The latter are far less sure and are chosen simply on the basis that, for a certain material in a certain regime, they give an accurate representation of the material response. And if experiment or microscopic theory should show them inadequate, then some other better choice should be found. But as noted in the Introduction, such failure is not (usually) reason to

reject the entire macroscopic approach. With this in mind, we again wish to emphasize that in this paper when we suggest some specific constitutive choices which seem to have some range of utility, they should be viewed, at best, as the equivalent of (3.6a–c) and, in this sense, provisional.

We next consider the constitutive equation for the stress tensor in the electron gas. The recoverable part of the stress, p^n , is the equation of state of the gas and, from (3.4)₂, it should depend primarily on n . In the simplest case, at “low” densities (relative to the effective density of states) the equation for the pressure is that descriptive of an ideal gas

$$(3.7a) \quad p^n = kTn,$$

where T is the lattice temperature assumed constant and uniform. At higher densities, a Fermi-Dirac equation of state becomes more appropriate [20]. Since in this paper we treat only “low” density situations we shall use (3.7a) exclusively. We again note that the use of this equation should not be taken as an assertion that the electron temperature is the same as that of the lattice and that therefore description of “hot” electron phenomena is excluded; (3.7a) is merely an approximation for the equation of state of the electron gas which, if necessary, could be modified. Concerning the dissipative part of the stress tensor we use the general linear expression of fluid mechanics (Newtonian fluid)

$$(3.7b) \quad \tau_d^n = \lambda_v \mathbf{I} \nabla \cdot \mathbf{u} + 2\mu_v \mathbf{d},$$

where μ_v and λ_v are the shear and bulk viscosities, respectively. As in fluid mechanics we assume the Stokes condition, $\lambda_v + \frac{2}{3}\mu_v = 0$, and, in this case, positive entropy production demands simply that μ_v be positive.³ Now, in kinetic theory viscosity depends linearly on mean-free-path according to the formula (for a gas like the electron gas with no internal degrees of freedom) [21]

$$(3.7c) \quad \mu_v = \frac{ncm\Lambda}{2},$$

where c is the microscopic (thermal) velocity of the electrons and Λ is the mean free path. In using this microscopic formula in (3.7b) it is important to be aware of a difference between the physics of an

³The Stokes condition is assumed for convenience only. For the one-dimensional problems we treat in this paper this condition is never actually required so whether or not it is valid for the electron gas (in the high-field regimes of interest) is not an issue here.

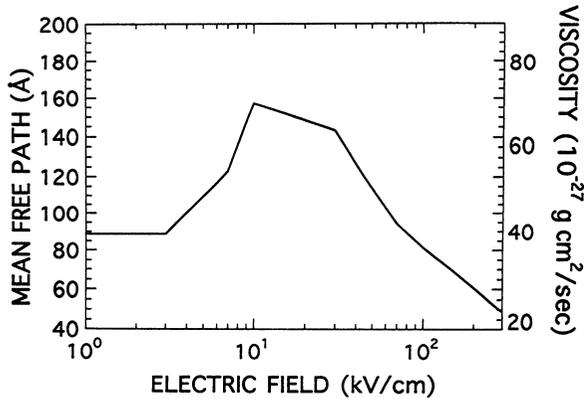


FIGURE 2 Electron mean free path as a function of electric field as obtained in Ref. 22. Also shown are the values of the kinematic viscosities which correspond to the mean free paths via (3.7c). These viscosities were used in the simulations of this paper.

electron gas in a solid and that of fluid mechanics. In the latter, the scattering is entirely between fluid particles so that Λ is inversely proportional to n and μ_v is therefore essentially constant. By contrast, in the electron gas Λ is generally determined by scattering other than electron-electron scattering and is therefore largely independent of density (except at high densities). As a result, for the electron gas in a semiconductor, μ_v is essentially proportional to n . And consequently, in electron transport applications the *kinematic viscosity*, $\nu_v \equiv \mu_v/n$, is closer to being a constant and hence is a better viscosity parameter to work with. Nonetheless, even the kinematic viscosity is not generally constant. In high-field transport the mean free path is known to vary significantly with electric field [22] as shown in Fig. 2 and this causes the kinematic viscosity to vary also via (3.7c) (Fig. 2). For this paper we assume the kinematic viscosity to be that given in Fig. 2. Our expectation is that this gives the correct viscosity to within, say, a factor of two. Finally, we note that viscosity terms are often [9, 17] (but not always [6]) neglected in conventional fluid-based high-field transport descriptions. On the basis of numerical evidence given below, we find that this neglect is unjustified and that the expectation of Ref. 6 that such effects “are likely to be more pronounced at the shorter device lengths” (p. 64) is in fact true (see also, Ref. 23).

The final constitutive equation requiring specification is that for the interaction force, \mathbf{E}^n , between the electron fluid and the lattice. In the simplest case of diffusion-drift theory this force is purely dissipative and is given by (3.5)₁. However, for the high field transport situations of interest here this is inadequate. Specifically, history dependence be-

comes important because the electron-lattice interaction or “scattering” (including Bragg reflection) does not take place instantaneously. Thus, one general form for \mathbf{E}^n would be as a linear functional as in (3.6c). Alternatively, we might consider a “rate theory” like (3.6b) with

$$(3.8) \quad \mathbf{E}^n = \frac{\mathbf{u}}{\mu} - \chi \frac{d\mathbf{u}}{dt}.$$

The sign of the rate term ($\chi > 0$) is set by the fact that \mathbf{E}^n should depend on the velocity at previous rather than future times (causality). As indicated previously the velocity-dependent term is dissipative while the first rate term is recoverable and of the form asserted earlier. Inserting (3.8) into (2.2b) results in

$$(3.9) \quad (m - q\chi) n \frac{d\mathbf{u}}{dt} = -\nabla p^n + \nabla \cdot \tau_d^n - qn \left(\mathbf{E} + \frac{\mathbf{u}}{\mu} \right),$$

from which we see that the first rate term is indeed purely recoverable, acting as an inertia and modifying the mass of the electron. Thus the electron-lattice interaction acts to produce an *effective mass* in the continuum description just as it does in semi-classical electron dynamics. We are therefore led to define the effective mass as $m^* \equiv m - q\chi$. Note that m^* is *smaller* than m because of causality ($\chi > 0$). The rest of \mathbf{E}^n originates in collisions and is purely dissipative from (3.5). Finally, we remark that non-parabolicity effects are expressed in the theory when χ is not constant, a topic not discussed further in this paper.

The form chosen for the dissipative part of \mathbf{E}^n —in effect, a functional form for the mobility μ —is an essential ingredient of the high-field description. Rather than attempt to devise a complicated form which effectively “curve-fits” various Monte Carlo simulations we instead have chosen to use a mobility model taken from the literature (for silicon) in order to see what is possible with simple assumptions. The model we use is [24]

$$(3.10a) \quad \mu = \frac{2\mu_{LF}}{1 + \sqrt{1 + \frac{\mu_{LF}E}{\sqrt{2}u_{sat}} + 4\left(\frac{\mu_{LF}E}{u_{sat}}\right)^2}},$$

where u_{sat} is the saturation velocity (1.1×10^7 cm/sec) and E is the magnitude of \mathbf{E} . The electric field dependences in this formula serve to provide velocity saturation [25]. For the high field regimes of

interest in this paper we are interested in situations with rapid variations and as usual this means that μ must become history dependent to reflect the non-instantaneous nature of scattering. For this history dependence we again use a rate theory like (3.6b), simply replacing the E dependence of μ by a dependence on

$$(3.10b) \quad \left| \mathbf{E} - \tau_r \frac{d\mathbf{E}}{dt} \right|,$$

where τ_r is an adjustment time which measures (roughly) how far into the past the electric field history has influence on the mobility. We find that a “good” value for the adjustment time of the flow, τ_r , is about 0.12 psec in silicon at room temperature. Lastly, we note that if one is interested only in steady-state, $d\mathbf{E}/dt$ may be replaced in (3.10b) by its steady-state value $\mathbf{u} \cdot \nabla \mathbf{E}$.

As a final remark for this subsection, we reiterate that the constitutive equations (3.7)–(3.10) needed for the theory are less reliable than the remainder of the field theory and may not be robust when applied in other situations, e.g., under transient conditions. This is especially true of (3.10a) which is of interest only because of its simplicity and its connection to previous work. We again emphasize that in this paper our main purpose is exposition of general principles and not devising optimal material response functions. In future work, we hope to make detailed comparisons with many Monte Carlo simulations and either confirm or correct the constitutive theory as specified above. For now, with the specification of (3.6a), (3.7), (3.8) and (3.10a) with (3.10b), boundary conditions) of the single-fluid theory of electron transport in semiconductors at high fields.

B. Connections With Other Theories

The theory developed above reduces directly to standard diffusion-drift theory under certain conditions. In particular, if the electron gas is ideal [(3.6a)], the history dependence of the electron-lattice interaction is negligible [(3.10a) without (3.10b)] and inertia and viscosity are unimportant, then (3.9) takes the form of the diffusion-drift equation

$$(3.11) \quad n\mathbf{u} = -n\mu\mathbf{E} - D\nabla n,$$

with $D = kT\mu/q$. Thus, when these conditions are met in a transport simulation, e.g., in a large-device, our single-fluid theory reduces to the traditional modeling approach which is known to be accurate in these regimes. This is, of course, as it should be.

The connection between the theory of this section and existing fluid-based models for high-field transport regimes is more tenuous. Of these models, we focus here on the relationship to the phenomenological approach proposed by Thornber [13]. Thornber’s idea was that one could describe high field transport simply by adding additional terms to the diffusion-drift equation (3.11). With his specific additions one has the so-called augmented diffusion-drift equation [26]

$$(3.12) \quad n\mathbf{u} = -n\mu\mathbf{E} - Dn_{,x} + BnE_{,t} + WnE_{,x},$$

where the subscripts indicate partial differentiation and μ , D , B and W are supposed to be functions of the electric field E and are to be determined from Monte Carlo simulation. Obviously this equation is appropriate only for one-dimensional problems although, if it worked well, the generalization to three dimensions would probably be straightforward [27].

There are both important similarities and differences between our single-fluid theory and (3.12). The first and most fundamental similarity is that use of (3.12) represents an assertion that one need not solve the energy equation in order to describe “hot” electron effects. A second important similarity is that the terms added by Thornber serve to introduce rate effects into the diffusion-drift description and, as we have emphasized, such history dependence is a key feature of high-field transport. Other than this, however, Thornber’s approach would seem to be quite limited and indeed this has been found in numerical simulation [28]. From a theoretical point-of-view the most serious flaw in Thornber’s approach is that it is purely *ad hoc* (although attempts at justification from the Boltzmann equation have appeared [29]). That is, it stems from the widely-held but misleading notion that the diffusion-drift equation is a constitutive equation for the current so that one can add to it whatever terms seem appropriate to fit the physics. But as the derivation of (3.11) from (2.2b) shows, the current equation is in fact an approximate expression of momentum balance in the electron gas and, if this is not taken into account, one mixes physical principles (balance laws) and material response (constitutive equations). The “material” coefficients of the theory, e.g., B and W , are then at best composite quantities mixing properties of the fluid such as inertia with properties of the electron-lattice interaction such as mobility. And, in a worst case, these coefficients need not even be solely dependent on the materials but can depend on extraneous factors such as geometry. As we have stressed, one of the key advantages of the classical field theory approach is the clear separation that is

made between general physical principles and specific material response.

A second flaw in (3.12) is that, even if regarded as a constitutive equation, it does not obey fundamental macroscopic principles. In particular, if one wishes to introduce time/space derivatives into such an expression, rotational invariance demands that these be in the form $dE/dt \equiv E_{,t} + uE_{,x}$ and not as in (3.12). The further flaws of the Thornber approach in actual usage have been previously analyzed [28] and we therefore do not discuss this further. Nevertheless, it should again be said that the Thornber approach does recognize that solving energy balance explicitly may not be necessary and it tries to capture the essential history-dependent character of high-field transport. In this sense, our work in this section may be viewed as providing a physically well-founded version of Thornber's *ad hoc* description.

C. Simulation Examples

The single fluid model is best suited to the description of transport in a semiconductor like silicon where the transport is usually confined to one (six-fold degenerate) valley. When transport in other valleys with markedly different transport properties becomes significant the single fluid model (with any constitutive theory) will become increasingly inaccurate. For this reason, as noted in the Introduction we focus on modeling high-field transport in silicon. The particular device we examine is the standard one for such studies, a one-dimensional $n^+ - n - n^+$ diode. We treat equilibrium and steady-state problems only; transient and ac problems will be investigated in future work.

The boundary value problem appropriate for the $n^+ - n - n^+$ diode is as follows. Combining (2.2a-c) with (3.6a), (3.7a, b), (3.8) and (3.10a, b) we have the differential equations to be solved inside the device

$$(3.13a) \quad n_{,t} + (nu)_{,x} = 0,$$

$$(3.13b) \quad m^* n \frac{du}{dt} = -kTn_{,x} + qn\varphi_{,x} + \frac{4}{3}(nv_v u_{,x})_{,x} - \frac{qnu}{\mu},$$

$$(3.13c) \quad \varphi_{,xx} = \frac{q}{\epsilon_d} [n - N_D(x)],$$

where m^* is $0.26m$, ν_v is the kinematic viscosity

which depends on E as shown in Fig. 2 and the function $N_D(x)$ is the prescribed donor profile through the device. The boundary conditions for this problem are applied at the metal contacts to the n^+ regions and are that φ is prescribed at both contacts, $u_{,x}$ is taken to be zero at both contacts, n is assumed known at the cathode and $n_{,x}$ is zero at the anode. We note that because viscosity is included, these conditions do not over-determine the problem as was the case in the work of Gardner [30].

The problem is now fully specified and its solution obviously requires numerical methods. For this purpose, we have used the flux-corrected transport algorithm [31], a powerful technique for solving systems of conservation equations in the presence of steep gradients, e.g., shocks. This monotone method has been widely used in fluid dynamics but, to our knowledge, has not been applied to semiconductor transport problems heretofore (although a related approach has been used in Ref. 32). Discussion of the details of this method and its application to semiconductor transport is outside the scope of the present work and will be left for future publication.

For this paper we have performed a number of simulations for $n^+ - n - n^+$ diodes with three different base widths of $2\mu\text{m}$, $0.4\mu\text{m}$ and $0.1\mu\text{m}$. In each diode the n^+ cathodes and anode regions were doped at $5 \times 10^{17} \text{ cm}^{-3}$ while the n -type base regions were doped at $5 \times 10^{15} \text{ cm}^{-3}$. The large diode was simulated just to confirm that at these geometries the theory does indeed give the expected diffusion-drift behavior and no results for this diode will be shown. For the smaller diodes comparisons were also made with diffusion-drift simulations which used precisely the same mobility model (3.10a) but, of course, without rate effects (3.10b). In addition, we compared the fluid simulations with self-consistent Boltzmann/Monte Carlo simulation results reported in Refs. 33 and 34.

We discuss first the $0.4\mu\text{m}$ diode whose donor concentration and self-consistent electric field profiles are shown in Fig. 3. In this Figure, the peaks in the electric field associated with the carrier spillover at the cathode and anode are evident. The velocity profile for this diode under a bias of 1.5V and computed by solving (3.13) in steady-state is plotted in Fig. 4 [labeled "rate" to indicate that the rate model (3.10b) has been used]. For comparison, velocity profiles computed using diffusion-drift theory (labeled DD) and by Monte Carlo (labeled MC) are also shown. Comparing the latter two profiles we see that diffusion-drift theory is entirely inadequate. The reason is that diffusion-drift theory does not include the essential history dependence of the scat-

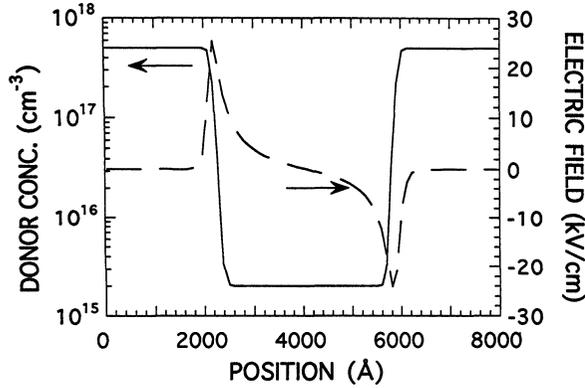


FIGURE 3 Donor concentration and electric field profile in the $0.4\mu\text{m } n^+ - n - n^+$ diode in equilibrium.

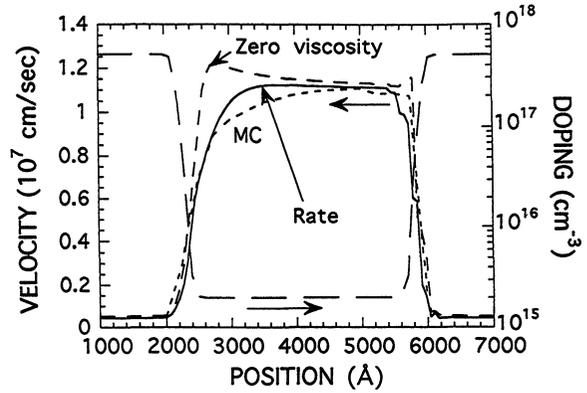


FIGURE 5 Velocity profile in the $0.4\mu\text{m}$ diode under a bias of 1.5V showing the rate theory result (labeled "Rate"), the Monte Carlo result of Ref. 34 (labeled "MC") and the rate theory result when viscosity is neglected (labeled "Zero viscosity"). The latter differs considerably from the result with viscosity included (especially in the low field region) showing the importance of this term in the theory.

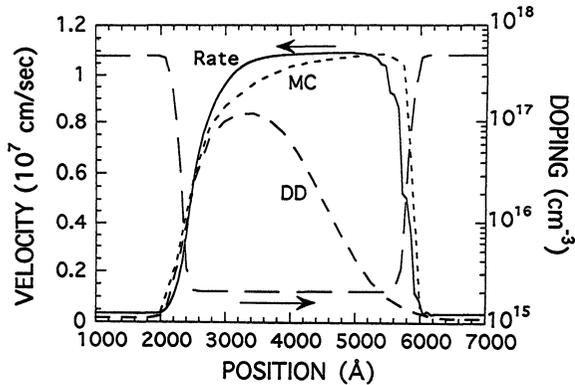


FIGURE 4 Velocity profiles as computed using the rate theory of Sec. 3 (labeled "Rate"), using diffusion-drift theory with the mobility model in (3.10a) (labeled DD) and a Monte Carlo result taken from Ref. 34 (labeled MC). The bias was 1.5V. The rate theory is seen to give good agreement with the Monte Carlo simulation while the diffusion-drift result is completely in error.

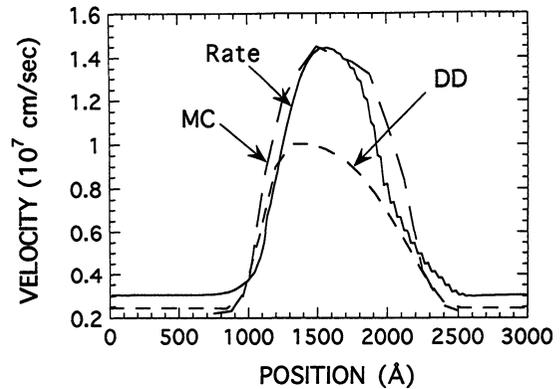


FIGURE 6 Same plot as Fig. 4 except for the $0.1\mu\text{m}$ diode (biased at 1V) with the Monte Carlo result taken from Ref. 33. The rate theory continues to do quite well in comparison with the microscopic simulation whereas the diffusion-drift result is even further off.

tering through (3.10b) and the effects of viscosity through (3.7b). (It also does not include inertia but the effect of this omission is unimportant as it effects only the rise and fall distances and then only slightly). When these effects are taken into account by solving (3.13), agreement with the Monte Carlo simulation is greatly improved although it is not perfect. To see the particular effect of viscosity in Fig. 5 we again show the full rate theory result (including the viscosity of Fig. 2) and the Monte Carlo result together with the result of a rate theory simulation in which viscous term has been set to zero. The viscosity is seen to have a significant effect especially in the low field region where the mean-free-path is longest and the viscosity largest.

The $0.1\mu\text{m}$ diode represents a more extreme test of the rate theory and especially of the simple electron-lattice interaction model represented by

(3.10a) and (3.10b). Nonetheless, we find remarkably good agreement with Monte Carlo simulation. In Fig. 6 we show the velocity profiles through the diode (the cathode edge is at 1000\AA and the anode edge is at 2000\AA) as computed with the rate theory, by Monte Carlo simulation and in the diffusion-drift approximation using (3.10a). The diffusion-drift result is even further off than in the longer device while the rate theory again does quite well. The largest errors seen in the rate theory result are in getting the exact locations of the rise of the velocity at the cathode and its fall at the anode. As a final plot we show the important influence of the history or rate dependence (3.10b) on these results. In

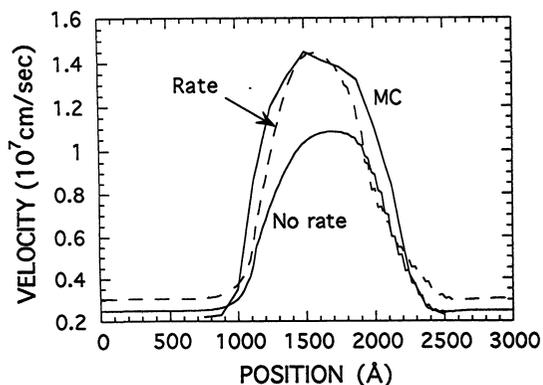


FIGURE 7 Similar plot to Fig. 6 showing the importance of history dependence in the mobility model. When this dependence is neglected ("no rate" curve) the disagreement with the Monte Carlo result is considerable.

Fig. 7, we have plotted the same rate theory and Monte Carlo simulation results along with the result when the adjustment time τ_r is set to zero. The latter is seen to be in sharp disagreement, actually being fairly close to the diffusion-drift result.

Overall, in these simulations we see that the simplest possible constitutive theory based on (3.7) and (3.10) gives quite good agreement with Monte Carlo simulation for velocity overshoot even in an ultra-small $0.1\mu\text{m}$ diode. The most important additions needed in the theory are seen to be the inclusion of rate effects in the mobility model [(3.10b)] and the inclusion of viscous effects [(3.7b)]. Although the agreement is not perfect (especially in the rise and fall of the velocity) these results suggest that with minor modification to the constitutive theory [especially (3.10)] agreement could be further improved. Efforts to develop a better mobility model for this purpose are currently underway. In any event, we conclude that a transport theory in which energy balance is not solved has promise as an approach for high-field transport simulation.

4. CONCLUSIONS

In this paper we have examined the foundations of macroscopic approaches to high-field electron transport description from the standpoint of classical field theory. This point-of-view is complementary to existing developments of such descriptions from the Boltzmann equation and we believe it provides insight into modeling methods for high-field transport problems. As we have stressed, the key simplification of the classical field theory formulation is its clear separation of underlying physical principles

from specific material response. This permits the transport theory to be understood in its simplest form without any discussion of explicit microscopic assumptions. The latter which are a central part of Boltzmann derivations can provide only *sufficient* conditions for the validity of the macroscopic description and therefore give little indication of *range* of applicability. To illustrate the classical field theory approach we presented two basic descriptions which are distinguished by whether or not the heat flux accompanying the electron flow is appreciable or not. When it is, a coupled energy balance equation must be solved for the associated temperature fields and the description is akin (if not identical) to the conventional energy transport description. In the more interesting case when the heat flux is negligible (a circumstance which frequently holds) the energy equation is no longer independent of the mechanical equations and need no longer be solved explicitly. In this case, we exhibited using numerical simulation how this simpler theory can describe velocity overshoot in a physically well-founded way and without needing to define an electron temperature. Our approach here may be viewed as providing a physical foundation for methods like Thornber's augmented diffusion-drift theory. And finally we believe the rational basis provided by this work can further the effort to develop succinct and computationally-efficient transport models needed for current and future semiconductor device analysis and engineering.

Acknowledgments

The author would like to thank Dr. C.R. Devore for his invaluable assistance with the flux-corrected transport algorithm, Mr. P. Dimtruk and Dr. J. Sôbehart for generously providing the Monte Carlo code, Ms. C.R. Brennan for her encouragement of this work and Prof. H.F. Tiersten for his insightful comments (over many years) on the classical field theory foundations. The author also thanks Dr. G.D. Doolen and the Center for Nonlinear Studies at Los Alamos for providing an environment conducive to the completion of this work and the Office of Naval Research for funding support.

References

- [1] C. Jacoboni and L. Reggiani, "The Monte Carlo method for the solution of charge transport in semiconductors with applications to covalent materials," *Rev. Mod. Phys.* 55, 645 (1983).
- [2] M. Shur, "Influence of nonuniform field distribution on frequency limits of GaAs field-effect transistors," *Elect. Letts.* 12, 615 (1976).
- [3] W.R. Curtice and Y.H. Yun, "A temperature model for the GaAs MESFET," *IEEE Trans. Elect. Dev.* ED-28, 954 (1981).

- [4] R. Cook and J. Frey, "Two-dimensional numerical simulation of energy transport effects in Si and GaAs MESFETs," *IEEE Trans. Elect. Dev.* ED-29, 970 (1982).
- [5] P.A. Sandborn, A. Rao and P.A. Blakey, "An assessment of approximate nonstationary charge transport models used for GaAs device modeling," *IEEE Trans. Elect. Dev.* 36, 1244 (1989).
- [6] H.L. Grubin, "Transport and material considerations for submicron devices," in *The Physics of Submicron Semiconductor Devices*, eds. H.L. Grubin, D.K. Ferry and C. Jacoboni, Plenum, New York (1989).
- [7] S.-C. Lee and T.-W. Tang, "Transport coefficients for a silicon hydrodynamic model extracted from inhomogeneous Monte Carlo calculations," *Solid-St. Electron.* 35, 561 (1992).
- [8] R. Stratton, "Diffusion of hot and cold electrons in semiconductor barriers," *Phys. Rev.* 126, 2002 (1962).
- [9] K. Bløtebjerg, "Transport equations for electrons in two-valley semiconductors," *IEEE Trans. Elect. Dev.* ED-17, 38 (1970).
- [10] R. Thoma, A. Emunds, B. Meinerzhagen, H.-J. Peifer and W.L. Engl, "Hydrodynamic equations for semiconductors with nonparabolic band structure," *IEEE Trans. Elect. Dev.* 38, 1343 (1991).
- [11] C. Truesdell and R. Toupin, *The Classical Field Theories*, Handbuch der Physik, Vol. III/1, Springer-Verlag, Berlin (1960); C. Truesdell and W. Noll, *The Non-linear Field Theories of Mechanics*, Handbuch der Physik, Vol. III/3, Springer-Verlag, Berlin (1965).
- [12] A.M. Anile and S. Pennisi, "Thermodynamic derivation of the hydrodynamical model for charge transport in semiconductors," *Phys. Rev.* B46, 13186 (1992); Y. Ohno, "Thermodynamic treatment of hot-carrier transport," *Semicond. Sci. Tech.* 7, B375 (1992).
- [13] K.K. Thornber, "Current equations for velocity overshoot," *IEEE Elect. Dev. Lett.* EDL-3, 69 (1982).
- [14] C.L. Wilson, "Hydrodynamic carrier transport in semiconductors with multiple band minima," *IEEE Trans. Elect. Dev.* 35, 180 (1988).
- [15] A.C. Eringen, *Mechanics of Continua*, Krieger, Huntington, New York (1980); S.R. deGroot and P. Mazur, *Non-equilibrium Thermodynamics*, North-Holland, Amsterdam (1962).
- [16] N. Dunwoody and I. Müller, "A thermodynamic theory of two chemically reacting ideal gases with different temperatures," *Arch. Rat. Mech. Anal.* 31, 344 (1968).
- [17] G. Baccarani and M.R. Wordeman, "An investigation of steady-state velocity overshoot effects in Si and GaAs devices," *Solid State Electron.* 28, 407 (1985).
- [18] M.R. Pinto, W.M. Coughran, Jr., C.S. Rafferty, R.K. Smith and E. Sangiorgi, "Device simulation for silicon ULSI," in *Computational Electronics*, eds., K. Hess, J.P. Leburton and U. Ravaioli, Kluwer, Boston (1990), p. 3.
- [19] D. Chen, E.C. Kan, U. Ravaioli, C.-W. Shu and R.W. Dutton, "An improved energy transport model including nonparabolicity and non-Maxwellian distribution effects," *IEEE Elect. Dev. Letts.* 13, 26 (1992).
- [20] V.C. Aguilera-Navarro, G.A. Estevez and A. Kostecki, "A note on the Fermi-Dirac integral function," *J. Appl. Phys.* 63, 2848 (1988).
- [21] W.G. Vincenti and C.H. Kruger, Jr., *Introduction to Physical Gas Dynamics*, Wiley, New York (1965).
- [22] M.V. Fischetti and S.E. Laux, "Monte Carlo analysis of electron transport in small semiconductor devices including band structure and space-charge effects," *Phys. Rev.* B38, 9721 (1988).
- [23] Y. Ohno, IEDM Tech. Digest, 319 (1989).
- [24] D. Roychoudhury and P.K. Basu, "A new mobility-field expression for the calculation of MOSFET characteristics," *Solid-St. Electron.* 19, 656 (1976).
- [25] S. Selberherr, *Analysis and Simulation of Semiconductor Devices*, Springer-Verlag, New York (1984).
- [26] J.A. Cooper, D.F. Nelson, S.A. Schwarz and K.K. Thornber, "Carrier transport at the Si-SiO₂ interface," in *VLSI Electronics: Microstructure Science*, 10, ed., N. Einspruch, Academic, New York (1985), Ch. 7.
- [27] I.C. Kizilyalli and M. Artaki, "Simulation of a GaAs MESFET including velocity overshoot: An extended drift-diffusion formalism," *IEEE Elect. Dev. Letts.* 10, 405 (1989).
- [28] M. Artaki, "Hot-electron flow in an inhomogeneous field," *Appl. Phys. Lett.*, 52, 141 (1988) and P.A. Blakey, S.A. Burdick and P.A. Sandborn, "On the use of Thornber's augmented drift-diffusion equation for modeling GaAs devices," *IEEE Trans. Elect. Dev.* 35, 1991 (1988); D. Chen, E.C. Kan and U. Ravaioli, "An analytical formulation of the length coefficient for the augmented drift-diffusion model including velocity overshoot," *IEEE Trans. Elect. Dev.* 38, 1484 (1991).
- [29] P.J. Price, "On the flow equation in device simulation," *J. Appl. Phys.* 63, 4718 (1988).
- [30] C.L. Gardner, J.W. Jerome and D.J. Rose, "Numerical Methods for the Hydrodynamic Device Model: Subsonic Flow," *IEEE Trans. Comp.-Aided Design CAD-8*, 501 (1989).
- [31] J.P. Boris and D.L. Book, "Flux-corrected transport. SHASTA, A fluid transport algorithm that works," *J. Comp. Phys.* 11, 38 (1973); S.T. Zalesak, "Fully multidimensional flux-corrected transport algorithms for fluids," *J. Comp. Phys.* 31, 335 (1979).
- [32] E. Fatemi, J. Jerome and S. Osher, "Solution of the hydrodynamic device model using high-order non-oscillatory shock capturing algorithms," *IEEE Trans. Comp.-Aided Design CAD-10*, 232 (1991).
- [33] D. Chen, E. Sangiorgi, M.R. Pinto, E.C. Kan, U. Ravaioli and R.W. Dutton, "Analysis of spurious velocity overshoot in hydrodynamic simulations," proceedings of NUPAD IV, p. 109 (1992).
- [34] E.C. Kan, U. Ravaioli and T. Kerkhoven, "Calculation of velocity overshoot in submicron devices using an augmented drift-diffusion model," *Solid-St. Electron* 34, 995 (1991); A. Gnudi, F. Odeh and M. Rudan, *Eur. Trans. Telecom. Relat. Technol.* 1, 307 (1990).

Biography

MARIO ANCONA is on the staff of the Naval Research Laboratory, Washington, DC. He is a member of IEEE and APS and his research interests include electronic device and process simulation, transport physics and numerical methods.

