PHYSICAL REVIEW B VOLUME 35, NUMBER 15 15 MAY 1987-II

Macroscopic physics of the silicon inversion layer

M. G. Ancona
Code 6813, Naval Research Laboratory, Washington D.C. 20375

H. F. Tiersten
Department of Mechanical Engineering, Rensselaer Polytechnic Institute, Troy, New York 12181

(Received 29 September 1986)

The diffusion-drift description of electrons and holes in a semiconductor is frequently used to obtain a detailed understanding of the physics and engineering of semiconductor devices. We show that, by generalizing the equation of state of the electron gas to include density-gradient dependences, this standard description can be extended to describe much of the quantum-mechanical behavior exhibited by strong inversion layers.

I. INTRODUCTION

For a detailed understanding of the physics and engineering of semiconductor devices a macroscopic (continuum) description involving the diffusion-drift current equations is often used. While this description is frequently employed to simulate transport along a strong inversion layer at the Si-SiO$_2$ interface, it is widely believed that for accurate modeling a quantum-mechanical description is required. In the present paper, we show this view to be incorrect by exhibiting how the macroscopic description can be extended to analyze the silicon inversion layer. In doing so, a description of a semiconductor in which a strong inversion layer is present is provided which is closely akin to that frequently employed in the numerical simulation of semiconductor devices.

The primary limitations of a physical description such as that involving the diffusion-drift current equations arise from the fact that it is a continuum description. That is, it is a description which involves quantities which are averages of microscopic variables over arbitrary volumes large compared to the spacing between discrete elements yet small with respect to the spatial scales of interest. The range over which such volumes exist defines the extent of validity of the continuum theory. Thus, for example, in the bulk of a semiconductor doped at 1×10^{15} cm^{-3}, use of the diffusion-drift description is questionable when spatial variations not much larger than 100 nm $[(1\times10^{-15})^{1/3}$ cm$]$ are of interest. Similarly, in a strong inversion layer (where doping is unimportant) with a typical electron density of 1×10^{18} cm^{-3}, a continuum theory is applicable only over space dimensions large compared to 10 nm. Because the inversion layer itself is on the order of 10 nm thick, it follows that the usual diffusion-drift description of the strong inversion layer violates its fundamental continuum assumption and is therefore invalid.

The violation of continuum assumptions by the standard diffusion-drift description when applied to the strong inversion layer suggests that no macroscopic description of the strong inversion layer is possible, i.e., a microscopic theory is essential.[1] However, by weakening the continuum assumptions appropriately, it is in fact possible to extend the range of the macroscopic description into this regime. In particular, instead of allowing the volumes over which microscopic averages are taken to be of arbitrary shape, we restrict the class of averaging volumes to be "pancakelike" regions lying in the plane of the inversion layer. In this way, the average spacing between electrons projected along the direction normal to the larger faces of the "pancake," is reduced thereby permitting continuum description of more rapid spatial variations in the normal direction. The price of this increased normal resolution is a decrease in the ability of the theory to describe rapid variations in the plane of the inversion layer. We shall refer to the assumption involved as a planar-averaged continuum assumption and, as an illustration, the usual application of the diffusion-drift description to the strong inversion layer[2] can be viewed as an implicit use of a planar-averaged continuum theory. Whether a generalized theory of this type is sensible and the extent to which one can push the normal resolution are set by the spatial scales of interest both perpendicular and parallel to the interface. For our work, a normal resolution of 1 nm is appropriate. This implies that, for an electron density of 1×10^{18} cm^{-3}, the lateral resolution relaxes to 32 nm.

Given that a generalized continuum theory may be constructed, there is then the question of whether such a description can account for the important quantum effects. As noted above, there is some doubt on this point in the literature; it is known, for example, that the diffusion-drift description in its usual form is inadequate in this regard.[2] Nonetheless, the existence of numerous macroscopic theories which successfully describe quantum phenomena, e.g., London's theory of superconductivity, belie the general statement. In the language of continuum physics, the reason the diffusion-drift description fails is not that its underlying balance laws are violated but rather that its constitutive theory is inadequate. Specifically, as we show here, the electrons in the inversion layer no longer act as the simple Maxwell gas assumed by the usual diffusion-drift description.

By permitting a more general equation of state for the electron gas, which contains a dependence on the density gradient as well as the density, much of the behavior of the strong inversion layer can be described. In particular, we demonstrate that for such a gas in static situations (no current flow), Newton's second law takes the form of a macroscopic "Schrödinger equation." This is obviously no longer a classical description—it involves a coefficient which would vanish if Planck's constant were zero—but it is not a fully quantum mechanical description either: It contains no phase information as this is lost in the continuum averaging. Consequently, there are transport phenomena, e.g., Aharonov-Bohm oscillations, which are beyond the range of this type of description. Nonetheless, within its range of applicability, the macroscopic theory has significant advantages. First, the description as formulated here is constructed so as to always satisfy the fundamental macroscopic laws of physics; there are no microscopic assumptions of questionable or uncertain validity involved. Secondly, the equations of electrostatics are coupled directly into the partial differential system rather than requiring a self-consistent iteration procedure. Also, boundary conditions are well-defined in the macroscopic approach, whereas microscopically boundary or interface effects arise in volumetric regions which themselves have quantum-mechanical descriptions. Finally, and perhaps most importantly, dissipative effects (scattering) are readily included in the approach developed here. In contrast, microscopic quantum transport theories, while applicable to a much larger class of problems, are far more complicated.

II. GOVERNING EQUATIONS FOR THE SEMICONDUCTOR-INSULATOR SYSTEM

The governing macroscopic equations are obtained, as in previous work,[3] by postulating the basic elements and interactions which together constitute a model or idealization of the semiconductor. Forcing the model to obey general conservation laws results in integral forms from which the governing differential equations and boundary conditions are then deduced. This system of equations is underdetermined and must be supplemented by constitutive equations descriptive of the specific materials under consideration. As noted in Sec. I, for the present work one of these equations—essentially the equation of state for the electron gas—is central. In the macroscopic description the electron-gas properties are set by a specification of the dependence of the internal energy density of the electron gas on the electron-gas density and other variables. The usual diffusion-drift description is obtained when a simple logarithmic dependence on density (Maxwell gas) is assumed.[3] More generally a series expansion in density (essentially the virial expansion of kinetic theory) could be chosen. In the present paper, we are concerned with gas properties in the vicinity of an interface; that is, in a region across which electron-gas properties change rapidly. In such a region it seems reasonable to suppose that the energy density depends not only on the density of the gas but also on the (large) gradient of the density. As shown in this work, the introduction of such a term along with the planar-averaged continuum assumption does indeed allow one to discuss the strong inversion layer.

As in our earlier investigation, we assume the semiconductor to be modeled by three interacting continua: inertialess electron and hole gases and a rigid lattice with charge density ρ^i. The two gases each possess a charge density, ρ^e or ρ^h, move with respect to the lattice with velocities v^e or v^h and are permitted to recombine through charge source densities, $\gamma^e = -\gamma^h$. Neighboring elements of gas interact across their surface of separation through fluid pressures p^e or p^h. In addition, for the electron gas we allow a more general interaction through a second-rank tensor "double force" Δ^e, and an associated surface traction $\mathbf{t}^e$, which are necessary in order that electron-density-gradient dependences enter the theory. The distinction between double force Δ^e and force (or pressure or traction) is that force does work when matter moves over a distance whereas Δ^e does work directly when the density of the matter changes. The flow of each gas through the lattice is impeded by (resistive) interaction forces $\mathbf{E}^e$ and $\mathbf{E}^h$. Since all constituents are charged they also interact electrostatically. Applying the conservation laws of charge, linear momentum, and energy and the laws of electrostatics to this three-continuum model results in the following integral forms:

$$\int_S \hat{\mathbf{n}}\cdot\mathbf{D}\,dS = \int_V (\rho^e+\rho^h+\rho^i)\,dV\ , \tag{2.1a}$$

$$\oint_C \mathbf{E}\cdot d\mathbf{r} = 0\ , \tag{2.1b}$$

$$\frac{\partial}{\partial t}\int_V \rho^e dV + \int_S \hat{\mathbf{n}}\cdot\mathbf{v}^e\rho^e dS = \int_V \gamma^e dV\ , \tag{2.1c}$$

$$\frac{\partial}{\partial t}\int_V \rho^h dV + \int_S \hat{\mathbf{n}}\cdot\mathbf{v}^h\rho^h dS = \int_V \gamma^h dV\ , \tag{2.1d}$$

$$-\int_S (\hat{\mathbf{n}}p^e+\mathbf{t}^e)\,dS + \int_V \rho^e(\mathbf{E}+\mathbf{E}^e)\,dV = 0\ , \tag{2.1e}$$

$$-\int_S \hat{\mathbf{n}}p^h dS + \int_V \rho^h(\mathbf{E}+\mathbf{E}^h)\,dV = 0\ , \tag{2.1f}$$

$$\begin{aligned}\frac{\partial}{\partial t}\int_V (\varepsilon+\rho^e\varepsilon^e+\rho^h\varepsilon^h)\,dV &= \int_S (-p^e\hat{\mathbf{n}}\cdot\mathbf{v}^e+\mathbf{t}^e\cdot\mathbf{v}^e+\Delta^e\nabla\mathbf{v}^e-p^h\hat{\mathbf{n}}\cdot\mathbf{v}^h \\ &\qquad -\hat{\mathbf{n}}\cdot\mathbf{v}^e\rho^e\varepsilon^e-\hat{\mathbf{n}}\cdot\mathbf{v}^h\rho^h\varepsilon^h)\,dS\ , \\ &\quad + \int_V \left[\rho^e\mathbf{E}\cdot\mathbf{v}^e+\rho^h\mathbf{E}\cdot\mathbf{v}^h+\mathbf{E}\cdot\frac{d\mathbf{P}}{dt}\right] dV\ .\end{aligned} \tag{2.1g}$$

The vector $\hat{\mathbf{n}}$ is the outwardly directed unit normal to the surface S enclosing the arbitrary volume V fixed in space and C is an arbitrary closed curve. In addition, ε is the stored internal energy per unit volume in the lattice, ε^e and ε^h are the internal energies per charge in the two gases, $\mathbf{P}$ is polarization and $\mathbf{D}$ is the electric displacement. The terms $\Delta^e\nabla\mathbf{v}^e$ and $\mathbf{t}^e\cdot\mathbf{v}^e$ in (2.1g) represent the rates at which work is done by the electron double force and its associated surface traction, respectively.[4]

Away from interfaces, the field variables are continuous and, by standard arguments,[3] differential forms can be deduced:

$$\nabla\cdot\mathbf{D}=\rho^e+\rho^h+\rho^i\ , \tag{2.2a}$$

$$\mathbf{E}\equiv-\nabla\phi\ , \tag{2.2b}$$

$$\frac{\partial\rho^e}{\partial t}+\nabla\cdot(\rho^e\mathbf{v}^e)=\gamma^e\ , \tag{2.2c}$$

$$\frac{\partial\rho^h}{\partial t}+\nabla\cdot(\rho^h\mathbf{v}^h)=\gamma^h\ , \tag{2.2d}$$

$$-\nabla p^e+\nabla\cdot\boldsymbol{\tau}^e+\rho^e(\mathbf{E}+\mathbf{E}^e)=0\ , \tag{2.2e}$$

$$-\nabla p^h+\rho^h(\mathbf{E}+\mathbf{E}^h)=0\ , \tag{2.2f}$$

$$\begin{aligned}\frac{d\varepsilon}{dt}+\rho^e\frac{d^e\varepsilon^e}{dt}+\rho^h\frac{d^h\varepsilon^h}{dt}-\left[\frac{p^e}{\rho e}+\frac{\boldsymbol{\eta}^e\cdot\nabla\rho^e}{(\rho^e)^2}\right]\frac{d^e\rho^e}{dt}-\frac{p^h}{\rho^h}\frac{d^h\rho^h}{dt}+\frac{\boldsymbol{\eta}^e}{\rho^e}\cdot\frac{d^e\nabla\rho^e}{dt}-\mathbf{E}\cdot\frac{d\mathbf{P}}{dt}\\=-\rho^e\mathbf{E}^e\cdot\mathbf{v}^e-\rho^h\mathbf{E}^h\cdot\mathbf{v}^h-\gamma^e\left[\varepsilon^e+\frac{p^e}{\rho^e}+\frac{\boldsymbol{\eta}^e\cdot\nabla\rho^e}{(\rho^e)^2}\right]-\gamma^h\left[\frac{p^h}{\rho^h}+\varepsilon^h\right]+\frac{\boldsymbol{\eta}^e}{\rho^e}\cdot\nabla\gamma^e\ ,\end{aligned} \tag{2.2g}$$

where in reaching (2.2e) and (2.2g) a standard argument due to Cauchy allows Δ^e and its associated $\mathbf{t}^e$ to be written in terms of an electron double-pressure vector $\boldsymbol{\eta}^e$ and an associated stress tensor $\boldsymbol{\tau}^e$ as $\hat{\mathbf{n}}\boldsymbol{\eta}^e$ and $\hat{\mathbf{n}}\cdot\boldsymbol{\tau}^e$, respectively. The specific form of the double-pressure is chosen because, as shown in the Appendix, it corresponds to the internal energy density of the electron gas depending on the density gradient. In addition, to obtain (2.2g) we have employed (2.2c)–(2.2f), the material derivatives,

$$\frac{d}{dt}=\frac{\partial}{\partial t},\ \frac{d^e}{dt}=\frac{\partial}{\partial t}+\mathbf{v}^e\cdot\nabla,\ \frac{d^h}{dt}=\frac{\partial}{\partial t}+\mathbf{v}^h\cdot\nabla\ , \tag{2.3a}$$

the fact that

$$\nabla\left[\frac{d^e\rho^e}{dt}\right]=\frac{d^e\nabla\rho^e}{dt}+\nabla\mathbf{v}^e\cdot\nabla\rho^e\ , \tag{2.3b}$$

and the assumption that

$$\boldsymbol{\tau}^e=-\rho^e\nabla\left[\frac{\boldsymbol{\eta}^e}{\rho^e}\right] \tag{2.3c}$$

which is made in order that ε^e depend only on volumetric deformation and not explicitly on the components of macroscopic strain.

Equation (2.2g) is commonly called the first law of thermodynamics for the macroscopic system under consideration. Just as in Appendix A of Ref. 3, the form of (2.2g) allows the second law of thermodynamics to be written as

$$\begin{aligned}\frac{d\varepsilon}{dt}+\rho^e\frac{d^e\varepsilon^e}{dt}+\rho^h\frac{d^h\varepsilon^h}{dt}-\left[\frac{p^e}{\rho^e}+\frac{\boldsymbol{\eta}^e\cdot\nabla\rho^e}{(\rho^e)^2}\right]\frac{d^e\rho^e}{dt}\\+\frac{\boldsymbol{\eta}^e}{\rho^e}\cdot\frac{d^e\nabla\rho^e}{dt}-\frac{p^h}{\rho^h}\frac{d^h\rho^h}{dt}-\mathbf{E}\cdot\frac{d\mathbf{P}}{dt}=T\frac{d\eta}{dt}\ ,\end{aligned} \tag{2.4}$$

where η is the entropy per unit volume. Equations (2.3) and (2.4) then imply that, for a uniform temperature state, the rate-of-entropy-production inequality is of the form

$$\begin{aligned}\frac{d\eta}{dt}=\frac{1}{T}\Bigg\{&-\rho^e\mathbf{E}^e\cdot\mathbf{v}^e-\rho^h\mathbf{E}^h\cdot\mathbf{v}^h\\&-\gamma^e\left[\varepsilon^e+\frac{p^e}{\rho^e}-\frac{\boldsymbol{\eta}^e\cdot\nabla\rho^e}{(\rho^e)^2}\right]\\&+\frac{\boldsymbol{\eta}^e}{\rho^e}\cdot\nabla\gamma^e-\gamma^h\left[\varepsilon^h+\frac{p^h}{\rho^e}\right]\Bigg\}\geq 0\ .\end{aligned} \tag{2.5}$$

Equations (2.2), (2.4), and (2.5) are the differential equations that describe a general three-constituent semiconductor. Being 13 equations in 33 unknowns, this system, as noted at the start of this section, is underdetermined and additional equations must be supplied. These auxiliary equations are the constitutive theory which describes the specific material properties of the particular system under consideration. Their functional forms may be deduced from the requirements imposed by the need to satisfy (2.4) and (2.5). The equations deriving from (2.4) are termed the recoverable or nondissipative constitutive equations while those from (2.5) are dissipative. The former are of primary concern for this work.

The form of (2.4) shows that the internal energies ε, ε^e, and ε^h have the desired functional dependences, i.e.,

$$\varepsilon=\varepsilon(\mathbf{P})\ , \tag{2.6a}$$

$$\varepsilon^h=\varepsilon^h(\rho^h)\ , \tag{2.6b}$$

$$\varepsilon^e=\varepsilon^e(\rho^e,\nabla\rho^e)\ , \tag{2.6c}$$

where the temperature (entropy) dependence has been left implicit. Inserting (2.6) into (2.4) and carrying out the time differentiation of ε^e and ε^h, we observe that since the time derivatives of ρ^e, $\nabla\rho^e$, $\mathbf{P}$, and η are independent and can hold arbitrary values, their respective coefficients mush vanish. The resulting equations are the recoverable constitutive equations, which may be written as follows:

$$\mathbf{D}=\mathbf{D}(\mathbf{E}) , \tag{2.7a}$$

$$\frac{p^h}{\rho^h}=\rho^h\frac{\partial\varepsilon^h}{\partial\rho^h} , \tag{2.7b}$$

$$\frac{p^e}{\rho^e}=\rho^e\frac{\partial\varepsilon^e}{\partial\rho^e}+\frac{\boldsymbol{\eta}^e\cdot\nabla\rho^e}{(\rho^e)^2} , \tag{2.7c}$$

$$\frac{\boldsymbol{\eta}^e}{\rho^e}=-\rho^e\frac{\partial\varepsilon^e}{\partial\nabla\rho^e} . \tag{2.7d}$$

The dissipative constitutive theory arises from (2.5) and as discussed in Ref. 3 gives rise to relations for $\mathbf{E}^e$, $\mathbf{E}^h$, and γ^e. These equations together with (2.7) make the differential system fully determinate and, when consistent boundary conditions are appended, allow for the formulation of well-posed semiconductor boundary-value problems. Before discussing the boundary conditions, however, we digress to discuss a more useful form of the governing equations.

As shown in Ref. 3, when the electron or hole gas is materially homogeneous (and gradient dependences are not important), it is possible and more convenient to transform the fluid pressure to a chemical potential. Since in the present work the hole gas meets these conditions, we have the relation

$$\nabla p^h=\rho^h\nabla\phi^h , \tag{2.8a}$$

where the chemical potential is defined by

$$\phi^h=\frac{\partial(\rho^h\varepsilon^h)}{\partial\rho^h} . \tag{2.8b}$$

Following Ref. 3, (2.8) permits (2.2f) to be transformed into an equation in which ρ^h does not appear explicitly:

$$-\nabla(\phi+\phi^h)+\mathbf{E}^h=0 . \tag{2.9}$$

From this, an integral form, valid even when field variables are discontinuous, may be postulated:

$$-\int_S\hat{\mathbf{n}}(\phi+\phi^h)dS+\int_V\mathbf{E}^h dV=0 . \tag{2.10}$$

For the electron gas, it turns out that a similar transformation may be effected also. Assuming material homogeneity, it can be shown that

$$-\nabla\cdot\boldsymbol{\tau}^e+\nabla p^e=\rho^e\nabla\left[\frac{\partial(\rho^e\varepsilon^e)}{\partial\rho^e}+\nabla\cdot\left[\frac{\boldsymbol{\eta}^e}{\rho^e}\right]\right] . \tag{2.11}$$

Consequently, a generalized chemical potential formulation can be implemented, with

$$\phi^{e*}=\phi^e+\nabla\cdot\left[\frac{\boldsymbol{\eta}^e}{\rho^e}\right] , \tag{2.12}$$

where

$$\phi^e=\frac{\partial(\rho^e\varepsilon^e)}{\partial\rho^e}$$

with which (2.2e) becomes

$$-\nabla(\phi+\phi^{e*})+\mathbf{E}^e=0 . \tag{2.13}$$

From (2.13) we can than postulate the integral form

$$-\int_S\hat{\mathbf{n}}(\phi+\phi^{e*})dS+\int_V\mathbf{E}^e dV=0 . \tag{2.14}$$

As an aside, we note that the existence of the form (2.12) which permits (2.13) and leads to a description of the strong inversion layer (as shown in Sec. IV) is equivalent to saying that a quasi-Fermi level (or electrochemical potential) description of the strong inversion layer is possible. Furthermore, it thus remains meaningful to speak of (2.13) as a *diffusion-drift* equation with a generalized diffusion "force" given by the gradient of the generalized chemical potential ϕ^{e*}.

A consistent set of boundary conditions is now obtained by applying the integral forms (2.1a)—(2.1d), (2.10), and (2.14) to an arbitrary pillbox region encompassing a portion of the interface and taking the limit as the volume collapses to the interface in the usual way. Performing this procedure we find the conditions,

$$\hat{\mathbf{n}}\cdot[\mathbf{D}]=\sigma , \tag{2.15a}$$

$$[\phi]=0 , \tag{2.15b}$$

$$\hat{\mathbf{n}}\cdot\left[\rho^e\mathbf{v}^e+\rho^h\mathbf{v}^h+\frac{\partial\mathbf{D}}{\partial t}\right]=0 , \tag{2.15c}$$

$$[\phi+\phi^{e*}]=f^{e*} , \tag{2.15d}$$

where

$$\int_S f^{e*}\hat{\mathbf{n}}\,dS=\lim_{V\to 0}\int_V\mathbf{E}^e dV ,$$

$$[\phi+\phi^h]=f^h , \tag{2.15e}$$

where

$$\int_S f^h\hat{\mathbf{n}}\,dS=\lim_{V\to 0}\int_V\mathbf{E}^h dV$$

and we employ the notation $[C]$ to denote C^+-C^- where $\hat{\mathbf{n}}$ is the unit normal directed from the minus to the plus side of the interface. Equation (2.15a) defines the surface charge density σ and f^{e*} and f^h are the forces per charge exerted by the interface on the electron and hole gases, respectively.[3,5] As was done in Ref. 5, a surface thermodynamics and a surface constitutive theory can be set up from which functional forms for f^{e*} and f^h are obtained. Lastly, in addition to the above conditions, two other conditions arise as a result of the introduction of electron-density-gradient dependence and consequently double force Δ^e into the theory. The first results from the fact that the normal components of the double forces exerted by each side of an interface must balance. Since $\Delta^e=\hat{\mathbf{n}}\boldsymbol{\eta}^e$ this immediately gives the condition

$$\hat{\mathbf{n}}\cdot[\boldsymbol{\eta}e]=0 . \tag{2.16a}$$

The second additional condition, which is obtained from the variational approach given in the Appendix, demands that the electron density be continuous across interfaces:

$$[\rho^e]=0 . \tag{2.16b}$$

To summarize then, the description of the semiconductor when electron density gradients are important consists of the differential equations (2.2a)—(2.2d), (2.9), and (2.13) with (2.8b), (2.12), and expressions for $\mathbf{E}^e$, $\mathbf{E}^h$, and γ^e together with the consistent set of boundary conditions, (2.15) and (2.16).

III. PARTIAL SPECIFICATION OF THE CONSTITUTIVE THEORY FOR THE Si-SiO$_2$ SYSTEM

As in our earlier work, it is reasonable to assume that (2.6a) takes the form usual for a linear dielectric and that (2.6b) describes the holes as ideal, i.e.,

$$\varepsilon^h = \frac{-E_v}{q} + \frac{kT}{q}\left[\ln\left[\frac{\rho^h}{qN_v}\right] - 1\right] , \tag{3.1}$$

where q is the charge on an electron and E_v and N_v are the valence band-edge energy and effective density of states, respectively.

From (2.7b) and (2.8b), we then have

$$p^h = \frac{kT}{q}\rho^h , \tag{3.2a}$$

$$\phi^h = \frac{-E_v}{q} + \frac{kT}{q}\ln\left[\frac{\rho^h}{qN_v}\right] . \tag{3.2b}$$

These forms can be expected to be valid at low hole densities ($\rho^h \le 10^{18} q\ \mathrm{cm}^{-3}$).

For the electron gas, we need instead to select both density *and* density-gradient dependences. Leaving the former unspecified for the moment, one would expect that for the latter the lowest-order form would be with the double-pressure $\boldsymbol{\eta}^e$ *linear* in the density gradient [just as (3.2a) is the lowest-order pressure-density relation]. It is readily shown that this is obtained if (2.6c) takes the form[6]

$$\varepsilon^e = \varepsilon_0^e(\rho^e) - \frac{b}{2}\left[\frac{d\ln\rho^e}{dn}\right]^2 , \tag{3.3}$$

where we have used the fact that the normal component of the gradient must dominate. As expected, inserting (3.3) into (2.7d) then yields

$$\boldsymbol{\eta}^e = b\frac{d\rho^e}{dn}\hat{\mathbf{n}} \tag{3.4a}$$

and (2.12) gives

$$\phi^{e*} = \phi_0^e + \frac{b}{\rho^e}\frac{d^2\rho^e}{dn^2} - \frac{b}{2}\left[\frac{d\ln\rho^e}{dn}\right]^2 , \tag{3.4b}$$

where

$$\phi_0^e = \frac{\partial(\rho^e\varepsilon_0^e)}{\partial\rho^e}$$

and the coefficient of proportionality b is a new macroscopic coefficient which characterizes the strength of the density-gradient dependence in the electron gas. Ideally, its value should be deduced from experimental measurement. The possible dependence of b on electron density will be considered below but for now we assume it constant. Finally, for the gradient-independent $\varepsilon_0(\rho^e)$ one could employ an ideal-gas form such as (3.1), or, for higher densities, a Fermi-Dirac form[7] or perhaps some more general equation.

Concerning the remainder of the constitutive theory, we shall assume that the electron and hole gases in the oxide are such that the oxide is an insulator, i.e., with $\rho^e = \rho^h = 0$. Because of this, (2.1b) requires $\rho^e = 0$ at the surface in the semiconductor. This, in turn, implies that in the insulating limit the condition (2.16a) is automatically satisfied and may be dropped.[8] Lastly, for this paper we leave $\mathbf{E}^e$, $\mathbf{E}^h$, γ^e, and the surface constitutive theory for f^{e*} and f^h unspecified.

IV. APPLICATION TO A STATIC ONE-DIMENSIONAL PROBLEM

As an illustration of the theory developed above, we examine the static situation in an idealized one-dimensional metal-oxide-semiconductor (MOS) capacitor. Applying the equations of Sec. II and III to such a structure, the following boundary-value problem can be formulated: Oxide, $-t_{\mathrm{ox}} \le x \le 0$,

$$\frac{d^2\phi}{dx^2} = 0 ; \tag{4.1}$$

semiconductor, $0 \le x < \infty$,

$$\frac{d}{dx}(\phi + \phi^{e*}) = 0 , \tag{4.2a}$$

$$\frac{d}{dx}(\phi + \phi^h) = 0 , \tag{4.2b}$$

$$\frac{d^2\phi}{dx^2} = \frac{-1}{\varepsilon_S}(\rho^e + \rho^h + \rho^i) , \tag{4.2c}$$

where

$$\phi^{e*} = \frac{\partial(\rho^e\varepsilon_0^e)}{\partial\rho^e} + \frac{b}{\rho^e}\frac{d^2\rho^e}{dx^2} - \frac{b}{2(\rho^e)^2}\left[\frac{d\rho^e}{dx}\right]^2 , \tag{4.2d}$$

$$\phi^h = \frac{kT}{q}\ln\left[\frac{\rho^h}{qN_v}\right] - \frac{E_v}{q} , \tag{4.2e}$$

and with boundary conditions

$$\phi = V_g \quad \text{at } x = -t_{\mathrm{ox}} \tag{4.3}$$

$$[\phi] = 0 , \quad x = 0 \tag{4.4a}$$

$$\rho^e = 0 , \quad x = 0 \tag{4.4b}$$

$$[\phi + \phi^{e*}] = f^{e*} , \quad x = 0 \tag{4.4c}$$

$$[\phi + \phi^h] = f^h , \quad x = 0 \tag{4.4d}$$

$$\phi \to 0, \ \rho^e \to \rho_b^e, \ \frac{d\rho^e}{dx} \to 0, \ \phi^h \to 0 \ \text{ as } x \to \infty . \tag{4.5}$$

As in Ref. 3, the conditions (4.4c) and (4.4d) are not independent and, in practice, can be replaced by a specification of the surface potential, $\phi(x=0)$. This serves to decouple the equations in the oxide from those in the semiconductor and, in this way, below, we discuss only the latter. The oxide equations can be solved subsequently and lead to an equation for the gate voltage required to produce the assumed surface potential.

Treating the semiconductor equations, we see that (4.2a) and (4.2b) can be integrated immediately. Employing (4.2d), (4.2e), and the conditions (4.5) we have,

$$\phi+\frac{kT}{q}\ln\left[\frac{\rho^h}{\rho_b^h}\right]=0\ , \tag{4.6}$$

$$\phi+\phi_0^e-\phi_b^{e*}+\frac{b}{\rho^e}\frac{d^2\rho^e}{dx^2}-\frac{b}{2(\rho^e)^2}\left[\frac{d\rho^e}{dx}\right]^2=0\ , \tag{4.7}$$

where

$$\phi_b^{e*}=\phi^{e*}(x\rightarrow\infty)$$

and

$$\rho_b^h=\rho^h\ \ (x\rightarrow\infty)\ .$$

The similarities between (4.7) and known equations can be made apparent by changing variables according to

$$S=(-\rho^e)^{1/2}\ , \tag{4.8}$$

which, as in Ref. 8, introduces the fact that $-\rho^e$ must be non-negative into the equations. The differential equations (4.7) and (4.2c) then become

$$2b\frac{d^2S}{dx^2}+S(\phi+\phi_0^e-\phi_b^{e*})=0\ , \tag{4.9a}$$

$$\frac{d^2\phi}{dx^2}=\frac{1}{\varepsilon_S}(S^2-\rho_b^h e^{-\mu\phi/D}-\rho^i) \tag{4.9b}$$

where ρ^h has been eliminated by using (4.6).

The correspondence between this system and the quantum-mechanical description of Stern[2] is evident: By viewing S as a "wave function," (4.9a) is recognizable as a "Schrödinger equation" with the energy eigenvalue replaced by a function of electron density and the usual coefficient containing Planck's constant replaced by the macroscopic constant $2b$.[9] In addition, the boundary conditions (2.16a) and (2.16b) and the fact that the former condition is dropped in the above problem are also in accord with quantum mechanics (infinite barrier problem). These connections in the static case with the usual one-electron effective-mass Hamiltonian approach imply that $b\simeq\hbar^2/(4qm^*)$ where m^* is a continuum average of the electron effective mass(es) normal to the interface. With this identification and an appropriate choice for the zero-gradient chemical potential ϕ_0^e, the system (4.9) must lead to results in agreement with quantum-mechanical calculation. That is, the electron density profiles calculated by solving (4.9) for particular surface potentials must agree with those computed by Stern.[2] Since (4.9) are macroscopic equations, they make no predictions about any microscopic quantities such as subband structure. If such things are of interest, one must employ a microscopic description.

The analysis of the static, one-dimensional problem given above demonstrates that much of the behavior of the strong inversion layer, heretofore described using quantum mechanics, can be understood within the framework of a diffusion-drift description. As shown in this work, this is achieved by permitting the internal energy density of the electron gas to depend upon the electron-density gradient as well as on the density itself. The general planar-averaged macroscopic description developed in this way is applicable not only to the static case but also to dynamic situations for which a quantum-mechanical description is far more complicated. In view of this, we believe that the approach outlined here will have value for analyzing transport in silicon MOS field-effect transistors as well as in various "quantum-well" devices.

Note added in proof. Numerical calculations [M. G. Ancona, in *Proceedings of the Conference on Numerical Modeling of Semiconductors,* (Boole, Dublin, in press)] have verified the expectations of Sec. IV concerning agreement with Ref. 2. Also, a more rigorous microscopic derivation has yielded the formula $b=\hbar^2/(12m^*q)$.

APPENDIX: DERIVATION OF THE GOVERNING EQUATIONS IN A NONDISSIPATIVE CASE FROM A VARIATIONAL PRINCIPLE

For nondissipative situations, one can deduce the governing equations of macroscopic systems from a variational principle (Lagrangian). For our situation this approach has two advantages over that used in Sec. II:[4] (1) the desired form of the energy density of the electron gas (i.e., with a gradient dependence) is specified at the start and (2) because we discuss a fluid it turns out that one of the boundary conditions is much more evident. Following Toupin[4] we postulate the variational principle (treating only the electron gas for simplicity and employing indicial notation with the Einstein summation convention):

$$\delta\int_V\rho^e\varepsilon^e dV=\int_V F_i^1\delta y_i dV+\int_S F_i^2\delta y_i dS+\int_S F_i^3 D(\delta y_i)dS\ , \tag{A1}$$

where S is a smooth surface enclosing the volume V, the F_i^j are generalized forces, the components y_i are the Eulerian coordinates of the electron gas, D signifies the normal derivative, and δ indicates variation holding the Lagrangian coordinates of the electron gas fixed.[4] Assuming $\varepsilon^e=\varepsilon^e(\rho^e,\nabla\rho^e)$, the left-hand side of (A1) can be rewritten as

$$\delta\int_V\rho^e\varepsilon^e dV=\int_V\rho^e\delta\varepsilon^e dV\ , \tag{A2}$$

where

$$\begin{aligned}\delta\varepsilon^e=&-\left[\rho^e\frac{\partial\varepsilon^e}{\partial\rho^e}+\rho^e{}_{,i}\frac{\partial\varepsilon^e}{\partial\rho^e{}_{,i}}\right](\delta y_j)_{,j}\\&-\rho^e\frac{\partial\varepsilon^e}{\partial\rho^e{}_{,i}}(\delta y_j)_{,ji}\\&-\rho^e{}_{,j}\frac{\partial\varepsilon^e}{\partial\rho^e{}_{,i}}(\delta y_j)_{,i}\ .\end{aligned} \tag{A3}$$

Inserting (A2) with (A3) into (A1), integrating by parts twice, and applying the divergence theorem or surface divergence theorem, we find

$$\int_V(p^e{}_{,i}-\tau^e_{ji,j}-F_i^1)\delta y_i dV-\int_S(\hat{\mathbf{n}}_i p^e-\hat{\mathbf{n}}_j\tau^e_{ji}+D_i(\hat{\mathbf{n}}_j\eta_i^e)+F_i^2)\delta y_i dS+\int_S(\hat{\mathbf{n}}_j\hat{\mathbf{n}}_i\eta_j^e-F_i^3)D(\delta y_i)dS=0\ , \tag{A4}$$

where p^e, η_i^e, and τ_{ij}^e are as given in (2.7c), (2.7d), and (2.3c).

Since the variations δy_j and $D(\delta y_j)$ are arbitrary and as usual are taken to be continuous across any boundary, their coefficients (in brackets) must vanish thereby leading, with appropriate identifications of the F_i^j, to (2.2e) and to boundary conditions similar to (2.15d) and (2.16a). In addition, the continuity of δy_j and $D(\delta y_j)$ implies the continuity of $(\delta y_j)_{,j}$ and thence the continuity of $\delta\rho^e/\rho^e$, and thus ρ^e which represents an additional boundary condition,

$$[\rho^e]=0\ , \tag{A5}$$

to be applied for the solution of semiconductor boundary-value problems. This was the condition asserted in (2.16b). Finally, for consistency with the variational principle, η_i^e does work on the electron gas according to

$$\int_S \hat{\mathbf{n}}_j \eta^e{}_i v_{j,i}^e dS \tag{A6}$$

as was assumed at the start of Sec. II. It should be noted that the expression

$$\int_S \hat{\mathbf{n}}_i \eta_i^e v_{j,j}^e dS \tag{A7}$$

is also consistent with the variational principle and, if this is assumed, the entire description is unchanged with the exception of (2.3c). The form (A6) is used in Sec. II because it yields a stress tensor in conformity with that calculable from equilibrium quantum-statistical mechanics.

---

[1] Note that a macroscopic description in which the strong inversion layer is considered to be part of the surface can still be formulated.

[2] F. Stern, CRC Crit. Rev. Solid State Sci. **5**, 499 (1974).

[3] M. G. Ancona and H. F. Tiersten, Phys. Rev. B **22**, 6104 (1980).

[4] The treatment in Sec. II of the double force follows more general work of A. E. Green and R. S. Rivlin, Arch. Ration. Mech. Anal. **16**, 325 (1964). The Appendix outlines a similar development from a variational principle following R. A. Toupin, Arch. Ration. Mech. Anal. **11**, 385 (1962). The connections between the two approaches and their relative merits have been discussed by H. F. Tiersten and J. L. Bleustein, in *R. D. Mindlin and Applied Mechanics*, edited by G. Herrmann (Pergamon, New York, 1974) p. 67.

[5] M. G. Ancona and H. F. Tiersten, Phys. Rev. B **27**, 7018 (1983).

[6] A linear gradient dependence in (3.3), which could arise only in crystals which can exhibit pyroelectricity, is readily shown to be unimportant since it leaves the differential equations unaffected. Its sole effect would be on the boundary condition (2.16a), the physical significance of which is unclear.

[7] For a Fermi-Dirac gas an appropriate form is given by W. B. Joyce and R. W. Dixon, Appl. Phys. Lett. **31**, 354 (1977).

[8] The reason for this can be seen by inserting (3.4a) into (2.16a) and expressing the fact that the electron density must be nonnegative by the change of variable $\rho^e \to -S^2$. Condition (2.16a) then demands the continuity $bS\,dS/dn$ which is satisfied for all dS/dn when $S=0$.

[9] In the case when b is a function of electron density the form of (4.10a) changes slightly, being replaced by

$$2\sqrt{b}\,\frac{d}{dx}\left[\sqrt{b}\,\frac{dS}{dx}\right]+S(\phi+\phi_0^e-\phi_b^{e*})=0\ .$$