PHYSICAL REVIEW B VOLUME 39, NUMBER 13 1 MAY 1989

Quantum correction to the equation of state of an electron gas in a semiconductor

M. G. Ancona
Code 6813, Naval Research Laboratory, Washington, D.C. 20375-5000

G. J. Iafrate
Army Electronics Technology and Devices Laboratory, Fort Monmouth, New Jersey

(Received 9 May 1988; revised manuscript received 16 September 1988)

In a recent paper [M. G. Ancona and H. F. Tiersten, Phys. Rev. B **35**, 7959 (1987)] a macroscopic description of conduction electrons in a semiconductor was presented in which the equation of state for the electron gas was generalized to include a dependence on the gradient of the density. This generalization led to a new transport equation—often expressible as a generalized diffusion–drift-current equation—which has been shown to accurately describe some important quantum mechanical effects occurring in semiconductor structures. In the present paper sufficient microscopic conditions under which the density-gradient–dependent equation of state does represent lowest-order effects of quantum mechanics are established using methods of quantum statistical mechanics. A microscopic derivation of the transport equation is also given.

I. INTRODUCTION

In a recent paper[1] a macroscopic description of electrons in a semiconductor was presented in which the equation of state of the electron gas was permitted to depend not only on the gas density but also on its gradient. The equation describing transport of such a gas was shown (under certain assumptions) to take the form of a generalized diffusion–drift-current equation, i.e., the usual diffusion–drift-current equation plus correction terms. In Ref. 1 it was argued and in subsequent work[2,3] shown that the generalized transport equation accurately describes some important quantum-mechanical phenomena including quantum confinement effects[2] and tunneling.[3] In this paper,[4] the underlying reason for this is made evident by a microscopic derivation of the macroscopic equation of state using Wigner function techniques. We demonstrate that, on a macroscopic scale, the lowest order effect of quantum mechanics on an electron gas is to make its equation of state density-gradient dependent. Among other assumptions in this demonstration, effects associated with quantum statistics and with many-body correlations are neglected and therefore the results are limited to "high" temperatures and "low" electron densities, conditions often characteristic of semiconductor situations.[5] In addition, the electron gas is assumed to be infinite in extent and thus effects due to boundaries (of relevance to semiconductor device applications) are not investigated.

II. MACROSCOPIC THEORY

In this section we briefly recount the key macroscopic results which for this paper serve as targets for the microscopic analysis. Reference 1 should be consulted for full details and for additional references. The equation of linear momentum balance in the electron gas under the assumption of negligible inertia may be written as

$$-\nabla(\varphi+\varphi^{e*})+\mathbf{E}^e=0\ , \tag{2.1}$$

where φ is electric potential, $\mathbf{E}^e$ is the drag force felt by the gas as it flows through the lattice and φ^{e*} is a generalized chemical potential. As discussed in Ref. 1, for (2.1) to be practically useful, specific (constitutive) expressions for $\mathbf{E}^e$ and φ^{e*} must be given. For $\mathbf{E}^e$ the lowest-order form is

$$\mathbf{E}^e=-\mathbf{v}^e/\mu^e\ , \tag{2.2}$$

which represents an assumption of linear drag where μ^e is the usual electron mobility. This is equivalent to treating scattering in a relaxation-time approximation and, for this paper, we make no effort to justify it on microscopic grounds. The constitutive expression for which we wish to provide microscopic justification is that for φ^{e*}; this is essentially the equation of state of the electron gas. In Ref. 1 we asserted that, to include lowest-order quantum effects, the equation of state must depend not only on the gas density but also on the density gradient. In effect, the inclusion of both density and density-gradient dependences means that the equation of state will contain some "nonlocality" which, of course, is a primary characteristic of quantum mechanics. Assuming the internal energy density of the electron gas, ϵ^e, to have these dependences, i.e., $\epsilon^e=\epsilon^e(\rho^e,\nabla\rho^e)$, it can be shown from thermodynamic arguments that the stress tensor τ^e in the gas must be of the form[1,6]

$$\tau^e=-p^eI+\rho^e\nabla\left[\rho^e\frac{\partial\epsilon^e}{\partial\nabla\rho^e}\right]\ , \tag{2.3a}$$

where I is the identity matrix and the gas pressure p^e is given by

$$p^e=\rho^e\left[\rho^e\frac{\partial\epsilon^e}{\partial\rho^e}+\nabla\rho^e\cdot\frac{\partial\epsilon^e}{\partial\nabla\rho^e}\right]\ . \tag{2.3b}$$

The expression for the generalized chemical potential φ^{e*} in terms of ϵ^e is[1]

$$\varphi^{e*}=\frac{\partial(\rho^e\epsilon^e)}{\partial\rho^e}-\nabla\cdot\left[\rho^e\frac{\partial\epsilon^e}{\partial\nabla\rho^e}\right] . \tag{2.4}$$

If we further assume that the functional dependences of ϵ^e on ρ^e and $\nabla\rho^e$ are both of lowest order,[7] then the internal energy density (apart from constants) takes the explicit form

$$\epsilon^e(\rho^e,\nabla\rho^e)=-\frac{1}{\beta}\ln\left[\frac{-\rho^e}{qN_C}\right]-\frac{b}{2}\frac{\nabla\rho^e\cdot\nabla\rho^e}{(\rho^e)^2} , \tag{2.5}$$

where $\beta\equiv 1/k_BT$, N_C is the effective density of states in the conduction band and the coefficient b is a new material parameter measuring the strength of the gradient effects in the gas. We term this the equation of state for an "ideal gradient gas." For such a gas, from (2.3), the stress tensor is given by

$$\tau^e=-\frac{\rho^e}{q\beta}I-b\left[\nabla\nabla\rho^e-\frac{\nabla\rho^e\nabla\rho^e}{\rho^e}\right] , \tag{2.6}$$

and from (2.4) we have the generalized chemical potential

$$\varphi^{e*}=-\frac{1}{\beta q}\ln\left[\frac{\rho^e}{-qN_C}\right]+b\left[\frac{\nabla^2\rho^e}{\rho^e}-\frac{1}{2}\frac{\nabla\rho^e\cdot\nabla\rho^e}{(\rho^e)^2}\right] , \tag{2.7a}$$

which may be written more compactly as[1]

$$\varphi^{e*}=-\frac{1}{\beta q}\ln\left[\frac{\rho^e}{-qN_C}\right]+2b\left[\frac{\nabla^2 s}{s}\right] ,$$

$$\text{where } s\equiv(-\rho^e)^{1/2} . \tag{2.7b}$$

By inserting (2.2) and (2.7b) into (2.1) we obtain the new transport equation

$$\mathbf{J}^e=\mu^e\rho^e\nabla\varphi-D^e\nabla\rho^e+2\mu^e b\rho^e\nabla\left[\frac{\nabla^2 s}{s}\right] , \tag{2.8}$$

which is of the form of a generalized diffusion-drift current equation. The first two terms on the right-hand side are the drift and diffusion terms, respectively, of the usual semiconductor current equation while the third term represents current arising as a result of gradient effects. This last component is a diffusion current in that it originates from a gradient of a chemical potential and it may be referred to as "quantum diffusion" since its inclusion in (2.8) yields a theory which contains both quantum confinement effects and quantum-mechanical tunneling.[1-3] In the present paper, this connection to quantum mechanics is made more direct by presenting sufficient microscopic conditions under which the new terms in (2.5)–(2.8) are explicitly quantum mechanical corrections.

III. MICROSCOPIC ORIGIN OF THE DENSITY-GRADIENT EFFECT

A microscopic derivation of macroscopic equations serves to establish sufficient microscopic conditions for the validity of such equations. In this section we find conditions under which quantum mechanics leads an electron gas to act as an ideal gradient gas.[8] Among these conditions are (1) the gas is infinite in extent, (2) quantum effects associated with Fermi-Dirac statistics are negligible, (3) many-body electron correlation effects are negligible, and (4) effective-mass theory applies. To some extent all of these conditions may be relaxed, however, since all but the first are typically well satisfied by the semiconductor applications of most interest, we proceed on this basis. Each condition is discussed further below as it arises.

Much of the microscopic derivation follows the pioneering work of Wigner,[9] although in detail we parallel the development of Ref. 10. In order to simplify the derivation we start from one-electron effective-mass theory, the microscopic underpinnings of which are well known, e.g., see Ref. 11. Among other things this assumes (self-consistent field approximation) that many-body electron correlations are negligible as is usually the case in "low"-density, "high"-temperature semiconductor situations.[5,8,11]

The effective-mass eigenstates ψ_n and energy eigenvalues E_n for the system under consideration (assuming a spherical constant-energy surface) are solutions of the effective-mass Schrödinger equation,

$$-\frac{\hbar^2}{2m^*}\nabla^2\psi_n(\mathbf{x})+U(\mathbf{x})\psi_n(\mathbf{x})=E_n\psi_n(\mathbf{x}) , \tag{3.1}$$

where $U(\mathbf{x})$ is the self-consistent electron potential energy and m^* is the effective mass. In thermal equilibrium at a temperature T the system may be described by the density matrix (mixed state),

$$\rho^{(1)}(\mathbf{x},\mathbf{x}')=\sum_n e^{-\beta E_n}\psi_n(\mathbf{x})\psi_n(\mathbf{x}') , \tag{3.2}$$

where, for the continuous spectrum, the sum is understood to be an integral and we have assumed a classical Boltzmann weighting of the states. To include quantum statistical effects in this one-electron picture a Fermi-Dirac weighting should be substituted into (3.2). We do not do this here because such effects on the equation of state have previously been considered[12] and because of our focus on low-density/high-temperature semiconductor situations. Choosing to employ a phase-space distribution function approach we define the one-particle Wigner function[9]

$$f^{(1)}_{W_0}(\mathbf{x},\mathbf{p})=\left[\frac{1}{\hbar\pi}\right]^3\int\exp\left[\frac{2i\mathbf{p}\cdot\mathbf{y}}{\hbar}\right]\rho^{(1)}(\mathbf{x}+\mathbf{y},\mathbf{x}-\mathbf{y})d\mathbf{y} . \tag{3.3}$$

Now, as Wigner showed, it is possible to gain information about a quantum system without solving (3.1), i.e., without determining the eigenstates or eigenvalues of the system. The technique is to obtain an asymptotic solution of an equation for (3.2) or (3.3) by using a perturbation procedure in which the ratio of the thermal deBroglie wavelength $\lambda_D\equiv\hbar[\beta/(2m^*)]^{1/2}$ to a characteristic length scale d, e.g., a screening length or a geometrical length, is assumed small. Such an approach will be valid for "high" temperature ($\lambda_D \rightarrow 0$). Following Ref. 11 we employ the Bloch equation satisfied by (3.3):

$$\frac{\partial f_{W_0}^{(1)}}{\partial \beta}=\left[\frac{\hbar^2}{8m^*}\nabla^2-\frac{p^2}{2m^*}\right]f_{W_0}^{(1)}-\theta' f_{W_0}^{(1)} , \tag{3.4}$$

where

$$\theta' f_{W_0}^{(1)}=\cos\left[\frac{\hbar}{2}\nabla_p\cdot\nabla_x\right]U(\mathbf{x})f_{W_0}^{(1)} ,$$

and ∇_x is understood to operate on $U(\mathbf{x})$ only. For purposes of clarity below, we put (3.4) into dimensionless form by rescaling β and U by a constant β_0, $\mathbf{p}$ by h/λ_D^0 [where $\lambda_D^0 \equiv \hbar(\beta_0/2m^*)^{1/2}$] and $\mathbf{x}$ by d so that (3.4) becomes

$$\frac{\partial f_{W_0}^{(1)}}{\partial \beta}=(\epsilon\nabla^2-p^2)f_{W_0}^{(1)}-\theta' f_{W_0}^{(1)} , \tag{3.5}$$

where

$$\theta' f_{W_0}^{(1)}=\cos(\sqrt{\epsilon}\nabla_p\cdot\nabla_x)U(\mathbf{x})f_{W_0}^{(1)} ,$$

$\epsilon \equiv [\lambda_D^0/(2d)]^2$ and the variables are now dimensionless.

For the perturbation procedure, we take $\epsilon \ll 1$ and then, expanding $f_{W_0}^{(1)}$ according to

$$f_{W_0}^{(1)}=\sum_{k=0}^{\infty}\epsilon^k\phi_k(\mathbf{x},\mathbf{p}) \tag{3.6}$$

and equating like powers of ϵ, we obtain

$$O(\epsilon^0):\ \ \phi_0=\exp[-\beta(p^2+U)] , \tag{3.7a}$$

$$O(\epsilon^1):\ \ \phi_1=\phi_0\left[-\beta^2\nabla^2 U+\frac{\beta^3}{3}(\nabla U\cdot\nabla U+2\mathbf{p}\cdot\nabla\nabla U\cdot\mathbf{p})\right] , \tag{3.7b}$$

and so on. As seen from (3.7a), when ϵ is vanishingly small (high-temperature limit), $f_{W_0}^{(1)}$ goes to the classical (unnormalized) Boltzmann distribution $f_{\mathrm{Cl}}^{(1)}(=\phi_0)$. The quantity ϕ_1 in (3.7b) represents the lowest-order quantum correction to $f_{W_0}^{(1)}$. In *dimensional* form, the expression for $f_{W_0}^{(1)}$ valid to $O(\epsilon)$ is

$$f_{W_0}^{(1)}=f_{\mathrm{Cl}}^{(1)}\left[1-\frac{\hbar^2\beta^2}{8m^*}\left[\nabla^2U-\frac{\beta}{3}\nabla U\cdot\nabla U-\frac{\beta}{3m^*}\mathbf{p}\cdot\nabla\nabla U\cdot\mathbf{p}\right]+O(\epsilon^2)\right] . \tag{3.8}$$

As a Wigner function, this approximate expression can now be used to compute various (approximate) expectation values by performing classical phase-space integrals. In particular, we have for the charge density

$$\rho^e(\mathbf{x})=\frac{-q}{(2\pi\hbar)^3}\int f_{W_0}^{(1)}(\mathbf{x},\mathbf{p})d\mathbf{p}=-q\left[\frac{m^*}{2\pi\beta\hbar^2}\right]^{3/2}e^{-\beta U}\left[1-\frac{\hbar^2\beta^2}{12m^*}\left[\nabla^2U-\frac{\beta}{2}\nabla U\cdot\nabla U\right]+O(\epsilon^2)\right] , \tag{3.9}$$

and for the stress tensor

$$\begin{aligned}\tau^e(\mathbf{x})&=\frac{1}{(2\pi\hbar)^3m^*}\int f_{W_0}^{(1)}(\mathbf{x},\mathbf{p})\mathbf{p}\mathbf{p}d\mathbf{p}\\&=\frac{1}{\beta}\left[\frac{m^*}{2\pi\beta\hbar^2}\right]^{3/2}e^{-\beta U}\left[I-\frac{\hbar^2\beta^2}{12m^*}\left[\nabla^2U-\frac{\beta}{2}\nabla U\cdot\nabla U-\nabla\nabla U\right]+O(\epsilon^2)\right] .\end{aligned} \tag{3.10}$$

The equation of state of the gas, i.e., the relation between τ^e and ρ^e, can now be obtained by combining (3.9) and (3.10) to eliminate the self-consistent potential U to order ϵ^2. We find

$$\tau^e=\frac{-\rho^e}{q\beta}I-\frac{\hbar^2}{12m^*q}\left[\nabla\nabla\rho^e-\frac{\nabla\rho^e\nabla\rho^e}{\rho^e}\right]+O(\epsilon^2) . \tag{3.11}$$

This equation is of the same form as (2.6), Ref. 6, thus demonstrating that to lowest order the quantum system does behave as an ideal gradient gas.[13] Moreover, a microscopic formula for the macroscopic coefficient b can now be identified:

$$b=\frac{\hbar^2}{12m^*q} . \tag{3.12}$$

That (2.6) and (3.11) are identical in form shows explicitly that the density-gradient correction introduced macroscopically into electron transport theory in Ref. 1 represents the lowest-order correction of quantum mechanics. To what extent does the lowest-order theory represent full effective-mass quantum mechanics? From our density-gradient expansion, the requirement is that ϵ be small, i.e.,

$$\epsilon\equiv\frac{\hbar^2\beta}{8m^*d^2}\cong(m_e/m^*)\frac{1.1\times10^{-12}\ \mathrm{cm^2K}}{Td^2}\ll 1 , \tag{3.13}$$

where m_e is the free-electron mass, and thus lowest-order density-gradient theory is best when the characteristic length is large, the temperature high and the effective mass close to its free-electron value. As a quantitative example, if the effective mass is $0.19m_e$ (silicon conduction band), the temperature is 300 K and the length scale is 10 nm, then ϵ is about 0.02 and one would expect the lowest-order theory to be useful. This is consistent with results from illustrative calculations made previously using the theory[1–3] and indicates that density-gradient theory should have a significant range of applicability.[14]

The derivation of (3.11) and the identification (3.12) achieve the main objective of this paper. Given these results—that the equation of state of an electron gas in which lowest-order quantum effects are considered is that of an ideal gradient gas—the macroscopic development of Ref. 1 then implies that the transport equation for this gas must be of the form (2.8). Nonetheless, it is of interest to provide a microscopic justification for this equation as well. This can be achieved in much the same way as the standard diffusion-drift current equation is derived from the Boltzmann transport equation (e.g., see Ref. 15). We start from the Wigner-Boltzmann transport equation which is the lowest-order equation of a quantum Bogoliubov-Born-Green-Kirkwood-Yvon (BBGKY) hierarchy. In a relaxation-time approximation and under near-equilibrium conditions this equation can be linearized to yield

$$f_W^{(1)}(\mathbf{x},\mathbf{p})=f_{W_0}^{(1)}(\mathbf{x},\mathbf{p})-\tau_W\left[\frac{\mathbf{p}}{m^*}\cdot\nabla f_{W_0}^{(1)}(\mathbf{x},\mathbf{p})+\theta f_{W_0}^{(1)}(\mathbf{x},\mathbf{p})\right], \tag{3.14}$$

where

$$\theta f_W^{(1)}=-\frac{2}{\hbar}\sin\left(\frac{\hbar}{2}\nabla_p\nabla_x\right)U(\mathbf{x})f_W^{(1)},$$

τ_W is the relaxation time and again ∇_x operates on $U(\mathbf{x})$ only. Inserting (3.8) into (3.14) then gives a first approximation to the *nonequilibrium,* one-particle Wigner distribution function $f_W^{(1)}$. From this, the current density is readily calculated to be identical to (2.6) with (3.12), thereby completing the demonstration.

Finally, we return to the point that these derivations are for an electron gas of infinite spatial extent [condition (1) above]. This would appear to be a significant restriction because semiconductor device situations in which quantum effects are important are typically strongly bounded, e.g., a quantum well. To relax this condition and give a microscopic treatment of a confined electron gas[16] is not straightforward particularly when transport is involved.[17] Not only is the derivation of the differential equations more complicated but boundary (or interface) conditions must be derived as well. This work is left to future investigation. For now, we note only that similar difficulties do not arise in the macroscopic approach[1] wherein boundary conditions arise naturally.

ACKNOWLEDGMENTS

One of us (M.G.A.) wishes to thank Dr. H. F. Tiersten, Dr. A. K. Rajagopal, Dr. S. Teitler, and Dr. A. M. Kriman for helpful discussions.

---

[1] M. G. Ancona and H. F. Tiersten, Phys. Rev. B **35-II**, 7959 (1987). The detailed derivations of this paper appear in an NRL Memorandum Report available from the authors.

[2] M. G. Ancona, COMPEL–Int. J. Comp. Math. Elect. Eng. **6**, 11 (1987).

[3] M. G. Ancona, Bull. Am. Phys. Soc. **33**, 786 (1988).

[4] This paper is an expanded version of work which appeared as M. G. Ancona and G. J. Iafrate, Bull. Am. Phys. Soc. **32**, 715 (1987).

[5] In silicon, "low" density is roughly below 3×10^{19} electrons/cm^3 and "high" temperature is above about 77 K.

[6] As noted in Ref. 1, macroscopically, τ^e is unique only to within an arbitrary tensor which in indicial notation takes the form $M_{[ki]j,k}$ where the bracketed indices indicate antisymmetry. Interestingly, the microscopic derivation from quantum mechanics appears to select this tensor to be zero when τ^e is written as in (2.3a), i.e., (2.6) is of the same form as (3.11) only in this case.

[7] As is well known, the lowest-order density dependence is that for which the gas pressure is linear in density (ideal gas). Similarly, for the density-gradient dependence the lowest-order form is obtained when the double-pressure (Ref. 1) is proportional to the density gradient.

[8] It should be noted that the macroscopic equations remain valid (given satisfaction of continuum assumptions) no matter how the gradient dependence arises and therefore can have wider applicability. An extreme example is that macroscopic equations similar to those presented in Sec. II apply also to completely classical situations of kinetic theory in which gradient effects result from two-body correlations. In that context the equations are known as Burnett equations, see S. Chapman and T. G. Cowling, *The Mathematical Theory of Non-uniform Gases,* 3rd ed. (Cambridge, New York, 1970), Chap. 15.

[9] E. P. Wigner, Phys. Rev. **40**, 749 (1932).

[10] I. Oppenheim and J. Ross, Phys. Rev. **107**, 28 (1957).

[11] W. Harrison, *Solid State Theory* (Dover, Toronto 1980).

[12] W. B. Joyce and R. W. Dixon, Appl. Phys. Lett. **31**, 354 (1977).

[13] It may be noted that this demonstration can also be reached directly from (3.9), if we identify $-U/q$ with the electric potential φ and use the fact that, when the system is in macroscopic equilibrium, $\varphi=-\varphi^{e*}$ apart from an additive constant. Equation (3.9) then becomes a relationship between φ^e and ρ^e and by a reversion of series an expression of the same form as (2.7a) is obtained. The derivation given in the text is preferable in that it requires no *a priori* identification of

$-U/q$ with φ.

[14] In contrast, the lowest-order terms of density-gradient expansions used previously in classical theory (Ref. 8) and in density functional theory [e.g., see J. Callaway and N. H. March, in *Solid State Physics,* edited by F. Seitz, D. Turnbull, and H. Ehrenreich (Academic, New York, 1984), Vol. 38, p. 186] accurately approximate the full expansions over very narrow ranges only and therefore are of limited usefulness.

[15] R. Stratton, IEEE Trans. Electron Devices **ED-19**, 1288 (1972).

[16] G. Bastard, Phys. Rev. B **24**, 5693 (1981).

[17] W. Frensley, Phys. Rev. B **36-II**, 1570 (1987).