Generalization of the Landauer conductance formula

S. Eränen
Semiconductor Laboratory, Technical Research Centre of Finland, Otakaari 7 B, SF-02150 Espoo, Finland

J. Sinkkonen
Electron Physics Laboratory, Helsinki University of Technolgy, Otakaari 7 A, SF-02150 Espoo, Finland
(Received 8 July 1986; revised manuscript received 7 October 1986)

We study the electrical current transport in conductor-insulator-conductor structures, where the charge carriers are assumed to traverse the insulating layer by tunneling. The current flow in the conductor contacts is treated by solving self-consistently the Poisson equation and the relaxation time form of the Boltzmann equation for an inhomogeneous electron system. The tunneling is taken into account by the appropriate boundary conditions for the electronic distribution functions in the two contacts. It is shown that the tunneling current density consists of two contributions, the first of which is a direct generalization of the celebrated Landauer conductance formula into the nonlinear voltage regime. The second contribution or the correction term originates from the screening of the electrical potential across the insulating layer and from the matching of the distribution functions on the opposite sides of the barrier. In the linear voltage regime for well-conducting contacts the tunneling current density is given by the ordinary Landauer result, but for semiconducting contacts the second contribution may become comparable to the first one. Moreover, it is shown that in the correction term the matching of the distribution functions is always negligible when compared to the screening effect. Finally, the limits of validity of our results are discussed.

I. INTRODUCTION

Today it is generally accepted that the electrical conductance of the tunneling current obeys the Landauer conductance formula. Since its first publication in 1957 (Ref. 1) the formula has gained widespread applications in the studies of localization and the scaling of electrical resistance in disorder systems, but, in addition, it has also been used in the calculations of the current-voltage characteristics of realistic structures in the field of the semiconductor device physics. Moreover, the initial form of the Landauer formula has later been generalized to account for energy-dependent transmission coefficients, finite temperatures, and, finally, the electron-electron interaction.[2,3] Also the validity of the formula has been the topic of several sophisticated treatments based on the quantum transport theory, the outcome of which generally agree with the original result.[4,5]

From the various derivations of the Landauer formula it is evident that the result can be exactly correct in the linear current-voltage regime, only. In addition, Landauer himself has on several occasions emphasized the existence of inhomogeneities in the electron density of the contacts near the reflecting boundary.[6,7] The authors feel that the exact nature and consequences of these inhomogeneities are not entirely explored. These facts serve as the motivation for the present article, which is an attempt to give a self-consistent description of the linear and nonlinear tunneling characteristics in one-dimensional electronic systems.

II. THEORY

Figure 1 gives a qualitative picture of the nonequilibrium energy-band diagram of our model system in a case, where the tunneling region is sandwiched between two metallic contacts. Here the tunneling region extends from $z=0$ to $z=L$, and the contacts are assumed to be sufficiently long on both sides of the tunneling junction. Moreover, the contacts are chosen to be of the same material, which implies, that in equilibrium the charge carrier concentrations n_0 and the chemical potentials μ_0 are equal in both contacts. In Fig. 1 the electrical current flows from the high electrical potential on the left-hand side of the junction to the low potential on the right. Evidently, even in a nonequilibrium state the electronic sys-

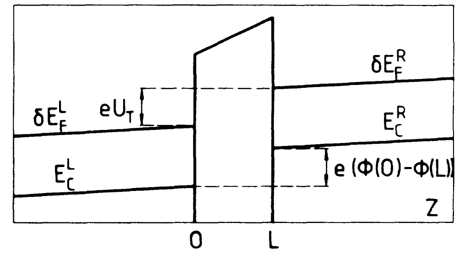

FIG. 1. A qualitative energy-band diagram of the nonequilibrium tunneling junction. δE_F^L (E_c^L) and δE_F^R (E_c^R) denote the quasi-Fermi-level (conduction-band edge) on the left- and right-hand side of the tunneling barrier.

tems of the contacts are uniform far from the junction, where the changes of the electrical potential are exactly balanced by the variations of the quasi-Fermi-level (electrochemical potential). However, near the junction our treatment will allow for the depletion or accumulation of electrons, although these effects have not been included in Fig. 1. It should be noted that due to the possible changes of the charge density near the junction the voltage drop $U_T(eU_T=\delta E_F^R(L)-\delta E_F^L(0))$ across the tunneling region is not necessarily equal to the difference of the electrostatic potential $\Delta\Phi=\Phi(0)-\Phi(L)$.

Consider now the distribution function of the electrons in the contact regions, where the velocity space of the charge carriers is treated as one-dimensional. In the relaxation-time approximation the solution of the Boltzmann equation for the distribution function of an inhomogeneous electron system may be written in the path integral form as

$$f(t)=f(t_0)\exp\left[-\int_{t_0}^{t}\frac{dt'}{\tau(t')}\right]+\int_{t_0}^{t}dt'\frac{f^{(0)}(t')}{\tau(t')}\exp\left[-\int_{t'}^{t}\frac{dt''}{\tau(t'')}\right]. \quad (1)$$

Here τ is the relaxation time, and $f^{(0)}$ stands for the local equilibrium distribution function.[8,9] In Eq. (1) the time integrals are taken along the phase-space trajectories, and t_0 denotes the point where the initial condition for the distribution function is specified. In order to continue the analysis we choose the relaxation time to be of the form $\tau=l/|v|$, where l is the scattering mean free path of the electrons. Also, we use the linearized version of the transport theory, which implies, that the integrals in Eq. (1) are taken along the free-particle trajectories.[10] Now, we may express the distribution function on the left-hand side for the electrons with positive velocities as

$$f_L(v_L)=\left[f_0^L-\left[\frac{\partial f_0^L}{\partial E}\right][\delta E_F^L(z)+e\Phi_L(z)]+\left[\frac{\partial f_0^L}{\partial E}\right]\int_{-\infty}^{z}dz'\frac{d\delta E_F^L}{dz'}e^{-(z-z')/l}\right]\Theta(v_L), \quad (2)$$

where f_0^L stands for the uniform equilibrium distribution and Θ for the unit step function. Equation (2) originates entirely (by partial integration) from the last term of Eq. (1) because the initial condition for the particles moving to the right is specified at $z=-\infty$, and thus the first term yields no contribution to this part of the distribution function. On the other hand, for the particles moving to the left the initial condition is given at the boundary of the tunneling region ($z=0$), whence the distribution function for the electrons with negative velocities becomes

$$f_L(-v_L)=\left[f_0^L-\left[\frac{\partial f_0^L}{\partial E}\right][\delta E_F^L(z)+e\Phi_L(z)]+f_1^Le^{z/l}-\left[\frac{\partial f_0^L}{\partial E}\right]\int_{z}^{0}dz'\frac{d\delta E_F^L}{dz'}e^{-(z'-z)/l}\right]\Theta(-v_L), \quad (3)$$

where f_1^L is an as yet undetermined function, which depends on the velocity, only. By using the same kind of reasoning as above, we obtain the distribution function of the electronic system on the right-hand side contact. Compared to Eqs. (2) and (3) the only difference is that now the initial condition for the electrons with positive velocities is given at the boundary of the tunneling junction ($z=L$) and for the electrons moving to the left at $z=\infty$. Consequently the distribution function reads

$$f_R(v_R)=\left[f_0^R-\left[\frac{\partial f_0}{\partial E}\right][\delta E_F^R(z)+e\Phi_R(z)]+f_1^Re^{-(z-L)/l}+\left[\frac{\partial f_0^R}{\partial E}\right]\int_{L}^{z}dz'\frac{d\delta E_F^R}{dz'}e^{-(z-z')/l}\right]\Theta(v_R), \quad (4)$$

$$f_R(-v_R)=\left[f_0^R-\left[\frac{\partial f_0^R}{\partial E}\right][\delta E_F^R(z)+e\Phi_R(z)]-\left[\frac{\partial f_0^R}{\partial E}\right]^{+}\int_{z}^{\infty}dz'\frac{d\delta E_F^R}{dz'}e^{-(z'-z)/l}\right]\Theta(-v_R), \quad (5)$$

where f_1^R is again an undetermined function of the velocity, only. For the further analysis it is noted that the velocities on the left and right contacts are related as $\frac{1}{2}mv_L^2=\frac{1}{2}mv_R^2+eU_T$.

It is obvious that the electrons on the left-hand side at $z=0$ with negative velocities have just reflected from the boundary of the junction or have traversed the barrier by tunneling from the right-hand side. Similarly, the electrons at $z=L$ with positive velocities have just been repelled by the barrier or have tunneled from the left-hand side. Now, because the transmission probability T is a function of energy, only, and the reflection probability $R=1-T$, we may write the above statements as

$$f_L(-v_L)|_{z=0}=f_L(v_L)|_{z=0}+T[f_R(-v_R)|_{z=L}-f_L(v_L)|_{z=0}], \quad (6a)$$

$$f_R(v_R)|_{z=L}=f_R(-v_R)|_{z=L}+T[f_L(v_L)|_{z=0}-f_R(-v_R)|_{z=L}]. \quad (6b)$$

Equations (6) connect the distribution functions of the electrons on the opposite sides of the tunneling region, and they essentially express the requirement of the particle conservation in the tunneling. These relations supply the conditions for the determination of the functions f_1^L and f_1^R. It should be noted that the relations in Eqs. (6) involve numbers in different parts of the distributions, and thus they do not include the step functions anymore.

Now the electrical current density j in our system may be written in several different forms

$$j=-\frac{em}{\pi A\hbar}\int_0^{+\infty}dv_L\,v_L[f_L(v_L)-f_L(-v_L)]=-\frac{em}{\pi A\hbar}\int_0^{+\infty}dv_R\,v_R[f_R(v_R)-f_R(-v_R)]$$
$$=-\frac{em}{\pi A\hbar}\int_0^{+\infty}dv_{L,R}\,v_{L,R}T[f_L(v_L)|_{z=0}-f_R(-v_R)|_{z=L}]\,, \tag{7}$$

where A is the cross-sectional area of the tunneling junction. The last form of Eq. (7) follows from the fact that $T(E)\equiv 0$, whenever $E<E_c^L(z=0)+eU_T$. We consider first the electrical current density in the left-hand side contact. By solving f_1^L from Eq. (6a) and substituting the distribution functions from Eqs. (2) and (3) into the first form of Eq. (7) we obtain

$$j=-j_0\int_{-\infty}^{z}dz'\frac{d\delta E_F^L}{dz'}e^{-(z-z')/l}+j_0e^{z/l}\int_{-\infty}^{0}dz'\frac{d\delta E_F^L}{dz'}e^{z'/l}+je^{z/l}-j_0\int_z^0 dz'\frac{d\delta E_F^L}{dz'}e^{-(z'-z)/l}\,, \tag{8}$$

where

$$j_0=\frac{em}{\pi A\hbar}\int_0^{+\infty}dv_L\,v_L\left[\frac{\partial f_0^L}{\partial E}\right]=\frac{em}{\pi A\hbar}\int_0^{+\infty}dv_R\,v_R\left[\frac{\partial f_0^R}{\partial E}\right]. \tag{9}$$

Because j must be constant throughout the system (independent of z), Eq. (8) is an integral equation for the gradient of the quasi-Fermi-level in the left contact. It is easily seen that the solution of this equation is

$$\frac{d\delta E_F^L}{dz'}=-\frac{j}{2j_0l}\,, \tag{10}$$

which implies that δE_F^L is a linear function of z as already anticipated in Fig. 1. Of course Fig. 1 exaggerates the magnitude of the gradient, which at normal current densities is a small positive quantity, because j_0 is large and negative. It is easily checked that the solution in Eq. (10) also satisfies the requirement of the density conservation in scattering $\int dv_L(f_L-f_L^{(0)})/\tau=0$. This guarantees the consistency of our treatment with the continuity equation. On the right-hand side we may proceed with the current density in the same fashion. It is evident that the solution for the gradient of the quasi-Fermi-level is now

$$\frac{d\delta E_F^R}{dz}=-\frac{j}{2j_0l}\,. \tag{11}$$

It is no surprise that the gradients of the quasi-Fermi-levels are equal on both sides of the tunneling region, because the contacts were chosen to be made of the same material. The surprising thing might be that the current density equation [like Eq. (8)] does really have such a simple solution.

After the substitution of Eqs. (10) and (11) into the distribution functions, we may apply the last form Eq. (7) in order to calculate the tunneling current density, which becomes

$$j=[1-\tfrac{1}{2}(\langle T_L\rangle+\langle T_R\rangle)]^{-1}\left[-\frac{em}{\pi A\hbar}\int_0^{+\infty}dv_R\,v_R T(f_0^L-f_0^R)+j_0\langle T_L\rangle[\delta E_F^L(0)+e\Phi_L(0)]-j_0\langle T_R\rangle[\delta E_F^R(L)+e\Phi_R(L)]\right], \tag{12}$$

where

$$\langle T_{L,R}\rangle=\frac{1}{j_0}\frac{em}{\pi A\hbar}\int_0^{+\infty}dv_{L,R}\,v_{L,R}T\left[\frac{\partial f_0^{L,R}}{\partial E}\right]. \tag{13}$$

In Eq. (12) the last two terms include the values of the electrostatic potential energy at both boundaries of the tunneling region, and thus the contribution of these terms to the tunneling resistance must be determined via the Poisson equation.

Since the distribution functions are known, we may directly calculate the electronic densities on the both sides of the tunneling region

$$n_{L,R}=\frac{m}{\pi A\hbar}\int_0^{+\infty}dv_{L,R}[f(v_{L,R})+f(-v_{L,R})]\,, \tag{14}$$

which become

$$n_L=n_0+\delta n_L=n_0-2\langle f_0\rangle\left[\delta E_F^L(0)-\frac{j}{2j_0l}z+e\Phi_L(z)\right]-q_Le^{z/l}\,, \tag{15}$$

$$n_R=n_0+\delta n_R=n_0-2\langle f_0\rangle\left[\delta E_F^R(L)-\frac{j}{2j_0l}(z-L)+e\Phi_R(z)\right]+q_Re^{-(z-L)/l}\,. \tag{16}$$

Here n_0 is the equilibrium carrier concentration and

$$\langle f_0\rangle = \frac{m}{\pi A\hbar}\int_0^{+\infty} dv_L \left[\frac{\partial f_0^L}{\partial E}\right] = \frac{m}{\pi A\hbar}\int_0^{+\infty} dv_R \left[\frac{\partial f_0^R}{\partial E}\right], \tag{17}$$

$$q_{L,R} = \frac{j}{j_0}\langle f_0\rangle + \frac{m}{\pi A\hbar}\int_0^{+\infty} dv_{L,R}\, T\left[f_0^L - \left[\frac{\partial f_0^L}{\partial E}\right]\left[\delta E_F^L(0)+e\Phi_L(0)\right] - \frac{j}{2j_0}\left[\frac{\partial f_0^L}{\partial E}\right] - f_0^R + \left[\frac{\partial f_0^R}{\partial E}\right]\left[\delta E_F^R(L)+e\Phi_R(L)\right] - \frac{j}{2j_0}\left[\frac{\partial f_0^R}{\partial E}\right]\right]. \tag{18}$$

Now the solutions of the Poisson equation for the carrier concentrations in Eqs. (15) and (16) are relatively simple, and after little algebra we obtain

$$\Phi_L = A_L e^{\lambda z} - \frac{l^2 e^2 q_R}{\epsilon + 2e^2 l^2\langle f_0\rangle} e^{z/l} + \frac{j}{2j_0 l e} z - \frac{\delta E_F^L(0)}{e} \tag{19}$$

$$\Phi_R = B_R e^{-\lambda(z-L)} + \frac{e^2 l^2 q_R}{\epsilon + 2e^2 l^2\langle f_0\rangle} e^{-(z-L)/l} + \frac{j}{2j_0 l e}(z-L) - \frac{\delta E_F^R(L)}{e}, \tag{20}$$

where ϵ is the permittivity of the contact lattice, $\lambda^2 = 2e^2|\langle f_0\rangle|/\epsilon$, and A_L and B_R are constants, which must be determined from the boundary conditions. Of course, the requirement that the carrier concentrations must be uniform far from the tunneling junction excludes the second, linearly independent solution of the homogeneous equation from Φ_L and Φ_R. Further, the solution of the Poisson equation within the barrier, which is here assumed to be electrically neutral, is simply $\Phi_T = A_T z + B_T$, where again A_T and B_T are constants. The boundary conditions for the electrostatic potential state that the potential and the electric displacement vector must be continuous across the boundaries at $z=0$ and $z=L$.[11] These conditions lead to the expressions for the constants A_L, A_T, B_T, and B_R, namely,

$$A_L = \frac{1}{\epsilon\lambda L + 2\epsilon_T}\left\{-\frac{\epsilon L j}{2j_0 l e} - \epsilon_T U_T + \frac{le}{(1-\lambda l)(1+\lambda l)}\left[q_L\left[\frac{\epsilon_T}{\epsilon} l + L + \frac{\epsilon_T}{\lambda\epsilon}\right] + q_R \frac{\epsilon_T}{\epsilon}\frac{\lambda l - 1}{\lambda}\right]\right\}, \tag{21}$$

$$A_T = \frac{1}{\epsilon\lambda L + 2\epsilon_T}\left[\frac{\epsilon j}{j_0 l e} - \epsilon\lambda U_T - \frac{el(q_R+q_L)}{1+\lambda l}\right], \tag{22}$$

$$B_T = \frac{1}{\epsilon\lambda L + 2\epsilon_T}\left\{-\frac{\epsilon L j}{2j_0 l e} - \epsilon\lambda L\frac{\delta E_F^L(0)}{e} - \frac{\epsilon_T}{e}[\delta E_F^L(0)+\delta E_F^R(L)] + \frac{le}{1+\lambda l}\left[q_L\left[L + \frac{\epsilon_T}{\epsilon\lambda}\right] - q_R\frac{\epsilon_T}{\epsilon\lambda}\right]\right\}, \tag{23}$$

$$B_R = \frac{1}{\epsilon\lambda L + 2\epsilon_T}\left\{\epsilon_T U_T + \frac{L\epsilon j}{2j_0 e l} + \frac{le}{(1-\lambda l)(1+\lambda l)}\left[q_L\frac{\epsilon_T}{\epsilon}\frac{1-\lambda l}{\lambda} - q_R\left[L + \frac{\epsilon_T}{\epsilon} l + \frac{\epsilon_T}{\epsilon\lambda}\right]\right]\right\}, \tag{24}$$

where ϵ_T denotes the permittivity of the tunneling region.

Now, when the potential functions are known, we may proceed to calculate the quantities $X = \delta E_F^L(0) + e\Phi_L(0)$ and $Y = \delta E_F^R(L) + e\Phi_R(L)$. Since q_L and q_R depend on X and Y, we finally end up with a pair of equations

$$X = X(b\langle T_R^L\rangle - a\langle T_L^L\rangle) + Y(a\langle T_L^R\rangle - b\langle T_R^R\rangle) + \alpha a - \beta b - c, \tag{25a}$$

$$Y = X(a\langle T_R^L\rangle - b\langle T_L^L\rangle) + Y(b\langle T_L^R\rangle - a\langle T_R^R\rangle) + \alpha b - \beta a + c, \tag{25b}$$

where

$$a = \frac{1}{\epsilon\lambda L + 2\epsilon_T}\frac{le^2}{(1+\lambda l)\epsilon\lambda}(\epsilon\lambda L + \epsilon_T), \tag{26a}$$

$$b = \frac{1}{\epsilon\lambda L + 2\epsilon_T}\frac{le^2\epsilon_T}{(1+\lambda l)\epsilon\lambda}, \tag{26b}$$

$$c = \frac{e}{\epsilon\lambda L + 2\epsilon_T}\left[\epsilon_T U_T + \frac{\epsilon L j}{2j_0 l e}\right], \tag{26c}$$

$$\alpha = \frac{j}{j_0}[\langle f_0\rangle - \tfrac{1}{2}(\langle T_L^L\rangle + \langle T_L^R\rangle)] + \frac{m}{\pi A\hbar}\int_0^{+\infty} dv_L\, T(f_0^L - f_0^R), \tag{26d}$$

$$\beta = \frac{j}{j_0}[\langle f_0\rangle - \tfrac{1}{2}(\langle T_R^L\rangle + \langle T_R^R\rangle)] + \frac{m}{\pi A\hbar}\int_0^{R} dv_R\, T(f_0^L - f_0^R), \tag{26e}$$

$$\langle T_k^p \rangle = \frac{m}{\pi A \hbar} \int_0^{+\infty} dv_k T \left[\frac{\partial f_0^p}{\partial E} \right], \quad k,p = L,R \ . \tag{26f}$$

In Eq. (26f) both k and p can independently take the values R and L. With these notations the solutions of Eqs. (25) read

$$\delta E_F^L(0) + e\Phi_L(0) = \frac{a\alpha - b\beta - c + (a^2-b^2)(\alpha\langle T_R^R\rangle - \beta\langle T_L^R\rangle) + c(a+b)(\langle T_L^R\rangle - \langle T_R^R\rangle)}{1 + a(\langle T_L^L\rangle + \langle T_R^R\rangle) - b(\langle T_R^L\rangle + \langle T_L^R\rangle) + (a^2-b^2)(\langle T_R^R\rangle\langle T_L^L\rangle - \langle T_R^L\rangle\langle T_L^R\rangle)} \tag{27a}$$

$$\delta E_F^R(L) + e\Phi_R(L) = \frac{\alpha b - \beta a + c + (a^2-b^2)(\alpha\langle T_R^L\rangle - \beta\langle T_L^L\rangle) - c(a+b)(\langle T_L^L\rangle - \langle T_R^L\rangle)}{1 + a(\langle T_L^L\rangle + \langle T_R^R\rangle) - b(\langle T_R^L\rangle + \langle T_L^R\rangle) + (a^2-b^2)(\langle T_R^R\rangle\langle T_L^L\rangle - \langle T_R^L\rangle\langle T_L^R\rangle)} \ , \tag{27b}$$

and, in order to obtain the exact current voltage characteristics of the tunneling junction, we must substitute Eqs. (27) into Eq. (12), which then gives our final result for the tunneling current density.

III. DISCUSSION

In our treatment of the current density in the tunneling junction we have closely followed the idea that in the resistivity problem the current flow is the causal agent and the voltage drop and the inhomogeneities of the carrier density develop from the continuous flow against obstacles. In the transport theory this approach was originally established by Landauer. The viewpoint clearly emphasizes the duality of the electric field and current, which is well known in the circuit theory, but it also lays stress on the operational realization of the four-point resistivity measurement, where the outer contacts are used to drive the current through the sample and the inner probes measure the voltage drop between any two points within the sample.

By looking at the result for the tunneling current density in Eq. (12) it is noted that the first term on the right-hand side is the direct generalization of the celebrated Landauer conductance formula into the regime of nonlinear current-voltage characteristics. This fact is easily seen by taking the linear term in the voltage across the barrier from our result and comparing it with the linear formula as expressed for instance in Ref. 2. The additional correction terms on the right-hand side of Eq. (12) give the contribution of the charge inhomogeneities of the contact regions to the tunneling current density. Clearly, if the electronic systems of the contacts were homogeneous, the contribution of these terms would be exactly zero, because $\delta E_F + e\Phi$ equals the change of the chemical potential $\delta\mu$. Also in the homogeneous case the voltage across the tunneling barrier would be the same as the difference of the electrostatic potential $\Phi(0) - \Phi(L)$. By Eqs. (19) and (20) it is evident that the charge accumulation or depletion near the tunneling region originates from the screening of the electric field inside the barrier and from the need to match the parts of the electronic distribution functions on the opposite sides of the tunneling barrier. The latter process requires a spatial distance of a few mean free paths, and this is also in principle the space occupied by the charge inhomogeneities, because the screening length is generally much shorter than the mean free path.

In order to obtain a more transparent view of the correction term in Eq. (12), we must consider the relative magnitudes of the various terms in Eqs. (27). Although our result is exactly correct in the one-dimensional velocity space only, it is obvious that the physics of the problem will remain essentially unchanged if the calculation is carried out in the three-dimensional velocity space.[9,12] Consequently, when considering the magnitudes in our results, we will use for the corresponding quantities the typical values of the three-dimensional velocity space, namely: $L = 50$ Å, $l = 500$ Å, $\epsilon = \epsilon_0, \ldots, 10\epsilon_0$, $\epsilon_T = 2\epsilon_0$, and $n_0 \approx 10^{21}$ cm^{-3}, which implies that $\lambda^{-1} = L_D \simeq 1$ Å and $\mu_0 \approx 4$ eV. Then the denominator in Eqs. (27) is to a very good approximation equal to unity. Moreover, it is found that in the numerators of Eqs. (27) c is always the dominant term. Thus, we have $\delta E_F^L(0) + e\Phi_L(0) \approx -\delta E_F^R(L) - e\Phi_R^L(L) \approx -c$, which is exactly the result that we would have obtained if we had set $q_L = q_R = 0$ in Eqs. (15) and (16). This is to indicate that the charge inhomogenities are essentially created by the screening effect, and the matching of the distribution functions gives a negligible contribution.

In order to continue the analysis we note that the magnitude of c is, in general, determined by $e\epsilon_T U_T / (\epsilon\lambda L + 2\epsilon_T)$, which is a positive quantity. Then $\delta\mu_L(0) = \delta E_F^L(0) + e\Phi_L(0)$ is negative, whereas $\delta\mu_R(L)$ is positive, which implies that on the right-hand side we have the accumulation of the electrons and the conduction-band edge E_c^R bends downwards near the tunneling barrier. On the other hand, on the left we have the depletion of the charge carriers and the conduction-band edge bends upwards near the point $z = 0$. Consequently, there is a narrow sheet of positive charge near $z = 0$ and of negative charge near $z = L$, which serve as the source and the sink of the high electric field within the tunneling region. Now, it is evident, that the general picture of the charge inhomogeneities and the band diagram in our result is in accord with the earlier discussion by Landauer.[6,7]

If we keep the approximation, that $q_L = q_R = 0$, the correction terms in Eq. (12) may simply be written as $-j_0 c(\langle T_R\rangle + \langle T_L\rangle)$. Now, because j_0 is negative and the expectation values of the transmission coefficient are positive, it is evident that the correction terms tend to increase the tunneling current and, even in the linear voltage regime, the correction is nonvanishing. Clearly, in the linear theory the energy integrals of the first term and the correction become identical, and the only difference between the terms is that the first term is multiplied by eU_T and the correction by $2c$. For well-conducting contact materials (metals and degenerate semiconductors) the screening length is nearly always much shorter than the tunneling barrier, whence the first term in Eq. (12) gives the dominant contribution to the tunneling current density, which is the original result by Landauer. Also in this case $\Delta\Phi$ and U_T are almost equal. On the other hand, in nondegenerate semiconductors the screening lengths may be quite large. For instance $n_0 \approx 10^{16}\,\mathrm{cm}^{-3}$ implies that at the room temperature $L_D \approx 400$ Å, which certainly exceeds the width of any tunneling barrier. Thus, if $L_D \gg L$, in the linear theory the correction becomes also proportional to eU_T, and the tunneling current density is twice the original Landauer result. Then also $\Delta\Phi$ is much smaller than U_T, and the charge inhomogeneities on both sides of the barrier extend deep into the contact regions. In passing it should be noted that our expression on the division of the change of the electrostatic potential across the tunneling barrier and the depletion and accumulation layers is a slight generalization of the old result by Ku and Ullman, who studied the capacitance of thin, nonconducting dielectric films.[13]

It is obvious, that the linearized treatment of the Boltzmann equation in the contract regions prevents our expressions from being completely general semiclassical results, because the linear theory requires small inhomogeneities and weak electric fields to give a correct picture about the transport processes. In our case this means that the accumulation and depletion of charge near the tunneling barrier and the electric fields within the contacts must remain small, but it does not exclude the possibility that the voltage across the barrier U_T may be large enough to cause nonlinear current-voltage characteristics. Usually, the major part of the applied voltage drops across the tunneling region, and thus, at normal voltages, the electric fields in the contact tend to be moderate. Then the actual requirement for the validity of our results is that charge inhomogeneities near the tunneling junction do not grow large, i.e., the screening effect inside the contacts must be sufficiently strong. Of course, this is equivalent to the statement that $U_T \approx \Delta\Phi$ or that the capacitance of the system is determined by the properties of the insulating layer. However, in the case, that the results are linear in U_T, the strength of screening does not effect the validity, and the expressions remain as explained in the previous paragraph.

Finally, as to the resistance of the total system (contacts and tunneling barrier), it is evident by Eqs. (10) and (11), that the charge inhomogeneities do not affect the resistivity of the contact materials, but the voltage drop across the contacts is directly proportional to the thickness of the contact layers. Consequently, the total resistance is equal to the conventional resistance of the contacts plus the tunneling resistance given by the generalized Landauer formula in Eq. (12).

In conclusion we have self-consistently derived the nonlinear current voltage characteristics for an arbitrary tunneling barrier. A particular emphasis has been put on the appropriate treatment of the electric transport and the charge inhomogeneities within the contact materials. The outcome of our theory shows that the charge carrier accumulation and depletion near the barrier is an essential feature of tunneling systems, and on certain occasions the charge inhomogeneities really have a relevant impact on the magnitude of the tunneling current density.

[1]R. Landauer, IBM J. Res. Dev. **1**, 223 (1957).

[2]H.-L. Engquist and P. W. Anderson, Phys. Rev. B **24**, 1151 (1981).

[3]W. Apel, J. Phys. C **16**, 2907 (1983).

[4]D. C. Langreth and E. Abrahams, Phys. Rev. B **24**, 2978 (1981).

[5]D. J. Thouless, Phys. Rev. Lett. **47**, 972 (1981).

[6]R. Landauer, in *Electrical Transport and Optical Properties of Inhomogeneous Media,—1978 (Ohio State University, Columbus),* Proceedings of the Conference on Electrical Transport and Optical Properties of Inhomogeneous Media, AIP Conf. Proc. No. 40, edited by J. C. Garland and D. B. Tanner (AIP, New York, 1978).

[7]R. Landauer, *Proceedings of the International Conference on Localization, Interaction and Transport Phenomena in Impure Metals,* (Springer, Berlin, 1984).

[8]N. D. Mermin, Phys. Rev. B **1**, 2362 (1970).

[9]S. Eränen and J. Sinkkonen, Phys. Rev. B **32**, 5447 (1985).

[10]N. W. Ashcroft and N. D. Mermin, in *Solid State Physics,* (Holt, Rinehart and Winston, New York, 1976), pp. 244—262.

[11]J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill, New York, 1941), p. 164.

[12]J. Sinkkonen and S. Eränen, Phys. Status Solidi B **134**, 391 (1986).

[13]H. Y. Ku and F. G. Ullman, J. Appl. Phys. **35**, 265 (1964).