Surface scattering in electron tunneling

A. V. Chaplik and M. V. Éntin

Institute of Semiconductor Physics, Siberian Division, USSR Academy of Sciences
(Submitted July 12, 1973)
Zh. Eksp. Teor. Fiz. **65**, 2505–2510 (December 1973)

A theory that takes into account electron scattering from irregularities of the interface between the media is developed for the tunnel current in the threshold region. It is shown that if the initial states corresponds to the top of the valence band of the semiconductor, then the surface scattering alters qualitatively the threshold behavior of the current. In the case of a metal, the functional dependence of the current on the voltage is the same as in an ideal structure, but the coefficient can decrease appreciably.

The elementary theories of phononless tunneling of electrons in solids presuppose that the electron-momentum components are conserved along the interface between the media. This means that scattering from the interface and in the volume of the barrier is neglected. Stratton's [1] phenomenological allowance for the diffuseness of the surface in the case of a metal-dielectric-metal system shows that the change in the tunnel current reduces to a numerical coefficient that ranges from 1 to 4, depending on the diffuseness coefficients of the two interfaces. It is obvious, however, that the role of the scattering processes can become more significant for threshold phenomena, when the tunnel current J undergoes abrupt changes at a certain value of the voltage V_0.

The results of the investigations of the threshold characteristics of tunnel systems are reported in a recently published collection [2]. It is assumed in the cited papers that there is no surface scattering.

In this paper we consider the influence of the geometrical irregularity of the interface on the behavior of the tunnel current as a function of the voltage. We shall show that in a number of cases allowance for the surface scattering leads to a qualitative change in the threshold dependences of the current. In addition, fluctuations in the barrier thickness can greatly increase the current in the tunnel structure in comparison with the equivalent smooth barrier (by equivalent is meant a barrier whose thickness is equal to the average thickness of the real barrier).

In all threshold problems, the final state of the electron is characterized by a small kinetic energy E, and consequently by a large wavelength. We shall assume that the electron wavelength in the final state is much larger than the height of the surface roughnesses.

The initial state in tunneling from the top of the valence band of the semiconductor also corresponds to a large wavelength. In tunneling from the metal, however, the electron wavelength prior to passing through the interface is of the order of the lattice constant. The dimensions of the surface roughnesses are assumed to be large in comparison with the lattice period, and we can therefore use the geometrical-optics approximation to find the wave function in the region of space corresponding to the initial state of the electron.

Let the barrier boundaries be described by the equations $z=u_1(\rho)$, $z=d+u_2(\rho)$, where ρ is a vector in the (xy) plane, d is the average thickness of the barrier, and $u_{1,2}(\rho)$ are random functions with zero mean value and identical, by assumption, binary correlation functions $W(\rho)$. In the regions $-\infty<z<u_1(\rho)$ (I) and $d+u_2(\rho)<z<\infty$ (III) the electron is described by effective masses m_1 and m_3, and the edges of the bands are located at the points $-U_0$ and 0 ($U_0>0$), respectively. In region II inside the barrier, we neglect the electric field F, a procedure justified if the parameter $m_e(eFd^2)^2$ is much smaller than the distance Δ to the boundary of the forbidden band of the dielectric (e and m_e are the electron charge and mass).

The traditional approach to tunneling theory is to regard the dielectric region as a potential barrier for the electrons. This model contains two parameters, the barrier height Δ and the electron effective mass m_2, in terms of which the transparency coefficient and the tunnel current is expressed. It is easy to show, however, that in the threshold region we can obtain the same results without resorting to these model concepts.

Indeed, the damping κ of the wave function in the z direction is determined by analytically continuing the dispersion law that holds in the dielectric into the region of complex p_z:

$$\mathrm{Im}\,p_z=\varkappa,\ \varepsilon(p_\parallel,\ p_z)=E.$$

The energy pertains in this case to the forbidden band. We assume that the given value E belongs to only one branch of the spectrum, i.e., there is no "intersection" of the forbidden bands of the dielectric. The main contribution to the tunnel current is made by electrons corresponding to the wave functions that attenuate least along the coordinate z. The quantity κ, regarded as a function of $p_\parallel$ (at a fixed energy), can therefore be expanded near its minimum:

$$\varkappa(\mathbf{p}_\parallel,\ E)=\varkappa_0(E)+\mathrm{const}\,p_\parallel^2.$$

For simplicity we assume the minimum to be isotropic and located at $p_\parallel=0$. In the Schrödinger equation

$$\varepsilon\left(-i\nabla_\parallel,\ -i\frac{\partial}{\partial z}\right)\psi=E\psi$$

for the envelope function ψ we can carry out expansion in the vicinity of the point 0, 0, $\kappa_0(E)$ in the Brillouin zone.

If the barrier is broad enough ($\exp[\kappa_0 d]\gg 1$), then we can confine ourselves in the expansion to the first power of the operator $\partial/\partial z$ (the diffraction approximation). The equation for the wave functions that attenuate in the direction of positive or negative z is

$$\frac{\partial\varepsilon(0,\ \varkappa_0)}{\partial\varkappa}\left(\frac{\partial}{\partial z}\pm\varkappa_0\right)\psi-\frac{1}{2}\frac{\partial^2\varepsilon(0,\ \varkappa_0)}{\partial\mathbf{p}_\parallel^2}\Delta_\rho\psi=0. \qquad (1)$$

It is recognized here that the energy of the tunneling electrons inside the barrier can be regarded near the tunneling threshold as coinciding with the bottom of the conduction band in region III (i.e., equal to zero for our choice of the origin).

The Green's functions of (1) are equal to

$$z^{-1}\exp\left(\mp\varkappa_0 z-\frac{b\rho^2}{2z}\right),\quad b\equiv\frac{\partial\varepsilon}{\partial\varkappa}\frac{\partial^2\varepsilon}{\partial p_\parallel^2}\Big|_{0,\varkappa_0}.$$

This enables us to explain the meaning of the parameters m_2 and Δ in the potential-barrier model. For the case of an isotropic mass we obtain

$$\varkappa_0=(2m_2\Delta)^{1/2},\quad m_2^{-1}=\frac{\partial^2\varepsilon(0,\varkappa_0)}{\partial p_\parallel^2},$$

with $b=\kappa_0$. The wave function and the current density should remain continuous at the boundaries of the regions. Under the assumptions made, this implies the requirement that ψ and $m^{-1}\partial\psi/\partial z$ be continuous.

The complete wave function inside the barrier is represented by a linear combination of the solutions of Eqs. (1). However, the assumption that the barrier has low transparency makes it possible to retain, with exponential accuracy, only one term of this combination in order to find the wave function in the region I. In other words, it is necessary to find the solution in the region $-\infty<z<u_1(\rho)$ by using the Huygens principle (the wavelength is smaller than the roughness), and to assume the barrier to be perfectly impenetrable.

We normalize the wave incident on the barrier by the condition $\psi_{inc}=\exp[i(\mathbf{p}_0\cdot\rho+qz)]$, where $\mathbf{p}_0$ is the tangential component of the momentum,

$$q^2=2m_1(U_0+E)-p_0^2\approx 2m_1U_0-p_0^2,$$

inasmuch as $E\ll U_0$ in the threshold region. To find the reflected wave, we consider a barrier section that is small in comparison with the characteristic diameter or confine ourselves for simplicity to the case of gently-sloping roughnesses with $|\nabla u(\rho)|\ll 1$. Then, accurate to terms of order ∇u^2, we obtain the reflected wave in the form

$$\psi_{ref}=(iqm_2+m_1\varkappa)(iqm_2-m_1\varkappa)^{-1}\exp[ip_0\rho-iq(z-2u(\rho))].$$

The complete solution in region I is

$$\psi=\psi_{inc}+\psi_{ref}.$$

In the same approximation, the wave function inside the barrier can be expressed with the aid of the Green's function:

$$\psi(\rho,z)=\int\frac{\mu_1(\rho')}{z}\exp\left[-\varkappa(z-u_1(\rho'))-\frac{\varkappa(\rho-\rho')^2}{2(z-u_1(\rho'))}\right]ds_1+\int\frac{\mu_2(\rho')}{z-d-u_2(\rho')}\exp\left[-\varkappa(z-d-u_2(\rho'))-\frac{\varkappa(\rho-\rho')^2}{2(z-d-u_2(\rho'))}\right]ds_2 \quad (2)$$

(the subscript 0 of κ_0 will henceforth be omitted; $s_{1,2}$ are the barrier boundaries).

The condition $e^{\kappa d}\gg 1$ enables us to neglect the second term of (2) at $z\sim u_1(\rho)$. The value of the "source density" μ_1 is expressed in terms of the derivative $\partial\psi/\partial z$ outside the layer and is equal to

$$\mu_1=\frac{\varkappa}{\pi}\exp[i(\mathbf{p}_0\rho+qu_1(\rho))]\left[1+\frac{i\varkappa m_1}{qm_2}\right]^{-1}.$$

The "source density" μ_2 at the second boundary is determined by the requirement $\partial\psi/\partial z=0$ at $z=d+u_2(\rho)$. This statement is a consequence of the threshold approximation, namely, when the total energy in region I tends to a value corresponding to the bottom of the conduction band in region III, the kinetic energy on the right of the barrier tends to zero. The electron wavelength corresponding to the z-component of the momentum then increases within limit, which is equivalent to vanishing of the normal derivative of the ψ function. Hence

$$\mu_2=\int\frac{\mu_1(\rho')}{d}\exp\left[-\varkappa(d+u_2(\rho')-u_1(\rho'))-\frac{\varkappa(\rho-\rho')^2}{2d}\right]d^2\rho'.$$

The wave function in the region III can be obtained from its value $z=d+u_2(\rho)$ on the surface, which is determined by Eq. (2), using the standard methods of potential theory. Leaving out the intermediate steps, we present an expression for the averaged partial current in the case of a Gaussian distribution of the random functions $u_{1,2}(\rho)$:

$$\langle j_z(\mathbf{p}_0,k)\rangle=\frac{\varkappa(m_2q)^2e^{-2\varkappa d}}{\pi^3m_2d[(m_2q)^2+(m_1\varkappa)^2]}\times\int(k^2-p^2)^{1/2}\exp\left[-\frac{\varkappa(\rho-\rho')^2}{4d}+(2\varkappa^2-q^2)W(0)+\varkappa^2W(\rho)+(\varkappa^2+q^2)W(\rho')+i(\mathbf{p}_0\rho'-p\rho)\right]dp\,d\rho\,d\rho'; \quad (3)$$

$k^2\equiv 2m_3E$. In order of magnitude we have $W(0)\equiv W_0\sim H^2$, where H is the characteristic height of the roughnesses. The wavelength in region I and II is much smaller than H, i.e., the inequalities κH, $qH\gg 1$ hold. The integrals in (3) can therefore be calculated by the saddle-point method.

We consider the case when $W(\rho)$ has a quadratic behavior as $\rho\to 0$, namely $W(\rho)-W(0)=-\gamma\rho^2$, where $\gamma>0$ (it is easy to show that $W(\rho)$ increases as $\rho\to 0$). The factors containing W in the integrand of (3) take the form

$$\exp(4\varkappa^2W_0)\exp[-\gamma\varkappa^2\rho^2-\gamma(\varkappa^2+q^2)\rho'^2]. \quad (4)$$

The first factor in (4) describes the barrier average-transparency increase due to the thickness fluctuations. It is equal to $\langle\exp[-2\kappa(u_2-u_1)]\rangle$ and should, of course, be smaller than $e^{2\kappa d}$. This limitation has a simple meaning. We describe the thickness fluctuations by a Gaussian ensemble. It is clear at the same time that $d+u_2-u_1$ is always larger than zero (the local thickness of the film is positive). Consequently, the values of u_2-u_1, which play an important role in the calculation of the mean values, should lie in the region of applicability of Gaussian statistics. From this we can readily derive the criterion $2\kappa d-4\kappa^2W_0\gg(8W_0)^{1/2}\kappa$. In the opposite case the fluctuations are so strong that the function is no longer a tunnel junction, since the entire current is determined by sections of practically zero thickness. The second factor in (4) describes effects connected with nonconservation of the tangential momentum.

In the integral of (3) we have competition between two characteristic dimensions in ρ and ρ', namely $(d/\kappa)^{1/2}$ (Fresnel zone) and $(\kappa^2\gamma)^{-1/2}$ (length over which coherence of the ψ function is destroyed by scattering from the surface). It follows therefore that the results will depend significantly on the ratio of these dimensions, i.e., on the value of the parameter $\gamma\kappa d$. In addition, a distinction should be made between tunneling to the bottom of the conduction band of region III from the Fermi level of the metal and from the top of the valence band of the semiconductor. Thus, a number of possibilities arise. We present below the final results for the total density J of the tunnel current at zero temperature (i.e., in the region $kT\ll e(V-V_0)$):

$$J=(2\pi)^{-3}\int\langle j_z(\mathbf{p}_0k)\rangle d^3p=(2\pi)^{-3}\int_0^{V-V_0}dE\int\langle j_z(\mathbf{p}_0k)\rangle\left(\frac{\partial E}{\partial p_z}\right)^{-1}dp_0. \quad (5)$$

1. Tunneling from the metal. If $\gamma\kappa d\ll 1$, the non-

conservation of the tangential momentum can be neglected. The current differs from the corresponding value for an ideal barrier only in the factor $\exp(4\kappa^2 W_0)$. The tangential momentum is conserved in practice likewise in the case $\gamma\kappa d \gg 1$ if the electron momentum to the right of the barrier greatly exceeds the uncertainty of the momentum due to the scattering, i.e., if the condition $V-V_0 \gg 4\kappa^2\gamma/m_3$ is satisfied. Thus, in the region $V-V_0 \ll \kappa/dm_3$, $\gamma\kappa d \ll 1$ we have

$$J=\frac{16\sqrt{2}}{15\pi^2}\frac{em_1\sqrt{m_3}\,m_2^2 p_F}{(m_1\varkappa)^2+(m_2p_F)^2}D(V-V_0)^{5/2}, \tag{6}$$

where p_F is the Fermi momentum in the metal and $D\equiv\exp(-2\kappa d+4\kappa^2W_0)$. In the region $V-V_0 \gg \max[\kappa/dm_3, 4\kappa^2\gamma/m_3]$ we have

$$J=\frac{4\sqrt{2}}{3\pi^2}\frac{em_1m_2^2\varkappa p_F}{\sqrt{m_3}\,d(m_1^2\varkappa^2+m_2^2p_F^2)}D(V-V_0)^{3/2}. \tag{7}$$

The scattering changes the value of the current only in the case when $\gamma\kappa d \gg 1$, $V-V_0 \ll 4\kappa^2\gamma/m_3$;

$$J=\frac{4\sqrt{2}}{15\pi^2}\frac{em_1\sqrt{m_3}\,m_2^2p_F}{\gamma\varkappa d[(m_1\varkappa)^2+(m_2p_F)^2]}D(V-V_0)^{5/2}. \tag{8}$$

2. Tunneling from the top of the valence band of the semiconductor. Far from the threshold, i.e., at $V-V_0 \gg \max[\kappa/m_3d, 4\kappa^2\gamma/m_3]$, we have

$$J=\frac{em_2^2}{2\pi\sqrt{m_1m_3}\,\varkappa d}D(V-V_0)^2. \tag{9}$$

At smaller excesses of the voltage over the threshold, the results depend strongly on the value of the parameter $\gamma\kappa d$: at $\gamma\kappa d \ll 1$, $V-V_0 \ll 4\gamma\kappa^2/m_3$ we have

$$J=e\sqrt{m_1m_3}m_2^2(48\pi\gamma\varkappa^4)^{-1}D(V-V_0)^4; \tag{10}$$

at $\kappa\gamma d \ll 1$, $4\kappa^2\gamma/m_3 \ll V-V_0 \ll \kappa/dm_3$ we have

$$J=e\sqrt{m_1m_3}m_2^2[4\pi\varkappa^2(m_1+m_3)]^{-1}D(V-V_0)^3; \tag{11}$$

and at $\gamma\kappa d \gg 1$, $V-V_0 \ll 4\kappa^2\gamma/m_3$

$$J=e\sqrt{m_1m_3}m_2^2(96\pi d\gamma^2\varkappa^3)^{-1}D(V-V_0)^4. \tag{12}$$

The number of possible limiting cases is larger for the semiconductor than for the metal, inasmuch as in the metal the region of integration with respect to p_0 is always much larger than $\gamma\kappa^2(\gamma\sim\nabla u^2\ll 1, \kappa\sim p_F)$, i.e., than the tangential-momentum uncertainty due to scattering from the surface. In a semiconductor this integration region is of the order of $m_1(V-V_0)$ and can be either larger or smaller than $\gamma\kappa^2$.

Thus we see from (6)–(12) that the behavior of J(V) in the metal, in the immediate vicinity of the threshold voltage, remains unchanged when surface scattering is taken into account, in the sense that $J\sim(V-V_0)^{5/2}$. However, the coefficient in this relation can contain γ, a parameter that characterizes the scattering. In this case the current is much smaller than in an ideal structure. In tunneling from the valence band of the semiconductor, the deviation of the surface from ideal changes the threshold behavior of the current qualitatively, namely $(V-V_0)^4$ instead of $(V-V_0)^3$.

We note finally that we have assumed throughout in this paper that direct transitions are possible. This is true if the structure of the bands to the left and to the right of the barrier is such that the transition does not call for a large change in the tangential momentum (on the order of the reciprocal-lattice vector). It is obvious that in the case of indirect transitions the surface scattering leads to the appearance of "tails" of the current in the region below threshold. The value of the below-threshold current is obviously determined by the value of the Fourier component of the function $\exp(W(\rho)-W(0))$ for momenta on the order of the critical change of the tangential momentum.

[1]R. Stratton, Phys. Rev. **136A**, 837 (1964).
[2]Translation in: Tunnel'noe yavlenie v tverdykh telakh (Tunneling in Solids), Mir, 1973, Chaps. 4, 5, 8, and 13.

Translated by J. G. Adashko
260