Emission of short-wavelength phonons in tunneling through Schottky barriers

L. S. Braginsky, E. M. Baskin, A. P. Kovchavtsev, G. L. Kuryshev, K. O. Postnikov, and I. M. Subbotin
Institute of Semiconductor Physics, 630090 Novosibirsk, Russia
(Received 16 December 1994)

It is shown that the short-wavelength phonons (the phonons whose wave vectors correspond to the edge points of the Brillouin zone) should be emitted in the ideal Schottky barriers in the direct-band-gap semiconductors. The reason for their not having been observed in the tunneling experiments is smoothing of the metal-semiconductor interface on the scale larger than the lattice constant. Two mechanisms of such a smoothing are analyzed. They are the influence of the image charge interaction at the interface and the mutual diffusion of atoms on the boundary between the noble or transition metal on the $A_{\rm III}B_{\rm V}$ semiconductor. The possible observations of short-wavelength phonons in the experiments are discussed. The theoretical results are verified in the experiments on Pt-InAs and Au-superthin-layer-InAs structures.

I. INTRODUCTION

It is well known that the short-wavelength phonons (both acoustical and optical ones, whose wave vector q corresponds to some particular points of the Brillouin zone) are observed in the inelastic tunneling spectroscopy (IETS) experiments in the metal-oxide-metal structures, while only the optical phonons with zero wave vectors are observed in the similar experiments in the metal-degenerate direct-band-gap semiconductor structures.[1] We are not aware of any papers where this difference has been explained. The possible explanation might be connected with the chance for an electron to emit the short-wavelength phonon ($q \sim \pi/a_0$, where a_0 is the lattice constant) in an electrode after the tunneling. The electron has a very strong change of its momentum then (by the phonon quasimomentum), remaining near the Fermi surface at the same time, because the phonon energy is usually much less than the Fermi one, ε_F. This is possible in metals (where $2p_F > \hbar q$, p_F is the Fermi momentum) and impossible in semiconductors (where $2p_F \ll \hbar q$). Nevertheless, how do we explain then the results of the experiments[2] where the short-wavelength phonons were observed in Au-superthin oxide-InAs, or[3] where such phonons were observed in n-type GaAs? A possible explanation is presented in this paper.

We show that the possibility of emission of the short-wavelength phonons under the tunneling is determined not only by the bulk properties of materials under study, but mainly by the properties of an interface between them. It is the shape of the potential barrier through which the tunneling takes place that is the most important. In the MIM (metal-insulator-metal) structures the potential barriers are high (a few eV), but thin (20–50 Å). On the contrary, in semiconductor structures the barriers are not high (a few tenths of an eV), but wide (100–1000 Å). Then the different imperfections at the interface[4] as well as an electrical image charge, which could not appreciably affect the barrier transparency, make its profile smooth and so the short-wavelength phonon emission becomes improbable.[5]

The problem of the short-wavelength phonon emission has been considered in Ref. 5. It has been shown that the potential barrier through which the tunneling occurs should be abrupt in order for the short-wavelength phonon emission to become possible. This signifies that the mean length for the potential deviation should be of the order of the phonon wavelength. The following qualitative argument is helpful to understand the reason. The classical force acting on the tunneling electron is $F = -\nabla U$ [where $U(z)$ is the potential barrier shape]. The force of the same magnitude acts on the crystalline lattice and might lead to the emission of the phonon whose quasimomentum should not exceed $\hbar q = F\Delta t$. Should we estimate the interaction time as $\Delta t \sim \hbar/U_0$ (where U_0 is the mean height of the barrier), then $q \sim |\nabla U|/U_0$. Let l be the mean length of the potential deviation, i.e., $|\nabla U| \sim U_0/l$. Then $q \sim l^{-1}$. In other words, only the phonons with wavelength $\lambda \geq l$ could be emitted under the tunneling. It is clear that l is of the order of the barrier width if the barrier is smooth. Otherwise, the magnitude of l might be smaller. So, for instance, for the barrier containing the short-range impurities the magnitude of l is of the order of impurity range or the scattering length.[5] Thus, in order to observe the emission of the phonons with wavelength λ of the order of a_0, a barrier with an essential potential change on the scale of about a_0 has to be created. The latter condition is not usually satisfied for the semiclassical barriers.

This paper is organized as follows. In Sec. II we present the general expression for the second derivative of the tunneling current from Ref. 5 and deduce I'' for the Schottky barrier supposing the single band approximation to be valid. In Sec. III we estimate an influence of an electrical image charge on I''. The results obtained in Sec. II and Sec. III are generalized in Sec. IV where we work outside the scope of the single band approximation. The experimental results on the IETS on Pt-InAs and Au-superthin-oxide-InAs structures are presented in Sec. V. The results are summarized in Sec. VI.

0163-1829/95/51(24)/17718(10)/$06.00 51 17 718

II. THE AMPLITUDE OF THE SECOND DERIVATIVE OF THE TUNNELING CURRENT FOR THE IDEAL SCHOTTKY BARRIER

The expression for the second derivative of the tunneling current with respect to bias d^2I/dV^2 has been obtained in Ref. 5:

$$\frac{d^2I}{d(eV)^2} = 2m_1m_2e\,\frac{\mathcal{V}^3}{(2\pi)^9\hbar^7} \times \int \frac{|T_{p_1p_2}|^2}{p_{1\perp}p_{2\perp}} \frac{\Gamma(\hbar\omega - eV)}{|\nabla_{\mathbf{q}}\omega|} \times d\omega\, d^2\mathbf{q_s}\, d^2\mathbf{p}_{1\|}\, d^2\mathbf{p}_{2\|} \; . \tag{1}$$

Subscripts 1 and 2 here and later on denote the different sides of the barrier. Subscript 1 corresponds to metal and 2 to semiconductor, m_1 and m_2 are the effective masses of an electron, $\mathcal{V}$ is the normalized volume, $T_{p_1p_2}(\mathbf{q})$ is the transfer matrix element (it will be determined afterwards), $p_{1\perp}$ and $p_{2\perp}$ are the normal-to-the-barrier plane components of an electron momentum, $\mathbf{p}_{1\|}$ and $\mathbf{p}_{2\|}$ are the parallel-to-the-barrier plane components: $p_{1\perp} = \sqrt{2m_1E_1 - p_{1\|}^2}$, $p_{2\perp} = \sqrt{2m_2E_2 - p_{2\|}^2}$, E_1 and E_2 are the electron energies, and $\omega = \omega(\mathbf{q})$ is the frequency of the phonon under study.

$$\Gamma(\hbar\omega - eV) = \frac{d^2}{d(eV)^2}\int_{-\infty}^{+\infty}\!\!\int f(E_1)\,[1 - f(E_2 + eV)] \times \frac{\gamma}{(E_1 - E_2 - \hbar\omega)^2 + \gamma^2} dE_1\, dE_2 \; . \tag{2}$$

$\gamma = \hbar/2\tau$, where τ is the free path time for an electron in the electrode. In the Schottky barriers τ is usually determined by the scattering on the impurities in the semiconductor.

$$f(E) = \left[e^{\frac{E-\mu}{k_B\Theta}} + 1\right]^{-1}$$

is the Fermi distribution function, μ is the Fermi level in the metal, k_B is the Boltzmann constant, and Θ is the temperature. Integration in Eq. (1) has to be done over the region inside the Fermi surfaces on $\mathbf{p}_{1\|}$,$\mathbf{p}_{2\|}$ and over the constant frequency surface $\omega(\mathbf{q}_s) = eV/\hbar$ on $\mathbf{q}_s$.

It is clear from Eq. (1) that $d^2I/d(eV)^2$ is not proportional to the phonon density of states $G(\omega)$,

$$G(\omega) = \frac{\mathcal{V}}{(2\pi)^3N}\int_0^\infty \frac{d^2\mathbf{q}_s}{|\nabla_{\mathbf{q}}\omega|},$$

N is the number of sites in the crystalline lattice. The appropriate expression could be obtained from (1) only if we assume that the main input to the second integral (1) comes from the vicinities of the singular points in the Brillouin zone where $\nabla_{\mathbf{q}}\omega(q) = 0$, and where the first integrand could be regarded as unchanged with respect to $\mathbf{q}$. Then we can write

$$\frac{d^2I}{d(eV)^2} = 2m_1m_2eN\,\frac{\mathcal{V}^2}{(2\pi)^6\hbar^7}\int \frac{|T_{p_1p_2}|^2}{p_{1\perp}p_{2\perp}} d^2\mathbf{p}_{1\|}\, d^2\mathbf{p}_{2\|} \times \int_0^{+\infty} G(\omega)\Gamma(\hbar\omega - eV)\, d\omega, \tag{3}$$

where the first integral is supposed to be determined in these particular points $\mathbf{q} = (\mathbf{q}_\|, q_\perp)$, and $\mathbf{q}_\|$, $q_\perp$ are its components along the barrier plane and perpendicular to it, respectively. Such a restriction is useful for the qualitative analyses of the situation under study. Nevertheless, in general the peak shape in the I'' spectrum should not be the same as that of the corresponding peak in the phonon density of states. The following assumptions had also been made when Eq. (1) was obtained.

(a) The parameters γ and $k_B\Theta$ are supposed to be the smallest parameters of the problem, i.e., they are much lower than U_0 and $\hbar\omega$.

(b) The barrier is supposed to be of low transparency. This means, first, that the transfer Hamiltonian approximation is valid, and second, that the bias occurs only across the barrier region, so that the electron gases are in equilibrium on both sides of the barrier. Therefore the Fermi distribution is valid for the electron on each side of the barrier, while its Fermi levels are separated by eV.

(c) In addition, it was supposed that I'' has a sharp peak at $eV = \hbar\omega$ and (1) is really valid when eV is in the vicinity of this peak. It becomes possible under this assumption to present (3) as a product of two integrals: the first of them determines the amplitude of the peak and the second one determines its shape. Indeed, it could be understood from (2) that $\Gamma(\hbar\omega - eV)$ as a function of eV has a narrow peak with a width equal to $\max(\gamma, k_B\Theta)$. If, in addition, $G(\omega)$ is a peaklike function near some ω then

$$\Phi\left(\frac{e}{\hbar}V\right) = \frac{\hbar}{\pi}\int G(\omega)\Gamma(\hbar\omega - eV)\, d\omega$$

is a peaklike function as well, and its width is equal to $\max[\gamma, k_B\Theta, W_{G(\omega)}]$, where $W_{G(\omega)}$ denotes the width of $G(\omega)$. Thus, $\Phi(eV) \approx G(eV)$, if $\gamma, k_B\Theta \to 0$; i.e., in this case I'' has the same shape as the peak in the density of phonon states.

The departure from the latter assumption should lead to an additional term in Eq. (1), which has no singularity at $eV = \hbar\omega$. It determines the $I''(eV)$ dependence in the region apart from the peaks. More precisely, this additional term outlines the background in the IETS spectrum, while Eqs. (1) and (3) determine the position of the peaks and their shape.

The amplitude of the peak is determined by the transfer matrix element

$$T_{p_1p_2} = g(\omega, q)\int \psi_r^*\psi_l e^{i\mathbf{q}\cdot\mathbf{r}}\, d^3r. \tag{4}$$

Here $g(\omega, q)$ is the constant of the electron-phonon interaction, dependent on its mechanism, and ψ_r and ψ_l are the wave functions. In the transfer Hamiltonian approximation they are determined in the following way: ψ_r should obey the Schrödinger equation in the barrier region and to the right of it and decay toward the left, then $z \to -\infty$; ψ_l should obey the Schrödinger equation in the barrier region and to the left side of it and decay toward the right, then $z \to +\infty$. We have adapted the Cartesian coordinate system where the plane XY is the barrier plane and set $z = 0$ in the metal-semiconductor interface (Fig. 1).

The smallness of $T_{p_1 p_2}$ and therefore the inelastic component of the tunneling current arise due to the oscillatory exponential function in the integrand (4). The integral (4) becomes exponentially small if the wave functions ψ_r and ψ_l do not contain sufficiently high-frequency Fourier components, as occurs in the smooth semiclassical barriers. We have shown in Ref. 6 that such high-frequency components should arise in the wave functions at the abrupt interface. It becomes possible if the potential $U(z)$ or an effective mass has a jump discontinuity at the interface. We shall see that this situation really occurs in the Schottky barriers as well.

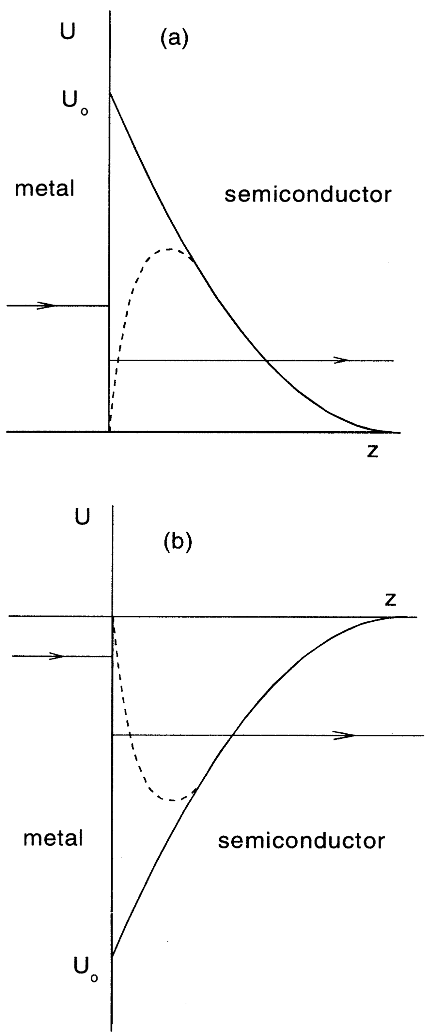

FIG. 1. The Schottky barrier profile at the interface (a) metal–n-type semiconductor and (b) metal–p-type semiconductor. The dotted line shows the alternation of the barrier due to the image charge interaction. The arrows schematically indicate the tunneling path of the electron which emits the short-wavelength phonon at the interface $z = 0$.

To obtain the solution of the Schrödinger equation in the contact region it is important to impose the proper matching conditions connecting the envelope wave functions from both sides of the interface. Usually, the so-called envelope-function approximation (EFA) is exploited. The envelope wave function as well as its derivative divided by effective mass are supposed to be continuous through the interface in this approximation. The EFA is shown to be valid if the interface is smooth, so that the effective band parameters of the electron are not changed substantially on the scale of about the lattice constant.[6] The applicability of this approximation was also numerically verified for a number of the abrupt heterojunctions of semiconductors with the same symmetry ($GaAs/Al_xGa_{1-x}As$, HgTe/CdTe, GaSb/InAs).[7] Nevertheless, it was shown that the EFA is not valid in the simple models if the contacted materials are rather different.[8,9] Thus the applicability of the EFA for the Schottky barriers and metal-oxide-semiconductor (MOS) structures is not clear yet if the interface is abrupt. This situation is most important when the short-wavelength phonon emission is studied.

For this reason we exploit the more general matching condition:

$$\begin{pmatrix} \psi(+0) \\ \psi'(+0) \end{pmatrix} = \widehat{T} \begin{pmatrix} \psi(-0) \\ \psi'(-0) \end{pmatrix}. \tag{5}$$

Here $\psi(\pm 0)$ and $\psi'(\pm 0)$ are the envelopes and their derivatives with respect to z on each side of the interface $z = 0$, $\widehat{T} = \|t_{ik}\|$ is the 2×2 matrix where t_{ik} are real, and $\det \|t_{ik}\| = m_2/m_1$. The latter condition is needed to ensure the flux conservation. In contrast to the EFA conditions (5) implies a jump discontinuity in the envelope wave function as well. The EFA follows from (5) when $t_{11} = 1$, $t_{12} = t_{21} = 0$, and $t_{22} = m_2/m_1$. The matching condition (5) is not the most general one. Nevertheless, it could be imposed if the effective-mass approximation is valid on both sides of the interface.[9] We shall see later that the matching conditions may affect the preexponent factor in the expression for the tunneling current only. So the qualitative results obtained in this paper are valid as well if the condition (5) is not applicable. Nevertheless, the applicability of the EFA for the Schottky barriers might become clear if the coefficients t_{ik} were to be measured in the tunneling experiments.

Let the potential barrier through which the tunneling takes place be of the form

$$U(z) = \begin{cases} 0 & \text{if } z < 0 \\ U_0 \left(1 - \frac{z}{d}\right)^2 & \text{if } 0 < z < d \\ 0 & \text{if } z > d. \end{cases} \tag{6}$$

The term $U(z)$ (6) has to be considered as the potential energy in the Schrödinger equation for the barrier with an n-type semiconductor if an energy is measured from the bottom of the conduction band [Fig. 1(a)] or with a p-type semiconductor if an energy is measured from the top of the valence band [Fig. 1(b)]. It is only important that the single band approximation is applicable. We suppose the semiconductor to be degenerate, so that the electron with a momentum equal to $\mathbf{p}_1$ ($p_1 \simeq p_F$) in the metal turns out to be in the semiconductor in the state with a much smaller momentum $\mathbf{p}_2$ ($p_2 \simeq p_0$). Here p_F and p_0 are Fermi momenta in the metal and semiconductor, respectively. The difference of the momentum components $\mathbf{q}_{\|} = \mathbf{p}_{1\|} - \mathbf{p}_{2\|}$ is carried out by the phonon. Only the phonons with the momenta corresponding to the peaks in the phonon density of states can be observed in the IETS.

Using the semiclassical approximation in the $z > 0$ region, we can write

$$\psi_l = \begin{cases} \sqrt{\frac{2}{\pi\mathcal{V}}}\sin\left(\frac{1}{\hbar}p_{1\perp}z + \delta\right)\exp\left(\frac{i}{\hbar}\mathbf{p}_{1\|}\cdot\mathbf{r}\right) & \text{if } z < 0 \\ \frac{m_2 p_{1\perp}\kappa_1}{m_1\sqrt{(t_{11}\kappa_1 + t_{21})^2 + (t_{12}\kappa_1 + t_{22})^2 p_{1\perp}^2}}\sqrt{\frac{2}{\pi\mathcal{V}|p_1|}}\exp\left[-\Gamma_1(z) + \frac{i}{\hbar}\mathbf{p}_{1\|}\cdot\mathbf{r}\right] & \text{if } z > 0, \end{cases}$$

$$\psi_r = \begin{cases} \frac{1}{2}\sqrt{\frac{2p_{2\perp}}{\pi\mathcal{V}|p_2|}}\exp\left[\Gamma_2(z) - \Gamma_2(b) + \frac{i}{\hbar}\mathbf{p}_{2\|}\cdot\mathbf{r}\right] & \text{if } z < b \\ \sqrt{\frac{2p_{2\perp}}{\pi\mathcal{V}p_2}}\cos\left[\frac{1}{\hbar}\int_{b_2}^{z} p_2\,dz - \frac{\pi}{4}\right]\exp\left(\frac{i}{\hbar}\mathbf{p}_{2\|}\cdot\mathbf{r}\right) & \text{if } z > b. \end{cases} \tag{7}$$

Here $p_{1,2} = \sqrt{2m_2[E_{1,2} - U(z)] - p_{1,2\|}^2}$ is the subbarrier momentum, $\mathbf{r}$ is the radius vector in the barrier plane XY,

$$\kappa_1 = \sqrt{2m_2(U_0 - E_1) + p_{1\|}^2},\ \Gamma_{1,2}(z) = \frac{1}{\hbar}\int_0^z |p_{1,2}|\,dz,$$

$$\tan\delta = -\frac{(t_{12}\kappa_1 + t_{22})p_{1\perp}}{t_{11}\kappa_1 + t_{21}}.$$

To obtain the transfer matrix element T the wave functions (7) should be inserted in Eq. (4). We must take into account that the semiclassical approximation becomes invalid in the vicinity of the turning point. Let T_1 be the contribution to (4) from the integration region before the turning point b and T_2 be that from the region after the turning point, i.e.,

$$T_1 = \frac{4\pi\hbar^2 g m_2 p_{1\perp}\sqrt{p_{2\|}\kappa_1}\,\delta(\mathbf{p}_{1\|} - \mathbf{p}_{2\|} - \hbar\mathbf{q}_{\|})}{m_1\sqrt{(t_{11}\kappa_1 + t_{21})^2 + (t_{21}\kappa_1 + t_{22})^2 p_{1\perp}^2}}\frac{1}{\mathcal{V}}\int_0^{b-\delta}\frac{\exp\left[\Gamma_2(z) - \Gamma_2(b) - \Gamma_1(z) + iq_\perp z\right]}{\sqrt{|p_1||p_2|}}\,dz,$$

$$T_2 = \frac{8\pi\hbar^2 g m_2 p_{1\perp}\sqrt{p_{2\|}\kappa_1}\,\delta(\mathbf{p}_{1\|} - \mathbf{p}_{2\|} - \hbar\mathbf{q}_{\|})}{m_1\sqrt{(t_{11}\kappa_1 + t_{21})^2 + (t_{21}\kappa_1 + t_{22})^2 p_{1\perp}^2}}\frac{1}{\mathcal{V}}\int_{b+\delta}^{\infty}\frac{\exp\left[-\Gamma_1(z) + iq_\perp z\right]}{\sqrt{|p_1|p_2}}\,dz.$$

The semiclassical approximation is not valid in the interval $[b - \delta,\ b + \delta]$, but the contribution to the integral from this region is negligible when $\delta \to 0$. This is because the integrands tend to infinity as $\delta^{-1/2}$ whereas the size of the interval is equal to 2δ. It could be shown by integrating T_1 and T_2 by parts that the contribution to $T = T_1 + T_2$ from the vicinity of the turning point vanishes. This is easy to understand. The turning point is distinguished only for the semiclassical method itself. The exact wave function does not have any singularity at this point, so that the contribution to the integral from its vicinity is absent as well. The only point where the short-wavelength phonon could be emitted is in the vicinity (about the phonon wavelength) of the interface $z = 0$ where the potential has a discontinuity, so that

$$T = -\frac{4\pi\hbar^3 g m_2 p_{1\perp}\sqrt{p_{2\|}}e^{-\Gamma_2(b)}}{\mathcal{V}m_1(\kappa_2 - \kappa_1 - i\hbar q_\perp)\sqrt{[(t_{11}\kappa_1 + t_{21})^2 + (t_{21}\kappa_1 + t_{22})^2 p_{1\perp}^2]\kappa_1}}\delta(\mathbf{p}_{1\|} - \mathbf{p}_{2\|} - \hbar\mathbf{q}_{\|}). \tag{8}$$

Here $\kappa_2 = \sqrt{2m_2\,(U_0 - E_2) + p_{2\|}^2}$.

Let us estimate the possibility for the short-wavelength phonon to be observed in the experiments. The expression for the transfer matrix element if the electron is tunneling elastically is as follows:

$$T_{\text{el}} = \frac{2\pi\hbar^3 p_{1\perp}\sqrt{\kappa_1 p_{1\perp}}e^{-\Gamma_1(b)}\delta(\mathbf{p}_{1\|} - \mathbf{p}_{2\|})}{\mathcal{V}m_1\sqrt{(t_{11}\kappa_1 + t_{21})^2 + (t_{12}\kappa_1 + t_{22})^2 p_{1\perp}^2}}.$$

The relative contribution of the inelastic component of the tunneling current could be estimated as

$$\eta = \frac{I_{\rm inel}}{I_{\rm el}} \sim \frac{|T|^2}{|T_{\rm el}|^2} N$$
$$\sim \frac{4|g|^2 N m_2^2}{\kappa_1^2[(\kappa_1-\kappa_2)^2+q_\perp^2]} e^{\frac{2}{\hbar}(\kappa_1-\kappa_2)b} .$$

Should we assume that the main mechanism of the electron-phonon interaction at the $q \sim \pi/a$ is the optical deformational potential, then

$$|g|^2 = \frac{\hbar D^2}{2\rho \mathcal{V} \omega_0},$$

where D is the constant of the optical deformational potential, $D \sim 10^8$–10^9 eV/cm, ρ is the density of the semiconductor, and ω_0 is the phonon frequency. Moreover, if $|\kappa_1 - \kappa_2| \ll \hbar q$ then

$$\eta \sim \frac{D^2 m_2}{\hbar\omega_0 M U_0 q_\perp^2} e^{\frac{2}{\hbar}(\kappa_1-\kappa_2)b}. \tag{9}$$

Here M is the mass of the unit cell of the semiconductor.

The preexponential factor of this expression for the known semiconductors reaches a value of the order of 10^{-3}–10^{-2}. The exponential factor is less than unity for the Schottky barriers with the n-type semiconductor, but is greater than unity if the semiconductor is of the p type. The difference becomes clear when the tunneling of the electrons to the n-type semiconductor [Fig. 1(a)] and p-type semiconductor [Fig. 1(b)] is compared. In both cases the electrons lose energy after the phonon has been emitted. This leads to the increase of the barrier height for the electron in the former case (when $\kappa_1 < \kappa_2$) but to its decrease in the latter case (when $\kappa_1 > \kappa_2$). It is known that the relative contribution of the inelastic component of the tunneling current is of the order of 10^{-3} in the metal-oxide-metal structures.[10] Therefore the short-wavelength phonons should be observed at least in the Schottky barriers with a p-type semiconductor where the effect is greater than that in MOS structures. The value of η is too small, so it is not the tunneling current itself, but its derivatives that are singular at $eV = \hbar\omega$ which are usually measured; in particular, I'', which in our case takes the form

$$\frac{d^2 I}{d(eV)^2} = \frac{2\mathcal{V}}{(2\pi)^7\hbar^3}\frac{e m_2^3 p_0^2}{m_1} S e^{-2\Gamma_2(b)} \int \frac{|g|^2 p_{1\perp}\Gamma(\hbar\omega - eV)}{\kappa_1[(t_{11}\kappa_1+t_{21})^2+(t_{12}\kappa_1+t_{22})^2 p_{1\perp}^2][(\kappa_1-\kappa_2)^2+\hbar^2 q_\perp^2]}\frac{d^2\mathbf{q}_s}{|\nabla_{\mathbf{q}}\omega|}. \tag{10a}$$

Here S is the contact area, $\mathbf{p}_{1\parallel} = \hbar\mathbf{q}_\parallel$, and $\mathbf{p}_{2\parallel} = 0$. If Eq. (3) holds then

$$\frac{d^2 I}{d(eV)^2} = \frac{e m_2^3 |g|^2 N S p_0^2 p_{1\perp}}{(2\pi)^3\hbar^4 m_1\kappa_1[(t_{11}\kappa_1+t_{21})^2+(t_{12}\kappa_1+t_{22})^2 p_{1\perp}^2][(\kappa_1-\kappa_2)^2+\hbar^2 q_\perp^2]} e^{-2\Gamma_2(b)} \cdot \Phi\left(\frac{e}{\hbar}V\right). \tag{10b}$$

$\mathbf{q}_\parallel$ and $q_\perp$ are the integration variables in Eq. (10a), whereas they are values corresponding to the singular points of the Brillouin zone in Eq. (10b). The contact area S has appeared in Eq. (10) as a result of substitution of T from Eq. (8) into Eq. (3). The squared δ function should be evaluated here in the following way:

$$\begin{aligned}\delta^2(\mathbf{p}_{1\parallel} - \mathbf{p}_{2\parallel} - \hbar\mathbf{q}_\parallel) &\equiv (2\pi\hbar)^{-4}\int e^{i/\hbar(\mathbf{p}_{1\parallel}-\mathbf{p}_{2\parallel}-\hbar\mathbf{q}_\parallel)(r'+r'')}\, d^2r'\, d^2r'' \\ &= (2\pi\hbar)^{-2}\delta(\mathbf{p}_{1\parallel}-\mathbf{p}_{2\parallel}-\hbar\mathbf{q}_\parallel)\int d^2r'' \\ &= (2\pi\hbar)^{-2} S\delta(\mathbf{p}_{1\parallel}-\mathbf{p}_{2\parallel}-\hbar\mathbf{q}_\parallel).\end{aligned}$$

Thus the δ function here leads to the conservation of the parallel-to-the-barrier momentum components and the current becomes proportional to the contact area S.

III. INFLUENCE OF THE IMAGE CHARGE ON THE SHORT-WAVELENGTH PHONON EMISSION UNDER TUNNELING

We have shown that the short-wavelength phonons should be emitted under tunneling through the Schottky barrier. However, they are not observed in the experiments.[1] Why? We believe that the various mechanisms which smooth out the potential profile at the interface are responsible for this. The image charge interaction is one such mechanism. Its influence will be estimated in this section.

The Schottky barrier profile (6) is parabolic if an interaction of the electron with its image in the metal is neglected. It is permissible only at a far distance from the metal surface $z = 0$. Otherwise, the image charge interaction has to be taken into account. This alters the barrier profile, so that it takes the form represented in Fig. 1 by the dotted line. The potential then could become too smooth, so that the probability of short-wavelength phonon emission gets exponentially small.

To consider the image charge interaction the additional term has to be involved in the Schottky potential, so that

$$U = U_0\left(1 - \frac{z}{d}\right)^2 - \frac{e^2}{4\epsilon_\infty z}. \tag{11}$$

There ϵ_∞ is the height-frequency dielectric constant of the semiconductor. It should be noted that the dielectric dispersion has to be taken into account when the electric image of the fast electron is being considered. This could be approximately realized if we suppose $\epsilon = \epsilon(\nu)$ in Eq. (11). The characteristic frequency ν can be estimated in the following way. The electric potential of the electron moving with velocity $\mathbf{v}$ in its intrinsic frame of reference is

$$\varphi(\mathbf{R}) = \frac{1}{2\pi^2} \int \frac{e^{i\mathbf{k}\cdot\mathbf{R}}}{k^2 \epsilon(\mathbf{k}\cdot\mathbf{v})} d^3k.$$

The main input to the integral comes from the region where $k \sim \pi/R$, so that $\nu \sim \pi v/R$. In our case $v \sim \sqrt{\frac{2U_0}{m_2}}$ and the distances of the order of $R \sim q^{-1} \sim a_0/\pi$ are the most important. Then $\nu \sim \frac{\pi^2}{a_0}\sqrt{\frac{2U_0}{m_2}}$. Usually the value $\hbar\nu$ does not exceed a few eV and so ϵ_∞ is close to the static dielectric constant. Nevertheless, if by chance the magnitude ν is sufficiently large then ϵ_∞ becomes smaller. This enlarges the influence of the image charge interaction in the problem considered. It is important to note that the distances of the order of the barrier width are distinctive when the problem of the barrier transmittency is being considered. The frequency ν is small enough then. This allows us to use the statistical dielectric constant in this case.

It is clear that inequality $U(z) > 0$ should be satisfied. So, the supposition is usually implied that $U(z)$ obeys Eq. (11) for large enough z only and comes to zero continuously when $z \to 0$. We assume that $U(z)$ is determined by (11) if $z \geq z_0 = \frac{e^2}{4\epsilon U_0}$ and $U(z) = 0$ if $z < z_0$. The approximation we have chosen for the true self-consistent potential leads to facilitation of short-wavelength phonon emission. Indeed, the self-consistent calculation where the electrons, which could be captured in the well near the interface where $U(z) < 0$, are taken into account should lead to $U(z)$ becoming smooth at $z = z_0$.

The derivative of $U(z)$ is maximal at the point z_0, where

$$\frac{dU}{dz} = \frac{4\epsilon_\infty U_0^2}{e^2} = \frac{U_0}{z_0}.$$

Then the short-wavelength phonon emission becomes probable if $\frac{1}{z_0} > q_{\max} \sim \frac{\pi}{a_0}$, i.e.,

$$\frac{4\epsilon_\infty U_0}{e^2} > q_{\max}. \tag{12}$$

Otherwise, the inelastic components of the tunneling current would contain the small factor $\exp\left(-\frac{2\pi}{a_0} z_0\right)$. A similar factor has to arise any time when the interface becomes smooth.[6] It is important to note that z_0 is independent of the profile of the potential barrier. Only its height in the interface region is significant. This concerns also the criterion (12).

Criterion (12) is not satisfied, usually, for the Schottky barriers and p-n-p structures where $U_0 \sim 0.1$ eV and is satisfied in the barriers with a dielectric involved, where U_0 is of the order of a few eV. The dielectric layer inserted into the metal-semiconductor interface should decrease the image charge influence and so result in the emission of the short-wavelength phonons. The potential has a shape as represented in Fig. 2 in this case. The sharp discontinuity of the potential at the dielectric-semiconductor interface should lead to the facilitation of the short-wavelength phonon emission as well. The magnitude of I'' could be obtained from Eq. (10) if its right-hand side were multiplied by $\exp(-w/\lambda_b)$, where λ_b is the subbarrier wavelength in the dielectric and w is its width.

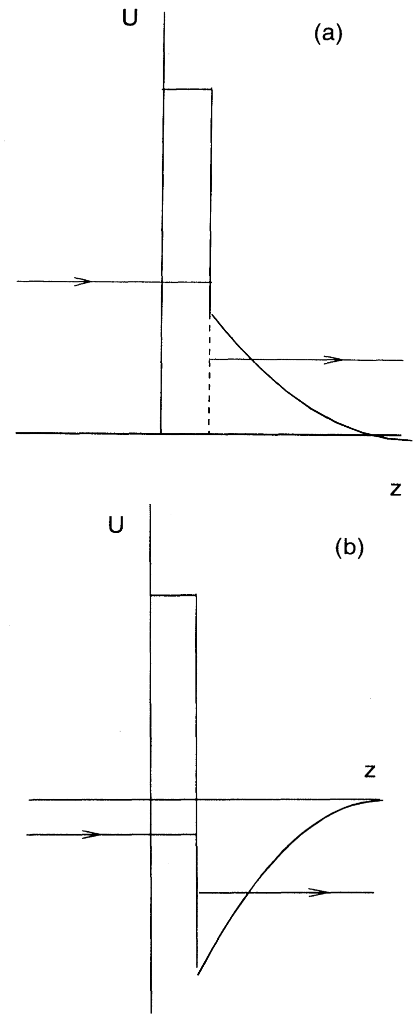

FIG. 2. The barrier profile in MOS structures with (a) n-type semiconductor and (b) p-type semiconductor.

IV. INFLUENCE OF THE SECOND BAND ON THE SHORT-WAVELENGTH PHONON EMISSION

We shall discuss here the emission of the short-wavelength phonons at the metal-p-type-semiconductor interface. The schematic potential versus distance diagram for the contact is presented in Fig. 3. We shall see later that the result substantially depends on the relative position of the band extrema at the contact and the Fermi level.

Let us suppose now that the Fermi level in the metal is situated near the bottom of the conduction band of the semiconductor at the interface [$|E_c(z=0) - \varepsilon_F| \ll E_g$, where E_g is the gap in the semiconductor]. The wave functions ψ_r and ψ_l could be determined from Eq. (7) where the more complicated z dependence of the subbarrier momentum $p_{1,2}$ has to be assumed. The electron then is found near the conduction band when the tunneling began and in the valence band after it. Therefore the influence of both conduction and valence bands should be taken into account.

The subbarrier momentum p_2 could be determined from the equation

$$E_n(p_2) + V_0(z) = E_2, \tag{13}$$

where $E_n(p_2)$ is the band electron energy in the bulk, $V_0(z)$ is the band bending near the interface, and n is the band number. We have assumed that energy is measured from the top of the valence band in the bulk (Fig. 3). The solutions of Eq. (13) are independent of the band number n. Otherwise, the different p_2 arising as the solutions of Eq. (13) for the different n (but for the same E_2) has to exist. This would mean an additional degeneration that is absent in the input Schrödinger equation for an electron in the crystalline lattice, from which Eq. (13) has been obtained. The solutions of Eq. (13) are imaginary if E_2 corresponds to the forbidden gap. The effective-mass approximation is applicable only if E_2 is close to the edge of a band. Then $E_n(p_2)$ could be expanded in power series where only the leading term could be retained. Then at the interface $z = 0$ we can choose $E_n = E_c$ (E_c is the conduction band electron energy). So

$$E_g + V_0(z) + \frac{p_2^2}{2m_e} = E_2,$$

i.e.,

$$|p_2(z)| = \sqrt{2m_e[E_g + V_0(z) - E_2]}. \tag{14}$$

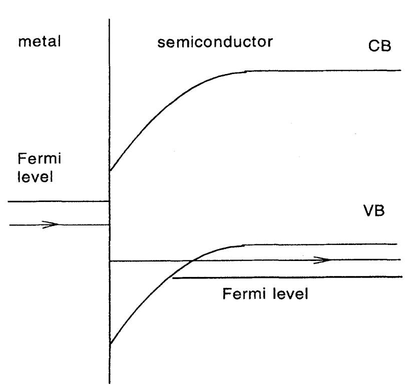

FIG. 3. Schematic potential versus distance diagram for the metal–p-type semiconductor contact.

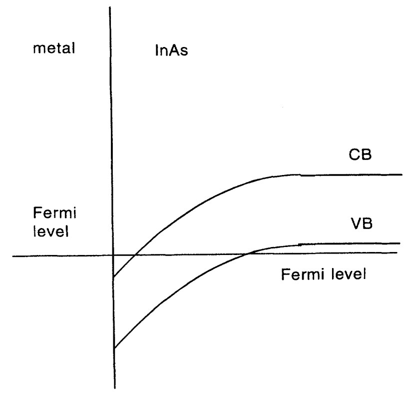

FIG. 4. Schematic potential versus distance diagram for the metal–p-type-InAs contact.

Here m_e is the electron effective mass in the conduction band. Near the turning point $z = b$ we can choose $E_n = E_v$ (E_v is the valence band electron energy), and so

$$V_0(z) - \frac{p_2^2}{2m_h} = E_2,$$

i.e.,

$$|p_2(z)| = \sqrt{2m_h[E_2 - V_0(z)]}. \tag{15}$$

Here m_h is the hole effective mass. The effective-mass approximation is not valid in the intermediate region inside the barrier. Nevertheless, $p_2(z)$ should be a smooth function of z if $V_0(z)$ were also smooth.

If $\varepsilon_F < E_c(z=0)$, then the integration of T_1 and T_2 immediately leads to (8), (10) with $m_2 = m_e$. This is because the main contribution to (8) comes from the vicinity of interface $z = 0$ where the potential has the jump discontinuity. This discontinuity is equal to $E_g + V_0(0) - \varepsilon_F$, and should become smooth due to the image charge. The necessary condition for short-wavelength phonons to be emitted is

$$\frac{1}{z_0} = \frac{2\epsilon_\infty[E_g + V_0(0) - \varepsilon_F]}{e^2} > q_{\max} \sim \frac{\pi}{a_0}. \tag{16}$$

If $\varepsilon_F > E_c(z=0)$ (Fig. 4, the situation really occurs in p-type InAs, for instance), then the potential is smooth everywhere and $p_2(z)$ is smooth too. This should lead to the inelastic tunnel matrix element (8) being exponentially small. The exponent should be of the order of $-q_{\max}w$, where w is the barrier width ($q_{\max}w \gg 1$), so it is the only distinctive parameter of the potential profile.

V. THE INFLUENCE OF THE POTENTIAL BARRIER PROFILE ON THE SHORT-WAVELENGTH PHONON EMISSION UNDER THE TUNNELING. THE EXPERIMENTAL TEST

The sharpness of the potential barrier profile is necessary for observation of the short-wavelength phonons. To verify this statement the tunneling spectra of two types of the p-type InAs specimens have been recorded (Fig. 5). In contrast to the standard technique the circuit with the constant amplitude (about 300 mV, 4.1 kHz) of the sinusoidal voltage on the background of the slowly varying sweep voltage was used for recording d^2I/dV^2. The peak position in the IETS spectra for the few specimens was reproduced with an accuracy of about 0.5 meV. The spectra were recorded at temperature $T = 4.2$ K in the

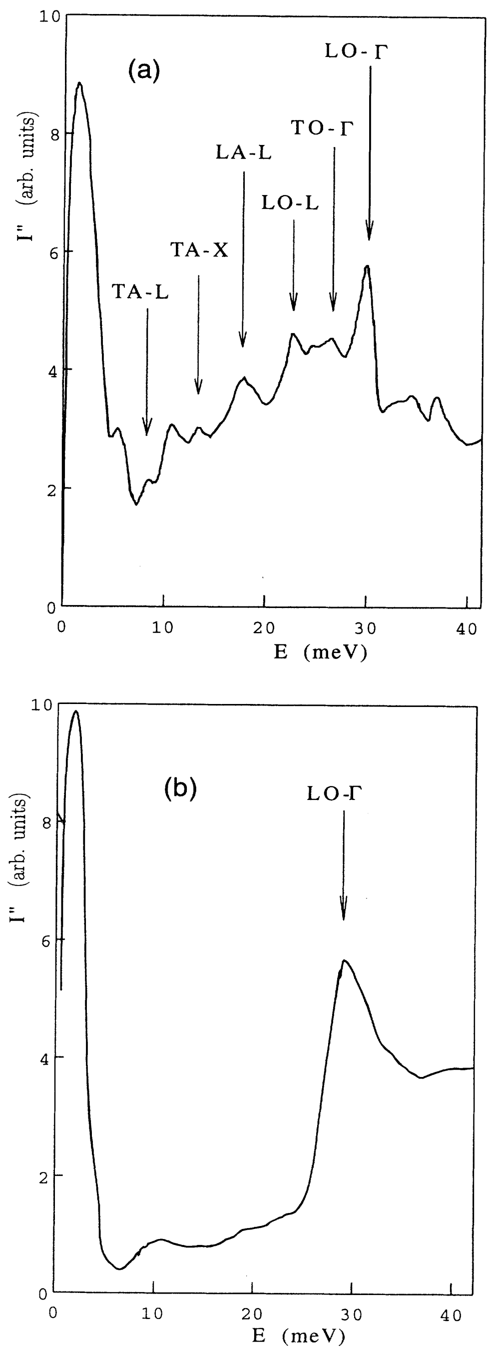

FIG. 5. IETS spectra for the Au-superthin-oxide-p-type-InAs structure (a) and Pt–p-type-InAs structure (b).

forward bias region (the bias of the metal relative to the substrate was negative).

Specimen 1 is the Au-superthin-oxide-InAs structure. The thickness of the oxide is about 3 nm. It was manufactured by oxidation of the p-type InAs substrates doped up to 2×10^{17} cm^{-3} in dry oxygen at 573 K for 2 h.

Specimen 2 is the Pt-InAs Schottky barrier. It was manufactured *in situ* in the ultrahigh vacuum installation Riber-250. The electrode was deposited with an electron beam by the autocrucible technique through the mask on the specimen, which had been cleaned before up to atomic purity. Unlike the barrier in specimen 1 the potential profile is continuous here due to lack of the dielectrical interlayer between the metal and the semiconductor.

The IETS spectrum for specimen 1 was obtained and identified in Ref. 2. It was shown that the peaks detected in the spectrum should be connected with the acoustical and optical phonons emitted in the semiconductor. The peak positions were also compared with the positions of absorption maxima in the IR spectra of InAs.[11]

We see that the peaks connected with the short-wavelength phonon emission in the tunneling spectra of specimen 1 are absent in the tunneling spectra of specimen 2. The existence of the superthin oxide in the former case, first, results in the potential discontinuity at the dielectric-metal interface (Fig. 2) and, second, makes inessential the smoothing of the potential profile at the interface due to the image charge interaction. The potential profile is continuous in specimen 2 (the band bending in this case is presented in Fig. 4). Nevertheless, the short-wavelength phonon might be emitted due to the discontinuity of the effective mass of the electron at the interface.[6] This does not happen because of the imperfection of the Pt-InAs interface. It is known[4] that the active mutual diffusion occurs at this interface. This leads to smoothing of the electron effective mass on the scale of about the lattice constant.

VI. DISCUSSION

The short-wavelength phonons should be observed under the tunneling in the direct-band-gap semiconductors if the effective parameters of the band structure have a discontinuity at the interface. The absence of such phonons in the experiments is connected with various mechanisms of smoothing of the effective parameters of the barrier at the interface.

The barrier could become smooth due to the image charge interaction at the metal-semiconductor interface. Usually, this interaction is disregarded when the tunneling in semiconductors is considered. Indeed, the influence of the image charge interaction on the barrier transparency is negligible for rather low and wide potential barriers, which is distinctive for the semiconductor structures. Nevertheless, this interaction results in the smoothing of the barrier profile and so the short-wavelength phonon emission becomes improbable. This has been proved in our experiments. A thin dielectric layer inserted between metal and semiconductor eliminates the smoothing of the potential and therefore the short-wavelength phonon emission becomes more effective.

The criterion (12) we have proposed is satisfied for the Schottky barriers if they are sufficiently high. This means that the smoothing of the barrier due to image charge interaction is insufficient and the short-wavelength phonons should be emitted. Really, the short-wavelength TA phonons have been observed in the IETS experiments on n-type GaAs/Pd.[3] We believe that the relatively high Schottky barrier in GaAs (0.78 eV) is the main reason for the observation of the short-wavelength phonons in Ref. 3. The additional mechanism of the electron-phonon interaction described in Ref. 3, i.e., the interaction of the electrons with the ionized impurities in the band bending region, also promotes the phonon emission. However, this effect occurs for any Schottky barrier, while the short-wavelength phonons have been observed in n-type GaAs only.

The imperfection of the metal-semiconductor interface is another reason for the Schottky barriers to become smoothed out. It is known[4] that an active mutual diffusion always occurs at the boundary of the noble and transition metals with the $A_{\rm III}B_{\rm V}$ semiconductors. Hence the main scale at the interface where the effective band parameters are changed, f (we shall call it the fuzzified parameter) becomes much larger than the lattice constant. In other words, $m(z)$, $U(z)$, and therefore $p_2(z)$ become nearly unchanged on the scale of about a_0, so that the matrix element (4) becomes exponentially small.[6] Perhaps the only exception is Ag. Deposition of Ag should lead to formation of an abrupt although not so flat interface,[4] so that the emission of the short-wavelength phonons becomes possible.

If the interface becomes smooth then the exponentially small factor has to arise in the expression for T. The exponent is of the order of $-qz_0$ if the smoothing is due to the image charge interaction or $-qf$ in the case of the fuzzified boundary. This means that only the phonons whose wave vectors q are not larger than $q_{\rm lim} = \min(z_0^{-1}, f^{-1})$ could be emitted. If the energy $\hbar\omega(q_{\rm lim})$ is situated apart from the relevant peak position in the phonon density of states, then those phonons could not be observed in the I'' spectrum. Nevertheless, the phonons with $q < q_{\rm lim}$ have to be emitted under the tunneling.

It seems to us that the influence of the sharpness of the barrier profile on the short-wavelength phonon emission is important for understanding of the anomalously high valley currents in the tunnel diodes. It is known[12] that these currents always increase with increasing doping density. It is established that elastic scattering centers also reduce the peak to valley ratio.[13] This fact cannot be connected with the localized states existing in the barrier. In our opinion it is important to remember that increase of the doping density always leads to a sharper potential barrier, so that more short-wavelength phonons could be emitted under tunneling. Moreover, the elastic scattering impurities inserted in the barrier could act as the additional origins where the short-wavelength phonons might be emitted if the impurity potential contains sufficiently high-frequency Fourier components.[5]

Our suggestions for the observation of the short-wavelength phonons in the experiments are as follows. It is preferable to use a p-type semiconductor. This would increase the relative contribution of the inelastic component to the tunneling current, i.e., the barrier height becomes lower for the tunneling electron when the phonon has been emitted (Fig. 1). It is necessary to make the boundary between the metal and semiconductor abrupt. This can be achieved in the following way.

(a) By choosing the materials which make the Schottky barriers sufficiently high that the criterion (12) becomes satisfied.

(b) By means of inserting the superthin dielectric layer between the metal and semiconductor. It should provide a high enough discontinuity of the potential (its magnitude then becomes of the order of a half band gap of the dielectric) and, in addition, it should decrease the influence of the image charge interaction.

(c) By means of inserting the elastically scattering impurities to create the additional origins where the short-wavelength phonons could be emitted.

(d) By means of manufacturing the contacts with the jump discontinuity of the electron effective mass (a so-called effective-mass interface[6]). Perhaps, this would be possible if Ag were used as the metal.

The first two possibilities were realized in Ref. 3 and in this work. The last one is the subject of further investigations.

ACKNOWLEDGMENTS

We thank Dr. M. Entin for helpful discussions, and Dr. Yu. Galitsyn and Dr. R. Sokolov for the preparation of specimens.

[1] R. Logan, in *Tunneling Phenomena in Solids*, edited by E. Burstein and S. Lundqvist (Plenum Press, New York, 1969).

[2] A. P. Kovchavtsev, G. L. Kuryshev, K. O. Postnikov, and S. A. Birukov, Fiz. Tekh. Poluprovodn. **19**, 2187 (1985) [Sov. Phys. Semicond. **19**, 1348 (1985)].

[3] P. Thomas and H. J. Queisser, Phys. Rev. **175**, 983 (1968).

[4] F. Bechstedt and R. Enderlein, *Semiconductor Surfaces and Interfaces. Their Atomic and Electronic Structures* (Akademie-Verlag, Berlin, 1988).

[5] E. M. Baskin and L. S. Braginski, Fiz. Tverd. Tela (Leningrad) **34**, 83 (1992) [Sov. Phys. Solid State **34**, 43 (1992)].

[6] E. M. Baskin and L. S. Braginsky, Phys. Rev. B **50**, 12 191 (1994).

[7] T. Ando, S. Wakahara, and H. Akera, Phys. Rev. B **40**, 11 609 (1989).

[8] Qi-Gao Zhu and Herbert Kroemer, Phys. Rev. B **27**, 3519

(1983).
[9] L. S. Braginsky and D. A. Romanov, Fiz. Tverd. Tela (St. Petersburg) (to be published).
[10] E. L. Wolf, *Principles of Electron Tunneling Spectroscopy* (Oxford University Press, New York, 1985).
[11] D. L. Sterwalt and R. F. Potter, Phys. Rev. A **137**, 1007 (1965).
[12] C. T. Sa, in *Tunneling Phenomena in Solids*, edited by E. Burstein and S. Lundqvist (Plenum Press, New York, 1969).
[13] E. Wolak *et al.*, Appl. Phys. Lett. **53**, 201 (1988).