1504 (1968).

[2]J. M. Stevels, in Non-Crystalline Solids, edited by V. D. Frechette (John Wiley & Sons, Inc., New York, 1960), p. 412.

[3]B. D. Josephson, Phys. Letters 1, 251 (1962).

[4]P. W. Anderson, in Lectures on the Many-Body Problem, edited by E. R. Caianiello (Academic Press, Inc., New York, 1964), Vol. 2, p. 113.

THEORY OF INELASTIC ELECTRON-SURFACE-PLASMON INTERACTIONS IN METAL-SEMICONDUCTOR TUNNEL JUNCTIONS*

K. L. Ngai, E. N. Economou, and Morrel H. Cohen

The James Franck Institute and Department of Physics, The University of Chicago, Chicago, Illinois 60637

(Received 25 April 1969)

The excess tunneling current due to electron-surface-plasmon interactions in semiconductor-metal tunnel junctions is calculated. The expression for the second derivative of this excess current, which corresponds to structure in d^2I/dV^2 as an increase in conductance at bias voltages near the surface plasmon energy in the semiconductor, agrees with experiment both in magnitude and line shape.

Recent studies of n-type GaAs-Pb surface-barrier tunnel junctions by Tsui[1] have shown evidence for the observation of surface-plasmon excitation in a degenerate semiconductor. Structure is observed in d^2I/dV^2 which corresponds to an increase in conductance at bias voltages near the surface-plasmon energy in GaAs. The bias position of the broad peak structures is strongly dependent on the electron concentration of the GaAs electrode. These measurements led Tsui to explain his observations as due to the excitation of surface plasmons (SP) in GaAs by tunneling electrons. Here we present a theoretical model based on Tsui's idea of SP excitation and amenable to quantitative calculation. The magnitude of the change in conductance induced by electron-SP interaction is in agreement with experiment as is the line shape. In addition, it helps to clarify some puzzling features of the observations.[1]

We idealize the semiconductor-metal junction as consisting of semiconductor (S) and metal (M), both semi-infinite, separated by a dielectric medium (I).[2] Their dielectric functions are taken to be of the form[3] $\epsilon_S(\omega)=\epsilon_\infty(1-\omega_S^2/\omega^2)$, $\epsilon_M(\omega)=1-\omega_M^2/\omega^2$, and $\epsilon_I(\omega)=\epsilon_\infty$ ($\epsilon_\infty=11.3$ and $m^*=0.07$ for GaAs[3]). The dispersion relation for surface-plasma oscillations has been calculated with inclusion of retardation effects (i.e., coupling of longitudinal Coulomb fields with transverse electromagnetic ones)[4] and is shown in Fig. 1. The mode described by the upper branch corresponds to oscillations in the M-I interface for large $\vec{p}_\parallel$ (the SP wave vector parallel to the interfaces). The initial sections of this branch are near the curves $\omega=c\epsilon_\infty^{-1/2}p_\parallel$ and $\omega^2=(c^2/\epsilon_\infty)p_\parallel^2+\omega_S^2$, and hence are oscillations of very small amplitude with no physical importance as has been discussed by one of us.[5] Consequently, for frequency range $\omega\sim\omega_S$, the upper mode does not exert any effect and can be neglected. The lower branch describes modes of primary importance. The dispersion is linear for $p_\parallel \ll k_S\equiv\omega_S/c$, i.e., $\omega=\bar{c}p_\parallel$ with

$$\bar{c}=c[d_i/(d_i+\lambda_M+\lambda_S)\epsilon_\infty]^{1/2},$$

$1/\lambda_M=\omega_M/c$, and $1/\lambda_S=\omega_S/c$. This is reminiscent of the SP modes that gave rise to Fiske steps in Josephson junctions.[6] When $p_\parallel \ll k_S$, the mode corresponds to oscillations in the S-I interface and can be adequately described by

$$\frac{\omega}{\omega_S}=\frac{\omega_1+i\omega_2}{\omega_S}=\frac{1}{\sqrt{2}}[1-\exp(-2p_\parallel d)]^{1/2}+i\frac{1}{2\tau\omega_S}, \quad (1)$$

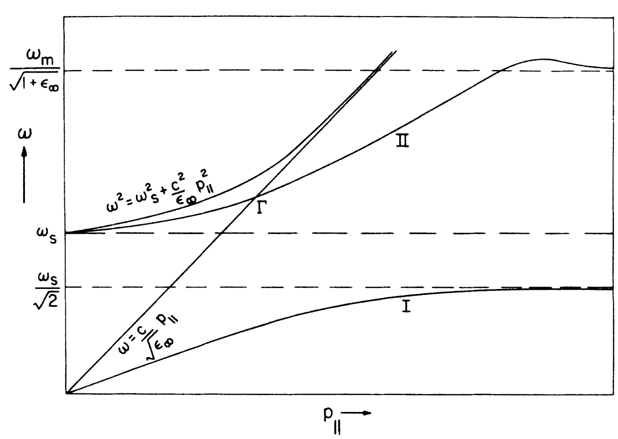

FIG. 1. Schematic dispersion relation for surface-plasmon oscillations in an idealized metal-semiconductor junction (see text). Γ corresponds to a frequency $\sim(\omega_M\omega_S)^{1/2}$.

to lowest order in (ω_S^2/ω_M^2). This expression is invalid for low frequencies because at these frequencies, polar lattice vibrations strongly modify[3] $\epsilon_S(\omega)$ from the form assumed. The error committed by omitting this contribution is not important in this consideration and we shall introduce a lower cutoff in frequency at somewhat above the maximum phonon energies. Equation (1) is inadequate also for large values of $p_{\parallel}$ such that the condition $p_{\parallel}v_F \ll \omega$, implicitly assumed to justify the form of the dielectric functions used, is violated. Quasiparticle-quasihole pair damping (i.e., Landau damping) becomes possible when $p_{\parallel} \sim \omega/v_F$ and a SP ceases to be a well-defined excitation. For the n-GaAs studied by Tsui, $k_c = \omega_S/v_F \approx k_F/3.5 \approx 1.5\times 10^6\ \mathrm{cm}^{-1}$, and the correction due to damping is significant at a much lower $p_{\parallel}$ than in the case of a typical metal, where $k_c = 0.48 r_s^{1/2} k_F \approx 10^8\ \mathrm{cm}^{-1}$ (r_s is the interelectron spacing measured in units of the Bohr radius). The contribution of Landau damping to the imaginary part of SP frequency has been calculated[7-10] and found to be linear in $p_{\parallel}/k_F$ for $p_{\parallel}/k_F < 1$, although its magnitude has not been agreed upon. On the other hand, the correction to the real part of the SP frequency due to the $\vec{p}$ dependence of the dielectric function is not well established.[11]

Our model is basically a one-electron picture in which we take the dynamics of the electron-SP interaction into account by including the SP wave fields in the potential which the electron sees in the insulating layer of the junction. Roughly speaking, when an electron is inside the electrodes its dynamical effects are well screened. This is no longer so when it tunnels into the dielectric region, and its long-range fields can excite SP. Thereby new tunneling channels are opened by the inelastic electron-SP interaction. The SP wave fields can be quantized in a standard manner. In the frequency region of our present interest, retardation effects can be shown to be negligible, and the SP wave fields can be approximated by the Coulomb potential $\varphi(\vec{r})$. We consider a continuum model of an electron gas with a rigid fixed background of positive charge, the Lagrangian of which is taken to be

$$L = \int \frac{1}{2nm}\vec{\pi}\cdot\vec{\pi}\,d^3r + \tfrac{1}{2}\int \delta\rho\,\varphi(\vec{r})\,d^3r,$$

where $\vec{\pi}$ is the momentum density, n the density, m the effective mass, and $\delta\rho$ the charge-density fluctuation of the electron gas.[12] The Lagrangian of the system is taken as the sum of two such contributions from the GaAs and Pb media. Expanding all relevant fields in eigenfunctions for SP oscillations, the method of quantum fields leads us to the creation and annihilation operators $c_{\vec{p}_{\parallel}}^{\dagger}$ and $c_{\vec{p}_{\parallel}}$ for SP. Then the interaction energy of an electron in the dielectric region $(0 < x < d)$ with the SP field is

$$e\varphi(\vec{r}) = e\sum_{\vec{p}_{\parallel}} \varphi_{\vec{p}_{\parallel}}(c_{\vec{p}_{\parallel}} + c_{\vec{p}_{\parallel}}^{\dagger})\exp(i\vec{p}_{\parallel}\cdot\vec{r}),$$

and

$$\begin{aligned}\varphi_{\vec{p}_{\parallel}} = \omega_{\vec{p}_{\parallel}}^{1/2}&\left\{\left(1+\frac{\epsilon_S}{\epsilon_\infty}\right)\exp(-p_{\parallel}x) + \left(1-\frac{\epsilon_S}{\epsilon_\infty}\right)\exp(-2p_{\parallel}d+p_{\parallel}x)\right\}\\ &\times\left\{\frac{p_{\parallel}}{\pi}\left[\left\{(\epsilon_\infty+1)\left(1+\frac{\epsilon_S}{\epsilon_\infty}\right)+(1-\epsilon_\infty)\left(1-\frac{\epsilon_S}{\epsilon_\infty}\right)\exp(-2p_{\parallel}d)\right\}\right.\right.\\ &\left.\left.\times\left\{\left(1+\frac{\epsilon_S}{\epsilon_\infty}\right)+\exp(-2p_{\parallel}d)\left(1-\frac{\epsilon_S}{\epsilon_\infty}\right)+\exp(-2p_{\parallel}d)\left(1+\frac{\epsilon_S}{\epsilon_\infty}\right)\right\}+(\epsilon_\infty-\epsilon_S)\exp(-2p_{\parallel}d)\right]\right\}^{-1/2}. \quad (2)\end{aligned}$$

The inelastic current due to SP excitation at temperature T and bias V is given by

$$\begin{aligned}\delta I(V,T) = \frac{-2e}{\hbar}\sum_{\vec{k}\vec{q}\vec{p}_{\parallel}}\int\frac{d\omega_1 d\omega_2}{\pi}&[f(\omega_1)-f(\omega_1-eV)][f(\omega_2)+N(\omega_2-\omega_1+eV)]\delta(\omega_1-\xi_{\vec{q}})\\ &\times \mathrm{Im}G^R(\vec{k},\omega_2)\,\mathrm{Im}D_{\vec{k}\vec{q}}^R(\omega_2-\omega_1+eV)\,\delta(\vec{k}_{\parallel}-\vec{q}_{\parallel}-\vec{p}_{\parallel}), \quad (3)\end{aligned}$$

using the general formula and notation of Bennett, Duke, and Silverstein,[13] and

$$\mathrm{Im}D_{\vec{k}\vec{q}}(\omega)=\frac{2m^*\pi e^2}{\varphi(\vec{k})\hbar^2}|\Lambda_{\vec{k}\vec{q}}^{(0)}|^2\sum_{\vec{p}_{\parallel}}|\theta_{\vec{p}_{\parallel}}|^2[\delta(\omega+\hbar\omega_{\vec{p}_{\parallel}})-\delta(x-\hbar\omega_{\vec{p}_{\parallel}})],$$

where $[(2m^*/\hbar^2)\varphi(\vec{k})]^{1/2}$ can be regarded as an average of the argument of the WKBJ exponent and

$$\theta_{\vec{p}_{\parallel}}=\int_0^d dx\,\varphi_{\vec{p}_{\parallel}}(x).$$

The summation over $\vec{p}_{\parallel}$ is over SP states of momentum $\vec{p}_{\parallel}$ and frequency $\omega_{\vec{p}_{\parallel}}$. Then as $T\to 0$,

$$\frac{d^2\delta I}{dV^2}=\frac{\pi^2m^*e^2}{2\varphi_{\mathrm{eff}}}\sum_{\substack{\vec{p}_{\parallel}\\ k_{c1}<p_{\parallel}<k_{c2}}}|\theta_{\vec{p}_{\parallel}}|^2\left\{\frac{dI}{dV}(eV-\hbar\omega_{\vec{p}_{\parallel}})\delta(eV-\hbar\omega_{\vec{p}_{\parallel}})+\frac{d^2I}{dV^2}(eV-\hbar\omega_{\vec{p}_{\parallel}})f(\hbar\omega_{\vec{p}_{\parallel}}-eV)\right\}, \tag{4}$$

when expressed in terms of the normal conductance dI/dV and its derivative and φ_{eff}, an effective barrier height. Both dI/dV and d^2I/dV^2 are obtainable from experiment. From Tsui's data we can show that the second term within the braces is two orders of magnitude smaller than the first and can be neglected. Here k_{c1} and k_{c2} are lower and upper cutoff wave numbers introduced for the physical reasons given earlier. However, real SP have finite lifetime as described by $\omega_2(p_{\parallel})=\mathrm{Im}\omega_{\vec{p}_{\parallel}}$. This can be accounted for by replacing the δ function in (4) by a Lorentzian $\hbar\omega_2(p_{\parallel})/\pi\{(\hbar\omega_1-eV)^2+\hbar^2\omega_2^2(p_{\parallel})\}$. $\omega_2(p_{\parallel})$ is the sum of the contributions due to collision broadening and Landau damping, i.e., $\omega_2(p_{\parallel})=\frac{1}{2}\Delta\omega_T+\alpha p_{\parallel}/k_c$. To include any correction to ω_1 due to the $\vec{p}$ dependence of the dielectric function we put[10] $\omega_1=2^{-1/2}\omega_S[1-\exp(-2p_{\parallel}d)]^{1/2}(1+\gamma p_{\parallel}/k_F)^{1/2}$. γ and α are left as adjustable parameters. We take $\Delta\omega_T\approx 8$ MeV as calculated from the electron mobility in the GaAs samples. The upper cutoff k_{c2} is determined by the requirement that at $p_{\parallel}=k_{c2}$, ω_2 is an appreciable fraction[14] of ω_1 so that the SP is no longer a well-defined mode of excitation. The theoretical line shapes for various α and γ are calculated from (4).

Useful information about γ, α, and hence k_{c2} can be derived by fitting the experimental line shape. We arrive at the result that line shapes inconsistent with Tsui's observations are obtained for the cases where (1) α is negligibly small and γ has any value, or (2) α/ω_S is in the range 0.01-1.0 as expected[15] and γ exceeds 0.1. These seem to imply that the estimations for $\omega_1(p_{\parallel})$ and $\omega_2(p_{\parallel})$ given by Ref. 10 are doubtful. For the purpose of the present Letter, we shall put $\gamma=0$, and the line shapes as calculated for different α (or k_{c2}) are shown in Fig. 2. We found good agreement between theory and experiment if $\alpha/\omega_S\approx 0.035$ ($k_{c2}/k_c\approx 2$) for all samples (see Table I). Moreover, the present theory explains quite well the various features of the experimental line shape. First, the long tail for low voltages is due to the fact that the dispersion for SP is given by $2^{-1/2}\omega_S[1-\exp(-2p_{\parallel}d)]^{1/2}$ instead of $\omega_S/\sqrt{2}$ for one interface. Second, the

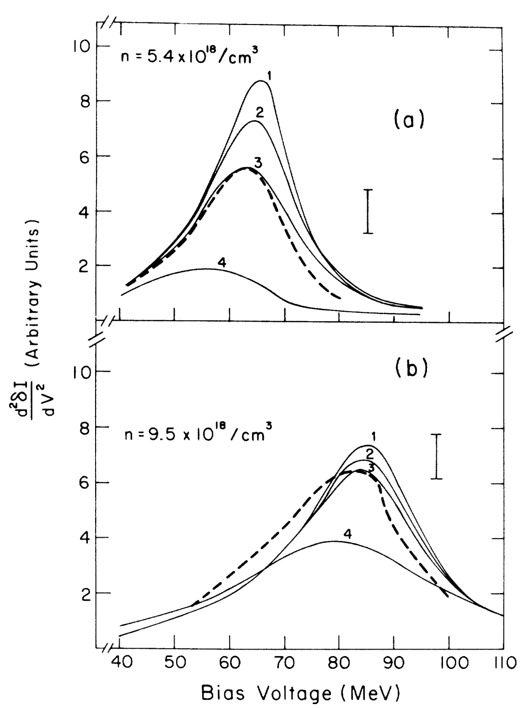

FIG. 2. The dashed curves are the experimental line shapes and the vertical bars indicates the estimated errors. (a) Curves 1, 2, 3, 4 are theoretical line shapes corresponding to upper cutoffs $k_{c2}=5.4$, 5.0, 3.0, 1.0 ($\times 10^6$ cm^{-1}), respectively. (b) Curves 1, 2, 3, 4 are theoretical line shapes corresponding to upper cutoffs $k_{c2}=4.6$, 4.0, 3.6, 3.0 ($\times 10^6$ cm^{-1}), respectively.

Table I. Comparison of line-shape parameters of surface-plasmon excitation between theory and experiment.

Electron density n (10^{18} cm^{-3})	Surface plasma frequency $\omega_S/\sqrt{2}$ (meV)	Peak position (meV)		Half-width (meV)	
		Theory	Experiment	Theory	Experiment
9.5	90	83	80	30	30
6.5	74	69.5	68	24.5	25
5.4	68	62	60	20	20
4.2	60	56	55	18	18

maximum does not occur at $\omega_S/\sqrt{2}$ (as Tsui assumed) but at a lower value (see Table I) because of the combined effects of dispersion and considerable Landau damping at $p_\parallel$ near k_{c2}. We have thus removed the discrepancy between the value of the observed "threshold"[1] and $\omega_S/\sqrt{2}$. Third, the broad line shape is mainly due to considerable Landau damping near k_{c2} and to a much less extent to dispersion. The strong dependence of the half-width on the electron concentration but not on electron mobility can be explained as follows. The half-width is determined roughly by the total broadening of the SP energy at $p_\parallel$ near k_{c2}. At such $p_\parallel$, Landau damping dominates over the collision broadening. Since the Landau damping is nearly linearly dependent upon ω_S or[16] $n^{1/2}$, the strong dependence on concentration is verified. Fourth, the reason for the absence of volume plasmon excitation by tunneling electrons is because the field of the bulk plasmon, unlike that of SP, is essentially zero in the dielectric film.

The magnitude of the $d^2\delta I/dV^2$ can be calculated by taking $d \approx 100$ Å, $\varphi_{\text{eff}} \approx \frac{1}{2}$ eV, and dI/dV as deduced from the experiment. Equation (4) gives a change in the second derivative of the tunneling current consistent with the experimentally[1] reported size of this change.

In summary, we have shown that the excess current due to electron-SP interaction can account not only quantitatively for the size but also for the line shape of Tsui's experiment.[17] The present analysis indicates that experiments like those of Tsui show promise for providing new quantitative information about the dispersion and damping of surface plasmons in degenerate semiconductors for wave numbers near k_c.

We would like to thank D. M. Newns for useful discussion. One of us (K.L.N.) wishes to express his gratitude to D. C. Tsui for many stimulating conversations and supplying the data.

*Research supported in part by the National Aeronautics and Space Administration, and aided through general support of materials science by the Advanced Research Projects Agency.

[1]D. C. Tsui, Phys. Rev. Letters 22, 293 (1969).

[2]Calculations we have performed (K. L. Ngai and E. N. Economou, to be published) indicate that the results as discussed in this Letter for our idealized model hold also for a realistic S-M junction.

[3]B. B. Varga, Phys. Rev. 137, A1896 (1965); A. Mooradian and G. B. Wright, Phys. Rev. Letters 16, 999 (1966).

[4]For a discussion of retardation effects, see J. C. Swihart, J. Appl. Phys. 32, 461 (1961); E. N. Economou, to be published.

[5]E. N. Economou, to be published.

[6]K. L. Ngai, to be published.

[7]Peter J. Feibelman, Phys. Rev. 176, 551 (1968).

[8]Ralph L. Guernsey, to be published.

[9]D. M. Newns, to be published.

[10]R. H. Ritchie and A. L. Marusak, Surface Sci. 4, 234 (1966).

[11]There are assertions (Feibelman, Ref. 7) that it is of higher order in $p_\parallel/k_F$ while others (Ritchie and Marusak, Ref. 10) contend otherwise, that it is of first order and quite appreciable.

[12]C. Kittel, Quantum Theory of Solids (John Wiley & Sons, Inc., New York, 1963).

[13]A. J. Bennett, C. B. Duke, and S. D. Silverstein, Phys. Rev. 176, 969 (1968).

[14]We have arbitrarily taken ω_2 is no greater than $0.2\omega_1$ as a criterion for cutoff.

[15]Ngai and Economou, Ref. 2.

[16]A plot of the measured half-width versus $n^{1/2}$ shows their relation is indeed linear.

[17]Recently C. B. Duke, M. J. Rice, and F. Steinrisser [Phys. Rev. (to be published)] have observed structures similar to those observed by Tsui in n-GaAs–In contacts. They have attributed their observations to bulk plasmon self-energy effects in the GaAs electrode. Their experimental results can be explained by our theoretical calculations as due to surface-plasmon excitation.