Theory of surface-plasmon excitation in metal-insulator-metal tunnel junctions

L. C. Davis
Research Staff, Ford Motor Company, Dearborn, Michigan 48121
(Received 30 March 1977)

A new calculation of surface-plasmon excitation in tunnel junctions is described. The tunnel junction is divided into three regions of complex dielectric function $\epsilon_L(\omega)$, $\epsilon_0(\omega)$, and $\epsilon_R(\omega)$ which correspond to the left electrode, the barrier, and the right electrode, respectively. Maxwell's equations are solved for the classical electromagnetic fields. The source terms are given by the quantum-mechanical transition current and charge, $\vec{J} = (ie\hbar/2m)(\psi^*_R \nabla\psi_L - \psi_L \nabla\psi^*_R)$ and $\rho = -e\psi^*_R\psi_L$, for an electronic transition from a state ψ_L in the left electrode to a state ψ_R in the right. The transition rate is given by $(-2/\hbar\omega)\mathrm{Re}\int \vec{E}^* \cdot \vec{J}\, d^3r$ where $\hbar\omega = E_L - E_R$. This new formulation avoids the need to quantize the electromagnetic fields and allows the use of complex dielectric functions. Current-carrying orthogonal eigenfunctions are used for ψ_L and ψ_R in place of the nonorthogonal basis set associated with the transfer-Hamiltonian theory of inelastic tunneling. The transition rate calculated from the latter theory differs from the present calculation by a factor as large as 10 to 100 in some instances. Numerical estimates of the rate of surface-plasmon excitation in Al-Al_2O_3-Ag junctions are given. The inelastic tunneling rate is found to be ~ 0.1 of the elastic rate (for electrons tunneling into Ag). Excitation of electromagnetic modes which can be made radiative by roughening the electrodes is discussed.

I. INTRODUCTION

In a recent letter,[1] Lambe and McCarthy (LM) reported emission of light from metal-insulator-metal (*M-I-M*) tunnel junctions. They interpreted the effect as being due to the excitation of the electromagnetic modes of the junction by tunneling electrons. These modes are called surface plasmons or slow waves.[2] Although they are not ordinarily radiative, surface roughness can cause them to radiate.[3] In the LM experiments, the metal electrodes were deliberately roughened to induce light emission.

In LM, the mechanism for excitation of the surface plasmons was likened to the excitation of molecular vibrations in tunnel junctions.[4] In this mechanism there is an inelastic channel for tunneling in which an electron makes a transition from one electrode to the other with the accompanying excitation of a surface plasmon. This is in contrast to an injection process[5] where the electron first tunnels elastically and then decays toward the Fermi level by the emission of surface or bulk plasmons (as well as by other scattering).

In a junction composed of dissimilar electrodes, such as LM's Al-Al_2O_3-Ag junctions, the plasma excitations of interest are primarily associated with one electrode—Ag in LM. For a bias voltage such that electrons are being injected into the Ag electrode, the excitation of plasma oscillations will be much stronger than in the opposite bias for the injection process. On the other hand, inelastic tunneling does not necessarily result in such asymmetric excitation and light emission. In the LM experiments, light was observed in both bias directions. It was concluded that inelastic tunneling plays a dominant role. Consequently, the inelastic tunneling excitation of surface plasmons in *M-I-M* junctions will be studied in this paper.

In metal-semiconductor junctions, inelastic tunneling due to surface-plasmon excitation has been observed by Tsui.[6] A theory for such junctions has been given by Ngai, Economou, and Cohen.[7] In their calculation, it is necessary to find the normal, electromagnetic modes of the junction and to quantize them. Since simple free-electron-like dielectric functions were assumed for both the metal and the semiconductor, it was straightforward to find the modes and to quantize them. For the problem considered in the present work, the dielectric functions are more complicated and it is not as easy to quantize the modes. In fact, for lossy dielectrics the modes may not be well defined, yet inelastic tunneling must still occur.

This difficulty can be avoided since there is actually no need to introduce normal modes or their quantization. The calculation of the transition rate can be cast into a form similar to an energy-loss calculation for electron beams.[8] The only difference is in the source terms for Maxwell's equations. In tunneling they are found from the quantum-mechanical transition current density[9] and transition charge density, whereas the source in the beam problem is an electron moving with uniform velocity (treated classically—except Ref. 10, which uses transition current and charge). In both cases the calculation of normal modes and their quantization is unnecessary. Only the response of the dielectric system to an external current and charge distribution is required and this can be found directly from Maxwell's equations. The transition rate can be calculated in a straightfor-

ward way even for lossy dielectrics where the normal modes are not well defined.

Another feature of the calculation by Ngai, Economou, and Cohen is the use of a nonorthogonal basis set. The set consists of states ψ_L (ψ_R) which are eigenfunctions of a Hamiltonian for the left (right) problem. This is the usual basis set for the transfer-Hamiltonian theory of inelastic tunneling.[11–13] The interaction between the tunneling electrons and the surface plasmons is represented by the matrix element $\langle\psi_R|-e\phi^{sp}|\psi_L\rangle$ where ϕ^{sp} is the surface-plasmon potential (retardation is neglected). The region of interaction is restricted to the barrier. However, from a study of inelastic tunneling due to vibration of molecular impurities in the barrier,[14] it has been concluded that the proper basis is a set comprised of orthogonal eigenfunctions of a single Hamiltonian for the entire system. A typical eigenfunction represents an electron incident upon the barrier from the left (or right) with transmitted and reflected components. In addition, it can be shown that the interaction should not be limited to just the barrier, but should also include the electrodes (at least the region within a mean free path of the barrier). For surface plasmons, the difference between the two methods of calculation can be large. Hence, the proper wave functions are used in this paper.

A general theory of the energy loss and the inelastic transition rate of a tunneling electron is given in Sec. II. This theory is applied to a barrier represented by a rectangular potential in Sec. III. Numerical results for Al-Al_2O_3-Ag are given in Sec. IV. The excitation of modes which are likely to be radiative is discussed for this system. In Sec. V, comparison is made to the inelastic transfer-Hamiltonian approach. A short symmary is given in Sec. VI.

II. GENERAL THEORY

Suppose that the junction consists of three regions of complex dielectric function, $\epsilon_L(\omega)$, $\epsilon_0(\omega)$, and $\epsilon_R(\omega)$, which correspond to the left electrode, the barrier, and the right electrode, respectively. Consider an electronic transition from a state ψ_L in the left electrode to a state ψ_R in the right. This represents the inelastic tunneling of an electron from the left electrode to the right. Associated with the transition is current density[9] (the symmetric form is used—see Ref. 10).

$$\vec{J} = (ie\hbar/2m)(\psi_R^* \nabla\psi_L - \psi_L \nabla\psi_R^*) . \tag{1}$$

The frequency dependence of $\vec{J}$ is $e^{-i\omega t}$ where ω is related to the difference in energy between the two states:

$$\hbar\omega = E_L - E_R . \tag{2}$$

The transition charge density can be found from the continuity equation

$$\frac{\partial\rho}{\partial t} + \nabla\cdot\vec{J} = 0 . \tag{3}$$

If ψ_L and ψ_R are exact eigenfunctions of the same single-particle Hamiltonian which is of the form

$$H_0 = (-\hbar^2/2m)\nabla^2 + U(\vec{r}) , \tag{4}$$

where $U(\vec{r})$ is the potential energy, then

$$\rho = -e\psi_R^*\psi_L \tag{5}$$

satisfies (3) with $\vec{J}$ given by (1). The quantity ρ is identified as the transition charge density.

Since the source terms ($\vec{J}$ and ρ) are known, it is possible to determine the electromagnetic fields from Maxwell's equations. From the calculated electric field $\vec{E}$, the rate of energy loss in the transition can be found:

$$\frac{dW}{dt} = -2\,\mathrm{Re}\int \vec{E}^*\cdot\vec{J}\,d^3r . \tag{6}$$

Re denotes the real part. Since the energy loss per transition must be $\hbar\omega$, the transition rate is given by

$$\frac{1}{\hbar\omega}\frac{dW}{dt} = \frac{-2}{\hbar\omega}\,\mathrm{Re}\int \vec{E}^*\cdot\vec{J}\,d^3r . \tag{7}$$

This transition rate is analogous to

$$(2\pi/\hbar)\,|\langle\psi_R|-e\phi^{sp}|\psi_L\rangle|^2\,\delta(\hbar\omega + E_R - E_L) \tag{8}$$

in the transfer-Hamiltonian theory.[7,11,12]

III. APPLICATION TO RECTANGULAR BARRIER POTENTIAL

In this section the general theory of the previous section is applied to a simple model of the barrier potential. If we divide the inelastic tunneling rate by the elastic rate, most of the details of the barrier tend to cancel, so for simplicity only a rectangular barrier potential will be considered, i.e.,

$$U(\vec{r}) = \begin{cases} U_0 , & 0 < z < b \quad \text{(9a)} \\ 0 , & \text{otherwise} . \quad \text{(9b)} \end{cases}$$

The z direction is taken to coincide with the normal to the barrier. The barrier width is b. Since most of the electrons which tunnel are incident on the barrier near the normal, we consider only normal incidence. To avoid unnecessary complications, the band structure of the electrodes will be ignored. For the purposes of calculating wave functions, the junction is perfectly symmetric. The asymmetry only comes into the dielectric functions.

A. Eigenfunctions

Assuming normal incidence the initial state can be written

$$\psi_L = A^{-1/2}\chi_L(z)e^{-iE_Lt/\hbar}, \tag{10}$$

where A is the junction area. The final state will in general have a component of wave vector in the plane of the junction. Let us take that component to be of magnitude q and in the negative x direction. This means that the electron transfers momentum $\hbar q$ (in the x direction) to the electromagnetic fields in the transition along with the energy transfer $\hbar\omega$. So

$$\psi_R = A^{-1/2}\chi_R(z)e^{-i(qx+E_Rt/\hbar)}. \tag{11}$$

As remarked previously, it has been shown[14] in the case of the excitation of a molecular vibrator the proper choice for χ_L and χ_R are current-carrying eigenfunctions. The same arguments apply to the excitation of surface plasmons. Therefore,

$$\chi_L = \begin{cases} (2L)^{-1/2}(e^{ik_Lz}+R_Le^{-ik_Lz}), & z<0 \quad (12a)\\ (2L)^{-1/2}(C_Le^{-K_Lz}+D_Le^{K_Lz}), & 0<z<b \quad (12b)\\ (2L)^{-1/2}T_Le^{ik_L(z-b)}, & b<z, \quad (12c)\end{cases}$$

where L is the length of either electrode. (Actually, this is just a normalization length since the electrodes are really treated as infinite.) From (4) and (9), we see that

$$k_L = (2mE_L/\hbar^2)^{1/2} \tag{13a}$$

and

$$K_L = [2m(U_0-E_L)/\hbar^2]^{1/2}. \tag{13b}$$

Likewise,

$$\chi_R = \begin{cases} (2L)^{-1/2}T_Re^{-ik_Rz}, & z<0 \quad (14a)\\ (2L)^{-1/2}(C_Re^{-K_R(b-z)}+D_Re^{K_R(b-z)}), & 0<z<b \quad (14b)\\ (2L)^{-1/2}(e^{-ik_R(z-b)}+R_Re^{ik_R(z-b)}), & b<z, \quad (14c)\end{cases}$$

where

$$k_R = (2mE_R)/\hbar^2 - q^2)^{1/2} \tag{15a}$$

and

$$K_R = [2m(U_0-E_L)/\hbar^2+q^2]^{1/2}. \tag{15b}$$

Note that k_R is the z component of the wave vector associated with ψ_R. The coefficients R, C, D, and T are determined by requiring continuity of χ and $d\chi/dz$ at $z=0$ and b. It is found that (dropping subscripts)

$$R = [(k-iK)/2k](1-e^{-2Kb})C, \tag{16a}$$

$$D = -[(k-iK)/(k+iK)]e^{-2Kb}C, \tag{16b}$$

$$T = [2iK/(k+iK)]e^{-Kb}C, \tag{16c}$$

and

$$C = [2k/(k+iK)]\{1-[(k-iK)/(k+iK)]^2e^{-2Kb}\}^{-1}. \tag{16d}$$

From these eigenfunctions, the transition current density $\vec{J}$ and charge density ρ can be calculated. The form of these quantities and the resultant fields is

$$F = F(z)e^{i(qx-\omega t)}. \tag{17}$$

Whenever $F(z)$ is written for any quantity F, the factor $e^{i(qx-\omega t)}$ has been omitted.

B. Electric field

Although retardation plays a role in the dispersion curve for surface plasmons and in the energy-loss function near the light line, it can be neglected for the region of the ω-q plane which is important in the present calculation. This approximation simplifies the algebraic manipulations significantly. The electric field is then given by

$$\vec{E} = -\nabla\phi, \tag{18}$$

where

$$-\nabla\cdot[\epsilon(z,\omega)\nabla\varphi] = 4\pi\rho \tag{19}$$

and

$$\epsilon(z,\omega) = \begin{cases} \epsilon_L(\omega), & z<0 \quad (20a)\\ \epsilon_0(\omega), & 0<z<b \quad (20b)\\ \epsilon_R(\omega), & b<z \quad (20c)\end{cases}.$$

Local dielectric functions have been assumed, i.e., the dependence on wave vector has been neglected. Both φ and ρ are of the form (17).

The transition charge density can be computed from (5) and (10)–(16):

$$\rho(z) = \begin{cases} \rho_{L+}e^{ik_+z}+\rho_{L-}e^{ik_-z}, & z<0 \quad (21a)\\ \rho_1e^{K_-z}+\rho_2e^{K_+z}+\rho_3e^{-K_+z}+\rho_4e^{-K_-z}, & 0<z<b \quad (21b)\\ \rho_{R+}e^{ik_+(z-b)}+\rho_{R-}e^{-ik_-(z-b)}, & b<z, \quad (21c)\end{cases}$$

where

$$\rho_{L+} = -eT_R^*/2AL, \tag{22a}$$

$$\rho_{L-} = -eT_R^*R_L/2AL \tag{22b}$$

$$\rho_1 = -eC_R^*C_Le^{-K_Rb}/2AL, \tag{22c}$$

$$\rho_2 = -eC_R^*D_Le^{-K_Rb}/2AL, \tag{22d}$$

$$\rho_3 = -eD_R^*C_Le^{K_Rb}/2AL, \tag{22e}$$

$$\rho_4 = -eD_R^*D_Le^{K_Rb}/2AL, \tag{22f}$$

$$\rho_{R+}=-eT_L/2AL\,, \tag{22g}$$

$$\rho_{R-}=-eR_R^*T_L/2AL\,, \tag{22h}$$

$$k_\pm=k_R\pm k_L\,, \tag{22i}$$

$$K_\pm=K_R\pm K_L\,. \tag{22j}$$

The boundary conditions are that φ and $\epsilon\,\partial\varphi/\partial z$ are continuous at $z=0$ and b and that φ is finite as $z\to\pm\infty$. The solution to (19) with charge density (21) which satisfies these boundary conditions is

$$\varphi(z)=\begin{cases} a_Le^{qz}+g_L(z)\,, & z<0 \quad (23a)\\ a_+e^{qz}+a_-e^{-qz}+g_0(z)\,, & 0<z<b \quad (23b)\\ a_Re^{-q(z-b)}+g_R(z)\,, & b<z\,, \quad (23c)\end{cases}$$

where the inhomogeneous part of the solution is

$$g_L(z)=\frac{4\pi}{\epsilon_L(\omega)}\left(\frac{\rho_{L+}e^{ik_+z}}{k_+^2+q^2}+\frac{\rho_{L-}e^{ik_-z}}{k_-^2+q^2}\right), \qquad z<0\,, \tag{24a}$$

$$g_0(z)=\frac{4\pi}{\epsilon_0(\omega)}\left(\frac{\rho_1e^{K_-z}}{q^2-K_-^2}+\frac{\rho_2e^{K_+z}}{q^2-K_+^2}+\frac{\rho_3e^{-K_+z}}{q^2-K_+^2}+\frac{\rho_4e^{-K_-z}}{q^2-K_-^2}\right), \quad 0<z<b\,, \tag{24b}$$

$$g_R(z)=\frac{4\pi}{\epsilon_R(\omega)}\left(\frac{\rho_{R+}e^{ik_+(z-b)}}{k_+^2+q^2}+\frac{\rho_{R-}e^{-ik_-(z-b)}}{k_-^2+q^2}\right), \quad b<z\,. \tag{24c}$$

The coefficients a_L, a_R, and $a_\pm$ are determined by applying the boundary conditions at $z=0$ and b. It is found that

$$a_+=-e^{-qb}[S_1(\epsilon_R-\epsilon_0)e^{-qb}-S_2(\epsilon_L+\epsilon_0)]/\eta\,, \tag{25a}$$

$$a_-=e^{-qb}[S_1(\epsilon_R+\epsilon_0)e^{qb}-S_2(\epsilon_L-\epsilon_0)]/\eta\,, \tag{25b}$$

$$a_L=a_++a_-+g_0(0)-g_L(0)\,, \tag{25c}$$

and

$$a_R=a_+e^{qb}+a_-e^{-qb}+g_0(b)-g_R(b)\,, \tag{25d}$$

where

$$S_1=\epsilon_L[g_L(0)-g_0(0)]+[\epsilon_0g_0'(0)-\epsilon_Lg_1'(0)]/q \tag{26a}$$

$$S_2=\epsilon_R[g_R(b)-g_0(b)]+[\epsilon_Rg_R'(b)-\epsilon_0g_0'(b)]/q \tag{26b}$$

$$\eta=(\epsilon_L+\epsilon_0)(\epsilon_R+\epsilon_0)-(\epsilon_L-\epsilon_0)(\epsilon_R-\epsilon_0)e^{-2qb}\,. \tag{26c}$$

The dispersion curve for surface plasmons can be obtained by setting $\eta=0$, i.e., for any value of ω, there is a (generally complex) value of q for which η vanishes.

C. Surface-plasmon excitation rate

Since the excitation of bulk plasmons must surely involve the injection process and other scattering not included in present model, let us only consider energy loss to surface plasmons. The surface and bulk contributing to φ can be separated by writing

$$\varphi(z)=\varphi_S(z)+\varphi_B(z)\,, \tag{27}$$

where the bulk-plasmon part is

$$\varphi_B(z)=\begin{cases} g_L(z)-g_L(0)e^{qz}\,, & z<0 \quad (28a)\\ 0\,, & 0<z<b \quad (28b)\\ g_R(z)-g_R(b)e^{-q(z-b)}\,, & b<z\,. \quad (28c)\end{cases}$$

The surface-plasmon portion is

$$\varphi_S(z)=\begin{cases} a_L'e^{qz}\,, & z<0 \quad (29a)\\ a_+e^{qz}+a_-e^{-qz}+g_0(z)\,, & 0<z<b \quad (29b)\\ a_R'e^{-q(z-b)}\,, & b<z\,, \quad (29c)\end{cases}$$

where

$$a_L'=a_L+g_L(0) \tag{30a}$$
$$=a_+e^{-qb}+a_-e^{qb}+g_0(0)$$

$$a_R'=a_R+g_R(b) \tag{30b}$$
$$=a_+e^{qb}+a_-e^{-qb}+g_0(b)\,.$$

It can be verified in several ways that Eqs. (28)–(30) give the proper separation. One is to consider the case where $\epsilon_L(\omega)$ and $\epsilon_R(\omega)$ are free-electron-like and $\epsilon_0(\omega)$ is a real constant. It can be shown that (29) leads to the same transition rate (calculated in subsequent paragraphs) as that calculated by quantizing normal modes and using expression (8) with the orthogonal eigenfunctions (10)–(16). Another way to verify the separation is to observe that no bulk loss is evident in any numerical evaluations of the transition rate, i.e., nothing which goes as $\mathrm{Im}[-1/\epsilon_L(\omega)]$ or $\mathrm{Im}[-1/\epsilon_R(\omega)]$. Im denotes imaginary part.

The rate of energy loss due to φ_S is

$$\frac{dW_S}{dt}=-2\,\mathrm{Re}\int(-\nabla\varphi_S)^*\cdot\vec{J}\,d^3r\,. \tag{31}$$

Noting that

$$(-\nabla\varphi_S)^*\cdot\vec{J}=\varphi_S^*\nabla\cdot\vec{J}-\nabla\cdot(\varphi_S^*\vec{J})\,, \tag{32}$$

we see that (31) can be rewritten as

$$\frac{dW_S}{dt}=-2\,\mathrm{Re}\int\varphi_S^*\nabla\cdot\vec{J}\,d^3r +2\,\mathrm{Re}\int\nabla\cdot(\varphi_S^*\vec{J})\,d^3r\,. \tag{33}$$

The first term can be simplified by the use of the continuity equation (3):

$$-2\,\mathrm{Re}\int\varphi_S^*\nabla\cdot\vec{J}\,d^3r=2\omega\,\mathrm{Im}\int\varphi_S^*\rho\,d^3r\,. \tag{34}$$

The second integral in (33) can be converted to a surface integral of $\varphi_S^*\vec{J}$ by Gauss's theorem. Since $\varphi_S^*\vec{J}$ vanishes as $z\to\pm\infty$ and is uniform in x and y, the surface integral vanishes. Hence,

$$\frac{dW_S}{dt} = 2\omega A \,\mathrm{Im}\int_{-\infty}^{\infty} \varphi_S^*(z)\rho(z)\,dz \tag{35}$$

and the transition rate is

$$\frac{1}{\hbar\omega}\frac{dW_S}{dt} = \frac{2}{\hbar} A\int_{-\infty}^{\infty} \varphi_S^*(z)\rho(z)\,dz\,. \tag{36}$$

Substituting $\rho(z)$ from (21) and φ_S from (29) into (35), we obtain

$$\frac{dW_S}{dt} = 2\omega A \,\mathrm{Im}\,(a_L'^* I_L + a_R'^* I_R + a_+^* I_+ + a_-^* I_- + I_0)\,, \tag{37}$$

where

$$I_L = \rho_{L+}/(q+ik_+) + \rho_{L-}/(q+ik_-)\,, \tag{38a}$$

$$I_R = \rho_{R+}/(q-ik_+) + \rho_{R-}/(q+ik_-)\,, \tag{38b}$$

$$I_\pm = \sum_{i=1}^{4} \rho_i F(K_i \pm q)\,, \tag{38c}$$

$$I_0 = \frac{4\pi}{\epsilon_0^*(\omega)} \sum_{i=1}^{4}\sum_{j=1}^{4} \rho_i\rho_j^* \frac{F(K_i+K_j)}{(q^2-K_j^2)}\,, \tag{38d}$$

$$F(K) = (e^{Kb}-1)/K\,,\quad F(0)=b\,, \tag{38e}$$

$$K_1 = K_-\,,\quad K_2 = K_+\,,\quad K_3 = -K_+\,,\quad K_4 = -K_-\,. \tag{38f}$$

Note that dW_S/dt is independent of the direction of $\vec{q}$ since there is complete rotational symmetry about the z axis.

In what follows, $\epsilon_0(\omega)$ will be taken as a real constant for simplicity. However, the ω dependence can be retained and $\epsilon_0(\omega)$ can be complex where required. In fact if one wished to study the longitudinal optical phonons of the barrier, the formulation given above would be suitable. In the simplest case $\epsilon_0(\omega)$ would be given by $\epsilon_\infty + (\epsilon_0 - \epsilon_\infty)\omega_T^2/(\omega_T^2 - \omega^2 - i\omega\gamma)$ where ϵ_∞ is the optical frequency dielectric constant, ϵ_0 is the static dielectric constant, ω_T is the transverse optical phonon frequency, and γ is the phonon width.

D. Differential probability

The probability of inelastically tunneling is the rate of tunneling divided by the attempt frequency ν_L:

$$P = \frac{1}{\hbar\omega}\frac{dW_S}{dt}/\nu_L\,, \tag{39}$$

where

$$\nu_L = (\hbar k_L/m)/2L\,. \tag{40}$$

The inelastic tunneling probability for energy transfer in the interval $\hbar\omega$ to $\hbar\omega + \delta(\hbar\omega)$ and momentum transfer (in the plane of the junction) in the annulus between $\hbar q$ and $\hbar q + \delta(\hbar q)$ is [noting that $\delta k_R = (m/\hbar^2 k_R)\,\delta(\hbar\omega)$]

$$\delta P^{\mathrm{in}} = \frac{1}{\hbar\omega\nu_L}\frac{dW_S}{dt}\frac{AL}{4\pi^3}\,2\pi q\,\delta q\frac{m}{\hbar^2 k_R}\,\delta(\hbar\omega)\,. \tag{41}$$

It is convenient to define

$$E_q = \hbar^2 q^2/2m\,, \tag{42}$$

which is the kinetic energy in the plane of the junction of an electron in state ψ_R. We can write (41) as

$$\delta P^{\mathrm{in}} = \frac{\partial^2 P^{\mathrm{in}}}{\partial(\hbar\omega)\partial E_q}\,\delta E_q\,\delta(\hbar\omega)\,, \tag{43}$$

where the differential inelastic tunneling probability is defined as

$$\frac{\partial^2 P^{\mathrm{in}}}{\partial(\hbar\omega)\partial E_q} = \frac{m^2 AL^2}{\pi^2\hbar^5 k_L k_R}\frac{1}{\hbar\omega}\frac{dW_S}{dt}\,. \tag{44}$$

Note that $dW_S/dt \propto 1/AL^2$ so that the normalization length and area cancel out.

The elastic tunneling probability is

$$P^{\mathrm{el}} = 16k_L^2 K_L^2 e^{-2K_L b}/(k_L^2+K_L^2)^2\,. \tag{45}$$

A convenient quantity to examine is the relative differential probability

$$r(E_q,\hbar\omega) \equiv \frac{1}{P^{\mathrm{el}}}\frac{\partial^2 P^{\mathrm{in}}}{\partial(\hbar\omega)\partial E_q}\,. \tag{46}$$

Most of the gross features of the barrier potential (including its dependence on bias voltage) tend to cancel in such a quantity so we expect that the use of a rectangular barrier does not introduce any appreciable error.

IV. NUMERICAL RESULTS FOR Al-Al_2O_3-Ag

In this section the results of Sec. III are evaluated numerically. The complex dielectric functions $\epsilon_L(\omega)$ and $\epsilon_R(\omega)$ are taken (approximately) to be the optical dielectric functions of Al,[15] and Ag,[16] respectively. The barrier dielectric constant is $\epsilon_0 = 3$ and the width is $b = 30$ Å. The barrier height is $U_0 = E_F + 1.0$ eV and $E_L = E_F = 10$ eV.

In Fig. 1, the relative differential probability $r(E_q,\hbar\omega)$ is shown for $E_q = 0.1$ eV as a function of $\hbar\omega$. If q is in units of Å^{-1} and E_q is in eV, then

$$q = 0.513\sqrt{E_q}\,. \tag{47}$$

So $q = 0.16$ Å^{-1} in this figure. A well-defined surface plasmon occurs for this q as evidenced by the pronounced and fairly narrow peak at $\hbar\omega = 3.4$ eV. The full width at half maximum is 0.1 eV.

By holding $\hbar\omega$ fixed (at 3.0 eV) and varying E_q, another representation of the surface plasmon is obtained in Fig. 2. The locus of peak positions in the $\omega - q$ plane gives the dispersion curve shown in Fig. 3. For well-defined surface plasmons, this

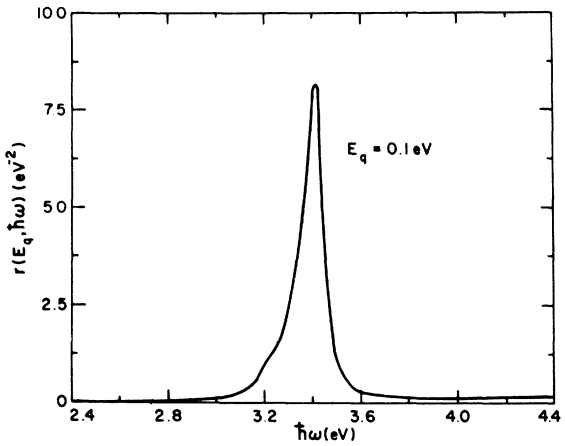

FIG. 1. Ratio of the differential probability for inelastic tunneling to the elastic probability vs $\hbar\omega$ for $E_q = 0.1$ eV ($q = 0.16$ Å^{-1}). The direction of tunneling is from Al to Ag in an Al-Al_2O_3-Ag junction.

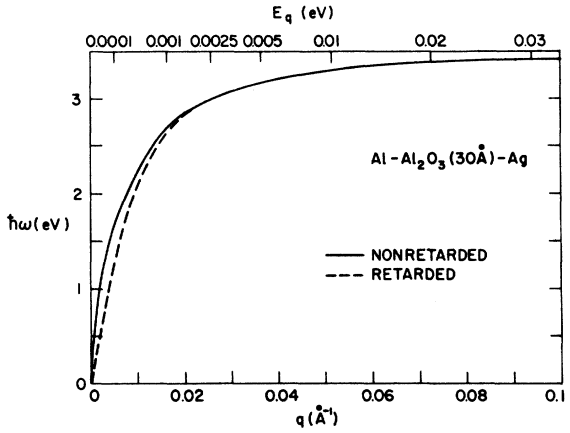

FIG. 3. Dispersion curve for surface plasmons in Al-Al_2O_3-Ag. The oxide thickness is 30 Å.

procedure gives the same results as setting $\eta = 0$ [see Eq. (26c)]. A full discussion of the various modes of a M-I-M structure is given in Ref. 2. Only the lowest mode (shown in Fig. 3) will be of importance here.

In Figs. 1 and 2, only tunneling in one direction, from Al to Ag, has been shown. Let us now examine the symmetry with respect to the interchange of left and right. This is equivalent to the symmetry with respect to bias voltage. If $\epsilon_L(\omega) = \epsilon_R(\omega)$ and the barrier potential is symmetric, the tunneling rate must be completely symmetric. If $\epsilon_L(\omega) \neq \epsilon_R(\omega)$ as in LM, some asymmetry is expected even if the barrier potential is symmetric (at zero bias). This is particularly true when $\hbar\omega$ and q are large. For example, for tunneling from left to right, an electron in the state ψ_L tunnels more readily than in ψ_R since $e^{-K_L b} \gg e^{-K_R b}$. If the surface plasmon field is large predominantly at the right side of the barrier ($z \simeq b$), the inelastic tunneling rate will be much larger than if the field were large on the left ($z \simeq 0$). For the Al-Al_2O_3-Ag system considered here, q must be $>1.6 \times 10^{-2}$ Å^{-1} ($E_q > 10^{-3}$ eV) before the surface plasmon field is sufficiently localized on the Ag side of the barrier to give a pronounced asymmetry. The relative differential probability for tunneling from Al to Ag and from Ag to Al for $E_q = 0.01$ eV is shown in Fig. 4. The integrated probability

$$y(E_q) = \int_{1\,\mathrm{eV}}^{4\,\mathrm{eV}} d(\hbar\omega)\, r(E_q, \hbar\omega) \tag{48}$$

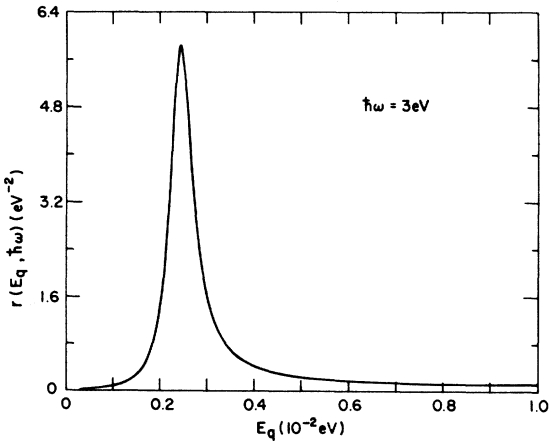

FIG. 2. Ratio of the differential probability for inelastic tunneling to the elastic probability vs E_q for $\hbar\omega = 3$ eV. The direction of tunneling is from Al to Ag.

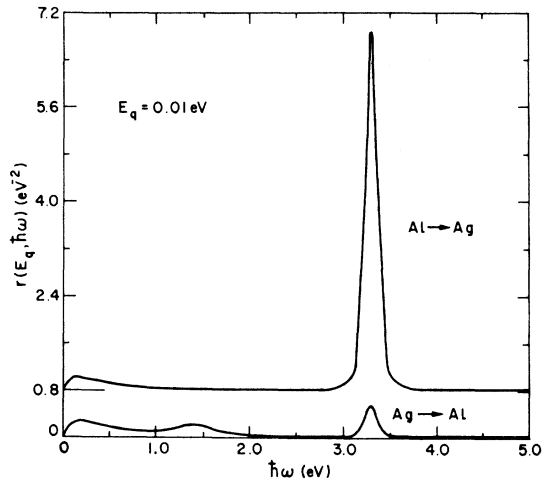

FIG. 4. Ratio of the differential probability for inelastic tunneling to the elastic probability vs $\hbar\omega$ for $E_q = 0.01$ eV ($q = 0.051$ Å^{-1}). The top curve is displaced by 0.8 eV^{-2} for clarity.

is shown in Fig. 5 as a function of E_q. The probability of inelastic tunneling from Al to Ag remains large for E_q up to ~0.1 eV before decreasing, whereas the probability for Ag to Al decreases appreciably when $E_q > 10^{-3}$ eV. In either case, the scattering is not nearly so peaked about the forward direction ($q = 0$) as for high-energy electron beams.[8]

If surface plasmons with large enough q radiate, bias asymmetry in the light emission can occur. Also, asymmetry in the barrier potential gives rise to asymmetry in the inelastic current and consequently in the emission of light. It should be noted that the effects of barrier asymmetry have not been investigated. At large bias voltage where surface plasmons are observed, the barrier is very asymmetric. In fact, one is usually in the Fowler-Nordheim regime where electrons tunnel into the oxide conduction band before reaching the electrode. Presumably, the ratio of inelastic to elastic tunneling (which is considered in this paper) is not affected significantly.

Another useful quantity to study is

$$R(E_{\max}, \hbar\omega_{\max}) = \int_0^{\hbar\omega_{\max}} d(\hbar\omega) \int_0^{E_{\max}} dE_q \, r(E_q, \hbar\omega) , \tag{49}$$

where $r(E_q, \hbar\omega) \equiv 0$ for $E_q > E_L - \hbar\omega$. This quantity is the relative probability of a normally incident electron (of energy E_L) to tunnel inelastically with energy transfer $\leq \hbar\omega_{\max}$ and momentum transfer $\leq (2mE_{\max})^{1/2}$. If all the electromagnetic modes of the junction for which $q \leq (2mE_{\max})^{1/2}/\hbar$ were radiative with 100% efficiency and the rest were nonradiative, $R(E_{\max}, \hbar\omega_{\max})$ would be the quantum efficiency for light emission of photons with energy $\leq \hbar\omega_{\max}$. From the nature of radiation induced by surface roughness,[3] there is some reason to believe this form is approximately correct since most of the light emitted is due to modes for which q is less than a cutoff. The cutoff is proportional to the reciprocal of the statistical correlation length of the surface roughness.

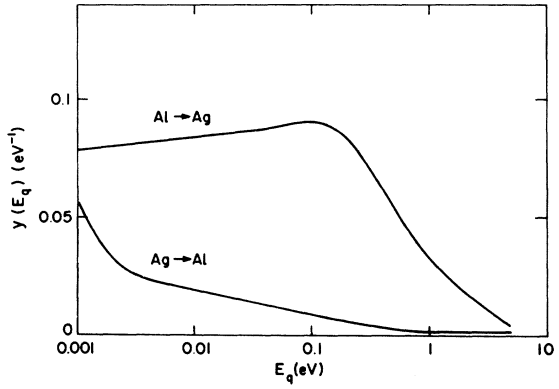

FIG. 5. Integrated probability $y(E_q) = \int_{1\,\mathrm{eV}}^{4\,\mathrm{eV}} d(\hbar\omega)\, r(E_q, \hbar\omega)$ vs E_q.

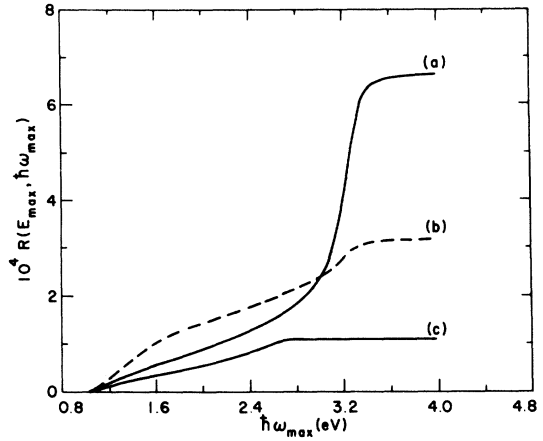

FIG. 6. Ratio of inelastic to elastic tunneling probability vs $\hbar\omega_{\max}$. (a) Al to Ag, $E_{\max} = \hbar^2 q^2_{\max}/2m = 0.01$ eV ($q_{\max} = 0.051$ Å^{-1}); (b) Ag to Al, $E_{\max} = 0.01$ eV; (c) $E_{\max} = 10^{-3}$ eV ($q_{\max} = 1.6 \times 10^{-2}$ Å^{-1}), no significant difference between Al to Ag and Ag to Al.

To determine the probability of exiting modes which are apt to radiate, $R(E_{\max}, \hbar\omega_{\max})$ is plotted as a function of $\hbar\omega_{\max}$ for various $E_{\max}$ in Figs. 6–8. The contribution from $\hbar\omega \leq 1$ eV has been omitted since photons of that energy are not detected in the LM experiments. In Fig. 6, the probability for values of $E_{\max} \leq 0.01$ eV ($q_{\max} \leq 0.051$ Å^{-1}) is shown. The modes involved here are likely to be the most radiative. For $E_{\max} \leq 10^{-3}$ eV, there is little difference between Al to Ag, and Ag to Al. In Figs. 7 and 8, larger values of $E_{\max}$ are considered. For Al to Ag and large $E_{\max}$ the inelastic rate is ~0.1 of the elastic rate, a substantial value. The values of

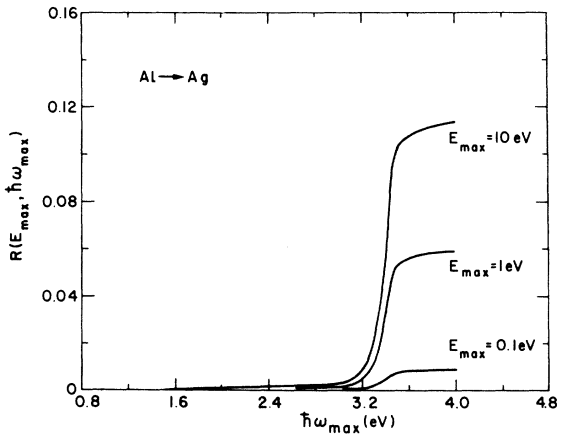

FIG. 7. Ratio of inelastic to elastic tunneling probability vs $\hbar\omega_{\max}$ for Al to Ag for $E_{\max} = 10, 1$ and 0.1 eV ($q_{\max} = 1.6$, 0.51, and 0.16 Å^{-1}).

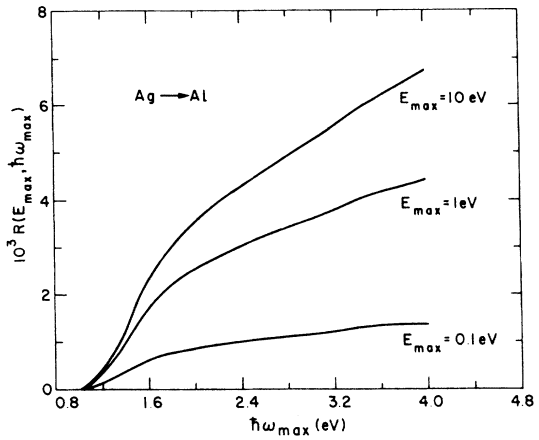

FIG. 8. Same as Fig. 7 except for tunneling from Ag to Al.

$R(E_{max}, \hbar\omega_{max})$ for $\hbar\omega_{max}=4$ eV are plotted as a function of E_{max} in Fig. 9. Clearly much of the scattering responsible for the large inelastic rate is at high values of q which may not radiate easily.

Since nonlocal effects have not been included (i.e., the wave vector dependence of the dielectric functions has been neglected), the contribution to the inelastic tunneling from large momentum transfers may be not be calculated entirely accurately. Also, surface-plasmon peaks may be broadened by Landau damping[7] and by the roughness of the electrodes,[3] which have not been included. However, it is not certain that any of these effects will diminish the total (integrated) inelastic probability, although they may redistribute the intensity in q and ω.

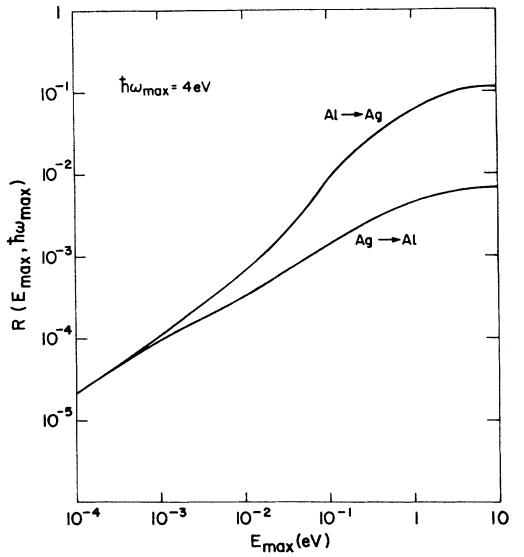

FIG. 9. Ratio of inelastic to elastic tunneling probability vs E_{max} for $\hbar\omega_{max}=4$ eV.

In the LM experiments, the Ag layer is thin (~300 Å), not infinitely thick as assumed. This introduces an additional plasma mode into the problem, namely, the free Ag surface plasmon. There is coupling between this mode and the one shown in Fig. 3 which shifts the dispersion curves.[2] Also, tunneling electrons can interact with the additional mode. However, these effects are not expected to be large in junctions of practical interest.

V. COMPARISON TO INELASTIC TRANSFER HAMILTONIAN

Since a common method of treating inelastic (assisted) tunneling involves the inelastic transfer Hamiltonian,[7,11,12] (ITH), it is worthwhile to examine it in more detail for surface-plasmon excitation. As remarked in Sec. I, for situations where the normal modes can be found and can be quantized, the transition rate for a mode with momentum $\hbar\vec{q}$ (in the plane of the junction) is

$$(2\pi/\hbar)\,|\langle\psi_R|-e\phi_{\vec{q}}^{sp}|\psi_L\rangle|^2\delta(\hbar\omega_{\vec{q}}+E_R-E_L)\,, \quad (50)$$

where $\phi_{\vec{q}}^{sp}$ is the potential for the mode and $\omega_{\vec{q}}$ is the frequency. Now in ITH, ψ_L and ψ_R are of the form (10) and (11), but χ_L and χ_R are given by

$$\chi_L=\begin{cases}(2/L)^{1/2}\sin(k_Lz-\gamma_L)\,, & z<0 \quad (51a)\\ -(2/L)^{1/2}\sin\gamma_L\, e^{-K_Lz}\,, & z>0 \quad (51b)\end{cases}$$

and

$$\chi_R=\begin{cases}(2/L)^{1/2}\sin\gamma_R e^{K_R(z-b)}\,, & z<b \quad (52a)\\ (2/L)^{1/2}\sin[k_R(z-b)+\gamma_R]\,, & z>b\,, \quad (52b)\end{cases}$$

where (dropping L and R indices)

$$\tan\gamma=k/K\,.$$

The z integration in $\langle\psi_R|-e\phi_{\vec{q}}^{sp}|\psi_L\rangle$ is over the region $0<z<b$.

When the modes are well defined, the energy-loss method reduces to the same form, Eq. (50), except that ψ_L and ψ_R are the eigenfunctions given by (10)–(15) and the region of integration includes the electrodes ($z<0$ and $z>b$) as well as the barrier.

It has been found for M-I-M junctions for small momentum transfer, the difference between the two approaches can be substantial, e.g., the transition rate can differ by a factor of 10–100 (ITH overestimates the rate). The reason for this difference is the large cancellation between the contribution to the matrix element from the electrodes and that from the barrier.

Another way to view this difference is to compute the transition current and charge for the wave

functions used in ITH [Eqs. (10), (11), (51), and (52)]. The transition current density is of the form

$$\vec{J}(z) = \begin{cases} \vec{J}^{(0)}(z), & 0 \le z \le b \quad (53a) \\ 0, & \text{otherwise}, \quad (53b) \end{cases}$$

where $\vec{J}^{(0)}(z)$ is given by (1) with the factor $e^{i(qx-\omega t)}$ omitted. By definition, there is no interaction in the electrodes and, hence, $\vec{J}(z)$ must vanish there. For the continuity equation (3) to be satisfied, the transition charge density must be

$$\rho(z) = \rho^{(0)}(z) + (i/\omega)\vec{J}_z^{(0)}(b)\delta(z-b) - (i/\omega)\vec{J}_z^{(0)}(0)\delta(z), \quad 0 \le z \le b, \quad (54a)$$

$$\rho(z) = 0, \quad \text{otherwise}, \quad (54b)$$

where

$$\rho^{(0)}(z) = -(e/LA)\chi_R(z)\chi_L(z). \quad (55)$$

In ITH, the surface charge densities $\pm(i/\omega)\vec{J}_z^{(0)}$ are omitted even though they are not small. Including the surface charges improves the agreement at small $\vec{q}$, yet this does not make it a correct theory.

VI. SUMMARY

A new theory of the excitation of surface plasmons by tunneling electrons has been given. The theory is similar to energy-loss calculations for electron beams,[8] except that the electrons are treated quantum mechanically, not classically. It represents an improvement in generality and accuracy over previous work.[7] Undoubtedly the methods discussed in this paper can be applied to other experiments (such as photoemission and field emission) where electrons interact with the plasma oscillations of a solid.

The system Al-Al_2O_3-Ag has been studied because of its potential as a light-emitting device.[1] It has been found that the quantum efficiency for the excitation of surface plasmons is approximately 10%. However, most of the modes excited are at q values which are too large to radiate efficiently unless surface roughness with extremely small wavelengths can be introduced. Most of the radiation comes from small q. Consequently, the bias asymmetry of the light emission is not very large.

ACKNOWLEDGMENTS

The author would like to thank Dr. J. Lambe and Dr. S. L. McCarthy for numerous discussions of their experiments and the interpretation of them. The assistance of Professor G. W. Ford is gratefully acknowledged.

[1] John Lambe and S. L. McCarthy, Phys. Rev. Lett. 37, 923 (1976).
[2] E. N. Economou, Phys. Rev. 182, 539 (1969).
[3] E. Kröger and E. Kretschmann, Z. Phys. 237, 1 (1970); J. M. Elson and R. H. Ritchie, Phys. Status Solidi B 62, 461 (1974); D. L. Mills, Phys. Rev. B 12, 4036 (1975); A. A. Maradudin and W. Zierau, Phys. Rev. B 14, 484 (1976).
[4] R. C. Jaklevic and John Lambe, Phys. Rev. Lett. 17, 1139 (1966); John Lambe and R. C. Jaklevic, Phys. Rev. 165, 821 (1968).
[5] T. L. Hwang, S. E. Schwarz, and R. K. Jain, Phys. Rev. 36, 379 (1976); D. P. Siu, R. K. Jain, and T. K. Gustafson, Appl. Phys. Lett. 28, 407 (1976).
[6] D. C. Tsui, Phys. Rev. Lett. 22, 293 (1969). Also see C. B. Duke, M. J. Rice, and F. Steinrisser, Phys. Rev. 181, 733 (1969); C. B. Duke, *ibid.* 186, 588 (1969); D. C. Tsui and A. S. Barker, Jr., *ibid.* 186, 590 (1969). Steinrisser *et al.* independently observed the effect but attributed it to bulk plasmons.
[7] K. L. Ngai, E. N. Economou, and M. H. Cohen, Phys. Rev. Lett. 22, 1375 (1969); K. L. Ngai and E. N. Economou, Phys. Rev. B 4, 2132 (1971).
[8] R. H. Ritchie, Phys. Rev. 106, 874 (1957); E. Kröger, Z. Phys. 216, 115 (1968).
[9] L. I. Schiff, *Quantum Mechanics*, 2nd ed. (McGraw-Hill, New York, 1955), p. 261.
[10] R. H. Ritchie and H. B. Eldridge, Phys. Rev. 126, 1935 (1962).
[11] A. J. Bennett, C. B. Duke, and S. D. Silverstein, Phys. Rev. 176, 969 (1968).
[12] C. B. Duke, *Tunneling in Solids* (Academic, New York, 1966).
[13] R. E. Prange, Phys. Rev. 131, 1083 (1963).
[14] A. D. Brailsford and L. C. Davis, Phys. Rev. B 2, 1708 (1970); L. C. Davis, *ibid.* 2, 1714 (1970).
[15] H. Ehrenreich, H. P. Philipp, and B. Segall, Phys. Rev. 132, 1918 (1963).
[16] P. B. Johnson and R. W. Christy, Phys. Rev. B 6, 4370 (1972).