

Hot-electron picture of light emission from tunnel junctions

J. R. Kirtley, T. N. Theis, J. C. Tsang, and D. J. DiMaria

IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598

(Received 8 November 1982)

Metal-insulator-metal tunnel junctions emit optical radiation when biased at voltages in the range 2–4 V. We argue that a complete picture of this radiation process includes hot electrons, which excite surface electromagnetic resonances, which are in turn coupled to external radiation through surface roughness. This picture is supported by measurements of the temperature and second-metal-electrode thickness dependence of the emission intensities, and by light emission from surface plasmons excited by optical pumping and by charge injection.

I. INTRODUCTION

Lambe and McCarthy¹ pointed out that tunnel junctions emitted optical radiation when biased at voltages V_0 in the range 2–4 V. The emitted radiation became more intense if the junctions were intentionally roughened. The emitted radiation from randomly roughened junctions was broadband with a characteristic linear onset below a critical frequency ν_c , related to the junction bias voltage by $h\nu_c = eV_0$. Light-emitting tunnel junctions have some of the attributes of an attractive display device: They are flat, operate at room temperature in air, require low dc bias voltages, have emission frequencies with an upper cutoff tunable through the visible, and are inexpensive to make. However, the best demonstrated external quantum efficiencies (the number of photons out divided by the number of electrons crossing the tunneling barrier which are approximately the external power efficiencies) are of order 10^{-4} .^{2–7} It is, therefore, of practical as well as fundamental interest to understand the emission processes.

Several papers have described light emission from tunnel junctions in terms of a two-step process: (1) inelastically tunneling electrons excite collective electromagnetic oscillations of the junction, and (2) the electromagnetic oscillations emit external radiation.^{1,8–11} There are two classes of electromagnetic oscillations relevant to the tunneling junction geometry.¹² The first is the junction or slow mode. This mode has fields and energies located primarily in the junction region, and can be described as primarily electrostatic. Because of the screening between the two metal-insulator interfaces, the junction mode has a speed of propagation much slower than that of light in free space. The second relevant mode is the Ag-air interface fast surface-plasmon polariton. It has fields and charges localized pri-

marily at the Ag-air interface, and propagates at a speed quite close to that of light in free space. Surface roughness is required for either mode to radiate light, since energy and momentum cannot be simultaneously conserved in a radiative transition without roughness scattering.

It has also been reported that radiation from tunnel junctions is dominated by the junction mode.^{10,11} In this view the coupling to surface plasmons is relatively efficient, since the slow mode has large field strengths in the tunneling region, but that the radiative step is relatively inefficient. The inefficiency of the radiative step could be attributed to the large difference between the wavelengths of the junction modes and light. Roughness on a scale of 1–10 nm would be required to efficiently couple the junction mode to light. Roughness on this scale, while possibly present, is difficult to characterize to test these ideas experimentally.

By fabricating tunnel junctions on holographically produced gratings with single Fourier components of roughness of order 800 nm, we were able to show that radiation from the fast mode was an important mechanism for light emission from tunnel junctions in general, and, in fact, dominated over radiation from the junction mode in our samples.^{6,7}

Since the electric fields associated with the fast surface mode extend throughout the metal-insulator-metal structure, and are strongest at the Ag-air interface, the coupling between the tunneling electrons and the surface plasmons may well occur outside the tunneling barrier region. We therefore make a distinction between inelastic tunneling and hot-electron coupling to surface plasmons: We describe a process in which the electrons lose energy in the barrier region as inelastic. Conversely, we describe a process in which the electrons first tunnel into one of the electrodes before losing energy as hot. A complete description of the tunneling-

electron—surface-plasmon coupling process would include both events as appropriate limits. While such a comprehensive description does not at present exist, important qualitative conclusions about the dynamics of the system result if we make this distinction.

As noted above, light emission from randomly roughened tunnel junctions has a linear onset below a critical frequency. This can be explained quite naturally in the inelastic tunneling picture in terms of the linearly increasing final density of states available to the tunneling electrons with decreasing energy loss.⁸ However, it is difficult to make this qualitative explanation fit the data quantitatively. As will be shown below, when the energy dependence of the tunneling matrix elements are taken into account, the predicted onset below the critical frequency is slower than linear, and is, in fact, slower than observed experimentally from tunnel junctions on gratings.

Further, we will present in this paper a number of experimental results which are difficult to explain using the inelastic tunneling picture: (1) The emission from junctions on gratings becomes exponentially less intense as the second-metal electrode is made thicker. The characteristic intensity falloff length does not match the optical screening length, and is dependent on sample-preparation procedures; (2) the emission-peak intensities increase dramatically as the junction temperatures decrease, even though the junction current-voltage characteristics and optical properties are not strongly temperature dependent; (3) similar external efficiencies and radiation patterns result if the surface plasmons are pumped by optical radiation or by charge injection. In neither the optical nor the charge-injection case is inelastic tunneling present.

All of these results can be understood using a three-step or hot-electron model: (1) Hot electrons are injected into the junction structure; (2) the injected hot-electron distribution relaxes primarily through the emission of phonons and cooler electrons, but also through the emission of surface plasmons; (3) the surface plasmons couple out to external radiation through surface roughness. The hot-electron model has important qualitative consequences which will be explored.

II. TUNNELING RESULTS

We have described light emission from tunnel junctions on gratings previously.^{6,7} In these papers the results were analyzed in terms of the inelastic tunneling picture. We will show in this paper how the old results and new results to be presented below are more consistent if viewed in the hot-electron picture. We will merely outline previously presented

experimental details here. Al-Al₂O₃-Ag tunnel junctions were fabricated on (80–1200)-nm periodicity, (0–100)-nm amplitude and holographically produced corrugated gratings. In some samples the photoresist was itself used as a substrate, while in other samples the pattern was transferred from the resist to a Si substrate by ion milling. Tunnel junctions were formed on the grating substrates. The junctions were formed by evaporating 2-mm-wide, 10-mm-long, and 40-nm-thick Al films through mechanical masks, oxidizing the films, and then completing the junctions with a long 2-mm-wide (15–80)-nm-thick Ag film. All electrical measurements were made with a four-terminal technique. The substrates were mounted on the cold finger of a closed-cycle refrigerator with optical windows, and run in vacuum. The emitted radiation was frequency analyzed using a single-pass monochromator and photon counting. Only the light through a narrow aperture 1.4° wide and 9.5° high, in the plane defined by the junction normal and the grating periodicity wave vector, was allowed to enter the spectrometer.

The emitted light was composed of narrow angle-tunable peaks superimposed on a broad background. Analysis of the peak energies as a function of emission angle showed that the radiation was dominated by emission from the Ag-air interface fast surface-plasmon polariton. It is possible that the broad background we observed had contributions from the junction or localized plasmon modes. All of the data presented here has had the broad emission subtracted from the fast-mode peaks. This allowed us to analyze only the radiation from well-characterized modes.

Our experimentally measured dispersion curves, linewidths, and dependence of peak intensities on grating amplitudes agreed well with a theory of Laks and Mills¹³ using the inelastic tunneling model. This should not be surprising since these properties depend only on the electrostatics of the system. However, there were serious discrepancies between theory and experiment for those properties which depended on the coupling between tunneling electrons and surface plasmons. For example, the observed intensities were at least 35 times stronger than predicted theoretically.⁶

One of the remarkable points of agreement between the theory of Laks and Mills¹³ and our experiments occurred when we compared the integrated peak intensities for a single sample for a series of bias voltages. Laks and Mills,¹³ following a suggestion by Hone *et al.*,⁸ wrote the radiation from a tunnel junction as

$$\frac{dP}{d\Omega d\omega} = |I(\omega)|^2 A(\omega, d, \epsilon), \quad (1)$$

where $dP/d\Omega d\omega$ is the power radiated per unit solid angle, per unit frequency interval, $A(\omega, d, \epsilon)$ is an "antenna factor" which depends on the dielectric properties of the junction structure, but does not depend on the bias conditions, and $|I(\omega)|^2$ is the power-spectral density of current fluctuations across the junction, which does depend on the details of the charge transport across the junction. We define the operator \hat{I} for current across the junction as

$$\hat{I} = \frac{ie}{\hbar} \sum_{k,q,s} (T_{kq} c_{ks}^\dagger c_{qs} + T_{kq}^\dagger c_{ks} c_{qs}^\dagger), \quad (2)$$

where T_{kq} is the tunneling matrix element from a state k in one metal electrode to a state q in the other, and c_{ks}^\dagger and c_{ks} are the electron annihilation and creation operators for a state with momentum k and spin s in one metal electrode. The power-spectral density of current fluctuations across the junction with frequency ω is defined by

$$|I(\omega)|^2 = \sum_f |\langle f | \hat{I} | 0 \rangle|^2 \delta(\omega - (E_f - E_0)/\hbar). \quad (3)$$

This expression has been evaluated in the limit of small frequencies by Hone *et al.*⁸:

$$|I(\omega)|^2 = \frac{eI_0}{2\pi} \left[1 - \frac{\hbar\omega}{eV} \right], \quad (4)$$

where I_0 is the current per unit area through the junction. When this simple expression is used to divide the current fluctuation factor from the observed light-emission intensities from tunneling junctions on gratings, the reduction is very good.⁶ That is, the antenna factor is relatively independent of bias conditions for a given junction, as it should be. This agreement is illustrated in Fig. 1 for an Al-Al₂O₃-Ag tunnel junction on an 800-nm periodicity photoresist grating for biases from 1.8–2.6 V.

The solid curve in Fig. 1 was obtained by fitting the logarithm of the experimental values to a fourth-order polynomial. If we define a reduced χ^2 value for the overlap of the curves as

$$\chi^2 = \frac{1}{n} \sum_{i=1}^n \left[\frac{Y(i) - Y_{\text{fit}}(i)}{Y(i)} \right]^2, \quad (5)$$

we obtain a χ^2 value of 0.0275.

However, the current fluctuation spectrum looks a good deal different if one takes into account the finite-energy losses present (2–3 eV) under experimental conditions. A physical interpretation of Eq. (3) is that electrons are transferred from a high-energy state in one electrode to a low-energy state in the other, and back again. Since more energetic

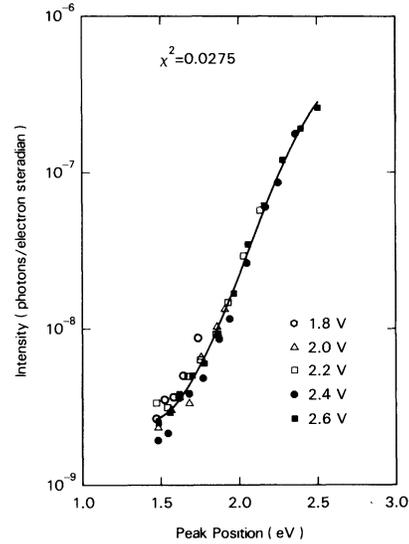


FIG. 1. Plot of the antenna factor for an Al-Al₂O₃-Ag tunnel junction on a grating obtained by dividing the integrated peak intensities by the expression for the power-spectral density of current fluctuations across the tunnel junction derived by Hone *et al.* [Eq. (4)]. The antenna factor should be independent of the bias conditions, so the curves for different bias voltages should overlap, as they do. The scatter between the experimental values and the smooth polynomial curve fitted to them has a χ^2 value of 0.0275.

electrons are more likely to penetrate the tunneling barrier, low-frequency fluctuations, in which the electrons on the average have more energy, are more likely to occur than high-frequency fluctuations. A simple technique for accounting for the dependence of the barrier penetration probabilities on energy is to split the tunneling matrix element $|T_{kq}|^2$ into two parts: One which transfers the electron from the left to the right electrode at the initial energy, and a second which transfers the electron from the right electrode to the left at the initial energy minus the energy associated with the fluctuation frequency. This is clearly an approximation to the real process, but is probably more correct than simply taking the low-frequency limit to the barrier penetration probabilities.

If we take the WKB approximation¹⁴ for the electronic wave functions, the low-temperature limit for the Fermi occupation functions, and use the notation and approximations that Simmons¹⁵ uses for the case of elastic tunneling, the power-spectral density of current fluctuations across the tunneling barrier per unit area is given by

$$|I(\omega)|^2 = \frac{2me^2}{h^3} \left[(eV - \hbar\omega) \int_0^{\epsilon_F - (eV - \hbar\omega)} dE_x \exp \left[- \int_0^1 dx [K(x) + Q(x)] \right] + \int_{\epsilon_F - (eV - \hbar\omega)}^{\epsilon_F} dE_x (\epsilon_F - E_x) \exp \left[- \int_0^1 dx [K(x) + Q(x)] \right] \right], \quad (6)$$

where ϵ_F is the Fermi energy of one of the metal electrodes, V is the voltage bias across the junction,

$$K(x) \equiv \left[\left[\frac{2m}{\hbar^2} \right] [U(x, V) - E_x] \right]^{1/2},$$

and

$$Q(x) \equiv \left[\left[\frac{2m}{\hbar^2} \right] [U(x, V) + \hbar\omega - E_x] \right]^{1/2}.$$

$U(x, V)$ is the position- and voltage-dependent barrier potential seen by the tunneling electrons, x is the Cartesian coordinate normal to the junction interfaces, l is the thickness of the barrier, and E_x is the kinetic energy of the tunneling electrons normal to the interface.

A comparison of the predictions of Eqs. (4) and (6) are plotted in Fig. 2 for an Al-Al₂O₃-Ag junction with $\phi_1=2.5$ eV, $\phi_2=3.5$ eV, and $l=1.3$ nm. The curves are normalized to the zero-frequency power-spectral density $|I(0)|^2 = eI_0/2\pi$. The solid curves are the results of the full expression [Eq. (6)]; the

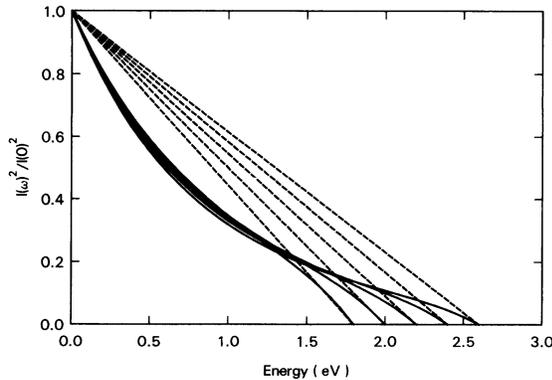


FIG. 2. Plot of the power-spectral density of current fluctuations for a tunnel junction with a trapezoidal barrier with heights $\phi_1=2.4$ eV, $\phi_2=3.5$ eV, and thickness $l=1.3$ nm. The dashed curves correspond to the predictions of the simple expression of Hone *et al.* [Eq. (4)]; the solid curves result from accounting for the energy dependence of the barrier penetration probabilities [Eq. (6)]. The more complex expression falls below the simple one, indicating that one might expect onsets in intensity below the critical frequencies that are slower than linear.

dashed curves are the low-frequency limit [Eq. (4)]. As expected, the high-frequency fluctuations fall below the curves predicted by the simple expression. As can be seen in Fig. 3, when the power-spectral-density term of Eq. (6) is divided out of the experimental data to obtain the antenna factor, the overlap between the curves for the different bias voltages is appreciably worse (a χ^2 of 0.0637) than that obtained using the simple expression (Fig. 1): The simple expression works better than it should. We argue that this is because the physics involved in the emission process is very different from the inelastic tunneling model, and that the good agreement of the simple inelastic model is probably fortuitous.

Another discrepancy was observed when the emission-peak energies were measured as a function of Ag-film thickness. These results have been detailed and supported by numerical results in Ref. 6. We outline them here to emphasize the close connection with the other results discussed in this paper. The observed peak widths and energies did not

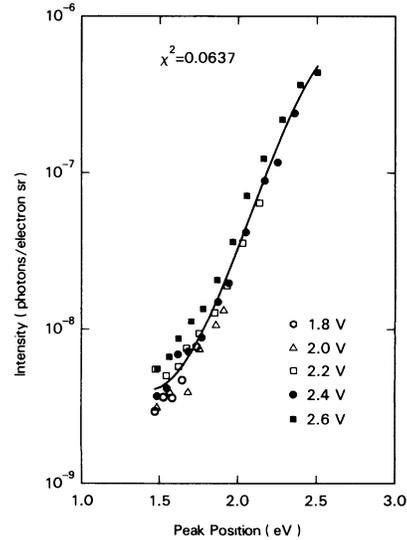


FIG. 3. Plot of the data of Fig. 1 for the antenna factor using the power-spectral density of fluctuations across the tunnel junction of Eq. (6). The curves for the different bias voltages do not overlap as well as in Fig. 1: Accounting for the energy dependence of the barrier penetration probability makes the agreement between the inelastic tunneling model and experiment worse, raising the χ^2 value to 0.0637.

change appreciably as the Ag was made thicker, but the emission efficiencies (normalizing out the bias currents) became exponentially smaller. Figure 4(a) shows a plot of the relative emission efficiencies (photons/electron sr) for a set of Al-Al₂O₃-Ag junctions on a 1.1- μ m Si grating. The emission intensities fell off with Ag-film thickness d with the approximate relation $I = Ce^{-d/t}$, where t was about 20 nm. There may also have been some weak energy dependence: The peaks at lower energies appeared to fall off more rapidly with Ag thickness than those at higher energies. Also plotted in Fig. 4(a) is the relative optical screening measured from Ag films on control quartz wafers, using the attenuation of the 632.8-nm line from a helium-neon laser. This screening length does not depend strongly on energy in this region. The experimentally determined fall-off length was significantly longer than the optical screening length for this sample. In contrast, Fig. 4(b) shows the dependence of emission-peak efficiencies on Ag-film thickness for a set of junctions on an 815-nm periodicity grating. In this case the characteristic falloff lengths were shorter than the optical screening length.

The radiative efficiency of surface plasmons is relatively insensitive to Ag-film thickness, as has been demonstrated by Moreland, Adams, and Hansma.¹⁶ Therefore, the dependence of emission intensities on film thickness that we observe must reflect the efficiency of production of surface plasmons by tunneling electrons. The theory of Laks and Mills for light emission from tunnel junctions on grat-

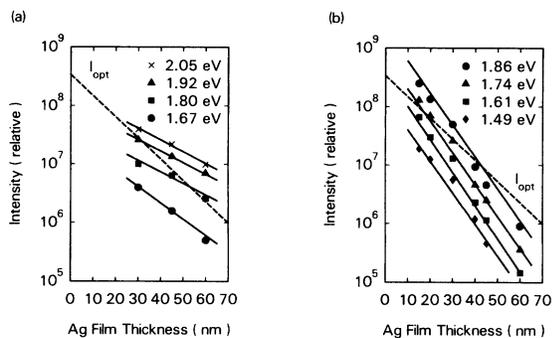


FIG. 4. Comparison of the Ag-film thickness dependence of emission intensities for Al-Al₂O₃-Ag light-emitting tunnel junctions on gratings. Junctions fabricated on a relatively smooth Si substrate and with fast Ag evaporation rates have intensities that decrease with Ag-film thickness at a slower rate than junctions fabricated on photoresist substrates with slow Ag evaporations. In neither case is the characteristic length the optical screening length, as one would expect if the interaction between the tunneling electrons and the surface plasmons occurred in the barrier region.

ings¹³ includes a current fluctuation driving term that models the spatial and frequency dependences of inelastic tunneling transitions in the junction. If the inelastic transitions occur only in the tunneling barrier region, the falloff in emission intensities with Ag-film thickness should be given by the optical screening length, since the surface-plasmon fields decay from the Ag surface with this characteristic length. If, however, the inelastic transitions extend throughout the junction volume, there should be no attenuation with Ag-film thickness. In neither case should coupling of the tunneling electrons to surface plasmons in the bulk of the junction result in an attenuation length shorter than the optical screening length. We, therefore, interpret the results of Fig. 4 to mean that the coupling of the tunneling electrons to surface plasmons occurs not in the junction volume, but primarily at the Ag-air interface. The results of Fig. 4 are then a measure of the hot-electron attenuation length in the Ag films.

The attenuation lengths for the two sets of junctions differ because of differences in sample fabrication: The first difference is that the samples of Fig. 4(a) were made on an etched Si wafer while those of Fig. 4(b) were made on a photoresist grating. Electron micrographs show that the Si wafer is significantly smoother. The small-scale roughness present in the photoresist films could act as agglomeration sites, tending to reduce the average film grain sizes. The second difference is that the Ag films in Fig. 4(a) were evaporated at 1–2 nm/sec while those of Fig. 4(b) were evaporated at 0.2–0.4 nm/sec. The slower evaporation rate might be expected to result in higher defect densities. The differences in substrate roughness and evaporation rates could both be expected to reduce the hot-electron mean free path for the sample of Fig. 4(b) relative to that of Fig. 4(a). The value of ~ 20 nm for the hot-electron mean free path at 2 eV for the sample of Fig. 4(a) agrees well with 18 nm reported by Crowell and Sze.¹⁷

It is reasonable to expect hot electrons to couple to surface plasmons more strongly at the Ag-air interface than in the bulk of the film. The argument goes as follows: The wavelengths for electrons at the Fermi surface of a free-electron metal are of order 0.1 nm. The surface-plasmon fields decay into the metal with an attenuation length of about 10 nm. Therefore, the overlap integrals involved in the hot-electron–surface-plasmon matrix elements due to the bulk of the metal tend to be small. On the other hand, the surface-plasmon fields at the Ag-air interface change rapidly on a scale of 0.1 nm, and the surface contributions to the matrix elements can be large. Another way of saying this is that (in the free-electron model) momentum cannot be conserved

in surface-plasmon emission from hot electrons in the bulk metal, but the surface can provide the required momentum. Large surface contributions have, in fact, been observed in photoemission from metals.¹⁸ We have developed a simple theory for surface coupling of tunneling electrons to surface plasmons which will be published separately.¹⁹ The important point for this discussion is that the dependence of emission efficiencies on Ag-film thickness in such a model does not follow the optical attenuation lengths. Therefore, the interaction between tunneling electrons and surface plasmons cannot be exclusively in the barrier region.

A second indication of hot-electron effects comes from measurements of the temperature dependence of the light emission from junctions on gratings. In Fig. 5(a) we plot the peak intensity from an Al-Al₂O₃-Ag junction on a 10-nm amplitude, 1225-nm periodicity grating as a function of temperature. The sample was rotated with respect to the optical

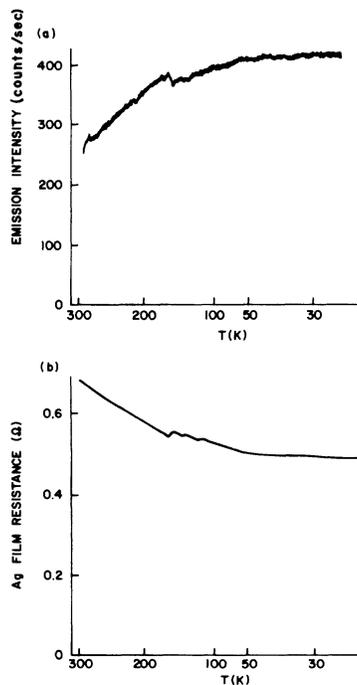


FIG. 5. Emission intensity vs temperature and Ag-film resistance vs temperature for an Al-Al₂O₃-Ag tunnel junction on a 1225-nm periodicity, 10-nm amplitude quartz grating substrate. The sample was rotated so that the surface-plasmon emission peak was observed at 1.77 eV. The junction bias voltage was held constant at 2.13 V with a nearly constant bias current. As the temperature was lowered the emission intensity increased, presumably because of an increase in the hot-electron mean free path in the Ag, as evidenced by the decrease in the Ag-film resistance.

aperture until the surface-plasmon emission peak was at 1.77 eV. The sample bias current was held constant at 44.1 mA; the sample bias voltage was approximately 2.13 V. The temperature was swept down over a period of about 90 min. The junction resistance changed slightly as the temperature was lowered such that the voltage across the junction increased by 5% in going from room temperature to 25 K. This slight change in the bias voltage was not nearly large enough to explain the $\sim 60\%$ increase in the emission intensity as the sample temperature was lowered. In addition, the emission-peak widths did not change appreciably as the temperature was lowered from room temperature to 25 K. This indicated that the damping of the surface plasmons was not strongly dependent on temperature. We determined that the optical attenuation lengths of the Ag films were not changing with temperature by passing the beam from a He-Ne laser through a 28-nm Ag film on a glass substrate and measuring the transmitted intensity. The optical transmission of the film at 632.8 nm was unchanged to within our experimental accuracy of 10% over a temperature range of 300–25 K.

In Fig. 5(b) we plot the Ag-counterelectrode film resistance as a function of temperature for the same sample as that of Fig. 5(a). The Ag-film resistance went down about 20% as the temperature decreased. The resistance of the film as a function of temperature had sharp structure, presumably due to morphological changes. This structure was quite reproducible but hysteretic: The changes occurred at higher temperatures when the temperature was swept up than when it was swept down. The salient point is that the structure in the Ag-film resistance was reproduced in the junction emission intensity. The changes in the strip resistance were too small to affect the bias of the junction significantly: The effective junction resistance was about 50 Ω at this bias voltage; the Ag-strip resistance was less than 1 Ω and changed less than 0.2 Ω . We therefore interpret these results to mean that in addition to the lowering of the Ag-film resistance due to a freezing of the metal phonons, there were also electron scattering mechanisms in the Ag film that changed as the temperature was lowered. The hot-electron mean free path increased as the temperature was lowered, increasing the probability for the hot electrons to get to the Ag-air interface to couple to surface plasmons, and thereby increasing the emission efficiency. These experimental results are difficult to explain in an inelastic tunneling model, since it is difficult to understand how a change in the scattering properties of the Ag film could affect the emission efficiency if the tunneling-electron-surface-plasmon interaction was in the tunneling region.

III. OPTICAL PUMPING

Further evidence in favor of hot-electron effects in light emission from tunnel junctions came from experiments in which the surface plasmons were pumped by other means. Laser radiation of a set energy incident on a tunnel junction on a grating exhibits a sharp reflectivity minimum at an angle corresponding to the angle at which the laser couples to surface-plasmon polaritons through the grating roughness. The decay of these surface plasmons must result in a distribution of hot electrons in the metal with energies up to the incident laser energy. The hot electrons can subsequently excite surface plasmons of lower energy, and plasma-mediated radiation at lower energies is indeed observed.^{20,21}

The laser- and tunneling-electron-pumped data show striking similarities. Figure 6 shows a plot of the Stokes-shifted radiation from a tunnel junction irradiated with 60 mw of laser radiation at 2.41 eV. The incident laser beam was held at the resonance angle so that about 90% of the incident energy was taken out of the specular beam to excite the surface wave. The scattered light was apertured as described above and frequency-analyzed using a double monochromator. The same optical system was used for both the laser-pumped and tunneling-pumped emission described in this section. The junction of Fig. 6 had a molecular monolayer of 4-nitrobenzoic acid absorbed on the aluminum oxide in the Al-Al₂O₃-Ag structure, and therefore showed sharp surface-enhanced Raman scattering peaks in addition to broader emission peaks corresponding to coupling out of the Ag-air fast surface-plasmon po-

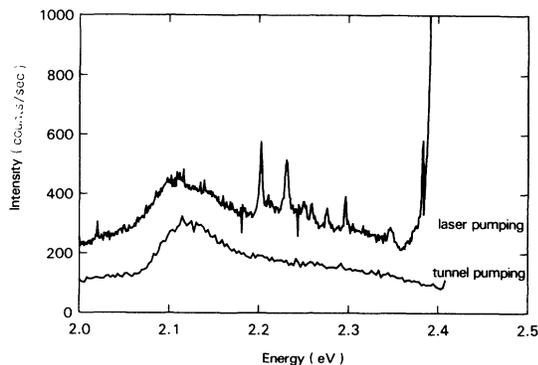


FIG. 6. Comparison of the surface-plasmon emission from an Al-Al₂O₃-Ag tunnel junction for laser pumping and tunneling pumping. The sharp lines in the laser-pumped case are surface-enhanced Raman-scattered light from a monolayer of 4-nitrobenzoic acid included in the junction region. The broad peaks at 2.12 eV are the surface-plasmon peaks.

lariton. The broad emission peaks moved in frequency and changed in intensity as the observation angle was changed, just as the emission peaks from tunneling pumping did. Figure 6 shows the emission spectrum of the same sample, with the same collection optics, but without an incident laser beam and with the junction biased at 2.41 V and 25 mA. The absence of sharp Raman scattering peaks in the tunneling-pumped emission spectrum is to be expected because the injected electron-energy distribution was relatively broad, but the broad emission peaks are present, regardless of how (optical pumping or tunnel current injection) the sample is excited.

Figure 7 contains a comparison of the plasmon emission-peak intensity as a function of peak energy (obtained simply by rotating the observation angle) for tunneling and optical pumping, with the total pumping power per unit area imaged on the spectrometer slit normalized out. Not only does the optically pumped data show a linear onset below the pump energy, but the total external quantum efficiency is very similar to that for the tunneling pumped case. The linear onset of emission intensities from light-emitting tunnel junctions below the bias voltage energy has been used as an argument for an inelastic tunneling process. The fact that this linear onset also occurs, with comparable efficiency, in the optical pumping case in which there are no tunneling electrons, indicates a hot-electron light-emission mechanism.

We argue that similar processes are occurring in the two cases: Hot-electron distributions are introduced by tunnel injection or optical absorption, the electron distributions relax in energy partially

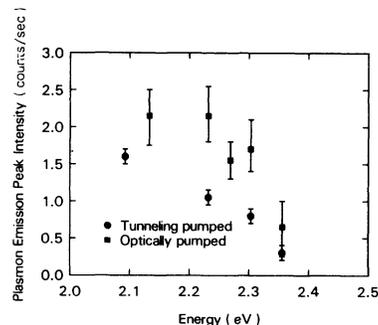


FIG. 7. Comparison between the plasmon emission-peak intensities from an Al-Al₂O₃-Ag tunnel junction under optical and tunneling pumping. The similar linear onset below the pump energy and similar total external quantum efficiencies indicate that similar processes are occurring in the two cases.

through the emission of surface plasmons, and the surface plasmons couple out to light through the grating roughness.

IV. CHARGE-INJECTION STRUCTURES

A third method for producing surface-plasmon emission from a thin metal film is by charge injection.²² A schematic of the charge-injection structures^{23,24} we used is shown in Fig. 8. A single-crystal degenerately *n*-type-doped (0.001 Ω cm) Si wafer was used as a substrate. A Si-rich ($\sim 15\%$ excess Si) SiO₂ layer (20–50)-nm thick was deposited on the substrate by chemical vapor deposition. This was followed by a stoichiometric SiO₂ layer of about 50 nm and then a gate-metal electrode of typically 20–25 nm was deposited by evaporation. The concentration of Si in the SiO_x was determined from the ratio of silane (SiH₄) to nitrous oxide (N₂O) in the gas phase during the chemical vapor deposition.^{23,24} The Si-rich SiO₂ had small ($< \sim 5$ -nm-diam) Si inclusions.²⁵ Electrons tunnel from Si inclusion to Si inclusion, until they reach the interface with the SiO₂ layer, where they are injected into the SiO₂ conduction band through field-enhanced Fowler-Nordheim tunneling. The electric field enhancement is caused by the nonplanar shape of the Si islands. This lowers the gate voltage required for current flow. Once in the conduction band, the electrons proceed rapidly through the SiO₂ and are injected into the gate-metal electrode. The principle advantage of the charge-injection structure over a planar Si-SiO₂-*M* structure is that it can inject large current densities ($> 10^2$ A/cm²) for long periods of time at moderate gate voltages without destructive

breakdown of the SiO₂ layer.

These structures had current-voltage characteristics that approximately followed the Fowler-Nordheim relation $I = AV^2 e^{-B/V}$. We typically ran at biases of 25 V with up to 100 μ A current through a sample area of 1 mm² (10^{-2} A/cm²). As current was passed through the devices, charge was trapped in the oxide, lowering the electric field at the injecting Si-rich SiO₂-SiO₂ interface so that more voltage was required to keep the current through the device constant. Eventually, sufficient charge was trapped and sufficiently large fields were built up to cause a destructive breakdown. We found that we could pass about 2 C/cm² through the devices before they failed.

The charge-injection devices had some intrinsic roughness because of the Si inclusions in the Si-rich oxide. We introduced additional roughness to some of the samples by depositing polysilicon on the Si substrates, doping it *n*-type degenerate with POCl₃, oxidizing the polysilicon to an SiO₂ thickness of about 50 nm, and then stripping the oxide. The oxide grows preferentially faster along grain boundaries and certain crystallite orientations, leaving a rough surface upon removal. Controlled roughness was also produced in some samples by etching holographically produced gratings into the Si substrates before the deposition of the SiO₂.

Figure 9 compares the emitted radiation from a polysilicon-roughened Al-gate charge-injection structure (LUM24-H) with that from an identical structure laid down on a single-crystal Si substrate (LUM24-X1). The Al-gate electrodes were 25 nm thick and were annealed at 400°C for 30 min in forming gas after deposition. The devices were pulsed with the gate at about 20–25 V positive with respect to the substrate, with an average current through the sample of 10 μ A. The pulses were 1-msec wide with a 5% duty cycle. The low duty cycle helped to discriminate against phototube dark counts. The spectra in Fig. 9 are normalized for the relative spectrometer throughput, with the throughput at 400 nm set arbitrarily to 1. The emission was collected with *F*/4.8 mirror collection optics. The radiation from the nominally smooth sample appears to be dominated by luminescence from the oxide, which is planned to be described in later publications. The additional emission resulting from roughness, indicated by the difference curve of Fig. 9, has a linear onset below a critical photon energy of about 4.5 eV, and falls off in intensity in the red, in a manner similar to that observed in light-emitting tunnel junctions. The fact that the emission spectrum changed dramatically with surface roughness indicates that the additional component is surface-plasmon mediated. The onset energy of 4.5

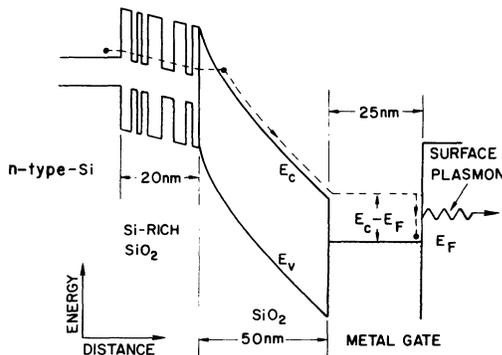


FIG. 8. Schematic energy-band diagram of a Si-Si-rich SiO₂-SiO₂-*M* charge-injection structure. The electrons tunnel through the Si-rich SiO₂ into the conduction band of the SiO₂, are injected into the gate-metal film at an energy corresponding to the step between the bottom of the SiO₂ conduction band and the Fermi level of the metal, and produce surface-plasmon radiation.

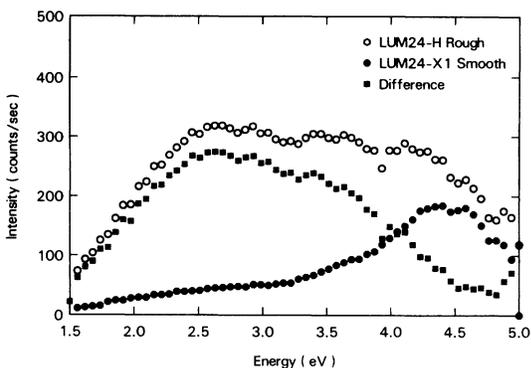


FIG. 9. Emission spectrum from a rough (LUM24-H, open circles) vs a smooth (LUM24-X1, closed circles) Al-gate charge-injection structure biased at $10 \mu\text{A}$. The difference spectrum, which is due to surface-plasmon-coupled radiation, has a linear onset at 4.5 eV, corresponding to the maximum energy, relative to the Fermi surface of the metal, of the injected electrons.

eV corresponds to the maximum energy (with respect to the Fermi level) of electrons injected into the metal electrode. This is larger than the potential step of 3.6 eV between the bottom of the conduction band of the SiO_2 and the Fermi level of the aluminum, obtained by Solomon and DiMaria²⁶ from similar devices using internal photoemission. Thus the spectral cutoff tells us that on the average, the injected electrons are heated about 1 V above the bottom of the oxide conduction band. This is the first direct evidence for such heating, although it has been predicted theoretically.²⁷

The total external quantum efficiency for LUM24-H was about 10^{-6} photons/electron, which is comparable to randomly roughened tunnel junction efficiencies of 10^{-4} , when one takes into account that the Al-gate electrode has much more highly damped surface-plasmon polaritons than the Ag and Au counter electrodes typically used for tunnel junctions.

A more dramatic proof that part of the emission from these charge-injection structures was surface-plasmon mediated was obtained by fabricating them on Si substrates with grating profiles. In Fig. 10 we show the light emission from a charge-injection device with a 25-nm-thick Al gate, fabricated on an 815-nm periodicity, 36-nm amplitude holographically produced grating. The emission was measured through a 7° wide and 16° high aperture. The spectrum shown in Fig. 10 was taken with an observation angle of 13° with respect to the sample normal. The position in energy of the sharp emission peak corresponded well to that predicted for the Al-vacuum interface fast surface-plasmon polariton.

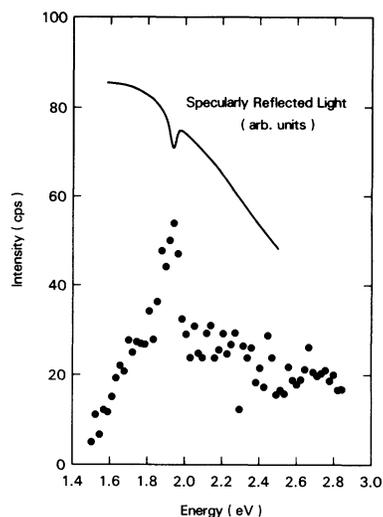


FIG. 10. Emission spectrum from an Al-gate charge-injection structure fabricated on a Si substrate with a grating profile. The sharp emission peak corresponds well to the dip in the specular reflection caused by coupling to Al-air interface surface plasmons. The total integrated efficiency of this peak emission is comparable to that observed from light-emitting tunnel junctions on gratings, when correction is made for the different dielectric properties of the two types of samples.

Also included in Fig. 10 is a plot of the specularly reflected radiation from a collimated beam from a tungsten lamp directed at the Al grating of the charge-injection structure, using the same geometry and collection optics as for the emission experiment. The reflectivity has a dip at the same energy as the peak in the emission, due to absorption of the incident radiation by surface plasmons. The specularly reflected light is plotted on a logarithmic scale. The dip corresponds to about a 90% decrease in the reflected intensity in *p* polarization. The emission peaks also varied in energy in the expected manner as the observation angle was changed.

There is a continuum background in Fig. 10 that was larger relative to the peak size than is typically observed for emission from Al- Al_2O_3 -Ag tunnel junctions on gratings. One reason was that the surface-plasmon resonance is weaker at the Al-air interface than at the Ag-air interface because of the relative dielectric properties of the two metals. The theory of Laks and Mills predicts a ratio of smooth background to peak height of about 2 to 1 for the Al-air interface plasmon emission at 2 V. We observe a ratio of about 1 to 1. There is probably also a contribution to the residual background from coupling caused by the random roughness of the

charge-injection layer, and also emission from the oxide itself.

The total integrated intensity under the sharp emission peak of Fig. 10 is 8.2×10^{-12} counts/electron sr. This is about 8 times smaller than would be predicted by analogy with our tunneling junction results using the theory of Laks and Mills¹³ (see, for example, Fig. 10 of Ref. 6) to take into account the differences in dielectric properties of an Al-air interface versus a tunnel junction structure, grating amplitude, and grating periodicity. This difference could well result from the fact that the charge-injection structure inject at 3.5 eV, higher than the ~ 2.5 eV for the tunnel junctions studied by Kirtley *et al.* Electrons at higher energies have shorter lifetimes, resulting in lower external efficiencies. Also, the charge-injection structures were run at room temperature, while the tunnel junctions were usually run at low temperatures. Emission efficiencies are higher at lower temperatures, as we have shown above.

We interpret our results from the charge-injection structures as follows: Electrons in the conduction band of the SiO₂ have relatively short mean free paths (2–6 nm),²⁸ and are therefore primarily near (~ 1 eV) the bottom of the SiO₂ conduction band at the field strengths we are using.²⁷ They are injected into the Al with a narrow energy distribution, but relax before emitting surface plasmons, giving the same characteristic linear onset observed in light emission from tunnel junctions. The very similar emission efficiencies and spectral properties in the charge-injection structures and the tunnel junctions indicate that similar physical processes are occurring in each. Since the oxide layer is about 50-nm thick in the charge-injection structure, direct tunneling through the SiO₂ is certainly not a conduction mechanism in these devices, and inelastic tunneling cannot be the driving factor behind the emission.

V. CONCLUSIONS

The experimental evidence presented above argues for a hot-electron mechanism, as opposed to an in-

elastic tunneling mechanism, for light emission from tunnel junctions, as well as from optically pumped thin films and charge-injection structures.

A very interesting puzzle arises from these results since light emission from tunnel junctions is symmetric, or nearly so¹ with respect to the bias voltage polarity (we have confirmed this observation for emission from junctions on gratings), but surface-plasmon-mediated emission from charge-injection structures only occurs when the metal gate is held positive. These observations can be reconciled with the other experiments reported in this paper if the tunneling process creates hot electrons (or holes) on both sides of the insulating region. The conventional one-electron picture of electron tunneling results in a hot electron being injected into one electrode, leaving a cold hole in the other. It may be that excitations on opposite sides of the insulating region are strongly coupled if the insulator is sufficiently thin, possibly through the slow surface-plasmon polariton mode.

Once suitable analysis methods have been developed, light emission from metal films will represent a probe of hot-electron dynamics in metals, and charge transport through insulators, in an energy regime and with an energy resolution unavailable to other techniques, and under a variety of different excitation conditions. Understanding the underlying mechanisms is therefore important for potential applications that extend well beyond the analysis of light emission from tunnel junctions. The hot-electron model is an aid in understanding these processes.

ACKNOWLEDGMENTS

We would like to acknowledge the assistance of A. M. Torreson, D. W. Dong, J. A. Tornello, and J. A. Calise in sample preparation, C. F. Alliota for electron microscopy, B. A. Ek for ion milling, and D. J. Scalapino and D. L. Mills for valuable discussions.

¹J. Lambe and S. L. McCarthy, Phys. Rev. Lett. **37**, 923 (1976).

²S. L. McCarthy and J. Lambe, Appl. Phys. Lett. **30**, 427 (1977).

³P. K. Hansma and H. P. Broida, Appl. Phys. Lett. **32**, 545 (1978).

⁴A. Adams, J. C. Wyss, and P. K. Hansma, Phys. Rev. Lett. **42**, 912 (1979).

⁵A. Adams and P. K. Hansma, Phys. Rev. B **23**, 3597 (1981).

⁶J. R. Kirtley, T. N. Theis, and J. C. Tsang, Phys. Rev. B **24**, 5650 (1981).

⁷J. R. Kirtley, T. N. Theis, and J. C. Tsang, Appl. Phys. Lett. **37**, 435 (1980).

⁸D. Hone, B. Muhlschlegel, and D. J. Scalapino, Appl. Phys. Lett. **33**, 203 (1978).

⁹R. W. Rendell, D. J. Scalapino, and B. Muhlschlegel, Phys. Rev. Lett. **41**, 1746 (1978).

¹⁰L. C. Davis, Phys. Rev. B **16**, 2482 (1977).

¹¹B. Laks and D. L. Mills, Phys. Rev. B **20**, 4962 (1979);

- 21, 5175 (1980).
- ¹²E. N. Economou, Phys. Rev. 182, 539 (1969).
- ¹³B. Laks and D. L. Mills, Phys. Rev. B 22, 5723 (1980).
- ¹⁴L. I. Schiff, *Quantum Mechanics* (McGraw-Hill, New York, 1968), p. 269.
- ¹⁵J. G. Simmons, J. Appl. Phys. 34, 1793 (1963).
- ¹⁶J. Moreland, A. Adams, and P. K. Hansma, Phys. Rev. B 25, 2297 (1982).
- ¹⁷C. R. Crowell and S. M. Sze, in *Physics of Thin Films*, edited by G. Hass and R. E. Thun (Academic, New York, 1967), Vol. IV, p. 325.
- ¹⁸H. J. Levinson, E. W. Plummer, and P. J. Feibelman, Phys. Rev. Lett. 43, 952 (1979).
- ¹⁹J. R. Kirtley, T. N. Theis, J. J. Chang, and D. L. Mills (unpublished).
- ²⁰J. C. Tsang, J. R. Kirtley, and T. N. Theis, Solid State Commun. 35, 667 (1980).
- ²¹A. Girlando, W. Knoll, and M. R. Philpott, Solid State Commun. 38, 895 (1981).
- ²²T. N. Theis, J. R. Kirtley, D. J. DiMaria, and D. W. Dong (unpublished).
- ²³D. J. DiMaria, R. Ghez, and D. W. Dong, J. Appl. Phys. 51, 4830 (1980).
- ²⁴D. J. DiMaria, K. M. DeMeyer, C. M. Serrano, and D. W. Dong, J. Appl. Phys. 52, 4825 (1981).
- ²⁵A. Hartstein, J. C. Tsang, D. J. DiMaria, and D. W. Dong, Appl. Phys. Lett. 36, 836 (1980).
- ²⁶P. M. Solomon and D. J. DiMaria, J. Appl. Phys. 52, 5867 (1981).
- ²⁷H.-J. Fitting and J.-U. Friemann, Phys. Status Solidi 69, 349 (1982).
- ²⁸R. Poirier and J. Oliver, Appl. Phys. Lett. 21, 334 (1972).